# Comparing Multivariate Distributions: A Novel Approach Using Optimal Transport-based Plots

Sibsankar Singha, Marie Kratz, Sreekar Vadlamani

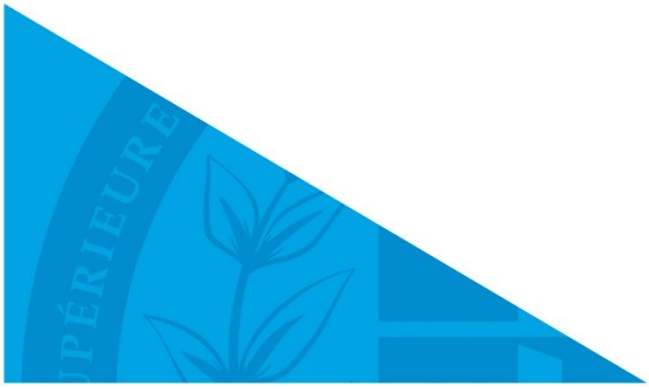
ESSEC BUSINESS SCHOOL

Enlighten. Lead. Change.

# COMPARING MULTIVARIATE DISTRIBUTIONS: A NOVEL APPROACH USING OPTIMAL TRANSPORT-BASED PLOTS

SIBSANKAR SINGHA, MARIE KRATZ AND SREEKAR VADLAMANI

**ESSEC RESEARCH CENTER**
WORKING PAPER 2401
MAY 13, 2024



# Comparing Multivariate Distributions: A Novel Approach Using Optimal Transport-based Plots

Sibsankar SINGHA[(a)]

Supervisors: Marie KRATZ[(b)] and Sreekar VADLAMANI[(c)]

[(a)] TIFR-CAM, Bangalore, India & ESSEC CREAR, France; Email: sibsankar@tifrbng.res.in

[(b)] ESSEC Business School, CREAR, Cergy-Pontoise, France; Email: kratz@essec.edu

[(c)] TIFR-CAM, Bangalore, India; Email: sreekar@tifrbng.res.in

April 30, 2024

**Abstract**

Quantile-Quantile (Q-Q) plots are widely used for assessing the distributional similarity between two datasets. Traditionally, Q-Q plots are constructed for univariate distributions, making them less effective in capturing complex dependencies present in multivariate data. In this paper, we propose a novel approach for constructing multivariate Q-Q plots, which extend the traditional Q-Q plot methodology to handle high-dimensional data. Our approach utilizes optimal transport (OT) and entropy-regularized optimal transport (EOT) to align the empirical quantiles of the two datasets. Additionally, we introduce another technique based on OT and EOT potentials which can effectively compare two multivariate datasets. Through extensive simulations and real data examples, we demonstrate the effectiveness of our proposed approach in capturing multivariate dependencies and identifying distributional differences such as tail behaviour. We also propose two test statistics based on the Q-Q and potential plots to compare two distributions rigorously.

*Keywords:* Q-Q plots; multivariate quantile; optimal transport; entropy regularisation; hypothesis testing; geometric quantile; tail behavior

# 1 Introduction

Univariate Quantile-Quantile (Q-Q) plot is a graphical tool used in statistics to assess whether two samples follow a similar distribution by comparing their respective quantiles. It is a simple yet powerful visualization technique that provides valuable insights into the shape, location, and scale of two samples.

While the univariate Q-Q plot is well defined, there is not much literature on multivariate Q-Q plots. This perhaps can be attributed to the absence of natural ordering. Easton and McCulloch [1990] introduced a permutation based multivariate Q-Q plot, where samples of equal length are matched against each other such that the total $\ell^2$ distance between them is minimised. Subsequently, the corresponding components of matched samples are plotted against one another in different bivariate plots. These plots then show visual evidence of similarity of underlying distributions of the two sets of samples. In Dhar et al. [2014], the authors used the crucial property of unique characterisation of underlying distributions by geometric quantile to develope componentwise Q-Q plots. Specifically, samples are matched according to their geometric rank, and then a componentwise plotting of matched samples provides insight into the underlying distributions.

In this context, we aim at extending this graphical tool using the optimal transport (OT) map, to which we refer as OT Q-Q plot. We prove that, as the sample size increases, the Q-Q plots are concentrating around the straight line passing through the origin and with slope 1, if and only if the two sets of samples are drawn from the same distribution. This characteristic enables us to visually compare two distributions.

Moreover, in the context of very large dimensions, the componentwise plot may not be ideal due to a large number of bivariate plots. So, we suggest an alternative approach based on OT potential function, which allows for the comparison of two sets of multivariate samples in a single bivariate plot, named OT potential plot. Similar to the componentwise plots, we establish that, as the sample size increases, the potential plot becomes arbitrarily close to the straight line with slope 1 and passing through the origin if and only if the underlying distributions of the two samples are identical.

Computing empirical OT maps can be costly, especially when sample sizes are large; see Hütter and Rigollet [2021]. Therefore, practical solutions have been proposed in the literature, one of the most popular approaches being the entropy regularisation, introduced in Cuturi [2013]. By selecting the regularisation parameter to be sufficiently small, the entropy regularised OT (EOT) map and EOT potential can closely approximate the OT map and OT potential, respectively. Moreover, we prove that the EOT map and EOT potential can also uniquely characterize the distribution.

These significant properties of EOT map and EOT potential serve as our driving force to also build Q-Q plots (we call it Q-Q, although EOT maps are not quantiles) and potential plots based on them. Finally, as given in Shapiro and Wilk [1965], Dhar et al. [2014], we propose test statistics to assess the relevance of the proposed Q-Q plots and potential plots.

We apply this (E)OT approach on examples from simulated and real data, then compare (E)OT Q-Q plots with geometric ones.

Our contributions are the following:

(i) We construct OT Q-Q plots for comparing multivariate distributions, as we can prove that these plots are concentrating, as the sample size increases, around the straight line passing through the origin and with slope 1, if and only if the two sets of samples are drawn from the same distribution.

(ii) We also propose OT potential plot, which is shown to share the same unique characterization property as for OT maps. This tool is quite interesting as it gives a single bivariate plot, whatever the dimension of the distribution.

(iii) We develop the same tools and characteristic properties as for OT when considering the EOT approach, widely used because less computationally challenging than the OT approach.

(iv) We propose test statistics to assess the relevance of the Q-Q and potential plots as tools for comparing multivariate distributions, for the EOT approach.

(v) We question the benefit of using (E)OT Q-Q or potential plots, to retrieve specific features of the distributions considered, specifically in the tail. We do so via an extensive simulation study.

(vi) We compare (E)OT Q-Q plots with geometric Q-Q plots developed in Dhar et al. [2014], on simulated and real data. We observe that the former is better performing visually in general (on our examples) than the latter to characterize tail distributions.

*Notation:* $\|\cdot\|$ denotes the $\ell^2$ norm, $P_d^{conv}$ the set of all probability measures on $\mathbb{R}^d$ with convex support, $\langle\cdot\rangle$ the usual inner product on $\mathbb{R}^d$, and $X \sim \nu$ means that $X$ has distribution $\nu$. $\pi \ll P$ implies that the measure $\pi$ is absolutely continuous with respect to the measure $P$. The measure $\mu$ is assumed to be uniform on the open unit ball $B^d$ unless otherwise specified. $\Pi(\mu, \nu)$ is the set of all couplings with first marginal $\mu$ and second marginal $\nu$, and $\nabla \equiv (\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_n})$. Let $\nu$ be a probability measure and $K$ be a $\nu-$measurable set with $\nu(K) > 0$. Then define the probability measure $\nu|_K(\cdot)$ such that $\nu|_K(A) = \frac{\nu(A \cap K)}{\nu(K)}$ for all $\nu-$measurable sets $A$.

*Structure of the paper:* In Section 2, we briefly recall the concepts related to OT, EOT maps, and potentials, together with their connections to the multivariate quantile function. Section 3 includes discussions and results related to constructions of multivariate Q-Q plots and potential plots based on OT and EOT. In Section 4, we propose two test statistics and study their asymptotics. In Section 5, we illustrate our (E)OT approach on simulated samples to compare multivariate distributions on various scenarios and experiments. In Section 6, we apply the method on real data. In Section 7, we compare results obtained with geometric and (E)OT Q-Q plots, respectively, first on simulated data, then on real data.

# 2 Background and preliminaries on multivariate quantile and potential function

## 2.1 Optimal transport approach

The motivation for defining multivariate quantile through optimal transport arises from certain inherent properties of the univariate quantile functions. One key property of the univariate quantile function is its nondecreasing nature. Another compelling aspect is that any probability measure is the pushforward of the uniform measure on the interval $(0, 1)$ by its quantile function. These two properties together uniquely characterize a quantile function; any function meeting these criteria must be a quantile. This property of transporting measure in high dimensions is at the heart of the theory of optimal transport and has a very rich literature (see Villani [2003, 2009], Santambrogio [2015] for an exhaustive review). In particular, the following theorem due to Brenier [1991] and McCann [1995], proves the existence and uniqueness of a map $T$ that pushes forward a measure $\mu$ to $\nu$ if $\mu$ is absolutely continuous with respect to the Lebesgue measure. Moreover, the map $T$ is the gradient of a convex function, i.e. $T = \nabla\phi$. The property that $T$ can be written as the gradient of a convex function, is analogous to the monotonicity property of univariate quantile function.

**Theorem 2.1 (Brenier–McCann)** *Let $\mu$ and $\nu$ be two probability measures on $\mathbb{R}^d$. Assume that $\mu$ is absolutely continuous with respect to the Lebesgue measure. Then, there exists a convex function $\varphi$ on $\mathbb{R}^d$, which is $\mu$ almost everywhere unique up to an additive constant, such that $\nabla\varphi$ pushes forward the measure $\mu$ to the measure $\nu$, notationally $\nabla\varphi\sharp\mu = \nu$. If $\nu$ is also absolutely continuous with respect to the Lebesgue measure, then $\nabla\varphi$ is invertible and $(\nabla\varphi)^{-1} = \nabla\varphi^*$, where $\varphi^*$ is the Fenchel–Legendre transform of $\varphi$. Additionally, if both $\mu$ and $\nu$ have finite second moment, then $\nabla\varphi$ uniquely solves the Monge problem:*

$$\underset{T:T\sharp\mu=\nu}{\arg\min} \int_{\mathbb{R}^d} \|x - T(x)\|^2 \mu(dx). \tag{2.1}$$

Let $\mu$ be the uniform probability measure on the unit ball $B^d$ and $\nu$ be any probabilty measure on $\mathbb{R}^d$, then Theorem 2.1 ensures the existence of a unique monotone map (the gradient of a convex function), which pushes forward $\mu$ to $\nu$. The following definition of multivariate quantile function is from Chernozhukov et al. [2017] with minor modification, as explained in Remark 2.4.

**Definition 2.2 (OT quantile; Chernozhukov et al. [2017])** *Let $\mu$ be the uniform probability measure on $B^d$. The function defined by $T_\nu = \nabla\varphi|_{B^d}$ such that $\nabla\varphi\sharp\mu = \nu$ for a convex function $\varphi$, is defined as the OT quantile function of the probability measure $\nu$.*

Although the OT quantile is almost everywhere unique, the convex functions in Theorem 2.1 are unique up to additive constants. Therefore setting its infimum to 0 will ensure their uniqueness.

**Definition 2.3 (OT Potential; Chernozhukov et al. [2017])** *Let $\mu$ be the uniform probability measure on the unit ball $B^d$. A convex function $\varphi_\nu$ satisfying $\nabla\varphi_\nu\sharp\mu = \nu$ and $\inf_{u\in B^d} \varphi_\nu(u) = 0$ is called the OT potential of $\nu$.*

**Remark 2.4**
1. *Since any convex function on bounded open set is bounded below by a finite number we can subtract the infimum to satisfy the condition $\inf_{u\in B^d} \varphi_\nu(u) = 0$. This will ensure that the potential function is non-negative everywhere.*
2. *In the definitions above, the fixed measure $\mu$ is assumed to be the uniform probability measure on the unit ball. However, one can choose a different reference distribution; for instance, in Chernozhukov et al. [2017], Hallin et al. [2021], $\mu$ is considered to be the spherically uniform measure. We refer the reader to [Ghosal and Sen, 2022, Remark 3.11] for discussion on the effect of choosing different reference measures.*

Later in this paper, we will extensively use the following properties of both OT quantile and OT potential, which are a straightforward consequence of Theorem 2.1.

**Properties 2.5 (Unique characterisation of measure)** *Let $\nu_1$ and $\nu_2$ be two probability measures. Then,*

*(i) $\nu_1 = \nu_2$ if and only if $\varphi_{\nu_1}(u) = \varphi_{\nu_2}(u)$, for $\mu-$almost every $u \in B^d$, where $\varphi_{\nu_1}, \varphi_{\nu_2}$ are OT potentials.*

*(ii) $\nu_1 = \nu_2$ if and only if $T_{\nu_1}(u) = T_{\nu_2}(u)$, for $\mu-$almost every $u \in B^d$, where $T_{\nu_1}, T_{\nu_2}$ are the OT quantiles.*

**Empirical OT quantile -** When a distribution on $\mathbb{R}^d$ is supported on finite points (i.i.d. samples), the empirical quantile function can be defined in either of the following ways:

- Discrete to discrete: In this case, the empirical quantile function is an optimal transport map between two sets of samples. Thus, clearly, the quantile function is supported on the finite sample, generated uniformly on $B^d$.
- Continuous to discrete or semi-discrete: In this case, the quantile function is the transport map between the uniform measure on $B^d$ and a finite sample. Therefore, the map is defined on the whole unit ball. See Chernozhukov et al. [2017] for more details.

For simplicity, we consider the discrete to discrete version of empirical OT quantiles; and refer the reader to Chernozhukov et al. [2017], Hallin et al. [2021], Ghosal and Sen [2022] for more properties on empirical OT quantiles.
For the purpose of this work, we shall consider i.i.d. samples $\mathcal{X}^n = \{X_1, X_2, ..., X_n\}$ with $X_i \sim \nu$ and $\mathcal{U}^n = \{U_1, U_2, \cdots, U_n\}$ with $U_i \sim \mu$, $\mu$ being the uniform measure on $B^d$. Then, we define the empirical measures corresponding to $\mathcal{X}^n$ and $\mathcal{U}^n$ as

$$\nu_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}(\cdot) \qquad \text{and} \qquad \mu_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \delta_{U_i}(\cdot). \tag{2.2}$$

**Definition 2.6 (Empirical OT quantile; Hallin et al. [2021])** *The empirical quantile function of the empirical measure $\nu_n$ with respect to the reference measure $\mu_n$ is a function $T_{\nu_n} : \mathcal{U}^n \to \mathcal{X}^n$ such that*

$$\sum_{i=1}^{n} \|U_i - T_{\nu_n}(U_i)\|^2 = \min_{\sigma_n \in \mathcal{S}_n} \sum_{i=1}^{n} \|X_i - U_{\sigma_n(i)}\|^2, \tag{2.3}$$

*where $\mathcal{S}_n$ denotes the set of permutations of the set $\{1, \ldots, n\}$.*

**Remark 2.7** *The minimization problem (2.3) resembles the Monge problem in discrete setting. Any solution to this problem satisfies the cyclical monotonicity property, which is a unique feature of the OT map (see Villani [2003, 2009]).*

**Remark 2.8** *Observe that the function $T_{\nu_n}$ is random, and is supported on the random set $\mathcal{U}^n$. Although in our notation the empirical OT quantile $T_{\nu_n}$ seems to depend only on the measure $\nu_n$, note that it is defined with respect to a reference $\mathcal{U}^n$. In Hallin et al. [2021], instead of taking a random sample $\mathcal{U}^n$, the authors construct a deterministic grid on $B^d$ to define empical OT quantiles.*

**Remark 2.9** *The map $T_{\nu_n}$ may not be unique. For example, let $\mathcal{U}^2 = \{(0,0), (.5,.5)\}$ and $\mathcal{X}^2 = \{(0,.5), (.5,0)\}$, and note that, there are only two possible bijective maps from $\mathcal{U}^2$ to $\mathcal{X}^2$ and the cost remains the same for each of them. Therefore there are two possibilities for the map $T_{\nu_2}$. However, since $\mathcal{U}^n$ is a sample drawn from an absolutely continuous distribution $\mu$, the event that $T_{\nu_n}$ is not unique has probability zero; see Hallin et al. [2021].*

For the empirical OT potential, we will adopt the definition from Chernozhukov et al. [2017], where it is defined such that its gradient coincides with the empirical quantile function on its finite support. The definition is based on the dual formulation of the *Monge problem* (2.1).

**Definition 2.10 (Empirical OT potential; Chernozhukov et al. [2017])** *Let $\mu_n$ and $\nu_n$ be as defined in (2.2). The empirical OT potential of $\nu_n$ with respect to $\mu_n$ (the reference measure) is a convex function $\varphi_n$ such that $(\varphi_n, \varphi_n^*) \in \Phi_0$ solves the following dual problem*

$$\int \varphi_n \, d\mu_n + \int \varphi_n^* \, d\nu_n = \inf_{(\psi,\psi^*)\in\Phi_0} \left( \int \psi \, d\mu_n + \int \psi^* \, d\nu_n \right), \tag{2.4}$$

*where $\Phi_0$ denotes the set of all pairs of convex conjugates $(\psi, \psi^*)$ such that $\inf_{u \in B^d} \psi(u) = 0$.*

Note that $\nabla\varphi_n$, when restricted to the set $\mathcal{U}^n$, coincides with the empirical quantile function $T_{\mathcal{X}^n}$.

**Lemma 2.11** *Let $\nu$ be an absolutely continuous probability measure with convex support. Let $T_\nu$ and $T_{\nu_n}$ be population and empirical quantile functions with respect to $\mu$ and $\mu_n$, respectively. Then, for any compact set $K \subset B^d$,*

$$\max_{\substack{U_i \in K \\ 1 \leq i \leq n}} |T_{\nu_n}(U_i) - T_\nu(U_i)| \xrightarrow{a.s.} 0, \qquad \text{as } n \to +\infty.$$

**Proof.** Since $\nu$ is absolutely continuous, the quantile function $T_\nu$ is bijective. The inverse of the quantile is defined as the multivariate distribution function or OT distribution function and denoted by $F_\nu$, i.e. $T_\nu = F_\nu^{-1}$. Similarly, the empirical distribution function $F_{\nu_n}$ is the inverse of the empirical quantile, i.e. $T_{\nu_n} = (F_{\nu_n})^{-1}$.
Note that the support of $\nu$ is convex, $T_\nu : B^d \to \mathbb{R}^d$ is a homeomorphism between $B^d$ and interior of $supp(\nu)$. Let $T_{\nu_n}(U_i) = X_{\sigma_n(i)}$. Then we can write,

$$|T_{\nu_n}(U_i) - T_\nu(U_i)| = \left|X_{\sigma_n(i)} - T_\nu(U_i)\right| = \left|T_\nu(F_\nu(X_{\sigma_n(i)})) - T_\nu(F_{\nu_n}(X_{\sigma_n(i)}))\right|,$$

which, combined with Proposition 2.4 in Hallin et al. [2021] and the fact that $T_\nu$ is uniformly continuous on compact sets, concludes the proof of Lemma 2.11. □

**Lemma 2.12** *Assume that the support of $\nu$ ($supp(\nu)$) is compact and the optimal transport map $T_\nu$ is a homeomorphism from the open unit ball $B^d$ to the interior of $supp(\nu)$. Then, for any compact $K \subset B^d$,*

$$\sup_{u \in K} |\varphi_{\nu_n}(u) - \varphi_\nu(u)| \xrightarrow{\mathbb{P}^*} 0, \qquad \text{as } n \to +\infty,$$

*where $\xrightarrow{\mathbb{P}^*}$ denotes the convergence in outer probability in the sense of van der Vaart and Wellner [1996].*

**Proof.** It is a straightforward adaptation of the proof of [Chernozhukov et al., 2017, Theorem 3.1] developed for OT maps, to OT potentials.

## 2.2 Entropy regularised optimal transport

While the optimal transport map provides a meaningful definition of the multivariate quantile function, it comes along with practical challenges. Estimating the optimal transport for large samples is costly, with $O(n^3 \log n)$ complexity for a sample size $n$. Cuturi [2013] proposed an entropy regularised optimal transport (EOT), which closely approximates the optimal transport, and reduces the complexity to $O(n^2)$. In the following, we shall briefly review the basic concepts and properties of EOT.

Given two measures $\mu$ and $\nu$, a weak version of the *Monge problem*, known as the *Kantorovich relaxation*, is given by

$$\underset{\pi\in\Pi(\mu,\nu)}{\arg\min} \int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{1}{2}\|x-y\|^2 d\pi(x,y), \tag{2.5}$$

where $\Pi(\mu,\nu)$ is the set of all couplings between $\mu$ and $\nu$. Since $\Pi(\mu,\nu)$ is compact in the weak topology and the cost function is lower semicontinuous, therefore the existence of a minimiser to the *Kantorovich problem* is guaranteed (see [Villani, 2009, Theorem 4.1]), whereas the minimiser does not always exist in the *Monge problem*. Additionally, the *Kantorovich problem* has a unique minimiser, denoted by $\pi^*$, whenever both $\mu$ and $\nu$ have finite second moment and $\mu$ is absolutely continuous with respect to the Lebesgue measure. Furthermore, the unique minimiser $\pi^*$ is supported on the graph of the OT map (Brenier map) between $\mu$ and $\nu$. In particular, $\pi^* = (\nabla\varphi, I)_\sharp\mu$, where $\nabla\varphi$ is the OT map and the convex function $\varphi$ is the OT potential. Alternatively, we can write

$$\nabla\varphi(u) = E_{\pi^*}[X|U=u], \text{ for } \mu \text{ - almost every } u, \tag{2.6}$$

where $(U,X)\sim\pi^*$. This establishes link between OT map ($\nabla\varphi$) and the solution ($\pi^*$) of the *Kantorovich problem*.

We will now discuss the entropy regularised version of the *Kantorovich problem*, which is defined by adding an extra penalty in the cost function $\int_{\mathbb{R}^d\times\mathbb{R}^d}\frac{1}{2}\|x-y\|^2 d\pi(x,y)$. Specifically, for $\varepsilon>0$ and probability measures $\mu$ and $\nu$ which have finite second moment, the entropy regularised *Kantorovich problem* is defined as

$$\underset{\pi\in\Pi(\mu,\nu)}{\arg\min}\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}\frac{1}{2}\|u-x\|^2 d\pi(u,x) + \varepsilon KL(\pi|P)\right), \tag{2.7}$$

where $P=\mu\otimes\nu$ (the product measure) and $KL(\pi|P)$ is the the Kullback–Leibler divergence given by:

$$KL(\pi|P) = \begin{cases} \mathbb{E}_\pi[\log\frac{d\pi}{dP}], & \text{if } \pi\ll P,\\ \infty, & \text{otherwise.}\end{cases}$$

This problem is known to have a unique minimiser $\pi^*_\varepsilon\in\Pi(\mu,\nu)$, as proven in [Nutz, 2022, Theorem 4.2].

Now, we recall the relation between OT map and the *Kantorovich solution* in (2.6), and analogously define the EOT map. In particular, the EOT map is defined as the conditional expectation with respect to $\pi^*_\varepsilon$.

**Definition 2.13 (EOT map; Pooladian and Niles-Weed [2022])** *Let $\nu$ be any probability measure defined on $\mathbb{R}^d$ with finite second moment. Let $\mu$ be the uniform probability measure on $B^d$. Consider $(U,X)\sim\pi^*_\varepsilon$, then the EOT map $T^\varepsilon_\nu(u)$ is defined as:*

$$T^\varepsilon_\nu(u) = E_{\pi^*_\varepsilon}[X|U=u], \text{ for } \mu-\text{almost every } u.$$

We shall now explore an alternative representation of the EOT map. Specifically, analogous to OT map, we observe that EOT map can also be represented as the gradient of some potential function (not convex in general). This is achieved in duality. The EOT problem (2.7) admits strong duality, shown in [Genevay, 2019, Theorem 7], in the following way: let $C_\varepsilon$ denote the cost of the regularised version of the *Kantorovich problem*, i.e.

$$C_\varepsilon = \min_{\pi \in \Pi(\mu,\nu)} \left( \int \frac{1}{2} \|u - x\|^2 d\pi(u,x) + \varepsilon KL(\pi|P) \right), \tag{2.8}$$

then,

$$C_\varepsilon = \max_{(\varphi^\varepsilon, \psi^\varepsilon) \in L^1(\mu) \times L^1(\nu)} \left( \int \varphi^\varepsilon d\mu + \int \psi^\varepsilon d\nu + \varepsilon \int e^{\frac{\varphi^\varepsilon(u) + \psi^\varepsilon(x) - \frac{1}{2}\|u-x\|^2}{\varepsilon}} d(\mu \otimes \nu) - \varepsilon \right). \tag{2.9}$$

If $\mu$ and $\nu$ have finite second moment, the maximiser of (2.9) is unique, $\mu \otimes \nu$–almost surely, up to an additive constant. More precisely, if $(\varphi_1^\varepsilon, \psi_1^\varepsilon)$ is another pair of solutions to the dual Kantorovich problem, then there exists a constant $a$ such that $\varphi_1^\varepsilon = \varphi^\varepsilon + a$ and $\psi_1^\varepsilon = \psi^\varepsilon - a$. Importantly, the EOT map $T_\nu^\varepsilon$ can be expressed as the gradient of an EOT potential function (see Section 3 in Pooladian and Niles-Weed [2022]), i.e.

$$T_\nu^\varepsilon(u) = u - \nabla \varphi^\varepsilon(u) = \nabla \left( \frac{\|u\|^2}{2} - \varphi^\varepsilon(u) \right). \tag{2.10}$$

**Definition 2.14 (EOT potential)** *Let $(\varphi^\varepsilon, \psi^\varepsilon)$ be a maximiser of (2.9), where $\mu$ is the uniform probability measure on the unit ball $B^d$ and $\nu$ is any probability measure defined on $\mathbb{R}^d$ with finite second moment. Then the function defined by*

$$\varphi_\nu^\varepsilon(u) = \frac{\|u\|^2}{2} - \varphi^\varepsilon(u) - \left[ \varphi^\varepsilon(u_0) - \frac{\|u_0\|^2}{2} \right], \ u \in B^d,$$

*where $u_0 \in B^d$ is a fixed point, is called the EOT potential corresponding to $\nu$.*

**Remark 2.15** *Whenever minimisers exist, we will set $u_0 = \arg\min_{u \in B^d} \left[ \varphi^\varepsilon(u) - \frac{\|u\|^2}{2} \right]$, which will ensure that the EOT potential is non-negative everywhere.*

Now, we move to the unique characterisation property of EOT maps and potentials.

**Proposition 2.16** *Let $\mu$ be the uniform measure on the unit ball $B^d$ and $\nu_1$, $\nu_2$ be two probability measures with finite second moment. Let $\varepsilon > 0$ be fixed.*

*(i) Let $\varphi_{\nu_1}^\varepsilon$ and $\varphi_{\nu_2}^\varepsilon$ be EOT potentials corresponding to $\nu_1$ and $\nu_2$, respectively. Then, $\nu_1 = \nu_2$ if and only if $\varphi_{\nu_1}^\varepsilon(u) = \varphi_{\nu_2}^\varepsilon(u)$, for $\mu$-almost every $u$.*

*(ii) Consider two EOT maps $T_{\nu_1}^\varepsilon$ and $T_{\nu_2}^\varepsilon$ corresponding to the measures $\nu_1$ and $\nu_2$. Then, $\nu_1 = \nu_2$ if and only if $T_{\nu_1}^\varepsilon(u) = T_{\nu_2}^\varepsilon(u)$, for $\mu$-almost every $u$.*

**Proof:** (i) If $\nu_1 = \nu_2$, the proof follows from the fact that the optimisation problem in (2.9) has a unique minimiser up to the addition of a constant.
Now let us prove the reverse implication by contradiction. Suppose $\varphi^{\varepsilon}_{\nu_1}(u) = \varphi^{\varepsilon}_{\nu_2}(u)$ for $\mu$ almost every $u$ and for some $\varepsilon > 0$, but $\nu_1 \neq \nu_2$. Since any pair of EOT potentials $(\varphi^{\epsilon}_{\nu}, \psi^{\epsilon}_{\nu})$ satisfies the Schrödinger system (see [Nutz, 2022, Section 1]),

$$\int e^{\frac{\varphi^{\varepsilon}_{\nu}(u)+\psi^{\varepsilon}_{\nu}(x)-\frac{1}{2}\|u-x\|^2}{\varepsilon}} d\mu(u) = 1, \qquad \text{for } \nu - \text{ almost every } x,$$

$$\int e^{\frac{\varphi^{\varepsilon}_{\nu}(u)+\psi^{\varepsilon}_{\nu}(x)-\frac{1}{2}\|u-x\|^2}{\varepsilon}} d\nu(x) = 1, \qquad \text{for } \mu - \text{ almost every } u,$$

we have

$$\psi^{\varepsilon}_{\nu_1}(x) = -\varepsilon \log \int e^{\frac{\varphi^{\varepsilon}_{\nu_1}(u)-\frac{1}{2}\|u-x\|^2}{\varepsilon}} d\mu(u) = \psi^{\varepsilon}_{\nu_2}(x) \tag{2.11}$$

where the second equality follows by replacing $\varphi^{\varepsilon}_{\nu_1}(u)$ with $\varphi^{\varepsilon}_{\nu_2}(u)$. Then, the proof follows from [Nutz, 2022, Theorem 4.2].
(ii) The proof is similar to (i) after observing that the EOT map can be written in terms of potential function, i.e. $T^{\varepsilon}_{\nu}(u) = u - \nabla\varphi^{\varepsilon}_{\nu}(u)$ ( see (2.10)). □

Let us now consider the empirical counterpart of EOT map and potential, which will be later used to develop Q-Q and potential plots.

**Empirical EOT map and its convergence** Let $\mathcal{X}_n = \{X_1, \cdots, X_n\}$ and $\mathcal{U}^n = \{U_1, U_2, \cdots, U_n\}$ be two sets of i.i.d. samples drawn from $\nu$ and $\mu$ (the uniform measure on $B^d$), respectively. Let $\nu_n$ and $\mu_n$ be empirical measures induced from $\mathcal{X}^n$ and $\mathcal{U}^n$, as defined in (2.2).

**Definition 2.17 (Empirical EOT map and potential)** *The empirical EOT map and potential are obtained by replacing $\nu, \mu$ with their empirical counterparts $\nu_n, \mu_n$ in Definitions 2.13 and 2.14, respectively. We denote the empirical EOT map by $T^{\varepsilon}_{\mathcal{X}_n}$ and the potential by $\varphi^{\varepsilon}_{\mathcal{X}^n}$.*

**Remark 2.18**

1. *Notice that, although the measure $\mu_n$ is supported on a finite set $\mathcal{U}^n$, the empirical EOT map $T^{\varepsilon}_{\mathcal{X}^n}$ is by default defined on $B^d$. However, recall that the empirical OT quantile is defined only on sample points $\mathcal{U}^n$.*

2. *If $\nu$ is supported on a compact set, then, for a fixed $\varepsilon > 0$, the empirical function $(T^{\varepsilon}_{\mathcal{X}^n}, \varphi^{\varepsilon}_{\mathcal{X}^n})$ converges almost surely to $(T^{\varepsilon}_{\nu}, \varphi^{\varepsilon}_{\nu})$ as $n \to \infty$; see [Goldfeld et al., 2023, Lemma 1(iii)].*

3. *Note that in Definition 2.17, the number of observations in $\mathcal{X}^n$ and $\mathcal{U}^n$ is the same. In the case of two different sample sizes, i.e. $\mathcal{X}^n$ and $\mathcal{U}^m$ with $n \neq m$, we can similary define emprical EOT map and potential by replacing $\mu_n$ with $\mu_m$ in the definition.*

# 3 Construction of multivariate Q-Q plots

Let $\mathcal{X}^n = \{X_1, \cdots, X_n\}$ and $\mathcal{Y}^n = \{Y_1, \cdots, Y_n\}$ be two sets of i.i.d. samples generated from $\nu_X$ and $\nu_Y$, respectively, and $\mathcal{U}^n$ generated from the uniform measure on $B^d$. Consider $T_{\mathcal{X}^n}$ and $T_{\mathcal{Y}^n}$ the empirical quantile functions with respect to the reference sample $\mathcal{U}^n$. For each $i \in \{1, \cdots, d\}$ and $K \subset B^d$, define the set

$$\Delta_i(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K) = \{(\langle T_{\mathcal{X}^n}(U_j), e_i\rangle, \langle T_{\mathcal{Y}^n}(U_j), e_i\rangle) : U_j \in K, j = 1, \cdots, n\}, \quad (3.1)$$

where $e_i$ is the $i-$th canonical basis vector. Therefore $\Delta_i(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K)$ is a subset of $\mathbb{R}^2$, where each element is a pair of the $i-$th components of the OT quantile functions $T_{\mathcal{X}^n}$ and $T_{\mathcal{Y}^n}$. Now consider the empirical optimal potentials $\varphi_{\mathcal{X}_n}$ and $\varphi_{\mathcal{Y}_n}$. Then, for $K \subset B^d$, define

$$\Gamma(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K) = \{(\varphi_{\mathcal{X}_n}(U_j), \varphi_{\mathcal{Y}_n}(U_j)) : U_j \in K,\ j = 1, \cdots, n\}. \quad (3.2)$$

In Theorem 3.2, we show that, as $n$ increases, $\Delta_i(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K)$ and $\Gamma(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K)$ become arbitrarily close to the straight line with slope 1 and intercept 0 if and only if $\nu_X(\cdot) = \nu_Y(\cdot)$. With this property, we can define the OT Q-Q plots and OT potential plots as follows:

**Definition 3.1** *Let $K \subset B^d$ be a compact set. An OT Q-Q plot is a collection of $d$ individual scatter plots, where the $i$–th one displays the set $\Delta_i(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K)$ defined in (3.1). Similarly, a scatter plot of the set $\Gamma(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K)$ defined in (3.2), is called OT potential plot.*

Let $L_\eta$ be the $\eta$-neighbourhood of the diagonal line $L = \{(u, u) : u \in \mathbb{R}\}$, i.e.

$$L_\eta = \{(x, y) \in \mathbb{R}^2 : (x - u)^2 + (y - u)^2 < \eta^2, \text{ for some } u \in \mathbb{R}\}. \quad (3.3)$$

**Theorem 3.2** *Let $\nu_X, \nu_Y$ be probability measures, defined in $\mathbb{R}^d$, with convex support, i.e. $\nu_X, \nu_Y \in \mathcal{P}_d^{conv}$. Let $\mathcal{X}^n = \{X_1, \cdots, X_n\}$ and $\mathcal{Y}^n = \{Y_1, \cdots, Y_n\}$ be two i.i.d. samples drawn from $\nu_X$ and $\nu_Y$, respectively. Let $\mathcal{U}^n = \{U_1, \cdots, U_n\}$, where $U_i$'s are i.i.d. with common distribution $\mu$. Then, $\nu_X = \nu_Y$ if and only if one of the following holds*

*(i)* $\mathbb{P}\left(\liminf_{n\to\infty} \bigcap_{i=1}^{d} \{\Delta_i(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K) \subseteq L_\eta\}\right) = 1$, *for all compact $K \subset B^d$ and all $\eta > 0$.*

*(ii)* $\lim_{n\to\infty} \mathbb{P}^*\left(\{\Gamma(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K) \subseteq L_\eta\}\right) = 1$, *for all compact $K \subset B^d$ and all $\eta > 0$, where $\mathbb{P}^*$ is the outer probability, in the sense of van der Vaart and Wellner [1996].*

**Proof.** (i) First assume $\nu_X = \nu_Y$. Then the corresponding quantile functions are also equal, $T_{\nu_X} = T_{\nu_Y}$. By the construction of the set $\Delta_i(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K)$ (see (3.1)), it follows

that, for any fixed $\varepsilon > 0$,

$$\{\Delta_i(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K) \subseteq L_\eta, \forall i = 1, \cdots, d\} \supseteq \left\{ \max_{\substack{1\le j\le n \\ U_j \in K}} \|T_{\mathcal{X}_n}(U_j) - T_{\mathcal{Y}_n}(U_j)\|^2 \le 2\eta^2 \right\}, \tag{3.4}$$

where $T_{\mathcal{X}_n}$ and $T_{\mathcal{Y}_n}$ are discrete to discrete optimal transport maps from $\mathcal{U}_n$ to $\mathcal{X}_n$ and $\mathcal{U}_n$ to $\mathcal{Y}_n$, respectively. To prove the first part of the theorem, it is enough to show that the probability of the event in the right side of (3.4) converges to 1 as $n$ increases to infinity. Now the sample quantile functions $T_{\mathcal{X}_n}$, $T_{\mathcal{Y}_n}$ converge uniformly to $T_{\nu_X}$ and $T_{\nu_Y}$, respectively (by Lemma 2.11), and $\lim_{n\to\infty} T_{\mathcal{X}_n} = \lim_{n\to\infty} T_{\mathcal{Y}_n}$ by assumption, hence the result.
Conversely, assume that (i) holds for all $\eta > 0$ and for any compact subset $K \subset B^d$. Therefore, almost surely, $\|T_{\mathcal{X}_n}(U_j) - T_{\mathcal{Y}_n}(U_j)\|$ converges to 0 as $n$ goes to infinity. Now, the convergence of empirical OT quantile to its population counterpart (see Lemma 2.11) implies that $T_{\nu_X}(U_j) = T_{\nu_Y}(U_j), \forall j \ge 1$. Since the random set $\{U_n : n \ge 1\}$ is almost surely dense in the unit ball $B^d$, and $T_{\nu_X}, T_{\nu_Y}$ are continuous, thus $T_{\nu_X}(u) = T_{\nu_Y}(u),\ \forall u \in B^d$. Now from Proposition 2.5, we conclude that $\nu_X = \nu_Y$.

(ii) If $\nu_X = \nu_Y$, the proof is similar to the first part of (i).
Conversely, assume that (ii) holds for all compact subsets $K$ and $\eta > 0$. Then we have,

$$\mathbb{P}^* \left( \max_{\substack{1\le j\le n \\ U_j \in K}} |\varphi_{\nu_X}(U_j) - \varphi_{\nu_Y}(U_j)| > \eta \right) \underset{n\to\infty}{\longrightarrow} 0. \tag{3.5}$$

Let $A_\eta = \{x : |\varphi_{\nu_X}(x) - \varphi_{\nu_Y}(x)| > 2\eta\}$. If $\mu(A_\eta) = 0,\ \forall \eta > 0$, then $\varphi_{\nu_X} = \varphi_{\nu_Y}$.
If there exists some $\eta > 0$ such that $\mu(A_\eta) > 0$, choose a compact subset $K$ such that $\mu(K \cap A_\eta) > 0$. Now,

$$\begin{aligned} \mathbb{P}^* \left( \max_{\substack{1\le j\le n \\ U_j \in K}} |\varphi_{\nu_X}(U_j) - \varphi_{\nu_Y}(U_j)| > \eta \right) &> \mathbb{P}^* \left( \max_{\substack{1\le j\le n \\ U_j \in K\cap A_\eta}} |\varphi_{\nu_X}(U_j) - \varphi_{\nu_Y}(U_j)| > \eta \right) \\ &= \mathbb{P}^* \left( U_j \in K \cap A_\eta \ \text{ for at least one } j \right). \end{aligned} \tag{3.6}$$

The last identity follows by the definition of the set $A_\eta$. Since $\mu(A_\eta) > 0$, it follows that $\mathbb{P}^* \left( U_j \in K \cap A_\eta \ \text{ for at least one } j \right) \underset{n\to\infty}{\longrightarrow} 1$. This contradicts (3.5), hence $\varphi_{\nu_X} = \varphi_{\nu_Y}$. Now the proof follows from Proposition 2.5 $(i)$ . □

Let $\varepsilon > 0$ be fixed and $K \subset \mathbb{R}^d$ be such that $\mathcal{X}_n \cap K$ and $\mathcal{Y}_n \cap K$ are non empty. Recall that $\mathcal{U}^n$ is a reference sample drawn from the uniform distribution $\mu$ on $B^d$. Let $T^\varepsilon_{\mathcal{X}_n \cap K}$ and $T^\varepsilon_{\mathcal{Y}_n \cap K}$ be two empirical EOT maps (for the subsamples $\mathcal{X}_n \cap K$ and $\mathcal{Y}_n \cap K$,

respectively), and $\varphi^{\varepsilon}_{\mathcal{X}^n\cap K}$ and $\varphi^{\varepsilon}_{\mathcal{Y}^n\cap K}$ be two EOT potentials, as defined in Section 2.2. Then, for $i = 1, \cdots, n$, define the sets,

$$\Delta^{\varepsilon}_i(\mathcal{X}^n \cap K, \mathcal{Y}^n \cap K : \mathcal{U}^n) = \left\{ \left( \left\langle T^{\varepsilon}_{\mathcal{X}_n\cap K}(U_j), e_i \right\rangle, \left\langle T^{\varepsilon}_{\mathcal{Y}_n\cap K}(U_j), e_i \right\rangle \right) : j = 1, \cdots, n \right\} \tag{3.7}$$

and

$$\Gamma^{\varepsilon}(\mathcal{X}^n \cap K, \mathcal{Y}^n \cap K : \mathcal{U}^n) = \left\{ (\varphi^{\varepsilon}_{\mathcal{X}_n\cap K}(U_j), \varphi^{\varepsilon}_{\mathcal{Y}_n\cap K}(U_j)) : \ j = 1, \cdots, n \right\}. \tag{3.8}$$

**Definition 3.3** *Let $K \subset \mathbb{R}^d$ be a compact set. An EOT Q-Q plot is a collection of $d$ individual scatter plots, where the $i$–th one displays the set $\Delta^{\varepsilon}_i(\mathcal{X}^n \cap K, \mathcal{Y}^n \cap K : \mathcal{U}^n)$. Similarly, a scatter plot of the set $\Gamma^{\varepsilon}(\mathcal{X}^n \cap K, \mathcal{Y}^n \cap K : \mathcal{U}^n)$ is called EOT potential plot.*

The next theorem proves similar results as in Theorem 3.2 but for EOT.

**Theorem 3.4** *Let $\nu_X$ and $\nu_Y$ be two probability measures with finite second moment. Let $\mathcal{X}^n = \{X_1, \cdots, X_n\}$ and $\mathcal{Y}_n = \{Y_1, \cdots, Y_n\}$ be two i.i.d. samples drawn from $\nu_X$ and $\nu_Y$, respectively. Let $\mathcal{U}^n = \{U_1, \cdots, U_n\}$, where $U_i$'s are i.i.d. with common distribution $\mu$. Then $\nu_X = \nu_Y$ if and only if one of the following holds.*

*(i)* $\mathbb{P}\left( \liminf_{n\to\infty} \{\Gamma^{\varepsilon}(\mathcal{X}^n \cap K, \mathcal{Y}^n \cap K : \mathcal{U}^n) \subseteq L_\eta\} \right) = 1$, *for all compact $K \subset B^d$ and all $\eta > 0$.*

*(ii)* $\mathbb{P}\left( \liminf_{n\to\infty} \bigcap_{i=1}^{d} \{\Delta^{\varepsilon}_i(\mathcal{X}^n \cap K, \mathcal{Y}^n \cap K : \mathcal{U}^n) \subseteq L_\eta\} \right) = 1$, *for all compact $K \subset B^d$ and all $\eta > 0$.*

**Proof.** $(i)$ First assume that $\nu_X = \nu_Y$. Let $\nu_X|_K$ be a probability measure obtained by restricting $\nu_X$ on the compact set $K$. Similarly, $\nu_Y|_K$ be the restriction $\nu_Y$ on $K$. Then the potentials for the restricted measures are equal, i.e. $\varphi^{\epsilon}_{\nu_X|_K} = \varphi^{\epsilon}_{\nu_Y|_K}$. Since both (random) empirical measures $\nu_{\mathcal{X}_n\cap K}$ and $\nu_{\mathcal{Y}_n\cap K}$ converge (weakly) to the limit $\nu_X|_K (= \nu_Y|_K)$ almost surely, therefore Lemma 1$(iii)$ of Goldfeld et al. [2023] implies that $\|\varphi^{\epsilon}_{\mathcal{X}_n\cap K} - \varphi^{\epsilon}_{\mathcal{Y}_n\cap K}\|_{\mathcal{C}^s}$ converges to 0 almost surely, for any integer $s \geq 1$. Here, $\|\cdot\|_{\mathcal{C}^s}$ defined as the usual Hölder norm with exponent $s$. The proof now follows through a similar argument as employed in Theorem 3.2.

$(ii)$ The proof is similar as for the first part, after observing that the map $\varphi \to \nabla\varphi$ is continuous and linear from $\mathcal{C}^s$ to $\mathcal{C}^{s-1}$. □

# 4 Test Statistics for comparing two distributions

As in Shapiro and Wilk [1965], Dhar et al. [2014], we propose two test statistics which are motivated from our Q-Q plots. Consider the i.i.d random variables $\mathcal{X}^n = \{X_1, \cdots, X_n\}$ and $\mathcal{Y}^n = \{Y_1, \cdots, Y_n\}$ with common distributions $\nu_X$ and $\nu_Y$, respectively, and assume the corresponding EOT maps $T^{\varepsilon}_{\mathcal{X}^n}, T^{\varepsilon}_{\mathcal{Y}^n}$ as defined in Section 2.2. We can measure the total deviation of the EOT Q-Q plots from the straight line $L$ by the quantity $\int_{B^d} \|T^{\varepsilon}_{\mathcal{X}^n}(u) - T^{\varepsilon}_{\mathcal{Y}^n}(u)\|^2 d\mu(u)$. Similarly, the total deviation in the EOT potential plot can be measured by the quantity $\int_{B^d} \|\varphi^{\varepsilon}_{\mathcal{X}^n}(u) - \varphi^{\varepsilon}_{\mathcal{Y}^n}(u)\|^2 d\mu(u)$. Therefore, we define the following two test statistics

$$E_n = n \int_{B^d} \|T^{\varepsilon}_{\mathcal{X}^n}(u) - T^{\varepsilon}_{\mathcal{Y}^n}(u)\|^2 d\mu(u) \quad \text{and} \quad F_n = n \int_{B^d} |\varphi^{\varepsilon}_{\mathcal{X}^n}(u) - \varphi^{\varepsilon}_{\mathcal{Y}^n}(u)|^2 d\mu(u). \tag{4.1}$$

Since the supports of $\nu_X$ and $\nu_Y$ are compact, it follows from [Goldfeld et al., 2023, Lemma 1] that for any $s \in \mathbb{N}$, $\|\varphi^{\varepsilon}_{\mathcal{X}^n}\|_{\mathcal{C}^s(B^d)} \vee \|\varphi^{\varepsilon}_{\mathcal{Y}^n}\|_{\mathcal{C}^s(B^d)} \leq R_s$, $\forall n \geq 1$, where $R_s$ is a constant depending on $\varepsilon$ and the support of the measures $\nu_X$, $\nu_Y$. In other words, the potentials and their derivatives are uniformly bounded on $B^d$. This implies that $E_n$ and $F_n$ are almost surely finite for all $n \geq 1$.

**Proposition 4.1** *Let $\nu_X$ and $\nu_Y$ be two probability measures supported on compact subsets of $\mathbb{R}^d$. Let $\nu_X = \nu_Y$, and $\mathcal{X}^n, \mathcal{Y}^n$ be two independent sets of samples drawn from $\nu_X$ and $\nu_Y$, respectively. Consider $F_n$ and $E_n$ as in (4.1), then we have*

$$F_n \xrightarrow[n\to\infty]{d} F \qquad \text{and} \qquad E_n \xrightarrow[n\to\infty]{d} E,$$

*where $E$ and $F$ are non-negative random variables with finite variance.*

**Proof.** The samples $\mathcal{X}^n$ and $\mathcal{Y}^n$ are independent of each other, hence $\varphi^{\varepsilon}_{\mathcal{X}^n} - \varphi^{\varepsilon}_{\nu_X}$ and $\varphi^{\varepsilon}_{\nu_Y} - \varphi^{\varepsilon}_{\mathcal{Y}^n}$ are also independent. Therefore, it follows from [Goldfeld et al., 2023, Theorem 1] that,

$$\sqrt{n}\left(\varphi^{\varepsilon}_{\mathcal{X}^n} - \varphi^{\varepsilon}_{\nu_X}, \varphi^{\varepsilon}_{\mathcal{Y}^n} - \varphi^{\varepsilon}_{\nu_Y}\right) \xrightarrow[n\to\infty]{d} (Z_X, Z_Y), \tag{4.2}$$

where $Z_X$ and $Z_Y$ are i.i.d mean zero Gaussian random elements, taking values in the space $\mathcal{C}^s(B^d)$. Observe that the map $f \mapsto \int_{B^d} |f(u)|^2 du$ is continuous from $\mathcal{C}^s(B^d)$ to $\mathbb{R}$, for any $s \in \mathbb{N}$. Hence, $F_n \xrightarrow[n\to\infty]{d} F := \int_{B^d} \|(Z_X + Z_Y)(u)\|^2 d\mu(u)$. Therefore, by Fernique's theorem, we have $var(F) < \infty$.
The result $E_n \xrightarrow[n\to\infty]{d} E$ follows by an application of the functional delta method, as proven in [Goldfeld et al., 2023, Corollary 1]. A similar line of argument holds for $F_n$. □

We can then deduce the following theorem from Proposition 4.1.

**Theorem 4.2** *Assume that $\nu_X$ and $\nu_Y$ are two probability measures with compact support. Consider the null hypothesis $H_0 : \nu_X = \nu_Y$ and the alternate hypothesis $H_1 : \nu_X \neq \nu_Y$. Let $E_n, F_n$ and their limits $E, F$ be defined as above, with $E$ and $F$ non-degenerate. Then the following holds:*

*(i) For $0 < \alpha < 1$ and $q_\alpha$ the $\alpha$-th quantile of $E$, we have,*
$\lim_{n\to\infty} \mathbb{P}(E_n > q_\alpha \,|H_0) = 1 - \alpha$ *and* $\lim_{n\to\infty} \mathbb{P}(E_n < q_\alpha \,|H_1) = 0.$

*(ii) For $0 < \alpha < 1$ and $q_\alpha$ the $\alpha$-th quantile of $F$, we have,*
$\lim_{n\to\infty} \mathbb{P}(F_n > q_\alpha \,|H_0) = 1 - \alpha$ *and* $\lim_{n\to\infty} \mathbb{P}(F_n < q_\alpha \,|H_1) = 0.$

**Proof.**

(i) Corollary 1 in Goldfeld et al. [2023] implies that, under the null hypothesis, $E_n$ converges in distribution to $E$. Therefore, we have $\lim_{n\to\infty} \mathbb{P}(E_n > q_\alpha | H_0) = \mathbb{P}(E > q_\alpha | H_0) = 1 - \alpha$.

Moving to the second part of the statement, first note that

$$\begin{aligned}
\frac{1}{n}E_n &= \int_{B^d} \|T^\varepsilon_{\mathcal{X}^n}(u) - T^\varepsilon_{\nu_X}(u) + T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u) + T^\varepsilon_{\nu_Y}(u) - T^\varepsilon_{\mathcal{Y}^n}(u)\|^2 d\mu(u) \\
&\geq \int_{B^d} \Big[ \|T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u)\|^2 + 2\langle T^\varepsilon_{\mathcal{X}^n}(u) - T^\varepsilon_{\nu_X}(u), T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u)\rangle \\
&\quad + 2\langle T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u), T^\varepsilon_{\nu_Y}(u) - T^\varepsilon_{\mathcal{Y}^n}(u)\rangle + 2\langle T^\varepsilon_{\mathcal{X}^n}(u) - T^\varepsilon_{\nu_X}(u), T^\varepsilon_{\nu_Y}(u) - T^\varepsilon_{\mathcal{Y}^n}(u)\rangle \Big] d\mu(u).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\{E_n > q_\alpha\} \supseteq \Bigg\{ &\int_{B^d} \Big[ \sqrt{n}\|T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u)\|^2 + 2\sqrt{n}\langle T^\varepsilon_{\mathcal{X}^n}(u) - T^\varepsilon_{\nu_X}(u), T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u)\rangle \\
&+ 2\sqrt{n}\langle T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u), T^\varepsilon_{\nu_Y}(u) - T^\varepsilon_{\mathcal{Y}^n}(u)\rangle \\
&+ 2\sqrt{n}\langle T^\varepsilon_{\mathcal{X}^n}(u) - T^\varepsilon_{\nu_X}(u), T^\varepsilon_{\nu_Y}(u) - T^\varepsilon_{\mathcal{Y}^n}(u)\rangle \Big] d\mu(u) > \frac{q_\alpha}{\sqrt{n}} \Bigg\}.
\end{aligned}$$

Now, notice that, under $H_1$ ,we have $\sqrt{n}\int \|T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u)\|^2 d\mu(u) \underset{n\to\infty}{\to} +\infty$. Further, it follows from Corollary 1 in Goldfeld et al. [2023] that:

- $\sqrt{n}\int \left\langle T^\varepsilon_{\mathcal{X}^n}(u) - T^\varepsilon_{\nu_X}(u), T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u)\right\rangle d\mu(u)$ and $\sqrt{n}\int \left\langle T^\varepsilon_{\mathcal{Y}^n}(u) - T^\varepsilon_{\nu_Y}(u), T^\varepsilon_{\nu_X}(u) - T^\varepsilon_{\nu_Y}(u)\right\rangle d\mu(u)$ converge in distribution to some random variables with finite variance;
- $\sqrt{n}\int \left\langle T^\varepsilon_{\mathcal{X}^n}(u) - T^\varepsilon_{\nu_X}, T^\varepsilon_{\mathcal{Y}^n}(u) - T^\varepsilon_{\nu_Y}(u)\right\rangle d\mu(u)$ converges to zero almost surely.

Combining all these observations, we conclude that $\mathbb{P}(E_n > q_\alpha \,|H_1) \underset{n\to\infty}{\to} 1$, from which we deduce the second statement.

(ii) The proof of this part follows similarly as in part (i). □

**Remark:**

1. For the OT maps and OT potentials, the asymptotic limit theorems are not known in full generality. Whereas, in case of EOT, the limit theorems are proven in Goldfeld et al. [2023]. This is why we provided the test statistics only through EOT maps and EOT potentials. Nevertheless, once limit theorems known in the OT case, test statistics can be defined for OT maps and potentials in the same way as we did for EOT.

2. Note that the variances of the limiting random variables $E$ and $F$ are finite, but not known explicitly. This is why we evaluate them numerically in the examples given in Section 5.

# 5 Experimental design on simulated data

In the preceding sections, we defined Q-Q plots and potential plots, for both OT and EOT approaches, and established the requisite theory to assess if two given sets of multivariate samples originate from a same distribution. In this section, we test how these theoretical results apply in practice, and check the effectiveness of the visual tools offered by OT and EOT plots on various simulated data. To this aim, we perform four first experiments, considering two samples drawn from distributions such that: (I) they are identical; (II) they have different dependence structures; (III) they are related by a scaling map; (IV) one of them exhibits outliers. Thereafter, we consider the same question, but now focusing on the tail of the distribution. Note that the potential function which gradient pushes forward a regularly varying (RV) probability measure into another regularly varying (RV) probability measure, is also (under some conditions) regularly varying; see [de Valk and Segers, 2018, Theorem 5.1]. Can we clearly distinguish between light and heavy tails with these (OT, EOT) Q-Q and potential plots, like in the case of geometric quantiles (or univariate quantiles)? To answer this question, we proceed to two additional experiments, where we compare multivariate Gaussian distribution with two different heavy-tailed distributions. All scenarios in the following examples are developed considering i.i.d and non i.i.d distributions.
More formally, using our previous notation, for each experiment, we proceed as follows:

(i) Consider two samples, $\mathcal{X}^n$ and $\mathcal{Y}^n$ of size $n$ each and a compact set $K_1$. In all the experiments we choose $K_1$ such that $\mathcal{X}^n \cup \mathcal{Y}^n \subset K_1$;

(ii) Draw another sample $\mathcal{U}^n$ from the uniform distribution on $B^d$, which serves as a reference for comparison. Let $K_2 \subset B^d$ be a compact set. In all the experiments we choose $K_2$ such that $\mathcal{U}^n \subset K_2$;

(iii) Q-Q plots: build the sets $\Delta_i(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K_2)$, $i = 1, \cdots, d$, for the OT Q-Q plot and $\Delta_i^\varepsilon(\mathcal{X}^n \cap K_1, \mathcal{Y}^n \cap K_1 : \mathcal{U}^n)$ for the EOT Q-Q plots;

(iv) Potential plots: build $\Gamma(\mathcal{X}^n, \mathcal{Y}^n : \mathcal{U}^n \cap K_2)$ for the OT potential plot and $\Gamma^\varepsilon(\mathcal{X}^n \cap K_1, \mathcal{Y}^n \cap K_1 : \mathcal{U}^n)$ for the EOT potential plot;

(v) Examine if the Q-Q and potential plots concentrate around the straight line denoted by $L$ with slope 1 and intercept 0;

(vi) If they do, infer from Theorems 3.2 and 3.4 that the samples share the same distribution. Otherwise, look for any discernible pattern present in the scatter plots and whether this pattern may unveil any distinct features. For instance, observe if the points appear to cluster around any specific nonlinear curve (see Figure 9);

(vii) Estimate the values of the test statistics $E_n$ and $F_n$, and estimate the corresponding p-values, to statistically assess the conclusions drawn from the plots. Note that the p-values are meaningful only when comparing distributions supported on compact sets, as Theorem 4.2 is applicable to compactly supported distributions. However, when dealing with fully supported distributions, a partial analysis can be conducted by restricting them to a large enough compact set. Two Borel probability measures are considered identical if and only if their restrictions across all compact sets are identical.

(viii) Finally we also study numerically the role of the regularisation parameter $\varepsilon$ for EOT.

In the following examples, we present two kinds of plots: Q-Q plot and potential plot. In each Q-Q plot example, we display the OT and EOT Q-Q plots together in a single frame: the first row for OT and the second for EOT. In each plot, the $X$-axis corresponds to the first sample $\mathcal{X}^n$ and the $Y$-axis to the second sample $\mathcal{Y}^n$. Also, the potential plots are presented together in a single frame: the left for OT and the right for EOT. Please note that all computations are performed using the POT (Python Optimal Transport) package; for detailed documentation, we refer to Flamary et al. [2021].

## 5.1 Example: Comparison between two samples drawn from a multivariate (non i.i.d.) Gaussian distribution

Starting with step (i), we consider two sets of i.i.d. samples, $\mathcal{X}^n$ and $\mathcal{Y}^n$, drawn from distributions $\nu_X$ and $\nu_Y$, respectively, where $\nu_X = \nu_Y$ is a trivariate normal distribution with mean zero and covariance matrix $(\sigma_{ij})$, where, for instance, $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$, $\sigma_{12} = 0.5$, $\sigma_{13} = 0.2$ and $\sigma_{23} = 0$. The number of observations in each sample is 1000. Following steps (ii) to (iii), we obtain the OT and EOT Q-Q plots for these two samples, as displayed in Figure 1.

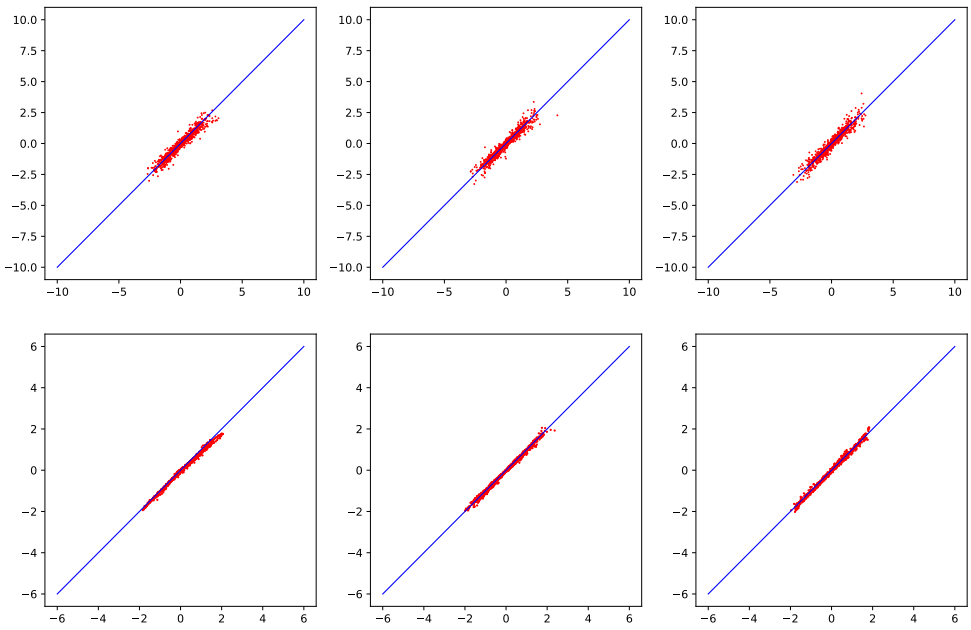


Figure 1: Q-Q plots for two samples, drawn from a trivariate normal distribution with mean zero and covariance matrix $(\sigma_{ij})$, where $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$, $\sigma_{12} = 0.5$, $\sigma_{13} = 0.2$, $\sigma_{23} = 0$. First row: OT Q-Q plot; second row: EOT Q-Q plot, with $\varepsilon = 10^{-2}$. The straight (blue) lines represent the line $L$.

Next, we provide the potential plots for this example following the step (iv); see Figure 2. Note that for EOT, we choose $\varepsilon = 10^{-2}$.

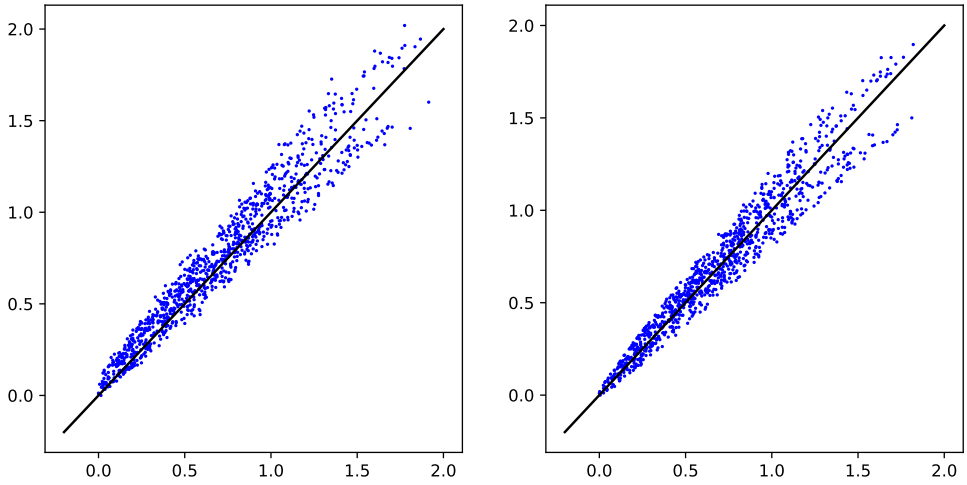


Figure 2: Potential plots for two samples, drawn from a trivariate normal distribution with mean zero and covariance matrix $(\sigma_{ij})$, where $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$, $\sigma_{12} = 0.5$, $\sigma_{13} = 0.2$, $\sigma_{23} = 0$. The left plot compares the OT potentials, the right one EOT potentials taking $\varepsilon = 10^{-2}$. The straight (black) lines represent the line $L$.

As expected, we observe that, for both types of comparisons, Q-Q and potential ones, the scatter plot is concentrated along the line $L$. Therefore, according to step (vi), we can infer that the two samples are drawn from the same distribution. Although the spread around the straight line $L$ appears larger in the case of potential plots than in Q-Q plots, notice that the two sets of plots are shown on different scales, and that the dispersion in absolute terms is comparable in both cases.
A relatively larger spread is observed in the extremal regions of the potential plots, which may be ascribed to poorer approximation of the potential function in the extreme region when having fewer observations.
Compared to the OT plots, we also observe sharper clustering around the straight line in case of EOT. A possible reason for this could be that EOT maps and potentials are more regular than their OT counterparts (Nutz [2022]).
Next, we perform statistical test checking the similarity of the underlying distributions of the two samples $\mathcal{X}^n$ and $\mathcal{Y}^n$. We set the null hypothesis as $H_0 : \nu_X|_{K_2} = \nu_Y|_{K_2}$, and empirically estimate the values of the test statistics $E_n$ and $F_n$ defined in (4.1). Additionally, the limiting distributions of $E_n$ and $F_n$ are empirically estimated, from which we deduce the $p$-values under $H_0$. We will proceed in the same way in all the illustrations. The values of the test statistics and $p$-values are reported in Table 1, varying the sample size $n$. Since the $p$-values are relatively high, whatever the sample size, it supports the assertion that the two samples have originated from the same distribution.

| $n$ | $E_n$ | $p$-value of $E_n$ | $F_n$ | $p$-value of $F_n$ |
|---|---|---|---|---|
| 250 | 45.19 | 0.1137 | 14.19 | 0.1559 |
| 500 | 32.17 | 0.6578 | 6.84 | 0.6723 |
| 1000 | 33.27 | 0.5941 | 6.84 | 0.6723 |

Table 1: Test statistics and $p$-values under the null hypothesis $H_0 : \nu_X|_{K_2} = \nu_Y|_{K_2}$.

## 5.2 Example: Characterising dependency

In multivariate setting, assessing dependency among marginals is an essential feature of multivariate statistical analysis. In this experiment, we compare two samples, one drawn from a distribution with independent marginals against one drawn from a distribution having dependent marginals, to see if there is any specific feature that can be observed from the Q-Q and potential plots. We consider two samples; the first one is drawn from $(\nu_X)$, a trivariate standard normal distribution, and the second from $(\nu_Y)$, a normal distribution with zero mean and covariance matrix $(\sigma_{ij})$, where $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$, $\sigma_{12} = 0.9$, $\sigma_{13} = 0$ and $\sigma_{23} = 0$. The Q-Q and potential plots are displayed in Figures 3 and 4, respectively.
In both Q-Q plots (OT and EOT), we observe that, although the points are concentrated along the line $L$, the concentration is stronger for the third component (especially for EOT), as compared to the plots for the first and second component. Since the second distribution exhibits a high correlation (0.9) between the first and second marginals, would this

pattern indicate some dependence? Answering this question will require further research and experiments, considering various dependence structures.

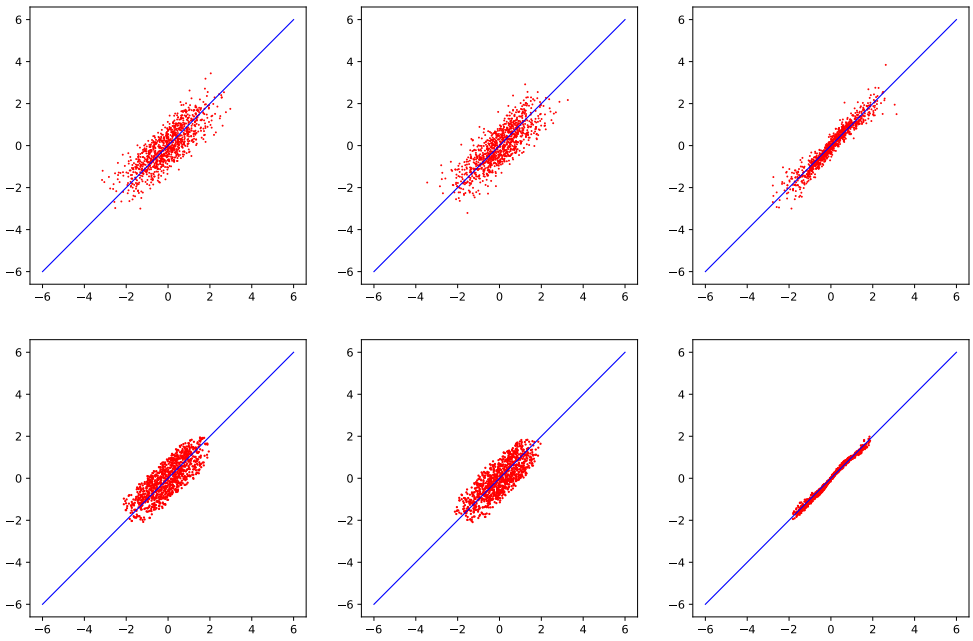


Figure 3: Q-Q plots for two samples, each of size 1000; the first sample is drawn from a trivariate standard normal distribution, the second from a normal distribution with zero mean and covariance matrix $(\sigma_{ij})$, where $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$, $\sigma_{12} = 0.9$, $\sigma_{13} = 0$, $\sigma_{23} = 0$. The first row displays OT Q-Q plots, the second EOT Q-Q plots with $\varepsilon = 10^{-2}$. The straight (blue) lines represent the line $L$.

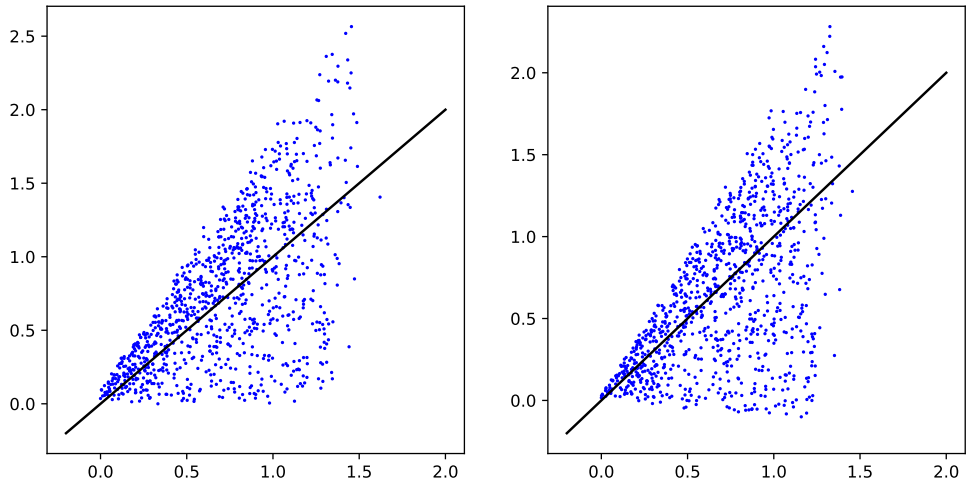



Figure 4: Potential plots for 2 multivariate samples, each of size 1000. The first sample is drawn from a trivariate standard normal distribution, the second from a normal distribution with mean 0 and covariance matrix $(\sigma_{ij})$, where $\sigma_{11} = \sigma_{22} = \sigma_{33} = 1$, $\sigma_{12} = 0.9$, $\sigma_{13} = 0$, $\sigma_{23} = 0$. Left: OT potential plot; Right: EOT potential plot with $\varepsilon = 10^{-2}$. The straight (black) lines represent the line $L$.

Turning to potential plots displayed in Figure 4, we can assess that the samples considered for the experiment have originated from different distributions. However, we cannot deduce any specific pattern in the plots that would suggest a high dependence among some components. Here too, further research needs to be developed.

Finally, we empirically estimate the values of the test statistics $E_n$ and $F_n$ defined in (4.1), as well as the corresponding $p$-values under $H_0 : \nu_X|_{K_2} = \nu_Y|_{K_2}$. The values are reported in Table 2. The small $p$-values lead to the statement that the two samples have originated from two different distributions.

| $n$ | $E_n$ | $p$-value of $E_n$ | $F_n$ | $p$-value of $F_n$ |
|---|---|---|---|---|
| 250 | 277.86 | 0 | 62.46 | 0 |
| 500 | 469.56 | 0 | 166.46 | 0 |
| 1000 | 877.14 | 0 | 304.53 | 0 |

Table 2: Test statistics and $p$-values under the null hypothesis $H_0 : \nu_X|_{K_2} = \nu_Y|_{K_2}$.

Observe that the values of $E_n$ and $F_n$ are getting larger as the sample size $n$ increases. This may indicate that the two underlying distributions are not the same, since we know that $E_n$ and $F_n$ tend to infinity in such a case. Also note that the $p$-values are exactly zero, which looks too good to be true. Although the limiting distributions of $E_n$ and $F_n$ are fully supported on the positive real line, our approximations of those are supported on bounded sets. Therefore, the large values of $E_n$ and $F_n$, which correspond to small probability regions of the limiting distributions, easily fall beyond the support of the approximated limiting distributions. To obtain non-zero $p$-values, either the approximations have to be

more accurate so that their supports are large enough to contain $E_n$ and $F_n$, or the limiting distributions have to be known precisely.

## 5.3 Example: Gaussian versus scaled Gaussian

We draw samples $\mathcal{X}^n$ from $\mu_X$, a standard normal distribution and $\mathcal{Y}^n$ from $\mu_Y$, a centered normal distribution with covariance matrix $\mathrm{diag}(1, 4, 1)$. The Q-Q plots are displayed in Figure 5. We observe that in the first and third columns in both rows, the points are concentrated around the line with slope 1 (blue line) whereas in the second column, the points are concentrated around the line with slope 2 (black line). This observation suggests that $T_{\mathcal{Y}n} = \mathrm{diag}(1, 2, 1)T_{\mathcal{X}_n}$, indicating that the second distribution differs from the first by a scaling factor represented by the matrix $\mathrm{diag}(1, 2, 1)$. Since $\epsilon$ is set very small ($10^{-2}$), the entropy regularised quantiles $T^{\epsilon}_{\mathcal{X}^n}$ closely approximate $T_{\mathcal{X}^n}$, and as a result, we see that in the second plot of the second row, the points are also approximately concentrated around the straight line with slope 2 (black line). However, it is important to recall that the EOT map is not necessarily scale equivariant in general.

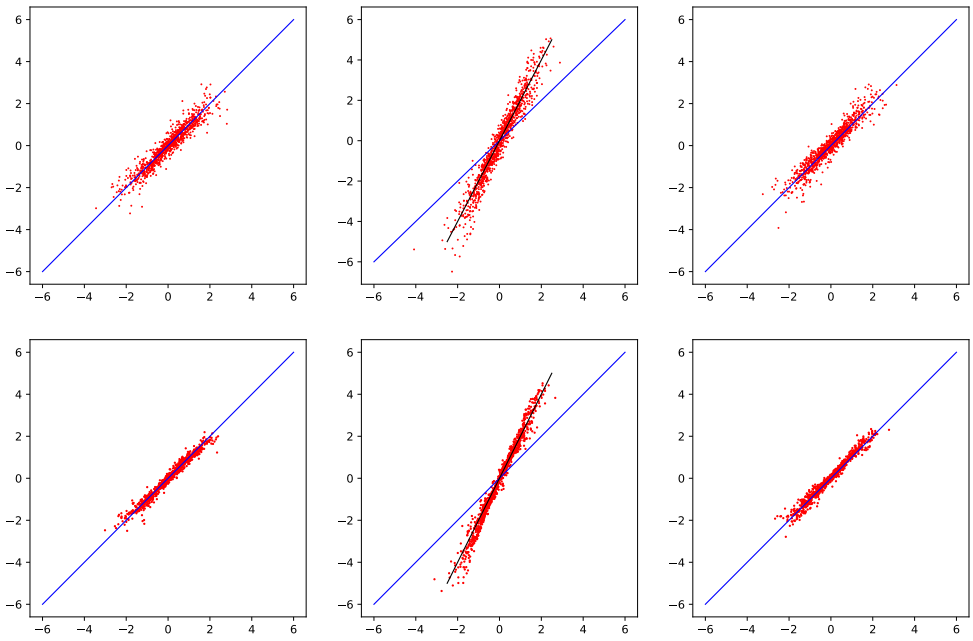


Figure 5: Q-Q plots for two samples, the first one is drawn from a trivariate standard normal distribution, the second one from a trivariate normal distribution with covariance matrix $= diag(1, 4, 1)$. The first row displays OT Q-Q plot and the second one EOT Q-Q plot with $\varepsilon = 10^{-2}$. The straight blue lines represent the line $L$ and the black line in the second column has slope 2 and intercept 0.

We present potential plots (left for OT and right for EOT) for the same pair of samples $\mathcal{X}^n$ and $\mathcal{Y}^n$ in Figure 6. Clearly, the scatter plots are spread out and not concentrated along the straight line $L$. This observation suggests that the underlying distributions corresponding

to the two samples are not identical. But, unlike the Q-Q plots (shown in Figure 5), the potential plots do not offer additional information about the underlying distributions, such as the scaling in the sample $\mathcal{Y}^n$.

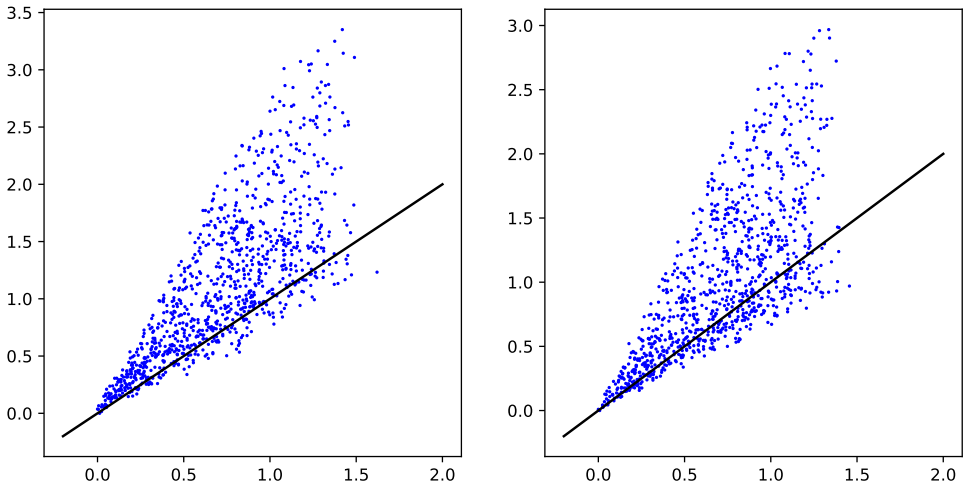



Figure 6: Potential plots for two samples. The first sample is drawn from a trivariate standard normal distribution and the second one from a trivariate normal distribution with mean zero and covariance matrix $= diag(1, 4, 1)$. The left plot is the OT potential plot and the right one EOT potential plot with $\varepsilon = 10^{-2}$. The straight (black) lines represent the line $L$.

Finally, we empirically estimate the values of the test statistics $E_n$ and $F_n$ defined in (4.1), as well as the corresponding $p$-values under $H_0 : \nu_X|_{K_2} = \nu_Y|_{K_2}$. The values are reported in Table 3. Since the $p$-values are relatively small, this supports the assertion that the two samples have originated from distinct distributions.

| $n$ | $E_n$ | $p$-value of $E_n$ | $F_n$ | $p$-value of $F_n$ |
|---|---|---|---|---|
| 250 | 222.95 | 0 | 63.58 | 0 |
| 500 | 394.71 | 0 | 121.40 | 0 |
| 1000 | 796.47 | 0 | 267.15 | 0 |

Table 3: Test statistics and $p$-values under the null hypothesis $H_0 : \nu_X|_{K_2} = \nu_Y|_{K_2}$.

## 5.4 Example: Outlier Detection

We show that outliers can be detected using OT and EOT Q-Q plots. We illustrate this with two simulated samples, $\mathcal{X}^n$ and $\mathcal{Y}^n$, drawn from trivariate standard normal distributions, each containing 1000 observations. Subsequently, we replace three observations in $\mathcal{Y}^n$ with outlier points $(8, 8, 8)$, $(9, 9, 9)$, and $(10, 10, 10)$, resulting in the transformed set $\bar{\mathcal{Y}}^n$. We follow steps (ii)–(iii) to build the Q-Q plots for $\mathcal{X}^n$ and $\bar{\mathcal{Y}}^n$, and display them in

Figure 7. The three black points, which are far from the rest of the observations, clearly reveal the presence of outliers. Next, we present the potential plots (step (iv)) for these two samples in Figure 8.
Notice that, although there are 3 outliers present in the second sample, we only see one outlying point in the potential plot (compare this with Q-Q plots in Figure 7). It appears that potential plots may be less informative in this case compared to the Q-Q plots.
It is worth noticing that we set the value of the regularisation parameter as $\varepsilon = 10^{-3}$ in this example, whereas in previous examples, we chose a larger value of $\varepsilon = 10^{-2}$. We observe that, with a higher value of $\varepsilon$, the EOT Q-Q plots do not distinctly separate the outliers. We shall study empirically the role of the regularisation parameter in visual analysis in Subsection 5.6.

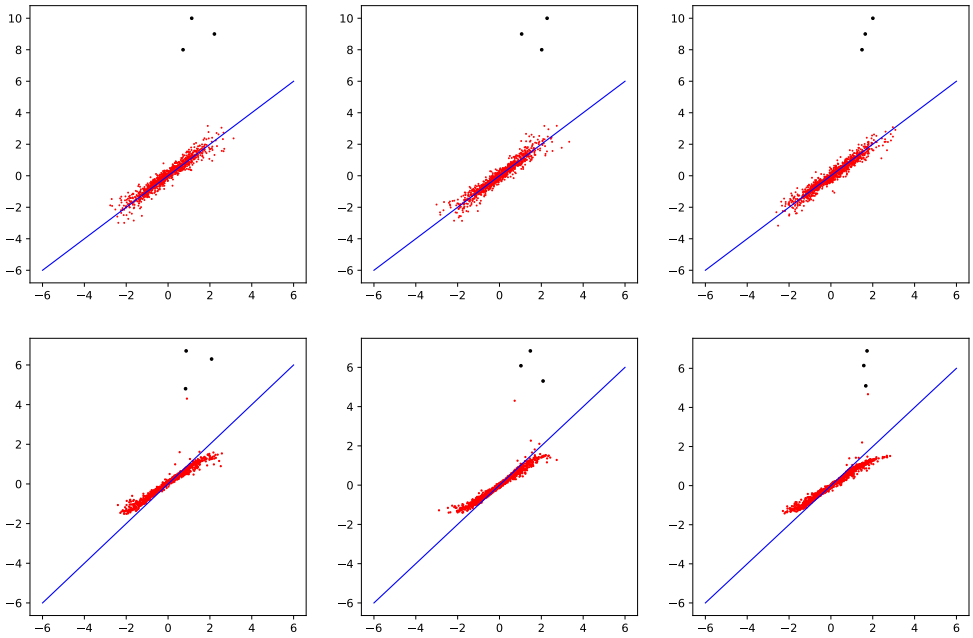


Figure 7: Q-Q plots for two multivariate samples, each of size 1000. Both samples are generated from a multivariate standard normal, while three observations in the second sample are replaced by three outliers. First and second rows display the OT and EOT Q-Q plots, respectively. The regularisation parameter for the EOT is chosen as $\varepsilon = 10^{-3}$. The straight (blue) lines represent the line $L$.

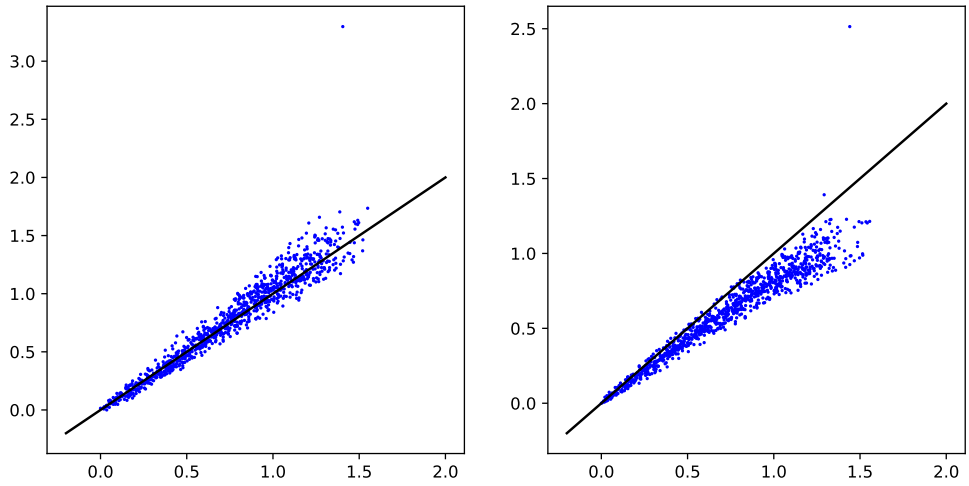



Figure 8: Potential plots (left for OT, right for EOT with $\varepsilon = 10^{-3}$) for two multivariate samples, each of size 1000. Both samples are generated from multivariate standard normal, while only the second sample is contaminated by some outliers. The straight (black) lines represent the line $L$.

## 5.5 Example: Light tail vs Heavy tail

To compare the heaviness of tail distributions, we provide a first example in dimension 3 considering two distributions: the standard normal distribution (for the light tail) and the Student's $t$–distribution (for the heavy one). Another example comparing the trivariate standard normal distribution with i.i.d. Pareto(3) marginals can be found in Appendix A. Recall that the density of the Student's $t$–distribution in dimension $d$ is given by

$$f(x) = \frac{\Gamma[(r+d)/2]}{\Gamma(r/2) r^{d/2} \pi^{d/2} |\Sigma|^{1/2}} \left[1 + \frac{1}{r}(x-m)^T \Sigma^{-1} (x-m)\right]^{-(r+d)/2}, \tag{5.1}$$

where $m \in \mathbb{R}^d$, $\Sigma$ is a $d \times d$ symmetric positive semidefinite matrix and $\Gamma(\cdot)$ is the gamma function. The parameter $r > 0$ determines the heaviness of the distribution in the sense that moments of order greater than $r$ are infinite. We follow step (i) and draw a sample $\mathcal{X}^n$ of size 1000 from the trivariate standard normal distribution ($\nu_X$). Similarly we draw $\mathcal{Y}^n$ of the same size from a trivariate Student's $t$–distribution, with parameters $m = 0$, $\Sigma = I_3$ (identity in $\mathbb{R}^3$) and $r = 3.2$ ($\nu_Y$). We go through steps (ii)–(iii), and display the Q-Q plots in Figure 9.

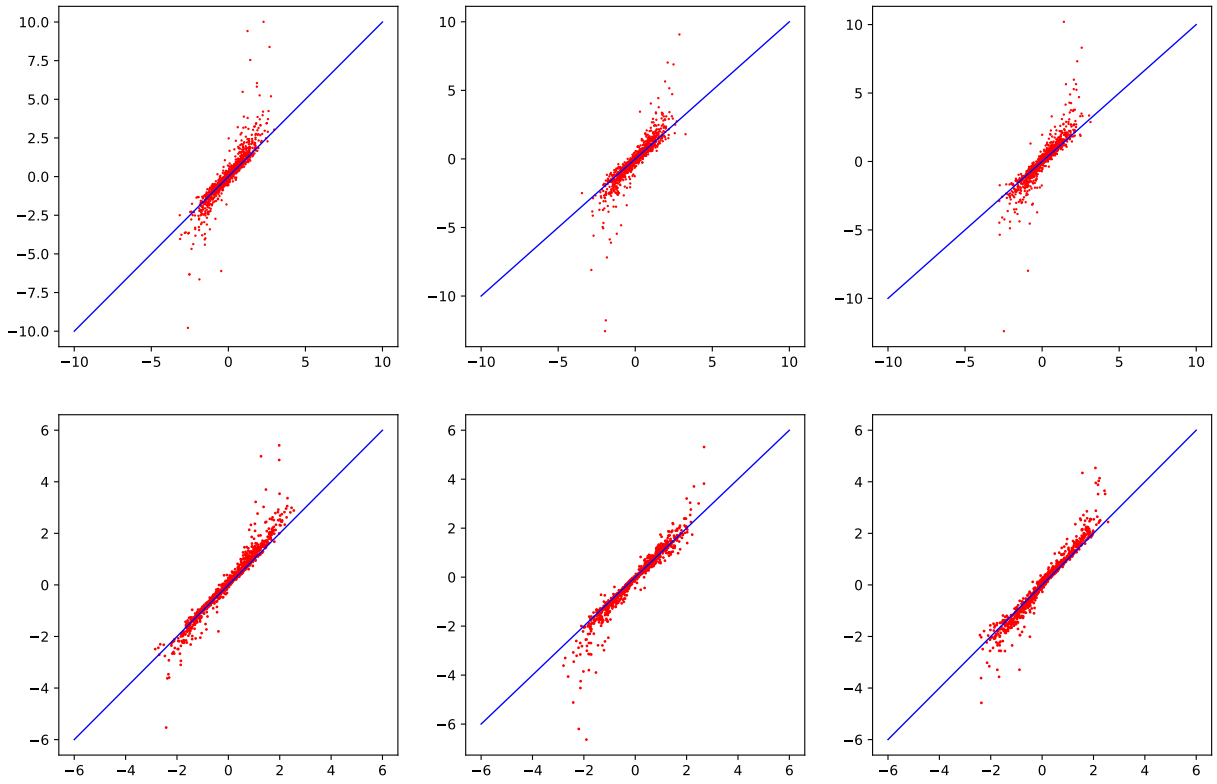


Figure 9: Q-Q plots for two samples each of size 1000, the first drawn from a trivariate standard normal distribution, the second from a multivariate Student's t-distribution. The first row display OT Q-Q plot and second EOT Q-Q plot, taking $\varepsilon = 10^{-3}$. The straight (blue) lines represent the line $L$.

Here, we observe that the OT and EOT quantiles for $\mathcal{Y}^n$ grow at a faster rate than those of $\mathcal{X}^n$. The peculiar shape (S-shape) of the scatter plot seen in Figure 9 is very familiar in the univariate Q-Q plots involving heavy vs light tail comparisons. This hints at a similar behaviour in the OT plots. This feature, which is very useful in the univariate analysis, needs further exploration in the (E)OT setting.

| $n$ | $E_n$ | $p$-value of $E_n$ | $F_n$ | $p$-value of $F_n$ |
|---|---|---|---|---|
| 250 | 670.64 | 0 | 13.30 | 0.33325 |
| 500 | 1410.92 | 0 | 33.52 | 0.003375 |
| 1000 | 2771.30 | 0 | 92.38 | 0.000125 |

Table 4: Test statistics and $p$-values under the null hypothesis $H_0 : \nu_X|_{K_2} = \nu_Y|_{K_2}$.

We then present potential plots (step (iv)) for this particular example in Figure 10. It can be seen from the plots that the points are scattered around a nonlinear (increasing) curve above the line $L$. This implies that the potential function associated with the second sample has a higher growth rate. Since the (E)OT maps are gradient of (E)OT potentials, a higher growth rate of potential implies that the corresponding distribution has a heavier

tail than that of first one.

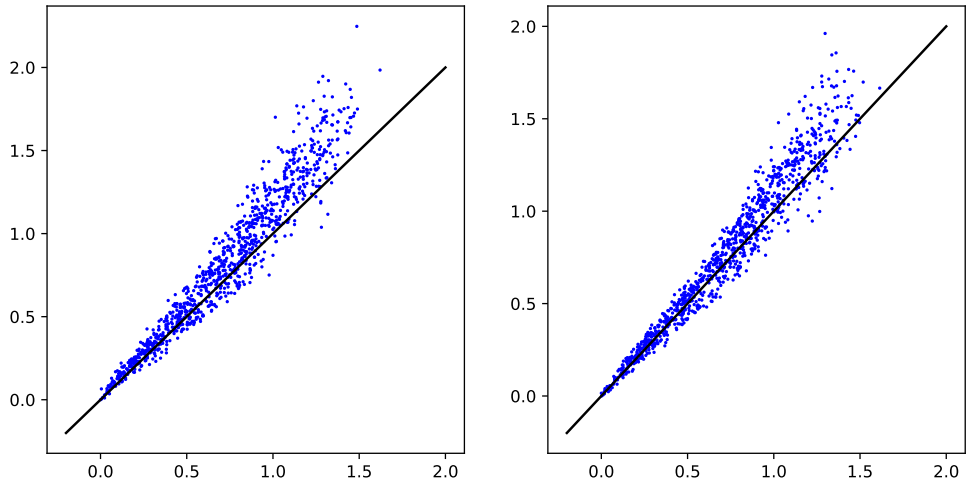



Figure 10: Potential plots (left for OT, right for EOT with $\varepsilon = 10^{-3}$) for two multivariate samples, each of size 1000; the first one is drawn from a trivariate standard normal distribution and the second from a multivariate Student's t-distribution. The straight (black) lines represent the line $L$.

The $p$-values corresponding to $E_n$ and $F_n$ are provided in Table 4, although not so informative given the question on tail behaviour.

## 5.6 The effect of the regularisation parameter in EOT

As stated in Proposition 2.16, both EOT map and potential uniquely characterise distributions, regardless of the value of the regularisation parameter $\varepsilon$. However, the regularisation parameter plays an important role in the visual analysis. For a reasonably small value of $\varepsilon$ (depending upon the distribution), the EOT map and potential closely approximate the OT map and potential, respectively. As a consequence, the EOT plots (Q-Q and potentials) look very similar to the OT plots (Q-Q and potentials); see for e.g. Figures 1–6.
On the other hand, the EOT map (resp. potential), with a large value of $\varepsilon$, is far from being a close approximation of the OT map (resp. potential).
Taking back the setup of Example 5.1, we compute and display in Figures 11 and 13, respectively, the EOT Q-Q plots and potential plots when varying $\varepsilon$. We consider three values of $\varepsilon$: $10^{-3}, 10^{-2}$ and $10^{-1}$, respectively. We observe that points are not only scattered around the line $L$, they also begin to concentrate around a point on the line as the value of $\varepsilon$ increases.

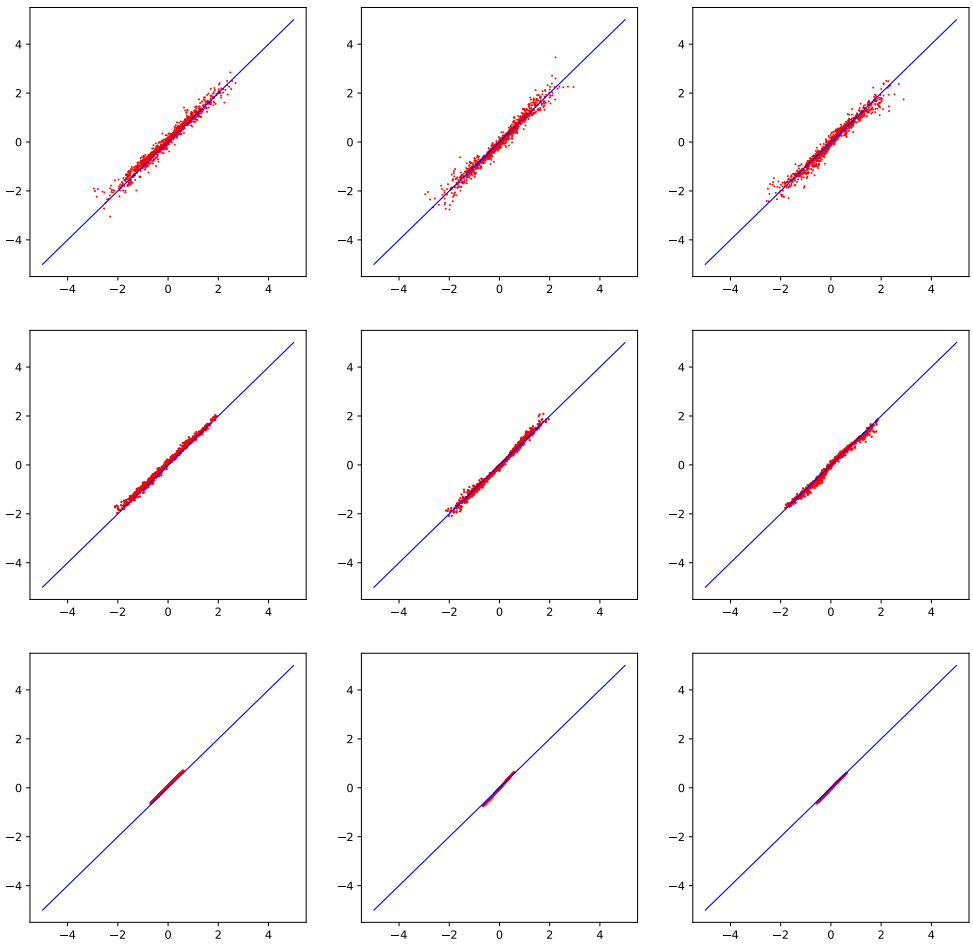


Figure 11: EOT Q-Q plots for two samples, both generated from a trivariate standard normal distribution. The value of the regularisation parameter ($\varepsilon$) is $10^{-3}$ for the first row, $10^{-2}$ for the second row and $10^{-1}$ for the third row. The straight (blue) lines represent $L$.

Recall that, as $\varepsilon$ increases to $+\infty$, the EOT map converges to a constant map, with the constant being the mean of the target distribution. Therefore, for larger $\varepsilon$ (e.g., in this case $\varepsilon = 10^{-1}$), the EOT map is close to a constant, resulting in concentration around a point in the Q-Q plot. We next consider the same samples as in Example 5.5. We display the EOT Q-Q plots in Figure 12, for three different values of $\varepsilon$, $10^{-3}, 10^{-2}$ and $10^{-1}$, respectively. Observe that, although the points are deviating from the line $L$, which indicates that the two samples are non similar, they do not reveal the tail heaviness (compared with Example 5.1).

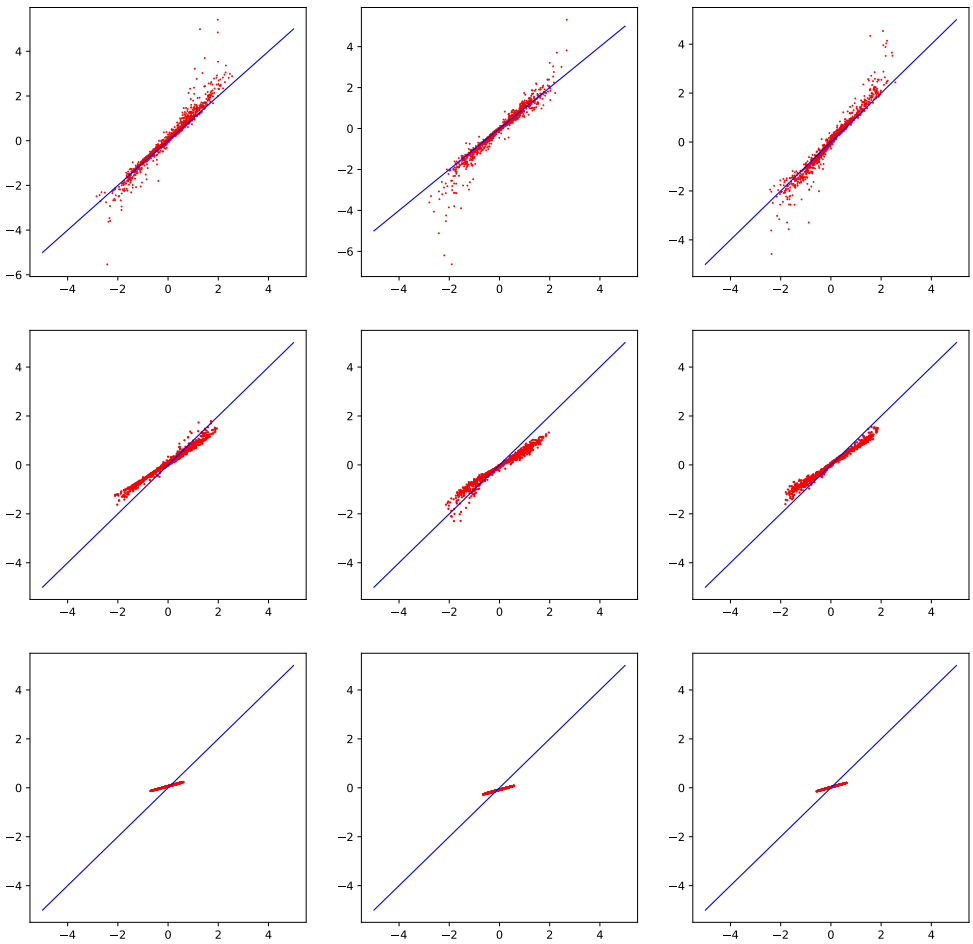


Figure 12: EOT Q-Q plots for two samples. The samples are chosen as in Example 5.5. The value of regularisation parameter ($\varepsilon$) is $10^{-3}$ for the first row, $10^{-2}$ for the second row and $10^{-1}$ for the third row. The blue lines represent $L$.

Therefore, for a large value of $\varepsilon$, while the EOT Q-Q plot can distinguish samples from different distributions, it may not reveal specific details such as the heaviness of the tails.

Let us now discuss the EOT potential plots for different values of $\varepsilon$. We first consider the setup in Example 5.1, let both samples are drawn from a trivariate standard normal distribution. We display the EOT potential plots in Figure 13, for three different values of $\varepsilon$, $10^{-3}, 10^{-2}$ and $10^{-1}$, respectively. Observe that, for large values of $\varepsilon$, the plots are not convincing enough to infer that the two sets of samples are generated from the same distribution.

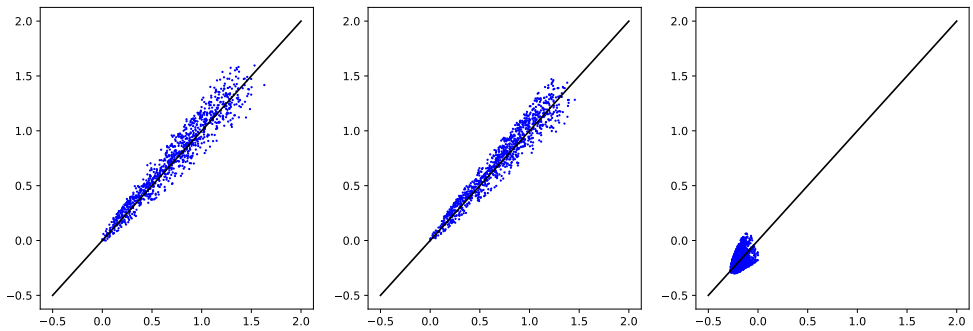


Figure 13: EOT potential plots for two samples, both generated from a trivariate standard normal distribution. The value of the regularisation parameter ($\varepsilon$) for the left plot is $10^{-3}$, middle $10^{-2}$ and right $10^{-1}$. The blue lines represent $L$.

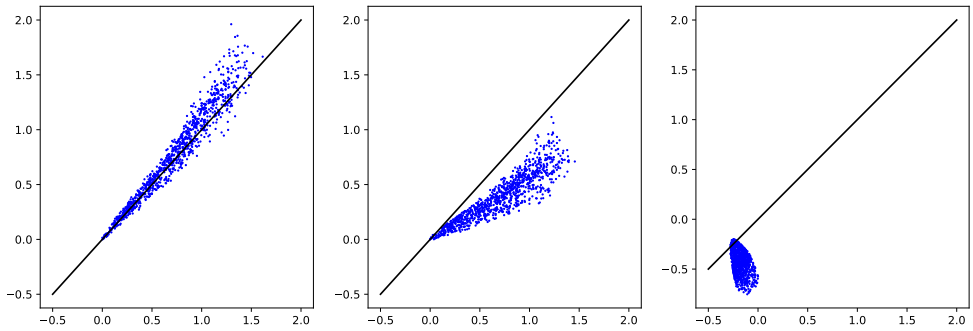


Figure 14: EOT potential plots for two samples. We consider the same sample as in part ($b$) of Example 5.5. The value of the regularisation parameter ($\varepsilon$) is $10^{-3}$ for the left plot, $10^{-2}$ for the middle one and $10^{-1}$ for the right one. The blue lines represent $L$.

We show another example of potential plots in Figure 14. The samples are the same as in Example 5.5. Observe that, although the plots for each value of $\varepsilon$ suggest that the samples are not very similar, it is difficult to infer visually (especially when $\varepsilon$ is large) that one of the samples has a heavier tail than the other (Unless the EOT plot is very close to the OT one as in Figure 10).
It seems therefore ideal to choose small values of $\varepsilon$ for better visual analysis. Nevertheless, it is important to note that smaller the value of $\varepsilon$, longer the algorithm takes to converge. Moreover, if $\varepsilon$ is chosen too small, the algorithm (Sinkhorn) need more iteration to converge.

## 5.7 Key takeaways

(i) We observe that both (OT and EOT) Q-Q plots and potential plots can efficiently identify the similarity between distributions. We experiment this with i.i.d and non i.i.d Gaussian distributions.

(ii) When the samples are generated from a same distribution (in our example, a Gaussian distribution) with one of them having a scaling or shifting, we can identify this effect with Q-Q plots. Potential plots do not identify the shift, but the scale.

(iii) We have shown with some examples that outliers can be detected with OT Q-Q plots and potential plots. For EOT, one needs to choose the value of the regularisation parameter to be small enough to observe them.

(iv) Considering two examples, we showed that both Q-Q and potential plots can be used as a visual tool to compare tail distributions. For instance, comparing a light tail sample (e.g. Gaussian sample) with a heavy tail one (e.g. multivariate Student's t-distribution), we could identify on the plots the heaviest tail between the two distributions.

(v) When comparing distributions in high dimension, potential plots give a 2-dimensional representation, while a large number of componentwise plots is needed in the case of OT (and also geometric) Q-Q plots. As such, potential plots can be very useful in applied fields, e.g. risk management, as they offer:

- A discriminating tool between light and heavy tail,
- A visual validation tool for multivariate modelling.

Given these characteristics, to compare visually two multivariate distributions, we would recommend to proceed as follows:

1. Plot the potential function to detect if the two samples are drawn from a same multivariate distribution;

2. To obtain the whole comparison with specific features, consider the (E)OT Q-Q plots.

As a last remark, if comparing two samples which one suspects to come from a same family of distributions, then one may standardize the data to compute the (E)OT potential and Q-Q plots.

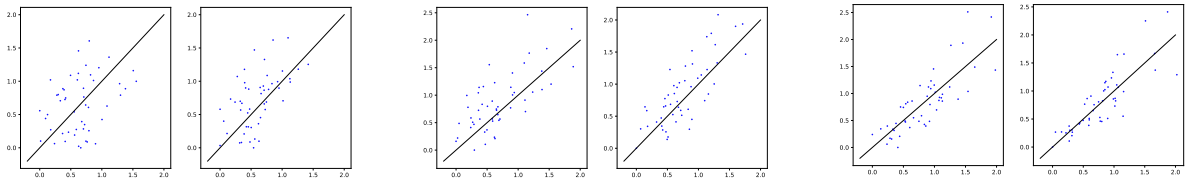
# 6 Application on real data

In the previous section, we considered many scenarios and experiments with simulated samples to have a better understanding of (E)OT Q-Q and potential plots as a visual tool to compare multivariate distributions. Now, we can turn to applications on real data. We consider two examples.
The first example is the Fisher's Iris dataset, which can be downloaded from https://archive.ics.uci.edu, a standard dataset used in statistics. It was also considered in Dhar et al. [2014] for analysing multivariate Q-Q plots based on geometric quantile, giving us a way to compare the results obtained when choosing two types of multivariate quantiles

(see Section 7). The dataset has 4 variables, and 50 observations for each of the 4 variables. Due to the relatively smaller size of this dataset, we consider another example offering a larger sample size. This second example is the Turkish rice Osmanic dataset, downloaded also from the link https://archive.ics.uci.edu. It has 5 variables, and 2180 observations for each of the 5 variables.

## 6.1 Example 1: Fisher's Iris data

This Fisher's Iris dataset is constituted of three multivariate samples corresponding to three varieties: Iris Setosa, Iris Versicolour and Iris Virginica. Each variety consists of 50 observations and each observation contains the measurement of 4 variables, namely: sepal length, sepal width, petal length and petal width. Therefore, each sample is 4-dimensional with a size of 50. Following the steps (ii) to (iv) described in Section 5 to build the (E)OT potential and Q-Q plots, we compare each of the three samples with a 4-dimensional Gaussian sample of size 50. This choice of comparison is also motivated by the fact that Dhar et al. [2014] showed a strong evidence of normality of those samples using the geometric Q-Q plots. For a fair comparison, we follow the procedure of standardising the dataset, like in Dhar et al. [2014], before computing various quantiles and potentials. With the (E)OT approach, we obtain the potential plots displayed in Figure 15 for the three varieties of iris, and the Q-Q plots in Figure 16, for the Iris Setosa, Iris Versicolour, and Iris Virginica, respectively. Note that we also computed the associated test statistics and $p$-values but choose not to report them here, as they heavily rely on the asymptotics of the test statistics, which is not compatible with such a small sample size. Looking at the potential plots, we observe that the points are very much spread out, which would hint at non–Gaussianity of the samples (Iris Setosa, Iris Versicolour and Iris Virginica). However, as already pointed out, the sample size is very small for any reasonable inference.



Iris Setosa data | Iris Versicolour data | Iris Virginica data

Figure 15: Potential plots for two multivariate samples, each of size 50; the first one is drawn from a 4-dimensional standard normal distribution and the second one is the standardised Iris Setosa data for the first (left) pair of plots, the standardised Iris Versicolour data for the second (middle) pair of plots, and the standardised Iris Virginica data for the third (right) pair of plots. For each pair, the left plot is OT potential plot and the right one is EOT potential, choosing $\varepsilon = 10^{-3}$. The straight (black) lines represent the line $L$.

We now look at the (E)OT Q-Q plots, to compare the data with a multivariate standard Gaussian distribution. In Figure 16, for each of the three samples, the points are loosely concentrated around the straight line $L$ in the OT Q-Q plots, while in EOT Q-Q plots, they appear a bit more concentrated around the line $L$.

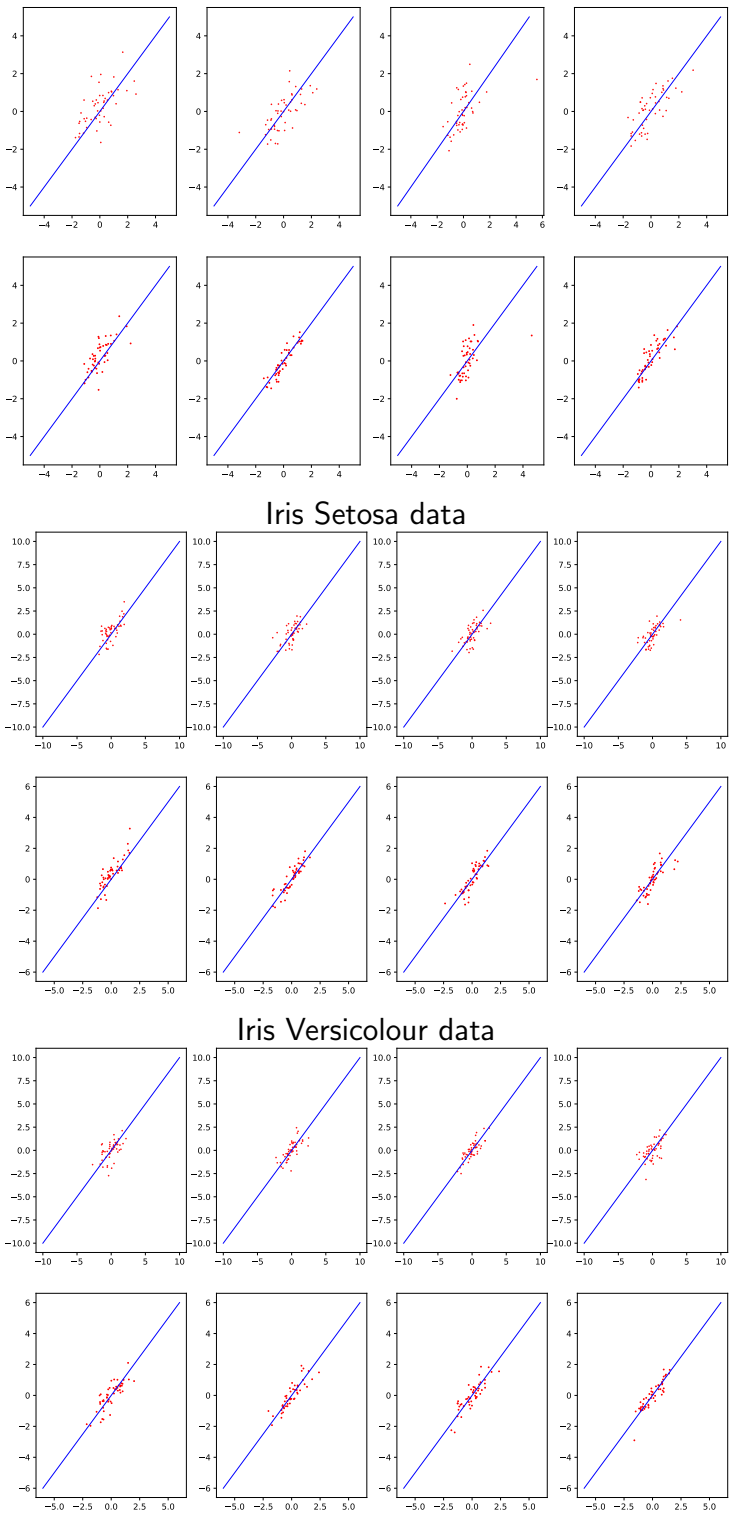



Figure 16: Q-Q plots for two samples, each of size $50$, the first one is generated from a $4$-dimensional standard Gaussian and the second one is, respectively, the standardised Iris Setosa, standardised Iris Versicolour and standardised Iris Virginica data. For each Iris variety considered, the first row displays OT Q-Q plots and the second one EOT Q-Q plots with $\varepsilon = 10^{-3}$. The straight (blue) lines represent the line $L$.

## 6.2 Example 2: Turkish rice Osmanic data

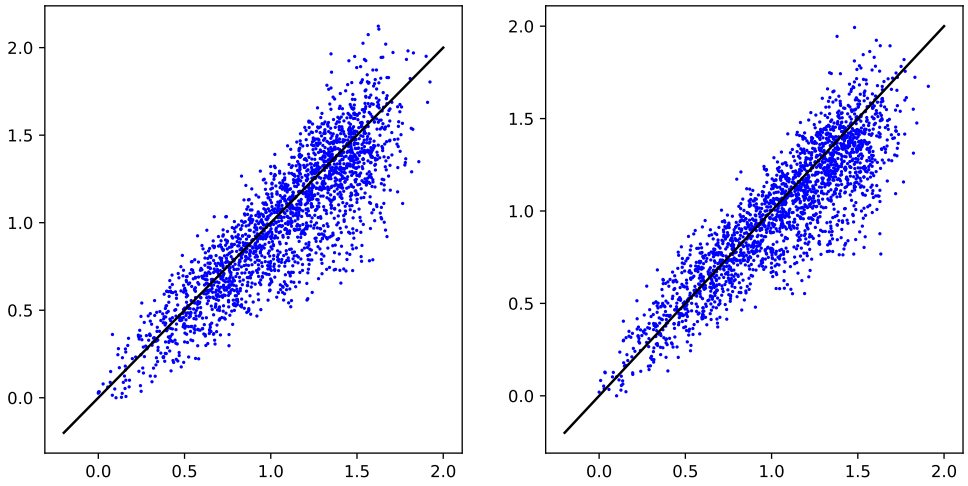


Figure 17: Potential plots for two multivariate samples, each of size $2180$; the first one is drawn from a $5$-dimensional standard Gaussian distribution, and the second one is the standardised Turkish rice Osmanic data. The left plot is OT potential plot and the right one EOT potential plot. For EOT, we take $\varepsilon = 0.5 \times 10^{-2}$. The straight (black) lines represent the line $L$.

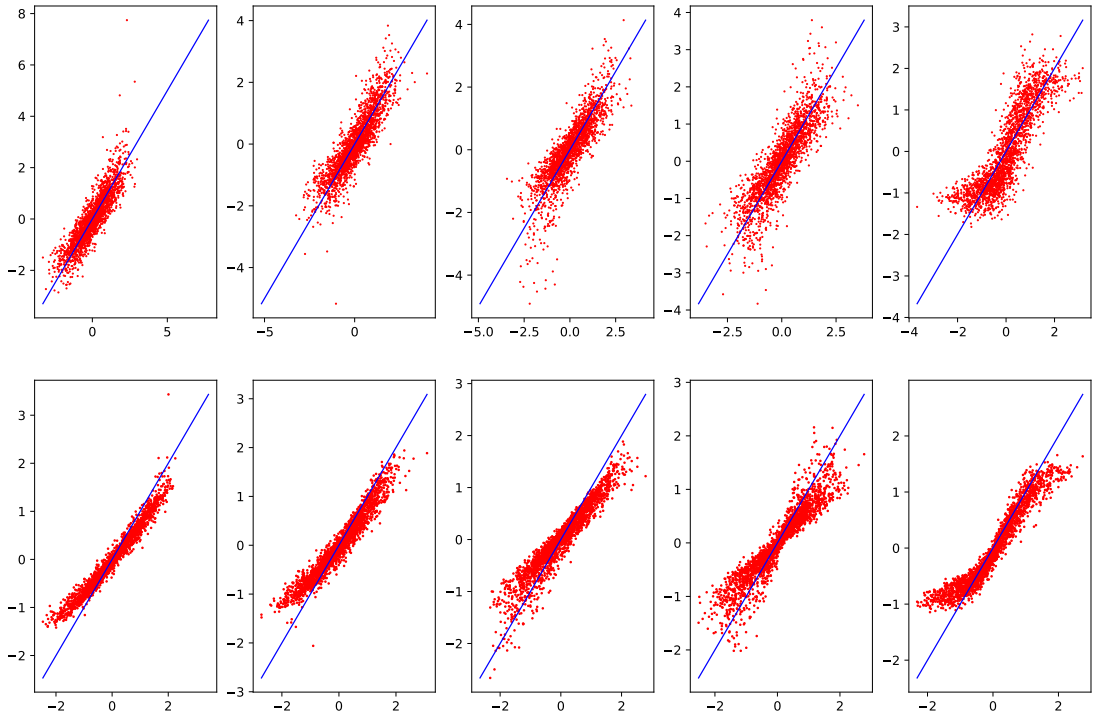


Figure 18: Q-Q plots (left for OT, right for EOT with $\varepsilon = 0.5 \times 10^{-2}$) for two multivariate samples, each of size $2180$; the first one is drawn from a $5$-dimensional standard Gaussian distribution and the second one is the standardised Turkish rice Osmanic data. The straight (black) lines represent the line $L$.

In this second example, we consider the Turkish rice Osmanic dataset of size 2180, with 5 variables that correspond to some features of the rice: Perimeter, Major Axis Length, Minor Axis Length, Convex Area, and Extent. As previously, we want to compare the 5-dimensional standardised empirical distribution of this data with a 5-dimensional standard Gaussian distribution. We draw the (E)OT potential plots (see Figure 17), followed by the (E)OT Q-Q plots (see Figure 18). For the EOT approach, we choose the regularisation parameter $\varepsilon$ to be $0.5 \times 10^{-2}$.

Looking at Figure 17, we observe that the points are spread out around the line $L$ (not as concentrated along $L$ as e.g. in Figure 13). So, we would deduce that the underlying distribution of the rice sample may not be well modelled by a multivariate standard Gaussian distribution. Let us move to the Q-Q plots in Figure 18 to gain more visual insights. There, both OT and EOT Q-Q plots clearly indicate that the rice sample is drawn from a non-Gaussian distribution. Moreover, we observe a heavy tail behavior, for each component.
Since the sample size is reasonable, let us complete this empirical analysis by computing the test statistics $E_n$ and $F_n$ (defined in (4.1)) and the corresponding $p$-values under the null hypothesis that the measure of the data (denoted $\nu_{\text{rice}}$) is Gaussian (denoted $\nu_G$), *i.e.* $H_0 : \nu_{\text{rice}}|_{K_2} = \nu_G|_{K_2}$.

| $n$ | $E_n$ | $p$-value of $E_n$ | $F_n$ | $p$-value of $F_n$ |
|---|---|---|---|---|
| 250 | 277.57 | 0.44825 | 76.73 | 0.00025 |
| 500 | 469.02 | 0 | 180.31 | 0 |
| 1000 | 950.45 | 0 | 312.65 | 0 |

Table 5: Test statistics and $p$-values under the null hypothesis $H_0 : \nu_{\text{rice}}|_{K_2} = \nu_G|_{K_2}$.

The $p$–values reported in Table 5 are small, as we could expect from what we observed on the various plots. This supports the assertion that the Turkish rice Osmanic data does not come from a Gaussian distribution.

# 7 Comparison with geometric Q-Q plot

In this section, we aim at comparing the QQ-plots as graphical tools (to compare two distributions) when adopting two main approaches, the (E)OT one and the geometric one. To do so, we display the OT, EOT, and geometric Q-Q plots for various simulated and real datasets, then, focus on comparing if one of the methods provides more informative or stronger visual insights about the two datasets being compared. For the procedure regarding geometric Q-Q plots, we refer the reader to Dhar et al. [2014].

## 7.1 Simulated data

We begin by considering two sets of samples, each containing 1000 observations. The first set is drawn from a 5-dimensional standard Gaussian distribution, while the second set is drawn from a 5-dimensional distribution, chosen as follows. The marginals are independent, with the first four following a univariate standard Gaussian distribution, and the fifth one following a univariate Pareto distribution with parameter 3.2. In Figure 19, we depict 3 types of multivariate Q-Q plots: the first row represents the OT Q-Q plots, the second row shows the EOT Q-Q plots when taking $\varepsilon = 0.5 \times 10^{-2}$, and the last row displays the geometric Q-Q plots.

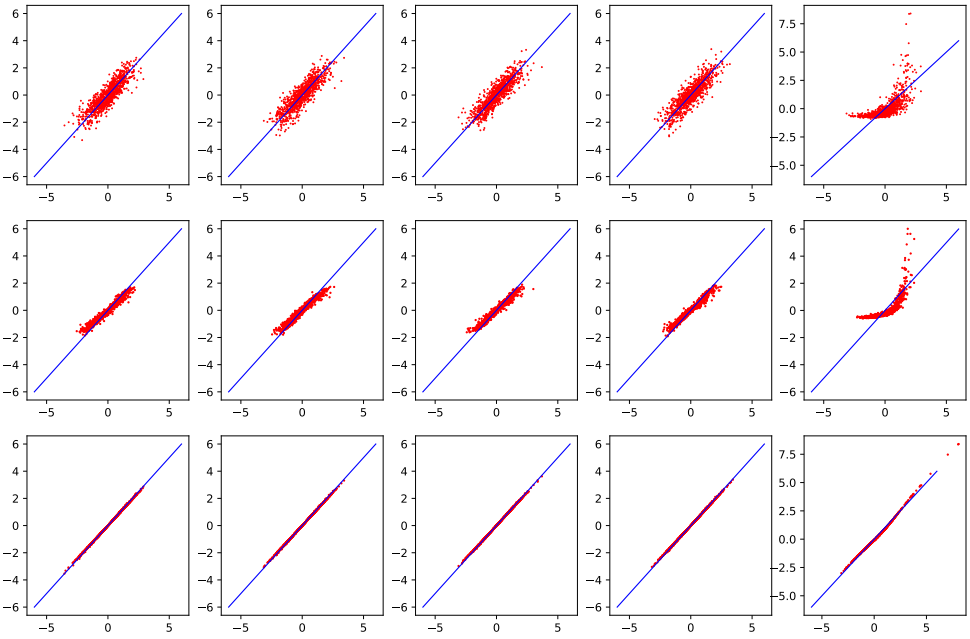


Figure 19: Q-Q plots for two samples, the first one is drawn from a 5-dimensional standard normal distribution, the second one from a 5-dimensional distribution whose first 4 marginals are standard normal and the last one is Pareto with parameter 3.2, all the marginals being independent of each other. The first row displays OT Q-Q plots, the second EOT Q-Q plots with $\varepsilon = 0.5 \times 10^{-2}$, and the last one geometric Q-Q plots. The straight (blue) lines represent the line $L$.

We observe in Figure 19 that OT and EOT Q-Q plots clearly reveal the presence of a heavy tail for the 5th component, whereas it is not so obvious from the geometric one. In the 5th component of the geometric Q-Q plot, we see some slight movement away from the straight line $L$, but very mild in comparison with the OT and EOT approaches.
Next, we perform a similar experiment but replace the 5th marginal distribution for the second sample by a Student's $t$-distribution with 3.2 degrees of freedom. The OT, EOT, and geometric Q-Q plots are displayed in Figure 20.
We observe that, while the OT and EOT Q-Q plots reveal the presence of heavy tails in the second distribution, the geometric Q-Q plot does not. Moreover, in the latter case,

the point cloud becomes highly aligned with the straight line $L$, suggesting that the two samples are drawn from the same distribution.

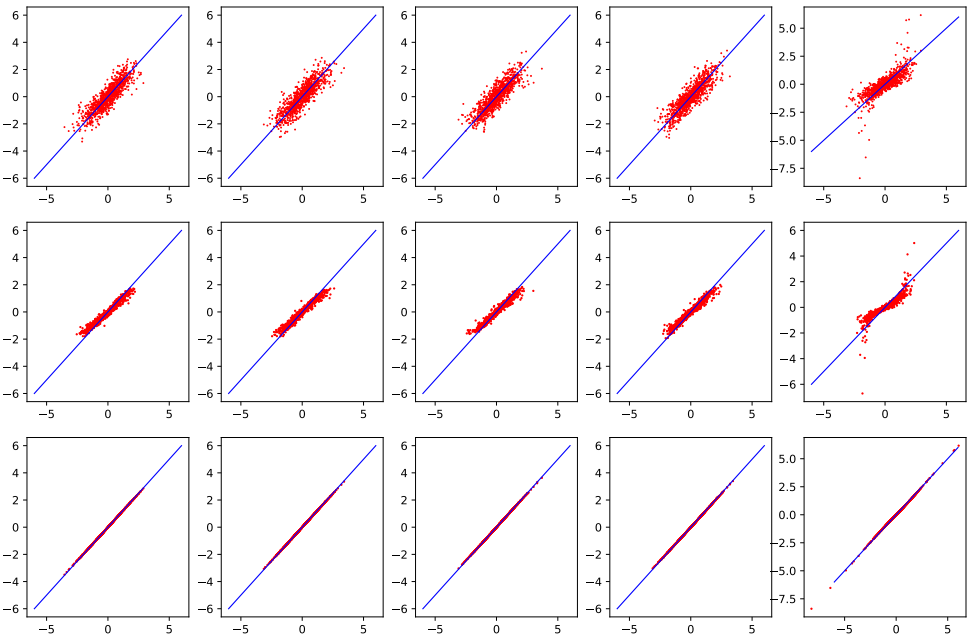


Figure 20: Q-Q plots for two samples each of size 1000, the first one is drawn from a 5-dimensional standard normal distribution, the second one from a 5-dimensional distribution whose first 4 marginals are standard normal and the last one is a Student's t-distribution with 3.2 degrees of freedom, all marginals being independent of each other. The first row displays OT Q-Q plots, the second one EOT Q-Q plots with $\varepsilon = 0.5 \times 10^{-2}$, and the last one geometric Q-Q plot. The straight (blue) lines represent the line $L$.

## 7.2 Real data

Here, we go back to Example 6.2, choosing the Turkish rice Osmanic dataset as second sample, while the first one is drawn from the 5-dimensional standard Gaussian distribution. We note that both OT and EOT Q-Q plots indicate that the second sample is drawn from a distribution that is different from the Gaussian distribution. Additionally, we observe distinct tail behavior in the third and fifth components. However, the geometric Q-Q plot fails to differentiate between the two samples.
In Appendix B, we perform a similar comparison, but now comparing the unaltered real data with a multivariate Gaussian distribution whose parameters are learned from the data.

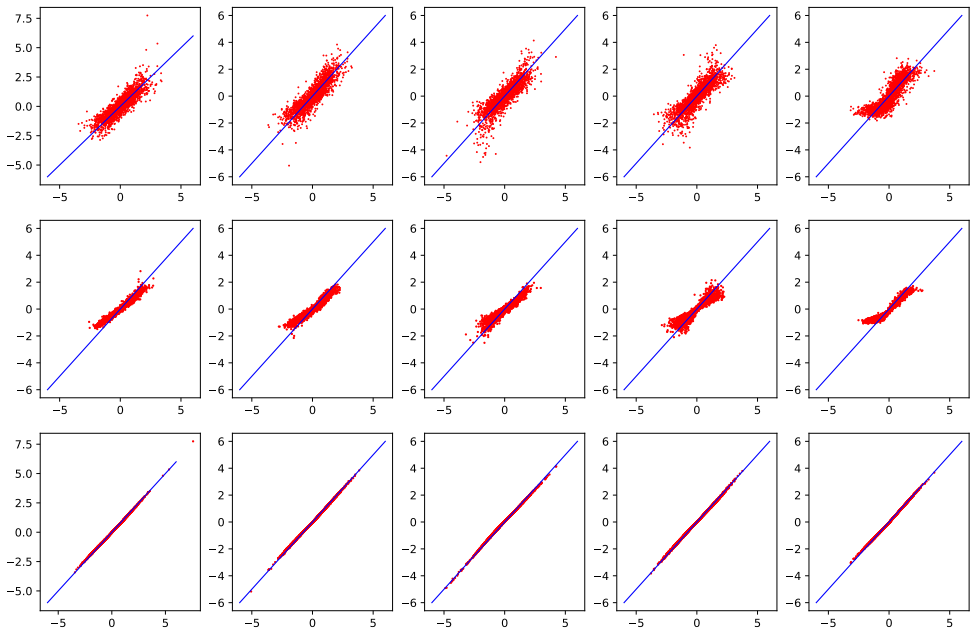


Figure 21: Q-Q plots for two samples, each of size 2180. The first sample is drawn from a 5-dimensional standard normal distribution, while the second sample consists of 2180 observations of standardised five features of Turkish rice Osmanic. The first row displays OT Q-Q plot, the second one EOT Q-Q plot with $\varepsilon = 0.5 \times 10^{-2}$ and the last one geometric Q-Q plot.

# 8 Conclusion

In recent times, multivariate OT based quantiles have become a very important object of research because it reflects many useful theoretical properties of univariate quantiles. Although OT is well understood theoretically, it is computationally very challenging for a large sample; it lacks statistical stability. Many possible solutions have been proposed to circumvent these issues, entropy regularisation (EOT) being one of them.

In this setting, considered both OT and EOT (maps and potentials) and obtained theoritical results that show that these (E)OT tools can be used to characterise distributions uniquely. Based on this property, we built multivariate Q-Q plots and potential plots by which one can visually compare two multivariate distributions.

(E)OT potential plots are interesting visual tools when comparing multivariate distributions, as they provide a 2-dimensional representation, which is an advantage w.r.t. Q-Q plots, especially in high dimension. However, Q-Q plots provide a visual comparison for specific features (scaling, shifting, outliers, tail heaviness) which is more obvious and readable than potential do.

Now, comparing OT versus EOT approach, we could observe that EOT Q-Q plots, with small regularization parameter, can determine well if two distributions are statistically similar or not, but they do not reveal informations about special features such as tail heaviness. Note that it is important to choose the value of $\varepsilon$ sufficiently small so that the

EOT map closely approximates the OT map, otherwise, for big values of $\epsilon$, the EOT map becomes close to a constant map. It is also observed that the smaller the values of $\varepsilon$, the longer the algorithm takes to execute. Also, if $\varepsilon$ is chosen very small, it happens sometimes that the code in Flamary et al. [2021] breaks down because of encountering NaN value.
Finally, we proposed statistical tests associated with the EOT Q-Q and potential plots, based on the available statistical stability results (central limit theorems) for EOT. Once limit theorems will be available for OT, we will be able to construct test statistics for OT as well, in a similar way as for EOT. Note that limit theory for OT is an area of active research, for recent developments on this topic, we refer e.g. to Sadhu et al. [2023], Manole et al. [2023].
Our future goal is to study the asymptotic of empirical estimators of the (E)OT quantile and potential function, specifically their tail behavior, as we did for geometric quantiles and half-space depths (see Singha et al. [2023]). It will be helpful to have a better understanding of such statistical tools in view of their applications.

# References


Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44:375–417, 1991.

V. Chernozhukov, A. Galichon, M. Hallin, and M. Henry. Monge-kantorovich depth, quantiles, ranks and signs. *Annals of Statistics*, 45(1):223–256, 2017.

M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transportation distances. *Advances in Neural Information Processing Systems*, 26:2292–2300, 2013.

C. de Valk and J. Segers. Tails of optimal transport plans for regularly varying probability measures. arXiv preprint:1811.12061, 2018.

S. Dhar, B. Chakraborty, and P. Chaudhuri. Comparison of multivariate distributions using quantile–quantile plots and related tests. *Bernoulli*, 20(3):1484–1506, 2014.

G.S. Easton and R.E. McCulloch. A multivariate generalization of quantile-quantile plots. *Journal of the American Statistical Association*, 85(410):376–386, 1990.

R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N.T.H. Gayraud, H. Janati, A. Rakotomamonjy, L. Redko, A. Rolet, A. Schutz, V. Seguy, D.J. Sutherland, R. Tavenard, A. Tong, and T. Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021. URL http://jmlr.org/papers/v22/20-451.html.

A. Genevay. Entropy-regularized optimal transport for machine learning. Ph.D. thesis, 2019. URL https://audeg.github.io/publications/these_aude.pdf.

P. Ghosal and B. Sen. Multivariate ranks and quantiles using optimal transport: Consistency, rates and nonparametric testing. *The Annals of Statistics*, 50(2):1012–1037, 2022.

Z. Goldfeld, K. Kato, G. Rioux, and R. Sadhu. Limit theorems for entropic optimal transport maps and the sinkhorn divergence. *arXiv preprint arXiv:2207.08683*, 2023.

M. Hallin, E. Del Barrio, J. Cuesta-Albertos, and C. Matrán. Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *Annals of Statistics*, 49:1139–1165, 2021.

J. C. Hütter and P. Rigollet. Minimax estimation of smooth optimal transport maps. *The Annals of Statistics*, 49(2):1166 – 1194, 2021.

T. Manole, S. Balakrishnan, J.N. Weed, and L. Wasserman. Central limit theorems for smooth optimal transport maps. *arXiv:2312.12407*, 2023.

R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80:309–323, 1995.

M. Nutz. Introduction to entropic optimal transport. Lecture notes, available at https://www.math.columbia.edu/~mnutz/docs/EOT_lecture_notes.pdf, 2022.

A.-A. Pooladian and J. Niles-Weed. Entropic estimation of optimal transport maps. arXiv:2109.12004, 2022.

R. Sadhu, Z. Goldfeld, and K. Kato. Stability and statistical inference for semidiscrete optimal transport maps. arXiv:2303.10155, 2023.

F Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Springer, 2015.

S. S. Shapiro and M. B. Wilk. An analysis of variance test for normality: Complete samples. *Biometrika*, 52:591–611, 1965.

S. Singha, M. Kratz, and S. Vadlamani. From geometric quantiles to halfspace depths: A geometric approach for extremal behaviour. arXiv:2306.10789 & ESSEC WP2307, 2023.

A. W. van der Vaart and J. A. Wellner. *Weak Convergence*. Springer New York, 1996.

C. Villani. *Topics in Optimal Transportation*. American Mathematical Society, 2003.

C. Villani. *Optimal Transport Old and New*. Springer, 2009.

# Appendix - Additional examples

## A Another example for the comparison of heavy versus light tail

Let $\nu_X$ be the push forward measure of the trivariate standard normal distribution (denoted $\nu$) by the function $f(x_1, x_2, x_3) = (1 + |x_1|, 1 + |x_2|, 1 + |x_3|)$, *i.e.* $\nu_X = f\sharp\nu$. Let $\nu_Y$ be a probability distribution on $\mathbb{R}^3$, with i.i.d. Pareto(3) marginals (with density function given by $p(x) = 3x^{-4}$, $x \geq 1$). Note that both the distributions $\nu_X$ and $\nu_Y$ are supported on the first orthant. We consider two samples, each of size 1000, drawn from $\nu_X$ and $\nu_Y$, respectively. Then, we follow the steps $(ii) - (iii)$ to provide the Q-Q plots corresponding to these samples, which are displayed in Figure 22. Clearly, the scatter plots do not cluster around the straight line $L$ (in blue). We further notice that in each plot, the points exhibit a nonlinear behaviour which shows significant deviation from $L$. Observe that, in each componentwise plots, the high (or extreme) quantiles corresponding to $\mathcal{Y}^n$ grow faster than those of $\mathcal{X}^n$, implying that the sample $\mathcal{Y}^n$ represent a distribution with relatively heavier tail.

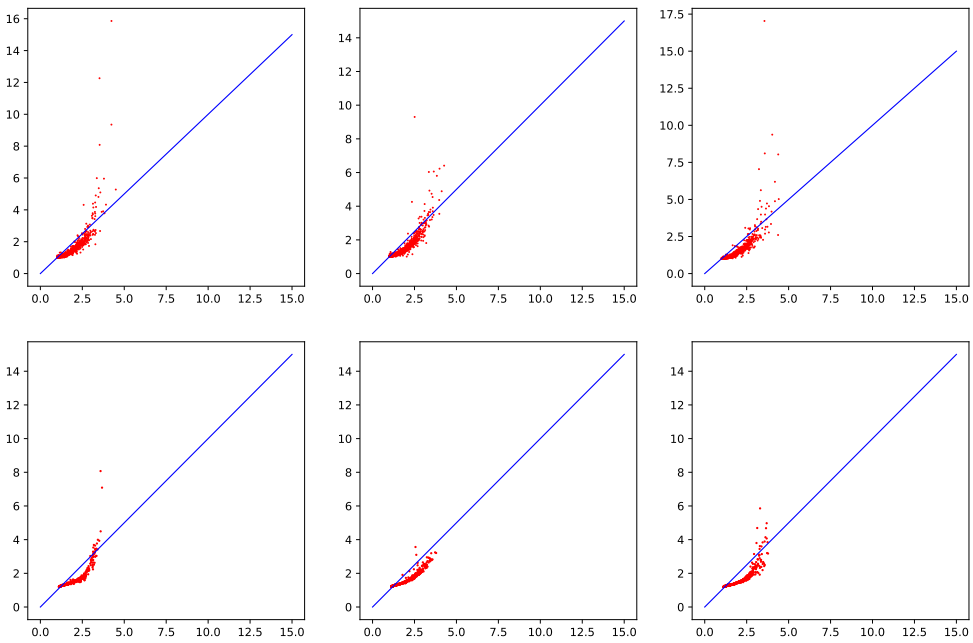


Figure 22: Q-Q plots for two samples, the first one is drawn from $\nu_X$ and the second from $\nu_Y$. The first row displays OT Q-Q plot and the second EOT Q-Q plot with $\varepsilon = 10^{-3}$. The straight (blue) lines represent the line $L$.

Next, the potential plots (step (iv)) for this particular example are displayed in Figure 23. It can be observed that the points are scattered around a nonlinear (increasing) curve above

the line $L$. This implies that the potential function associated with the second sample has a higher growth rate. Since the quantile is the potential gradient, higher growth rate of potential implies that, in absolute term, the quantiles are bigger, thereby suggesting that the corresponding distribution is heavy tailed.

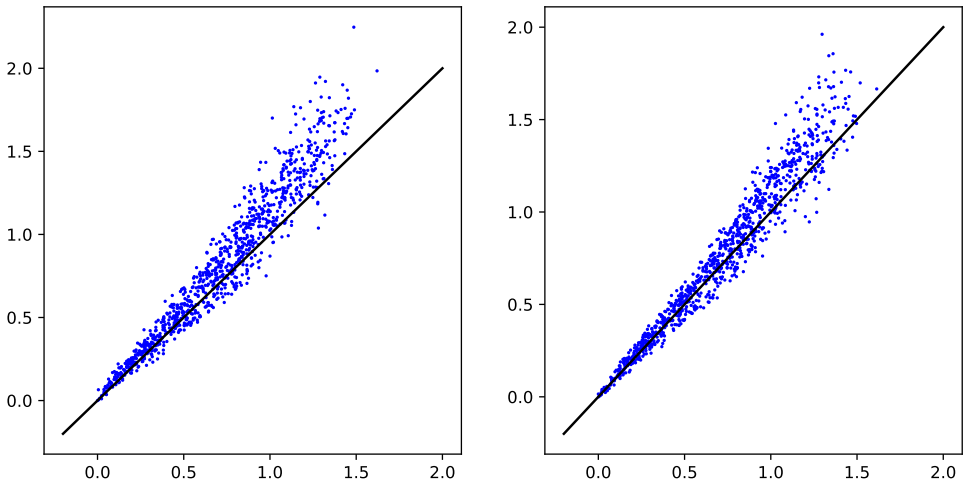



Figure 23: Potential plots (left for OT, right for EOT with $\varepsilon = 10^{-3}$) for two multivariate samples, each of size 1000; the first sample is drawn from $\nu_X$ and the second from $\nu_Y$. The straight (black) lines represent the line $L$.

# B Comparison among OT, EOT and geometric Q-Q plots

We consider the same examples on simuated and real data as in Section 7, but now we perform the comparison of the raw data with an appropriate multivariate Gaussian distribution whose parameters (mean and covariance for instance) are learned from the raw data.

## B.1 Simulated data

We begin by considering two samples, each containing 1000 observations.
Using the notation set forth in previous sections, $\mathcal{Y}^n$ is the sample to be compared; it is drawn from a 5-dimensional distribution with independent marginals, such that the first four marginals follow a univariate standard Gaussian distribution, and the fifth marginal follows a univariate Pareto distribution with parameter 3.2. The sample $\mathcal{X}^n$ is drawn from a 5-dimensional Gaussian distribution with mean $\hat{m}$ and covariance $\hat{\Sigma}$, where $\hat{m}$ and $\hat{\Sigma}$ are the sample mean and the sample covariance, respectively, of $\mathcal{Y}^n$.
In Figure 24, we depict various types of multivariate Q-Q plots: the first row represents the OT Q-Q plot, the second row shows the EOT Q-Q plot, and the last row displays the

geometric Q-Q plot.

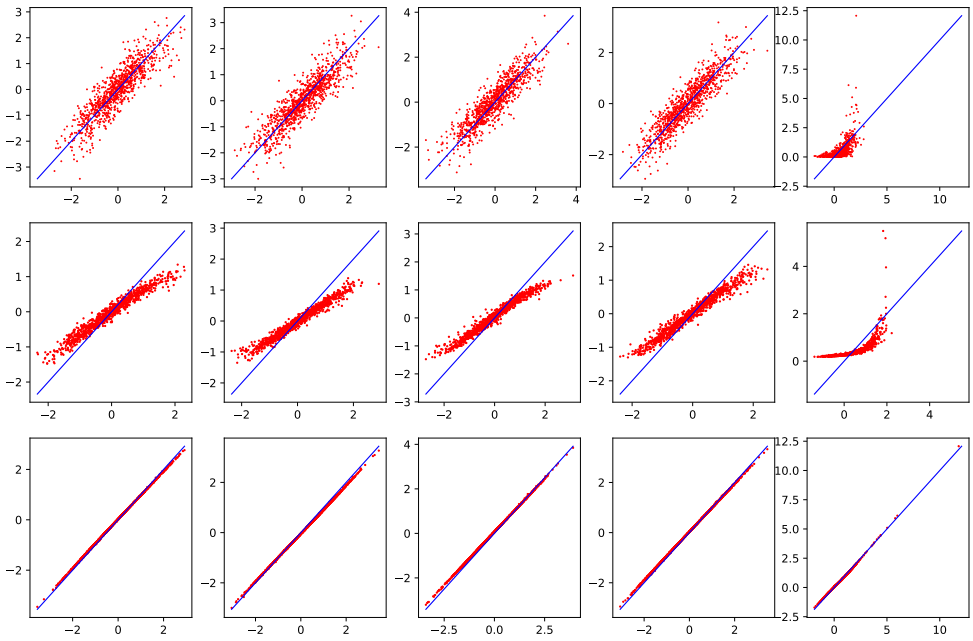


Figure 24: Q-Q plots for two samples, the first one is drawn from a 5-dimensional normal $(\hat{m}, \hat{\sigma})$ distribution, the second one from a 5-dimensional distribution whose first 4 marginals are standard normal and the last one is Pareto with parameter 3.2, all the marginals being independent of each other. The first row displays OT Q-Q plots, the second one EOT Q-Q plot with $\varepsilon = 0.5 \times 10^{-2}$, and the last one geometric Q-Q plots. The straight (blue) lines represent $L$.

We observe from Figure 24, that the OT and EOT Q-Q plots clearly reveal about the presence of a heavy tail in the fifth component, whereas it is not so obvious from the geometric one. In the fifth component of the geometric Q-Q plot, we see some slight deviation from the straight line $L$, but it is really mild in comparison to the OT and EOT.

Next, we perform a similar experiment but with different marginal distribution. $\mathcal{Y}^n$ is drawn from a 5–dimensional distribution, where the marginals are independent of each other. The first four marginals follow a univariate standard Gaussian distribution, while the last marginal follows a Student's $t$-distribution with degree of freedom 3.2. $\mathcal{X}^n$ is drawn from the 5–dimensional multivariate Gaussian distribution with mean $\hat{m}$ and covariance $\hat{\Sigma}$, where $\hat{m}$ and $\hat{\Sigma}$ are the sample mean and sample covariance of $\mathcal{Y}^n$, respectively. The OT, EOT, and geometric Q-Q plots are displayed in Figure 25.

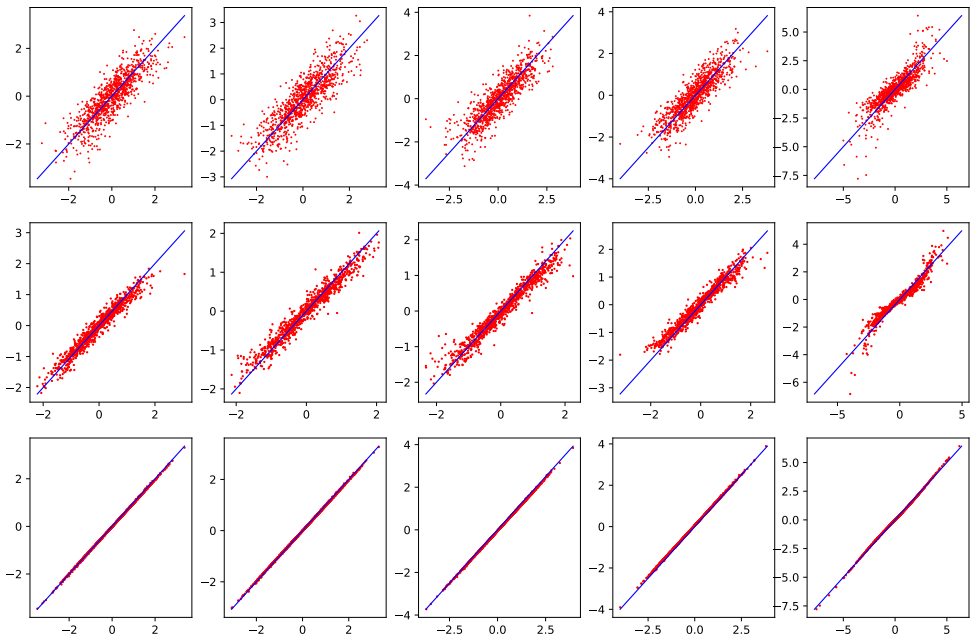


Figure 25: Q-Q plots for two samples each of size 1000, the first one is drawn from a 5-dimensional normal $(\hat{m}, \hat{\Sigma})$ distribution, the second one from a 5-dimensional distribution whose first 4 marginals are standard normal and the last one is Studen's t distribution with parameter 3.2 and all the marginals are independent of each other. The first row displays OT Q-Q plot, the second one EOT Q-Q plot with $\varepsilon = 0.5 \times 10^{-2}$, and the last one geometric Q-Q plot. The straight lines in blue represent the line $L$.

We observe that, while the OT and EOT Q-Q plots reveal the presence of heavy tails in the 5 th margnal, the geometric Q-Q plot does not. Moreover, in the latter case, the point cloud becomes highly aligned with the straight line $L$, suggesting that the two samples are drawn from the same distribution.

## B.2 Real data

Now let us move to real data, considering the same data as in Section 7.2. The sample $\mathcal{Y}^n$ consists of 2180 observations of five features of Turkish rice Osmanic (Perimeter, Major Axis Length, Minor Axis Length, Convex Area, and Extent). $\mathcal{X}^n$ is the sample drawn from the 5–dimensional Gaussian distribution with mean $\hat{m}$ and covariance $\hat{\Sigma}$, where $\hat{m}$ and $\hat{\Sigma}$ are the sample mean and the sample covariance, respectively, of $\mathcal{Y}^n$. The Q-Q plots are displayed in Figure 26.

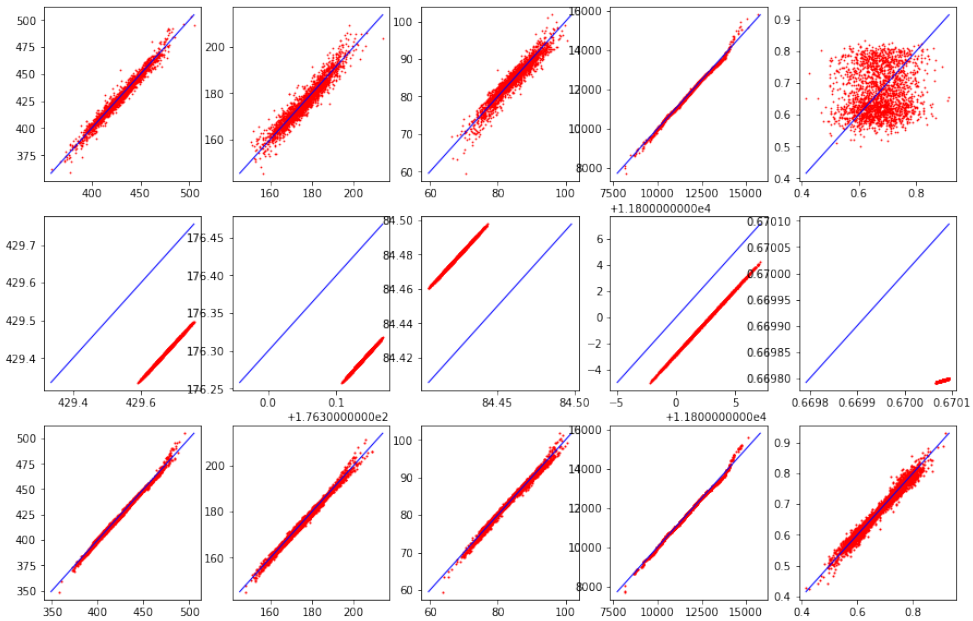


Figure 26: Q-Q plots for two samples, each of size 2180. The first sample is drawn from the 5-dimensional Gaussian distribution with mean $\hat{m}$ and covariance matrix $\hat{\Sigma}$, while the second sample consists of 2180 observations of five features of Turkish rice Osmanic. The first row displays OT Q-Q plots, the second one EOT Q-Q plots with $\varepsilon = 0.5 \times 10^{-2}$ and the last one geometric Q-Q plots. The straight (blue) lines represent $L$.

We observe that all the Q-Q plots indicate that the second sample is drawn from a distribution different from the Gaussian distribution with mean $\hat{m}$ and covariance matrix $\hat{\Sigma}$. Notably, the EOT Q-Q plots look very different as compared to the OT and geometric Q-Q plots, possibly due to the regularization parameter $\varepsilon$ not being sufficiently small. Due to computational limitations, we were unable to select $\varepsilon$ smaller than $0.5 \times 10^{-2}$. On the other hand, the OT Q-Q plots more prominently highlight the dissimilarities between the two samples, particularly evident in the 5-th component (note that the scaling is very different when comparing with other components) of the plots. It is important to note that none of the plots exhibit clear features of the distribution, like heavy tails. Compared to the Q-Q plots conducted after standardizing both samples in Example 6.2 and Section 7.2, this drawback indicates that standardizing the samples when working with real data might be a better approach.

**ESSEC Business School**
3 avenue Bernard-Hirsch
CS 50105 Cergy
95021 Cergy-Pontoise Cedex
France
Tél. +33 (0)1 34 43 30 00
**www.essec.edu**

**ESSEC Executive Education**
CNIT BP 230
92053 Paris-La Défense
France
Tél. +33 (0)1 46 92 49 00
**www.executive-education.essec.edu**

**ESSEC Asia-Pacific**
5 Nepal Park
Singapore 139408
Tél. +65 6884 9780
**www.essec.edu/asia**

ESSEC | CPE Registration number 200511927D
Period of registration: 30 June 2017 - 29 June 2023
Committee of Private Education (CPE) is part of SkillsFuture Singapore (SSG)

**ESSEC Afrique**
Plage des Nations - Golf City
Route de Kénitra - Sidi Bouknadel (Rabat-Salé)
Morocco
Tél. +212 (0)5 37 82 40 00
**www.essec.edu**

## CONTACT

**Research Center**
research@essec.edu