Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2011, Vol. 47, No. 2, 358–375
DOI: 10.1214/09-AIHP354

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Uniqueness and approximate computation of optimal incomplete transportation plans

P. C. Álvarez-Esteban[a,1,2], E. del Barrio[a,1,2], J. A. Cuesta-Albertos[b,1] and C. Matrán[a,1,2]

[a] *Universidad de Valladolid, Prado de la Magdalena S/N, Valladolid, 47005, Spain. E-mail: tasio@eio.uva.es*
[b] *Universidad de Cantabria, Facultad de Ciencias, Santander, 39005, Spain*

Received 22 June 2009; revised 4 December 2009; accepted 11 December 2009

Abstract. For $\alpha \in (0, 1)$ an α-trimming, P^*, of a probability P is a new probability obtained by re-weighting the probability of any Borel set, B, according to a positive weight function, $f \le \frac{1}{1-\alpha}$, in the way $P^*(B) = \int_B f(x) P(\mathrm{d}x)$.

If P, Q are probability measures on Euclidean space, we consider the problem of obtaining the best L_2-Wasserstein approximation between: (a) a fixed probability and trimmed versions of the other; (b) trimmed versions of both probabilities. These best trimmed approximations naturally lead to a new formulation of the mass transportation problem, where a part of the mass need not be transported. We explore the connections between this problem and the *similarity* of probability measures. As a remarkable result we obtain the uniqueness of the optimal solutions. These optimal incomplete transportation plans are not easily computable, but we provide theoretical support for Monte-Carlo approximations. Finally, we give a CLT for empirical versions of the trimmed distances and discuss some statistical applications.

Résumé. Pour $\alpha \in (0, 1)$, une α-coupe P^* d'une probabilité P selon une fonction positive f majorée par $1/(1-\alpha)$ est la probabilité obtenue pour tout ensemble de Borel B par $P^*(B) = \int_B f(x) P(dx)$.

Si P, Q sont deux probabilités sur l'espace euclidien, on considère le problème de minimiser la distance de Wasserstein L^2 entre (a) une probabilité et ses versions coupées (b) les versions coupées de deux probabilités. Ce problème mène naturellement à une nouvelle formulation du problème de transport de masse, où une partie de la masse ne doit pas être transportée. Nous explorons les liaisons entre ce problème et la *similitude* des mesures de probabilité. Un de nos résultats remarquables est l'unicité du transport de masse. Ces plans de transport optimal incomplets ne sont pas facilement calculables mais nous fournissons un appui théorique pour des approximations de Monte-Carlo. Enfin, nous donnons un TCL pour les versions empiriques des distances coupées et discutons certaines applications statistiques.

MSC: Primary 49Q20; 60A10; secondary 60B10; 28A50

Keywords: Incomplete mass transportation problem; Multivariate distributions; Optimal transportation plan; Similarity; Trimming; Uniqueness; Trimmed probability; CLT

1. Introduction

This paper considers a modified version of the classical Mass Transportation Problem (MTP in the sequel). Broadly speaking, the MTP can be formulated as trying to relocate a certain amount of mass with a given initial distribution to another target distribution in such a way that the transportation cost is minimized. This seemingly simple problem has

[1] Supported in part by the Spanish Ministerio de Educación y Ciencia and FEDER, Grants MTM2008-06067-C02-01 and 02.
[2] Supported in part by the Consejería de Educación de la Junta de Castilla y León.

a long history which dates back to Monge. The initial formulation of the problem can be summarized in present-day language as follows. Let P_1, P_2 be two probability measures on the Euclidean space $\mathbb{R}^k$ with norm $\|\cdot\|$ and Borel σ-field β. Consider the set, $\mathbb{T}(P_1, P_2)$, of maps transporting P_1 to P_2, that is, the set of all measurable maps $T:\mathbb{R}^k \to \mathbb{R}^k$ such that, if the initial space is endowed with the probability P_1, then the distribution of the random variable T is P_2. Then Monge's problem consists of finding a transportation map, T_0, from P_1 to P_2 such that

$$T_0 := \mathop{\arg\min}_{T\in\mathbb{T}(P_1,P_2)} \int_{\mathbb{R}^k} \|x - T(x)\| P_1(\mathrm{d}x).$$

A later, fundamental generalization of this problem is the so-called Kantorovitch–Rubinstein–Wasserstein (KRW) formulation which consists in finding

$$\mathcal{W}_2^2(P_1, P_2) := \inf_{\pi\in\mathcal{M}(P_1,P_2)} \left\{ \int \|x-y\|^2 \,\mathrm{d}\pi(x, y) \right\}, \tag{1}$$

with $\mathcal{M}(P_1, P_2)$ the set of finite, positive measures on $\beta\times\beta$ with marginals P_1 and P_2.

Apart from the consideration of different cost functions, the main difference between the Monge and the KRW problems is that the later is not related to transportation maps. We mean that in the KRW formulation masses sharing the same initial position may end up in different locations. The KRW minimization allows us also to consider the L_2-Wasserstein distance, $\mathcal{W}_2(P_1, P_2)$, between probability measures with finite second moment (see, e.g., Bickel and Freedman [3] for details). Remarkably, the Monge and the KRW formulations turn out to be equivalent under some smoothness assumptions.

Existence, uniqueness or regularity of mappings $T\in\mathbb{T}(P_1, P_2)$ satisfying $\int_{\mathbb{R}^k} \|x-T(x)\|^2 \,\mathrm{d}P_1(x) = \mathcal{W}_2^2(P_1, P_2)$ are problems that have attracted the attention of mathematicians from very different points of view. Fluid Mechanics, Partial Differential Equations, Optimization, Probability Theory and Statistics are in the very broad range of applications of this and related MTP's justifying the interest and also the different technical approaches for their study. To avoid a formidable amount of references we refer to the books by Rachev and Rüschendorf [20] and by Villani [24] for an updated account of the interest and implications of the problem, as well as to recent works illustrating the permanent actuality of the topic, as Ambrosio [2], Caffarelli et al. [6], or Feldman and McCann [15].

Here we will analyze a variant of the KRW problem involving incomplete mass transportation. Let us introduce it through a motivating example. Gangbo and McCann consider in [17] the problem of identifying a leaf l_0 by comparing it with a catalog. They analyze the approach based on minimizing the transportation cost between the uniform distribution on the outline of l_0 and its counterparts in the catalog. To avoid technicalities, we assume that we are dealing with black and white pictures of l_0 and the leaves in the catalog, rather than their outlines. We identify the grey-levels with the density of a measure, compute the associated L_2-Wasserstein distances and identify l_0 with the closest leaf in the catalog.

Now, let us assume that, as it often happens, the picture of l_0 is corrupted at some spots. It seems reasonable to delete those spots before making the comparisons. However, it is not always easy to tell a corrupted spot from a distinctive feature. A reasonable procedure would be to transport only a part of the initial mass, dismissing a small fraction, to minimize the transportation cost. If the leaves in the catalog are also corrupted by noise, then we should also allow some fraction of the target picture to remain unmatched.

Turning to a more abstract formulation, we observe some random object X with law P_1. Ideally, P_1 should equal P_2 (one of the underlying laws in the catalogue); the presence of noise means that, in fact,

$$P_1 = (1-\varepsilon)P_2 + \varepsilon N, \quad \varepsilon \le \alpha, \tag{2}$$

for some unspecified N if we assume that the noise level does not exceed α. If we compare P_1 to a noisy standard, P_2, our goal should be to assess whether

$$P_i = (1-\varepsilon_i)P_0 + \varepsilon_i N_i, \quad \varepsilon_i \le \alpha, i = 1, 2. \tag{3}$$

We say that P_1 is *similar* to P_2 at level α if (2) holds and, likewise, we say that P_1 and P_2 are *similar* (at level α) if (3) holds. In a natural way we end up in the problem of dismissing a (small) fraction of the masses represented by P_1 and

P_2. This resembles the trimming procedures employed in Statistics where, very often, outlying observations have to be deleted. A probabilistic approach to trimming was first considered in Gordaliza [18] in the setup of robust location estimation. We employ the following definition.

Definition 1.1. *Given $0 \le \alpha \le 1$ and Borel probability measures P, P^* on $\mathbb{R}^k$, we say that P^* is an α-trimming of P if P^* is absolutely continuous with respect to P, and $\frac{dP^*}{dP} \le \frac{1}{1-\alpha}$. The set of all α-trimmings of P will be denoted by $\mathcal{R}_\alpha(P)$.*

This definition of trimming is more general than other usual alternatives in that it allows points to be partially trimmed. It has been considered in Cascos and López-Díaz [8] and [9] for the construction and estimation of central regions and in our work [1] in a particular problem of essential model validation, which can be reformulated as follows: given (univariate) data from two random generators P_1, P_2, test whether $P_i = \mathcal{L}(\varphi_i(Z))$, $i = 1, 2$, for some r.v. Z and nondecreasing functions φ_i such that $\mathbb{P}(\varphi_1(Z) \neq \varphi_2(Z)) \le \alpha$.

In this paper we show the usefulness of Definition 1.1 for the assessment of similarity in the sense given in (2) and (3). In fact, with our definition, (2) holds if and only if $P_2 \in \mathcal{R}_\alpha(P_1)$, while (3) is equivalent to $\mathcal{R}_\alpha(P_1) \cap \mathcal{R}_\alpha(P_2) \neq \emptyset$, see Proposition 2.1 below. Thus, natural measures of the deviation from (2) or (3) are given by

$$\mathcal{W}_2\big(\mathcal{R}_\alpha(P_1), P_2\big) := \min_{P_1^* \in \mathcal{R}_\alpha(P_1)} \mathcal{W}_2\big(P_1^*, P_2\big),$$

$$\mathcal{W}_2\big(\mathcal{R}_\alpha(P_1), \mathcal{R}_\alpha(P_2)\big) := \min_{P_1^* \in \mathcal{R}_\alpha(P_1), P_2^* \in \mathcal{R}_\alpha(P_2)} \mathcal{W}_2\big(P_1^*, P_2^*\big).$$

We will refer to this minimal distances as *one-sided* and *two-sided trimmed distances* and to the minimization problems as *optimal incomplete transportation problems*. The corresponding minimizers will give a good approximation to the common part in P_1 and P_2 and, automatically, will determine the noisy spots. The main goal of this work is the study of these minimizers, showing existence, uniqueness and some qualitative properties, and to provide asymptotic results (LLN's and CLT's) for the empirical versions obtained from i.i.d. realizations of P_1 and P_2.

The paper is organized as follows. In Section 2, after giving some general background on trimmings, we show that, as in the classical optimal transportation problem, the KRW and Monge formulations are equivalent under absolute continuity. We also prove the uniqueness of the optimal transportation plan. From a technical point of view, the most remarkable (and difficult) result concerns the uniqueness in the two-sided problem (Theorems 2.11 and 2.16). Our definition of trimming allows us to partially trim some points. In fact Theorem 2.15 shows that only the mass placed on non-trimmed points is transported, while the mass on partially trimmed points must remain fixed. We complete Section 2 with a Law of Large Numbers for empirical optimal incomplete trimmings, which justifies the use of Monte-Carlo approximation. This is of primary interest, since there are not general explicit expressions for optimal incomplete transportation plans, even on the real line. Finally, Section 3 gives CLT's for empirical versions of the one- and two-sided trimmed distances. Our approach is based on an *empirical trimming process* for which we prove convergence to a certain Gaussian process. We show how these CLT's can be applied for testing null hypotheses related to (2) and (3). We provide a small simulation study showing the quality of the approximation for finite samples.

Once this work was completed we learned about recent work by Caffarelli and McCann [7] and Figalli [16] where the problem of transporting a fraction of the whole mass is also considered. Although their motivation is very different, a main goal of these works is the analysis of the uniqueness of the optimal transportation plan. Our uniqueness results improve those in these references in that our smoothness assumptions are minimal and allow to handle singular measures. This is of crucial importance in statistical applications in which one often has to deal with empirical measures (on the other hand [7] and [16] are concerned with regularity of the solutions and smoothness assumptions are needed for that purpose). Our proofs are purely probabilistic and we believe they are of independent interest, giving a new perspective on the topic.

We end this introduction with some notation. We write ℓ^k for Lebesgue measure on the space $(\mathbb{R}^k, \beta)$, while $\mathcal{F}_2(\mathbb{R}^k)$ will denote the set of probabilities on $\mathbb{R}^k$ with finite second moment. Given probabilities P, Q, by $P \ll Q$ we will denote absolute continuity of P with respect to (w.r.t.) Q, and by $\frac{dP}{dQ}$ the corresponding Radon–Nykodym derivative. By $\mathrm{supp}(P)$ we mean the support of P and by $P(\cdot|B)$ the conditional probability given the set B. With a slight abuse of notation, given Θ, $\Theta^* \subset \mathcal{F}_2(\mathcal{R}^k)$, we will often write

$$\mathcal{W}_2(P, \Theta) = \inf_{Q \in \Theta} \mathcal{W}_2(P, Q) \quad \text{and} \quad \mathcal{W}_2\big(\Theta, \Theta^*\big) = \inf_{(P,Q) \in \Theta \times \Theta^*} \mathcal{W}_2(P, Q).$$

Unless otherwise stated, the random vectors will be assumed to be defined on the same probability space (Ω, σ, ν). Weak convergence of probabilities will be denoted by $\to_w$ and $\mathcal{L}(X)$ will denote the law of the random vector X.

2. Trimmings and best trimmed approximations

We begin presenting some general background on trimmed probabilities. Further properties can be found in [1]. From the definition of $\mathcal{R}_\alpha(P)$ it is obvious that $P^* \in \mathcal{R}_\alpha(P)$ if and only if $P^* \ll P$ and $\frac{\mathrm{d}P^*}{\mathrm{d}P} = \frac{1}{1-\alpha} f$ with $0 \le f \le 1$. Thus, trimming allows us to downplay the weight of some regions of the measurable space without completely removing them from the feasible set.

The following proposition contains some useful facts about trimmings (see also [9]). Its proof is elementary and, hence, omitted.

Proposition 2.1. *For probabilities $P, \{P_n\}$ on $\mathbb{R}^k$:*

(a) *$P_2 \in \mathcal{R}_\alpha(P_1)$ iff $(1-\alpha)P_2(A) \le P_1(A)$ for every Borel set A iff (2) holds.*
(b) *$\mathcal{R}_\alpha(P_1) \cap \mathcal{R}_\alpha(P_2) \ne \varnothing$ iff (3) holds.*
(c) *If $\alpha < 1$ then $\mathcal{R}_\alpha(P)$ is compact for the topology of weak convergence.*
(d) *If $\alpha < 1$, $\{P_n\}_n$ is a tight sequence and $P_n^* \in \mathcal{R}_\alpha(P_n)$ for every n, then $\{P_n^*\}_n$ is tight. Moreover, if $P_n \to_w P$ and $P_n^* \to_w P^*$, then $P^* \in \mathcal{R}_\alpha(P)$.*

Next, we show that trimmings of a probability can be parametrized in terms of trimmings of another one.

Proposition 2.2. *If T transports Q to P, then*

$$\mathcal{R}_\alpha(P) = \{P^* \in \mathcal{P}(\mathcal{X}, \beta):\ P^* = Q^* \circ T^{-1},\ Q^* \in \mathcal{R}_\alpha(Q)\}.$$

Proof. If $\alpha = 1$ and $Q^* \ll Q$, then $P^* := Q^* \circ T^{-1} \ll P$, because $P(B) = 0$ implies $Q(T^{-1}(B)) = 0$, thus $P^*(B) = Q^*(T^{-1}(B)) = 0$. On the other hand, if $P^* \ll P$, we can define $w(y) = \frac{\mathrm{d}P^*}{\mathrm{d}P}(T(y))$ and $Q^*(B) = \int_B w(y) Q(\mathrm{d}y)$, hence, the change of variable formula shows for any set B in β

$$Q^* \circ T^{-1}(B) = \int_{T^{-1}(B)} \frac{\mathrm{d}P^*}{\mathrm{d}P}\big(T(y)\big) Q(\mathrm{d}y) = \int_B \frac{\mathrm{d}P^*}{\mathrm{d}P}(x) P(\mathrm{d}x) = P^*(B).$$

Let us assume that $\alpha < 1$. If $Q^* \in \mathcal{R}_\alpha(Q)$, then for any B in β

$$Q^* \circ T^{-1}(B) = \int_{T^{-1}(B)} \frac{\mathrm{d}Q^*}{\mathrm{d}Q}(x) Q(\mathrm{d}x) \le \frac{1}{1-\alpha} Q\big(T^{-1}(B)\big) = \frac{1}{1-\alpha} P(B),$$

thus $Q^* \circ T^{-1} \in \mathcal{R}_\alpha(P)$.

If we assume that $P^* \in \mathcal{R}_\alpha(P)$, by defining Q^* as above: $Q^*(B) = \int_B \frac{\mathrm{d}P^*}{\mathrm{d}P}(T(y)) Q(\mathrm{d}y)$, we have $Q^* \ll Q$, and $Q^* \circ T^{-1} = P^*$. Moreover, since $\frac{\mathrm{d}P^*}{\mathrm{d}P}(x) \le \frac{1}{1-\alpha}$ a.s. (P) and $P = Q \circ T^{-1}$, also $\frac{\mathrm{d}P^*}{\mathrm{d}P}(T(y)) \le \frac{1}{1-\alpha}$ a.s. (Q) hence $Q^* \in \mathcal{R}_\alpha(Q)$. □

Example 2.3. *If Q is the uniform distribution on the interval $(0, 1)$ then the set of distribution functions of elements of $\mathcal{R}_\alpha(Q)$ equals the set of absolutely continuous functions $h : [0, 1] \to [0, 1]$ with $h(0) = 0$, $h(1) = 1$ and such that $0 \le h'(x) \le 1/(1-\alpha)$ for almost every x. We write $\mathcal{C}_\alpha$ for this class of functions. Now, if P is a probability on the real line, a little thought yields that in this case Proposition 2.2 can be rewritten*

$$\mathcal{R}_\alpha(P) = \{P_h:\ h \in \mathcal{C}_\alpha\},$$

where P_h is the probability measures defined by $P_h(-\infty, x] = h(P(-\infty, x])$, $x \in \mathbb{R}$. This parametrization will be useful for the results in Section 3.

For our choice of metric, $\mathcal{W}_2$, it is well known (see, e.g., [3]) that for $P, Q \in \mathcal{F}_2(\mathbb{R}^k)$ the inf in (1) is attained, so that to find $\mathcal{W}_2^2(P, Q)$ it is enough to obtain a pair (X, Y) with $\mathcal{L}(X) = P$ and $\mathcal{L}(Y) = Q$ such that

$$\int \|X - Y\|^2 \,\mathrm{d}\nu = \inf\left\{\int \|U - V\|^2 \,\mathrm{d}\nu, \mathcal{L}(U) = P, \mathcal{L}(V) = Q\right\}.$$

Such a pair (X, Y) is called an L_2*-optimal transportation plan* (L_2-o.t.p.) for (P, Q). (L_2*-optimal coupling* for (P, Q) is an alternative, sometimes used, terminology.)

In Cuesta-Albertos and Matrán [11] (see also Brenier [4,5], Rüschendorf and Rachev [21] and McCann [19]) it was proved that, under continuity assumptions on the probability P, the L_2-o.t.p. (X, Y) for (P, Q) can be represented as $(X, T(X))$ for some suitable optimal map T. This map coincides with the (essentially unique) cyclically monotone map transporting P to Q (see [19]). In the sequel we will use the term o.t.p. for the pair (X, Y) which will also apply to the map T. For later use we summarize some properties in the following statement. The interested reader can find the proofs in Cuesta-Albertos et al. [11–13], and Tuero [22]. A different approach, involving more analytical proofs, is summarized in [24].

Proposition 2.4. *Assume $P, Q \in \mathcal{F}_2(\mathbb{R}^k)$ and $P \ll \ell^k$, and let (X, Y) be an o.t.p. for (P, Q). Then*:

(a) *The cardinal of the support of a regular conditional distribution of Y given $X = x$ is one, P-a.s.*
(b) *There exists a P-probability one set D and a Borel measurable cyclically monotone map $T : D \to \mathbb{R}^k$ such that $Y = T(X)$ a.s.*
(c) *If T is an o.t.p. for (P, Q), then T is a.e. continuous on* $\operatorname{supp}(P)$.
(d) *Assume $Q_n \in \mathcal{F}_2(\mathbb{R}^k)$ and $Q_n \to_w Q$. Let T_n be o.t.p.'s for (P, Q_n). Then $T_n \to T$, P-a.s.*

Turning back to trimmed probabilities, from Propositions 2.2 and 2.4(b) the following parametrization arises naturally.

Corollary 2.5. *If $P_0, Q \in \mathcal{F}_2(\mathbb{R}^k)$, and $P_0 \ll \ell^k$, then $\mathcal{R}_\alpha(Q) = \{P_0^* \circ T^{-1}\colon\ P_0^* \in \mathcal{R}_\alpha(P_0)\}$, where T is the* (*essentially*) *unique o.t.p. between P_0 and Q.*

Remark 2.6. *Once we have chosen a particular $P_0 \in \mathcal{F}_2(\mathbb{R}^k)$, $P_0 \ll \ell^k$, Corollary 2.5 suggests to consider the* common trimming *of probabilities*: *if $P_i \in \mathcal{F}_2(\mathbb{R}^k)$ and T_i is the o.t.p. between P_0 and P_i, $i = 1, 2$, given $P_0^* \in \mathcal{R}_\alpha(P_0)$, we say that the pair $(P_0^* \circ T_1^{-1}, P_0^* \circ T_2^{-1}) \in \mathcal{R}_\alpha(P_1) \times \mathcal{R}_\alpha(P_2)$ is a common trimming of P_1 and P_2 because it is induced by the same trimming of P_0. This suggests the consideration of a new measure of dissimilarity between P_1 and P_2 according to the shape of P_0, namely the minimal distance between common trimmings*

$$\min_{P_0^* \in \mathcal{R}_\alpha(P_0)} \mathcal{W}_2\bigl(P_0^* \circ T_1^{-1}, P_0^* \circ T_2^{-1}\bigr).$$

That was the approach adopted in [1] *for probabilities on the real line, taking the uniform law on the interval* (0, 1) *as the reference distribution.*

Proposition 2.4(d) and Corollary 2.5 allow also to show that any trimming of a weak limit of probabilities can be expressed as a weak limit of trimmings of those probabilities.

Lemma 2.7. *Assume $\alpha \in (0, 1)$. Let $Q, \{Q_n\}_n$ be probabilities on $\mathbb{R}^k$ such that $Q_n \to_w Q$. Then, if $Q^* \in \mathcal{R}_\alpha(Q)$, there exist $Q_n^* \in \mathcal{R}_\alpha(Q_n)$, $n \in \mathbb{N}$, such that $Q_n^* \to_w Q^*$.*

Proof. Fix $P \ll \ell^k$ and consider the sequence $\{T_n\}_n$ of o.t.p.'s between P and Q_n. If T is the o.t.p. between P and Q, Proposition 2.4(d) implies that $T_n \to T$, P-a.s.

By Corollary 2.5 $Q^* = P^* \circ T^{-1}$ for some $Q^* \in \mathcal{R}_\alpha(Q)$. Define now $Q_n^* = P^* \circ T_n^{-1}$. Then $Q_n^* \in \mathcal{R}_\alpha(Q_n)$ by Corollary 2.5. Since $T_n \to T$, P-a.s., and $P^* \ll P$, also $T_n \to T$, P^*-a.s. Therefore $Q_n^* = P^* \circ T_n^{-1} \to_w P^* \circ T^{-1} = Q^*$. □

2.1. *One-sided trimming*

We turn now to the optimal incomplete trasportation problems of the introduction. We consider first the one-sided case. From Definition 1.1, if $P \in \mathcal{F}_2(\mathbb{R}^k)$ and $P^* \in \mathcal{R}_\alpha(P)$ then

$$\int \|x\|^2 \,\mathrm{d}P^*(x) \le \frac{1}{1-\alpha}\int \|x\|^2 \,\mathrm{d}P(x).$$

Thus, $\mathcal{R}_\alpha(P) \subset \mathcal{F}_2(\mathbb{R}^k)$ if $P \in \mathcal{F}_2(\mathbb{R}^k)$. Our next result is a version of Proposition 2.1(c) for the metric $\mathcal{W}_2$.

Proposition 2.8. *For $\alpha \in (0,1)$ and $P \in \mathcal{F}_2(\mathbb{R}^k)$, $\mathcal{R}_\alpha(P)$ is compact in the $\mathcal{W}_2$ topology.*

Proof. Convergence in $\mathcal{W}_2$ is equivalent to weak convergence plus convergence of second moments ([3], Lemma 8.3). Since $\mathcal{R}_\alpha(P)$ is tight (Proposition 2.1(c)), given an infinite set $\mathcal{R} \subset \mathcal{R}_\alpha(P)$ we can extract a sequence $\{Q_n\}_n \subset \mathcal{R}$ that converges weakly. Let Q be its weak limit. Then $\mathcal{W}_2(Q_n, Q) \to 0$ iff $\|x\|^2$ is uniformly Q_n-integrable. Fix $t > 0$. Then

$$\int_{\|x\|>t} \|x\|^2 \,\mathrm{d}Q_n(x) = \int_{\|x\|>t} \|x\|^2 \frac{\mathrm{d}Q_n}{\mathrm{d}P}(x)\,\mathrm{d}P(x) \le \frac{1}{1-\alpha}\int_{\|x\|>t} \|x\|^2 \,\mathrm{d}P(x),$$

from which the uniform integrability of $\|x\|^2$ is immediate. □

A trivial consequence of this result is the existence of minimizers for the incomplete transportation problems (both one- and two-sided). Combined with Proposition 2.1 it also yields that similarity in the sense of (2) is equivalent to $\mathcal{W}_2(\mathcal{R}_\alpha(P_1), P_2) = 0$, while (3) holds iff $\mathcal{W}_2(\mathcal{R}_\alpha(P_1), \mathcal{R}_\alpha(P_2)) = 0$.

We will obtain uniqueness of the one-sided best trimmed approximation from strict convexity of $\mathcal{W}_2^2$. It is easy to check that $\mathcal{W}_2^2(P, Q)$ is a convex function of (P, Q), but under some smoothness, property (a) in Proposition 2.4 leads to a sharper result:

Theorem 2.9. *Consider $P_i, Q_i \in \mathcal{F}_2(\mathbb{R}^k)$ with $P_i \ll \ell^k$, $i = 1, 2$. If $Q_1 \ne Q_2$ and there is not a common o.t.p. T such that $Q_1 = P_1 \circ T^{-1}$ and $Q_2 = P_2 \circ T^{-1}$, then, for every γ in $(0,1)$,*

$$\mathcal{W}_2^2\big(\gamma P_1 + (1-\gamma)P_2, \gamma Q_1 + (1-\gamma)Q_2\big) < \gamma \mathcal{W}_2^2(P_1, Q_1) + (1-\gamma)\mathcal{W}_2^2(P_2, Q_2).$$

Proof. Write f_i for the density of P_i, and let $(X_i, T_i(X_i))$, $i = 1, 2$, be o.t.p.'s for (P_i, Q_i), $i = 1, 2$. If $P_\gamma := \gamma P_1 + (1-\gamma)P_2$ and $Q_\gamma := \gamma Q_1 + (1-\gamma)Q_2$, then $f_\gamma := \gamma f_1 + (1-\gamma) f_2$ is a density function for P_γ. Let us define on the support of P_γ the following random function:

$$T(x) = \begin{cases} T_1(x) & \text{with probability } \gamma f_1(x)/\big(\gamma f_1(x) + (1-\gamma) f_2(x)\big), \\ T_2(x) & \text{with probability } (1-\gamma) f_2(x)/\big(\gamma f_1(x) + (1-\gamma) f_2(x)\big). \end{cases}$$

If X_γ is any r.v. with probability law $\mathcal{L}(X_\gamma) = P_\gamma$, we have

$$\begin{aligned}
\nu\big[T(X_\gamma) \in A\big] &= \int \nu\big[T(X_\gamma) \in A \mid X_\gamma = x\big] P_\gamma(\mathrm{d}x) \\
&= \int I_A\big[T_1(x)\big] \frac{\gamma f_1(x)}{\gamma f_1(x) + (1-\gamma) f_2(x)} P_\gamma(\mathrm{d}x) \\
&\quad + \int I_A\big[T_2(x)\big] \frac{(1-\gamma) f_2(x)}{\gamma f_1(x) + (1-\gamma) f_2(x)} P_\gamma(\mathrm{d}x) \\
&= \gamma \int I_A\big[T_1(x)\big] f_1(x)\,\mathrm{d}x + (1-\gamma) \int I_A\big[T_2(x)\big] f_2(x)\,\mathrm{d}x \\
&= \gamma \nu\big[T_1(X_1) \in A\big] + (1-\gamma)\nu\big[T_2(X_2) \in A\big] \\
&= \gamma Q_1(A) + (1-\gamma) Q_2(A) = Q_\gamma(A).
\end{aligned}$$

Since $\mathcal{L}(T(X_\gamma)) = Q_\gamma$, by the same argument, we have

$$\begin{aligned}\mathcal{W}_2^2(P_\gamma, Q_\gamma) &\le \int \|X_\gamma - T(X_\gamma)\|^2 \, d\nu \\ &= \gamma \int \|X_1 - T_1(X_1)\|^2 \, d\nu + (1-\gamma) \int \|X_2 - T_2(X_2)\|^2 \, d\nu \\ &= \gamma \mathcal{W}_2^2(P_1, Q_1) + (1-\gamma)\mathcal{W}_2^2(P_2, Q_2).\end{aligned}$$

This shows that $\mathcal{W}_2^2(P_\gamma, Q_\gamma) < \gamma \mathcal{W}_2^2(P_1, Q_1) + (1-\gamma)\mathcal{W}_2^2(P_2, Q_2)$ unless T is an o.t.p. for (P_γ, Q_γ). If this were the case, (a) in Proposition 2.4 implies that T should be non-random, leading to

$$T(x) = \begin{cases} T_1(x) & \text{if } x \in \operatorname{supp}(P_1) \setminus \operatorname{supp}(P_2), \\ T_1(x) & (= T_2(x)) \text{ if } x \in \operatorname{supp}(P_1) \cap \operatorname{supp}(P_2), \\ T_2(x) & \text{if } x \in \operatorname{supp}(P_2) \setminus \operatorname{supp}(P_1). \end{cases}$$

But this would contradict the assumptions because it implies that T would be a common o.t.p. for (P_1, Q_1) and (P_2, Q_2). □

Taking $P_1 = P_2$ in Theorem 2.9, we obtain the strict convexity of $\mathcal{W}_2^2(P, \cdot)$.

Corollary 2.10. *Let P, Q_1, Q_2, be probability measures in $\mathcal{F}_2(\mathbb{R}^k)$. Assume $P \ll \ell^k$. If $Q_1 \neq Q_2$, then, for every γ in $(0,1)$,*

$$\mathcal{W}_2^2\big(P, \gamma Q_1 + (1-\gamma)Q_2\big) < \gamma \mathcal{W}_2^2(P, Q_1) + (1-\gamma)\mathcal{W}_2^2(P, Q_2).$$

Now, we trivially have uniqueness of the best one-sided trimmed approximation under smoothness.

Theorem 2.11. *Given $P_1, P_2 \in \mathcal{F}_2(\mathbb{R}^k)$ with $P_2 \ll \ell^k$ and $\alpha \in (0,1)$ there exists an unique $P_{1,\alpha} \in \mathcal{R}_\alpha(P_1)$ such that*

$$\mathcal{W}_2(P_{1,\alpha}, P_2) = \mathcal{W}_2\big(\mathcal{R}_\alpha(P_1), P_2\big).$$

The following example shows that uniqueness can be lost if P_2 does not have a density.

Example 2.12. *Set $P_1 = \frac{1}{2}\delta_{\{-1\}} + \frac{1}{2}\delta_{\{1\}}$ and $P_2 = \delta_{\{0\}}$. Obviously, every $P^* \in \mathcal{R}_\alpha(P_1)$ satisfies that $\mathcal{W}_2(P^*, P_2) = 1$, hence, the set of best trimmed approximations is $\mathcal{R}_\alpha(P_1)$.*

Uniqueness leads to a trivial proof of the following convergence result.

Theorem 2.13. *Consider $\{P_n\}_n, P, Q \in \mathcal{F}_2(\mathbb{R}^k)$, with $\mathcal{W}_2(P_n, P) \to 0$ and $\alpha \in (0,1)$. Then*

(a) *If $Q \ll \ell^k$ and $P_{n,\alpha} := \arg\min_{P_n^* \in \mathcal{R}_\alpha(P_n)} \mathcal{W}_2(P_n^*, Q)$, then*

$$\mathcal{W}_2(P_{n,\alpha}, P_\alpha) \to 0, \quad \text{where } P_\alpha := \operatorname*{arg\,min}_{P^* \in \mathcal{R}_\alpha(P)} \mathcal{W}_2\big(P^*, Q\big).$$

(b) *If $P \ll \ell^k$ and $Q_{n,\alpha} \in \mathcal{R}_\alpha(Q)$ satisfies that $\mathcal{W}_2(P_n, Q_{n,\alpha}) = \mathcal{W}_2(P_n, \mathcal{R}_\alpha(Q))$, then*

$$\mathcal{W}_2(Q_{n,\alpha}, Q_\alpha) \to 0, \quad \text{where } Q_\alpha := \operatorname*{arg\,min}_{Q^* \in \mathcal{R}_\alpha(Q)} \mathcal{W}_2\big(P, Q^*\big).$$

Proof. Both statements have similar proofs, so we consider only (a). By Proposition 2.1(a) the sequence $\{P_{n,\alpha}\}_n$ is tight and by the same argument as in the proof of Proposition 2.8, the function $\|x\|^2$ is uniformly integrable for $\{P_n\}_n$ thus also for $\{P_{n,\alpha}\}_n$. Therefore to show $\mathcal{W}_2(P_{n,\alpha}, P_\alpha) \to 0$ it suffices to guarantee that if $\{P_{r_n,\alpha}\}_n$ is any weakly convergent subsequence then $P_{r_n,\alpha} \to_w P_\alpha$.

By Proposition 2.1(b), if $P_{r_n,\alpha} \to_w P^*$, then $P^* \in \mathcal{R}_\alpha(P)$ and, therefore

$$\mathcal{W}_2(P_\alpha, Q) \le \mathcal{W}_2\big(P^*, Q\big) = \lim \mathcal{W}_2(P_{r_n,\alpha}, Q) \le \liminf \mathcal{W}_2\big(P^*_{r_n,\alpha}, Q\big) \tag{4}$$

for any choice $P^*_{r_n,\alpha} \in \mathcal{R}_\alpha(P_{r_n})$. Lemma 2.7 and the uniform integrability argument allow to choose this last sequence verifying $\mathcal{W}_2(P^*_{r_n,\alpha}, P_\alpha) \to 0$, hence $\mathcal{W}_2(P^*_{r_n,\alpha}, Q) \to \mathcal{W}_2(P_\alpha, Q)$, which joined with (4) and with the uniqueness of the best trimmed approximation P_α given by Theorem 2.11 shows that $P^* = P_\alpha$. □

2.2. *Two-sided trimming*

Uniqueness of the minimizers in the two-sided problem will require some additional notation and basic results. Given $v_0 \in \mathbb{R}^k$ with $\|v_0\| = 1$, we will consider H_0 an hyperplane orthogonal to v_0. The orthogonal projection on H_0 will be denoted by π_0 and for every $y \in \mathbb{R}^k$, we will denote $r_y = \langle y - \pi_0(y), v_0\rangle$. Given a measurable set $B \subset \mathbb{R}^k$, and $z \in H_0$, we will also denote

$$B_z := \big\{y \in B\colon \pi_0(y) = z\big\} \quad \text{and} \quad z_{v_0} := \{r_y\colon y \in B_z\},$$

Given the probability distribution P, we will denote with P° the marginal distribution of P on H_0 and with P_z a regular conditional distribution given z, where $z \in H_0$. This conditional probability induces in an obvious way a probability on the real line through the isometry $\mathcal{I}_z$ between $(\mathbb{R}^k)_z$ and $\mathbb{R}$, given by $y \to r_y$. This probability will be denoted λ_z and its distribution (resp. quantile) function will be denoted $F(x|z)$ (resp. $q_z(t)$). We stress on the joint measurability of these functions in the following lemma, that we include for future reference.

Lemma 2.14. *The maps $(x, z) \to F(x|z)$ and $(t, z) \to q_z(t)$ are jointly measurable in their arguments.*

Proof. Note that if $F(x, y)$ is a joint distribution function on $\mathbb{R} \times \mathbb{R}^{k-1}$ and $G(z)$ is the marginal on $\mathbb{R}^{k-1}$, then they are measurable (for probabilities supported on finite sets it is obvious and the generalization carries over through standard arguments). On the other hand, let us consider the measures η_x and μ respectively associated to the increasing functions $F(x, \cdot)$ and $G(\cdot)$. As a consequence of the Differentiation Theorem for Radon Measures (see, e.g., Sections 1.6.2 and 1.7.1 in Evans and Gariepy [14]), if we consider for any $z = (z_1, \dots, z_{k-1}) \in \mathbb{R}^{k-1}$, the sequence of rectangles $A_n(z) := \{(y_1, \dots, y_{k-1})\colon z_i - \frac{1}{n} < y_i \le z_i + \frac{1}{n}, i = 1, \dots, k-1\}$, we have the following a.s. convergence, leading to the measurability:

$$F(x|z) = \lim_{n\to\infty} \frac{\eta_x(A_n(z))}{\mu(A_n(z))}.$$

The measurability of $q_z(t)$ follows from the key property $x \le q_z(t)$ if and only if $F(x|z) \le t$. □

Our next result is a nice property of optimal incomplete transportation plans in the two-sided problem. We show that the best trimming functions are basically indicator functions of appropriate sets except for, maybe, points that remain untransported. In particular, partial trimming is impossible on $\mathrm{supp}(P)\backslash\mathrm{supp}(Q)$.

Theorem 2.15. *Consider $P, Q \in \mathcal{F}_2(\mathbb{R}^k)$ and $\alpha \in (0, 1)$. Assume that P has density f w.r.t. ℓ^k. If $P_1 \in \mathcal{R}_\alpha(P)$ and $Q_1 \in \mathcal{R}_\alpha(Q)$ satisfy*

$$\mathcal{W}_2^2(P_1, Q_1) = \mathcal{W}_2^2\big(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)\big) > 0,$$

and T is an o.t.p. for (P_1, Q_1), then $T(x) = x$ P-a.s. on the set $\mathcal{A} := \{x \in \mathbb{R}^k\colon a_1(x) \in (0, 1)\}$, where $a_1 := (1 - \alpha) f_1$ and f_1 is the density function of P_1 with respect to P.

Proof. Assume, on the contrary, that $P(\mathcal{A} \cap \{x \in \mathbb{R}^k\colon \|T(x) - x\| > 0\}) > 0$ and let us denote by $\hat{P}$ the conditional distribution of P given this set.

From (c) in Proposition 2.4 we have that T is a.e. continuous. Let x_0 be a point in the support of $\hat{P}$ in which T is continuous. Then, for every $\epsilon > 0$ there exists $\delta > 0$ such that $T(B(x_0, \delta)) \subset B(T(x_0), \epsilon)$. Let us denote $A = B(x_0, \delta) \cap \mathcal{A}$.

Let $v_0 = (T(x_0) - x_0)/\|T(x_0) - x_0\|$ and H_0 be the hyperplane orthogonal to v_0 which contains x_0. With the notation at the beginning of this subsection, taking ϵ small enough, we can assume that $m := \inf_{y \in B(T(x_0), \epsilon)} r_y$ is greater than $M := \sup_{y \in B(x_0, \delta)} r_y$. Therefore,

$$\big\|T(y) - \pi_0\big[T(y)\big]\big\| > r_y \quad \text{for every } y \in A. \tag{5}$$

On the other hand, we have

$$P(A) = \int_{H_0} P_z(A_z) P^\circ(\mathrm{d}z) = \int_{H_0} \lambda_z(z_{v_0}) P^\circ(\mathrm{d}z). \tag{6}$$

Since x_0 belongs to the support of $\hat{P}$, then $P(A) > 0$, thus

$$P^\circ\big\{z \in H_0\colon \lambda_z(z_{v_0}) > 0\big\} > 0. \tag{7}$$

Let $z \in H_0$ such that $\lambda_z(z_{v_0}) > 0$. If $y_1, y_2 \in A_z$ satisfy that $r_{y_1} < r_{y_2}$, the orthogonality between $(\pi_0(y) - x_0)$ and $(y - \pi_0(y))$ for every $y \in \mathbb{R}^k$ and (5) leads to

$$\begin{aligned}
\big\|y_1 - T(y_1)\big\|^2 &= \big\|T(y_1) - \pi_0\big[T(y_1)\big] + \pi_0(y_1) - y_1 + \pi_0\big(T(y_1)\big) - \pi_0(y_1)\big\|^2 \\
&= (r_{T(y_1)} - r_{y_1})^2 + \big\|\pi_0\big[T(y_1)\big] - z\big\|^2 \\
&> (r_{T(y_1)} - r_{y_2})^2 + \big\|\pi_0\big[T(y_1)\big] - \pi_0(y_2)\big\|^2 \\
&= \big\|y_2 - T(y_1)\big\|^2.
\end{aligned} \tag{8}$$

Now, we consider the partition of the set $A = A^- \cup A^+$ given by

$$\begin{aligned}
&A^- := \big\{y \in A\colon F\big(r_y|\pi_0(y)\big) \le 1/2\big\} \quad \text{and} \\
&A^+ := \big\{y \in A\colon F\big(r_y|\pi_0(y)\big) > 1/2\big\}.
\end{aligned}$$

From Lemma 2.14 we have that these sets are measurable. For almost every $z \in H_0$ satisfying $\lambda_z(z_{v_0}) > 0$ they define a value R_z, such that the sets

$$\begin{aligned}
&A_z^- := \{y \in A_z\colon r_y < R_z\}, \qquad A_z^+ := \big\{y \in A_z\colon r_y > R_z\big\}, \\
&z_{v_0}^- := \{r_y\colon y \in A_z^-\}, \qquad z_{v_0}^+ := \big\{r_y\colon y \in A_z^+\big\}
\end{aligned}$$

verify $\lambda_z[z_{v_0}^-] = \lambda_z[z_{v_0}^+] > 0$. Let λ_z^- and λ_z^+ be the probability λ_z conditioned to the sets $z_{v_0}^-$ and $z_{v_0}^+$ respectively, and let their corresponding distribution (resp. quantile) functions be $F^-(x|z)$ and $F^+(x|z)$ (resp. $q_z^-(t)$ and $q_z^+(t)$). Then, recalling the isometry $\mathcal{I}_z$ and the way to obtain o.t.p.'s in the real line, the map $\Gamma : A^- \to A^+$ defined by

$$\Gamma(y) = \mathcal{I}_{\pi_0(y)}^{-1}\big[q_{\pi_0(y)}^+\big[F^-\big(r_y|\pi_0(y)\big)\big]\big]$$

is an o.t.p. between P_z^- and P_z^+ for almost every $z \in H_0$ satisfying $P_z(z_{v_0}) > 0$. To end the construction, let us consider the function $a^* : \mathbb{R}^k \to \mathbb{R}$ defined as follows:

$$a^*(y) = \begin{cases} a_1(y), & \text{if } y \notin A, \\ a_1(y) - \min\big\{1 - a_1\big[\Gamma(y)\big], a_1(y)\big\}, & \text{if } y \in A^-, \\ a_1(y) + \min\big\{1 - a_1(y), a_1\big[\Gamma^{-1}(y)\big]\big\}, & \text{if } y \in A^+. \end{cases}$$

From this point, the proof involves three steps:

Step 1. *$f^* := a^*/(1-\alpha)$ is a density with respect to P that defines a probability $P^* \in \mathcal{R}_\alpha(P)$.*

Obviously $a^*(\mathbb{R}^k) \subset [0,1]$. On the other hand,

$$\int_{\mathbb{R}^k} a^*(y)P(\mathrm{d}y) = \int_{\mathbb{R}^k} a_1(y)P(\mathrm{d}y) - \int_{A^-} \min\{1 - a_1[\Gamma(y)], a_1(y)\}P(\mathrm{d}y) + \int_{A^+} \min\{1 - a_1(y), a_1[\Gamma^{-1}(y)]\}P(\mathrm{d}y). \tag{9}$$

For almost every $z \in H_0$ satisfying $P_z(A_z) > 0$, by construction, the law of a_1 under P_z^+, $P_z^+ \circ a_1^{-1}$, coincides with the law $P_z^- \circ (a_1(\Gamma))^{-1}$, while $P_z^+ \circ (a_1(\Gamma^{-1}))^{-1} = P_z^- \circ a_1^{-1}$. Therefore the last term verifies

$$\begin{aligned}
&\int_{A^+} \min\{1 - a_1(y), a_1[\Gamma^{-1}(y)]\}P(\mathrm{d}y)\\
&\quad= \int_{H_0}\left(\int_{A_z^+} \min\{1 - a_1(y), a_1[\Gamma^{-1}(y)]\}P_z(\mathrm{d}y)\right)P^\circ(\mathrm{d}z)\\
&\quad= \int_{H_0}\left(\int_{A_z^-} \min\{1 - a_1(\Gamma(y)), a_1(y)\}P_z(\mathrm{d}y)\right)P^\circ(\mathrm{d}z)\\
&\quad= \int_{A^-} \min\{1 - a_1[\Gamma(y)], a_1(y)\}P(\mathrm{d}y),
\end{aligned} \tag{10}$$

what, joined to (9) leads to $\int_{\mathbb{R}^k} a^*(y)P(\mathrm{d}y) = \int_{\mathbb{R}^k} a_1(y)P(\mathrm{d}y) = 1-\alpha$, which proves this step.

Step 2. *There exists a random map, T^*, transporting P^* to Q_1.*

Let us consider the random map T^* defined by $T^*(y) = T(y)$ on the complementary of A^+ and, for $y \in A^+$, taking the values $T(y)$ or $T[\Gamma(y)]$ with probabilities $f_1(y)/f^*(y)$ $(= a_1(y)/a^*(y))$ and $[f^*(y) - f_1(y)]/f^*(y)$ $(= [a^*(y) - a_1(y)]/a^*(y))$, respectively. These values are positive because, by construction, $a^*(y) > a_1(y)$ on A^+.

The argument to show that T^* transports P^* to Q_1 is analogous to that developed in Theorem 2.9, taking into account that $P_z^+ \circ a_1^{-1} = P_z^- \circ (a_1(\Gamma))^{-1}$.

Step 3. $\mathcal{W}_2^2(P_1, Q_1) > \mathcal{W}_2^2(P^*, Q_1)$.

By construction of T^* and inequality (8), we have

$$\begin{aligned}
\mathcal{W}_2^2(P^*, Q_1) &\le \int_{\mathbb{R}^k} \|y - T^*(y)\|^2 P^*(\mathrm{d}y)\\
&= \int_{(A^+)^c} \|y - T(y)\|^2 P^*(\mathrm{d}y)\\
&\quad+ \int_{A^+}\left(\|y - T(y)\|^2 \frac{f_1(y)}{f^*(y)} + \|y - T[\Gamma^{-1}(y)]\|^2 \frac{f^*(y) - f_1(y)}{f^*(y)}\right) f^*(y)P(\mathrm{d}y)\\
&< \int_{(A^-\cup A^+)^c} \|y - T(y)\|^2 f_1(y)P(\mathrm{d}y) + \int_{A^-} \|y - T(y)\|^2 f^*(y)P(\mathrm{d}y)\\
&\quad+ \int_{A^+}\left(\|y - T(y)\|^2 f_1(y) + \|\Gamma^{-1}(y) - T[\Gamma^{-1}(y)]\|^2 (f^*(y) - f_1(y))\right)P(\mathrm{d}y).
\end{aligned}$$

Moreover, by construction of the map Γ, recalling the relation $P_z^+ \circ (a_1(\Gamma^{-1}))^{-1} = P_z^- \circ (a_1)^{-1}$, we obtain that

$$\int_{A^+} \|\Gamma^{-1}(y) - T[\Gamma^{-1}(y)]\|^2 (f^*(y) - f_1(y))P(\mathrm{d}y) = -\int_{A^-} \|y - T(y)\|^2 (f^*(y) - f_1(y))P(\mathrm{d}y),$$

what, by construction of f^*, gives

$$\mathcal{W}_2^2\big(P^*, Q_1\big) < \mathcal{W}_2^2(P_1, Q_1),$$

contradicting the optimality of the pair (P_1, Q_1). □

We are ready now for the main result in this subsection.

Theorem 2.16 (Uniqueness). *Consider $P, Q \in \mathcal{F}_2(\mathbb{R}^k)$ with $P \ll \ell^k$ and $\alpha \in (0,1)$. If $\mathcal{W}_2^2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) > 0$, then there exists a unique pair of probabilities $P_1 \in \mathcal{R}_\alpha(P)$ and $Q_1 \in \mathcal{R}_\alpha(Q)$ such that*

$$\mathcal{W}_2^2(P_1, Q_1) = \mathcal{W}_2^2\big(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)\big). \tag{11}$$

Proof. Assume that (P_1, Q_1) and (P_2, Q_2) are two different pairs fulfilling (11), and let $a_i := (1-\alpha) f_i$, $i = 1, 2$, where f_i is the density function of P_i with respect to P. By using convex combinations $P_{\delta_i} = \delta_i P_1 + (1-\delta_i) P_2$ and $Q_{\delta_i} = \delta_i Q_1 + (1-\delta_i) Q_2, i = 1, 2$, with $\delta_1 \neq \delta_2$, from Theorem 2.9, we can assume that P_1 and P_2 have common support, and that T is the common o.t.p. for both solutions. That is, $Q_i = P_i \circ T^{-1}$, for $i = 1, 2$. Moreover, in the set $\{a_1 \neq a_2\}$ it is satisfied that $0 < a_1(y) < 1$, so that Theorem 2.15 implies that $T(x) = x$ on this set. But then it is easy to show that there exist sets $A \subset \{a_1 = a_2\}$ and $B \subset \{a_1 < a_2\}$ such that, defining

$$a^*(x) = \begin{cases} 0, & \text{if } x \in A, \\ a_2(x), & \text{if } x \in B, \\ a_1(x), & \text{if } x \notin A \cup B, \end{cases}$$

thus, $f^* := a^*/(1-\alpha)$ is the density function of a probability, say P^*, in $\mathcal{R}_\alpha(P)$, $Q^* := P^* \circ T^{-1}$ belongs to $\mathcal{R}_\alpha(Q)$ and

$$\begin{aligned} \mathcal{W}_2^2\big(P^*, Q^*\big) &= \int_{\mathbb{R}^k} \big\|x - T(x)\big\|^2 f^*(x) P(\mathrm{d}x) \\ &= \int_{\{a_1 = a_2\} - A} \big\|x - T(x)\big\|^2 f_1(y) P(\mathrm{d}x) \\ &< \int_{\{a_1 = a_2\}} \big\|x - T(x)\big\|^2 f_1(x) P(\mathrm{d}x) = \mathcal{W}_2^2(P_1, Q_1). \end{aligned}$$

□

With the uniqueness result in Theorem 2.16, the generalization of Theorem 2.13 to the two-sided case is straightforward.

Theorem 2.17. *Consider $\{P_n\}_n, \{Q_n\}_n, P, Q \in \mathcal{F}_2(\mathbb{R}^k)$, such that*

$$\mathcal{W}_2(P_n, P) \to 0, \qquad \mathcal{W}_2(Q_n, Q) \to 0 \quad \text{and} \quad P \ll \ell^k.$$

If $P_n^* \in \mathcal{R}_\alpha(P_n)$ and $Q_n^* \in \mathcal{R}_\alpha(Q_n)$ satisfy

$$\mathcal{W}_2\big(P_n^*, Q_n^*\big) = \mathcal{W}_2\big(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_n)\big),$$

then $\mathcal{W}_2(P_n^*, P^*) \to 0$ and $\mathcal{W}_2(Q_n^*, Q^*) \to 0$, where $P^* \in \mathcal{R}_\alpha(P)$, $Q^* \in \mathcal{R}_\alpha(Q)$ and $W_2(P^*, Q^*) = W_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q))$.

The Law of Large Numbers and the Glivenko–Cantelli Theorem assure (through a uniform integrability argument) that when $\{P_n^\omega\}_n$ is the sequence of empirical measures based on a sequence $\{X_n\}_n$ of i.i.d. random vectors, with law $P \in \mathcal{F}_2(\mathbb{R}^k)$, then $\mathcal{W}_2(P_n^\omega, P) \to 0$ for a.e. ω. This inmediately gives the following Law of Large Numbers for empirical best trimmed approximations. For the sake of brevity we state only the result for the two-sided case, but it can be trivially adapted to one-sided versions. These results allow the use of Monte-Carlo simulations to approximate the optimal incomplete transportation plans considered in this paper.

Theorem 2.18 (LLN for best empirical trimmings). *Let $\{X_n\}_n$, $\{Y_n\}_n$ be i.i.d. r.v.'s with $\mathcal{L}(X_n) = P$, $\mathcal{L}(Y_n) = Q$, $P, Q \in \mathcal{F}_2(\mathbb{R}^k)$, and let P_n^ω, Q_m^ω be the empirical distributions based on $\{X_1(\omega), \dots, X_n(\omega)\}$ and $\{Y_1(\omega), \dots, Y_m(\omega)\}$. Assume P or $Q \ll \ell^k$ and write*

$$(P_\alpha, Q_\alpha) := \arg\min\{\mathcal{W}_2(P^*, Q^*):\ P^* \in \mathcal{R}_\alpha(P), Q^* \in \mathcal{R}_\alpha(Q)\}.$$

If $P_{n,\alpha}^\omega \in \mathcal{R}_\alpha(P_n^\omega)$ and $Q_{m,\alpha}^\omega \in \mathcal{R}_\alpha(Q_m^\omega)$ satisfy $\mathcal{W}_2(P_{n,\alpha}^\omega, Q_{m,\alpha}^\omega) = \mathcal{W}_2(\mathcal{R}_\alpha(P_n^\omega), \mathcal{R}_\alpha(Q_m^\omega))$, and $\min(n, m) \to \infty$ then

$$(P_{n,\alpha}^\omega, Q_{m,\alpha}^\omega) \to (P_\alpha, Q_\alpha)$$

in the $\mathcal{W}_2 \times \mathcal{W}_2$ topology for almost every ω.

3. CLT for empirical trimmed distances and applications

We will show in this section the asymptotic normality of trimmed empirical distances. We restrict ourselves to the case of univariate data. A different, more involved approach can be used to prove a similar result in higher dimension and will be presented in subsequent work. Hence, we assume that P and Q are probabilities on the real line with distribution functions F and G, respectively, and write P_n and Q_m for the empirical measures based on $X_1, \dots, X_n$ (i.i.d. r.v.'s with common law P) and $Y_1, \dots, Y_m$ (i.i.d. r.v.'s with common law Q). We will consider the following technical assumptions:

$$P \text{ and } Q \text{ have finite moment of order } 4+\delta \text{ for some } \delta > 0. \tag{12}$$

F has a continuously differentiable density $F' = f$ such that

$$\sup_{t \in (0,1)} \left| \frac{t(1-t) f'(F^{-1}(t))}{f^2(F^{-1}(t))} \right| < \infty. \tag{13}$$

If Q has a density and $P_\alpha \in \mathcal{R}_\alpha(P)$ is the unique trimming of P such that $\mathcal{W}_2(P_\alpha, Q) = \mathcal{W}_2(\mathcal{R}_\alpha(P), Q)$, then the o.t.p. between P_α and Q is given by $G^{-1} \circ F_\alpha$, F_α being the distribution function associated to P_α. We set then $\varphi_{1,\alpha}(x) := x^2 - 2\tilde{\varphi}_{1,\alpha}(x)$, where $\tilde{\varphi}_{1,\alpha}$ is a primitive of $G^{-1} \circ F_\alpha$. Finally we define

$$\sigma_{1,\alpha}^2(P, Q) = \frac{1}{(1-\alpha)^2} \left(\int_0^1 \varphi_{1,\alpha}^2 \, \mathrm{d}P - \left(\int_0^1 \varphi_{1,\alpha} \, \mathrm{d}P \right)^2 \right) = \mathrm{Var}_P \left(\frac{\varphi_{1,\alpha}}{1-\alpha} \right). \tag{14}$$

Observe that $\varphi_{1,\alpha}$ is defined up to a constant, hence $\sigma_{1,\alpha}^2(P, Q)$ is well defined. Note also that $\sigma_{1,\alpha}^2(P, Q) = 0$ if $\mathcal{W}_2(\mathcal{R}_\alpha(P), Q) = 0$. Similarly, if (P_α, Q_α) is the unique pair in $\mathcal{R}_\alpha(P) \times \mathcal{R}_\alpha(Q)$ such that $\mathcal{W}_2(P_\alpha, Q_\alpha) = \mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q))$, we set $\varphi_{2,\alpha}(x) := x^2 - 2\tilde{\varphi}_{2,\alpha}(x)$, where $\tilde{\varphi}_{2,\alpha}$ is a primitive of the o.t.p. between P_α and Q_α, and define

$$\sigma_{2,\alpha}^2(P, Q) = \mathrm{Var}_P \left(\frac{\varphi_{2,\alpha}}{1-\alpha} \right). \tag{15}$$

We set $\sigma_{2,\alpha}^2(P, Q) = 0$ if $\mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) = 0$ and note that $\sigma_{2,\alpha}^2(P, Q)$ is not symmetric in P and Q.

We can state now the main result in this section.

Theorem 3.1. *If* (12) *and* (13) *hold then*

$$\sqrt{n}\left(\mathcal{W}_2^2(\mathcal{R}_\alpha(P_n), Q) - \mathcal{W}_2^2(\mathcal{R}_\alpha(P), Q)\right) \to_w N(0, \sigma_{1,\alpha}^2(P, Q)). \tag{16}$$

If Q satisfies also (13) *and $\frac{n}{n+m} \to \lambda \in (0,1)$ then*

$$\sqrt{\frac{nm}{n+m}}\left(\mathcal{W}_2^2(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_m)) - \mathcal{W}_2^2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q))\right) \to_w N(0, (1-\lambda)\sigma_{2,\alpha}^2(P, Q) + \lambda\sigma_{2,\alpha}^2(Q, P)). \tag{17}$$

Before giving a proof of Theorem 3.1 we briefly discuss its application to the assessment of the similarity models (2) and (3). Both cases can be dealt with in an analogous fashion, so we focus on the two-sided model. We recall (see the comments after Proposition 2.8) that P and Q are α-similar in the sense of (3) iff $\mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) = 0$. With the aid of Theorem 3.1, given $X_1, \dots, X_n$ i.i.d. P and $Y_1, \dots, Y_m$ i.i.d. Q we can test the related null hypotheses

$$H_1 : \mathcal{W}_2\big(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)\big) \le \Delta_0 \quad \text{vs.} \quad \mathcal{W}_2\big(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)\big) > \Delta_0,$$
$$H_2 : \mathcal{W}_2\big(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)\big) \ge \Delta_0 \quad \text{vs.} \quad \mathcal{W}_2\big(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)\big) < \Delta_0$$

for a given threshold $\Delta_0 > 0$ to be chosen by the practitioner. Observe that rejecting the null hypothesis H_2 allows us to conclude that, with high confidence, the unknown random generators P and Q are not far from similarity.

We should reject H_2 if we observe small values of $\mathcal{W}_2^2(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_m))$. For a proper choice of how small it should be for rejection, we note that the empirical version of $(1-\lambda)\sigma^2_{2,\alpha}(P, Q) + \lambda\sigma^2_{2,\alpha}(Q, P)$, namely,

$$\sigma^2_{n,m} = (1 - \lambda_{n,m})\sigma^2_{2,\alpha}(P_n, Q_m) + \lambda_{n,m}\sigma^2_{2,\alpha}(Q_m, P_n),$$

where $\lambda_{n,m} = n/(n + m)$, is a consistent estimator of the asymptotic variance (this can be proved with elementary techniques, we omit details) and, therefore,

$$Z_{n,m} = \frac{\sqrt{nm/(n+m)}}{\sigma_{n,m}}\big(\mathcal{W}_2^2\big(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_m)\big) - \Delta_0^2\big) \to_w N(0, 1)$$

under the assumptions of Theorem 3.1 if $\mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) = \Delta_0$. Clearly $Z_{n,m} \to \infty$ if $\mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) > \Delta_0$ and $Z_{n,m} \to -\infty$ if $\mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) < \Delta_0$. Thus, rejecting when

$$\mathcal{W}_2^2\big(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_m)\big) \le \Delta_0 + \frac{\sigma_{n,m}}{\sqrt{nm/(n+m)}}\Phi^{-1}(\beta),$$

where $\beta \in (0, 1)$ and Φ is the standard normal d.f., gives a test of asymptotic level β for H_2.

We show in Table 1 the results of a small simulation study about the performance of the latter test. We have generated data from two distributions differing in location ($P_1 = N(0, 1)$, $Q_1 = N(2, 1)$) or one being a contaminated version of the other ($P_2 = N(0, 1)$, $Q_2 = 0.8N(0, 1) + 0.2N(5, 1)$). We have considered different values of the sample sizes $n = m$ and different values of the threshold Δ_0^2. In each case we have generated 1000 replicates of $Z_{n,m}$ and we show the observed rejection frequencies for the above described test with a nominal level $\beta = 0.05$. We include in the upper row the value of the true trimmed distance $\mathcal{W}_2^2 = \mathcal{W}_2^2(\mathcal{R}_\alpha(P_i), \mathcal{R}_\alpha(Q_i))$. We observe that when $\Delta_0 = \mathcal{W}_2^2$ the observed frequencies show a good agreement to the nominal level $\beta = 0.05$, even very good for larger values of $\mathcal{W}_2^2$. We also see that the rejection frequency is very low when $\mathcal{W}_2^2$ exceeds the threshold, which means that, even for samples of small size, we are very unlikely to conclude approximate similarity of P and Q (rejection of H_2) when it is not close to being true.

For illustration on the extensions of Theorem 3.1 to higher dimension we include in Table 2 a second simulation study in which data come from $P = N_k(0, I_k)$, the standard k-dimensional normal law, and Q, obtained from P by a shift in location (of length 2). Here $k = 2, \dots, 6$, $n = m = 100, 200, 500$ and $\alpha = 0.05$.

In higher dimension the centering constant $\mathcal{W}_2^2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q))$ in Theorem 3.1 has to change, and $\hat{\mathcal{W}}_2^2$ denotes the bias-corrected centering. Again the nominal level is $\beta = 0.05$. We see that our conclusions are not much affected by the dimensionality, and the agreement with respect to the nominal level is good. We should say that the dimensionality does have an impact on the order of the bias, but we expect that a bootstrap correction could help to extend our procedure for testing H_2 with multivariate data.

We turn now to the proof of Theorem 3.1. With respect to the two sided case, if $\mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) = 0$ then we can take $P_0 \in \mathcal{R}_\alpha(P) \cap \mathcal{R}_\alpha(Q)$ and, then

$$\mathcal{W}_2\big(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_n)\big) \le \mathcal{W}_2\big(\mathcal{R}_\alpha(P_n), P_0\big) + \mathcal{W}_2\big(\mathcal{R}_\alpha(Q_n), P_0\big).$$

Hence, if we prove (16), the statement (17) when $\mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) = 0$ will be also proved. On the other hand, the CLT for the one-sided and the two-sided trimmed distances when $\mathcal{W}_2(\mathcal{R}_\alpha(P), \mathcal{R}_\alpha(Q)) > 0$ can be proved with

Table 1
Simulated power of $\mathcal{W}_2^2(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_n))$ for H_2 with $\beta = 0.05$

$P_1, Q_1; \alpha = 0.05$ $\mathcal{W}_2^2 \simeq 3.21$			$P_1, Q_1; \alpha = 0.1$ $\mathcal{W}_2^2 \simeq 2.64$			$P_2, Q_2; \alpha = 0.05$ $\mathcal{W}_2^2 \simeq 1.58$			$P_2, Q_2; \alpha = 0.1$ $\mathcal{W}_2^2 \simeq 0.74$		
Δ_0^2	n	Freq.	Δ_0^2	n	Freq.	Δ_0^2	n	Freq.	Δ_0^2	n	Freq.
2.50	50	0.002	2.00	50	0.005	0.50	50	0.008	0.25	50	0.052
	100	0.001		100	0.001		100	0.001		100	0.016
	200	0.000		200	0.000		200	0.000		200	0.003
	500	0.000		500	0.000		500	0.000		500	0.000
3.00	50	0.033	2.50	50	0.042	1.00	50	0.031	0.50	50	0.094
	100	0.024		100	0.040		100	0.009		100	0.055
	200	0.013		200	0.023		200	0.003		200	0.024
	500	0.006		500	0.008		500	0.000		500	0.003
3.21	50	0.058	2.64	50	0.061	1.58	50	0.083	0.74	50	0.140
	100	0.068		100	0.065		100	0.079		100	0.100
	200	0.062		200	0.066		200	0.091		200	0.096
	500	0.054		500	0.054		500	0.060		500	0.069
3.50	50	0.114	3.00	50	0.155	2.00	50	0.157	1.00	50	0.211
	100	0.155		100	0.194		100	0.193		100	0.200
	200	0.221		200	0.309		200	0.273		200	0.245
	500	0.347		500	0.511		500	0.405		500	0.326
4.00	50	0.276	3.50	50	0.357	2.50	50	0.276	1.50	50	0.346
	100	0.427		100	0.538		100	0.386		100	0.417
	200	0.686		200	0.800		200	0.583		200	0.630
	500	0.950		500	0.986		500	0.865		500	0.886

Table 2
Simulated power of $\mathcal{W}_2^2(\mathcal{R}_\alpha(P_n), \mathcal{R}_\alpha(Q_n))$ in $\mathbb{R}^k$ for H_2 with $\beta = 0.05$

$k = 2$		$k = 3$		$k = 4$		$k = 5$		$k = 6$	
Δ_0^2	Freq.	Δ_0^2	Freq.	Δ_0^2	Freq.	Δ_0^2	Freq.	Δ_0^2	Freq.
2.5	0.001	2.5	0	3.0	0	3.5	0	4.0	0
	0		0		0.001		0		0
	0		0		0		0		0
3.0	0.010	3.0	0.001	3.5	0.002	4.0	0	4.5	0
	0.008		0		0.001		0.001		0
	0.003		0.001		0.004		0.004		0.001
$\hat{\mathcal{W}}_2^2$	0.064	$\hat{\mathcal{W}}_2^2$	0.071	$\hat{\mathcal{W}}_2^2$	0.070	$\hat{\mathcal{W}}_2^2$	0.074	$\hat{\mathcal{W}}_2^2$	0.053
	0.051		0.070		0.055		0.064		0.058
	0.058		0.052		0.051		0.059		0.059
4.0	0.323	4.5	0.413	5.0	0.369	6.0	0.566	7.0	0.661
	0.568		0.781		0.841		0.976		0.992
	0.919		0.997		1		1		1
4.5	0.668	5.0	0.723	5.5	0.698	6.5	0.832	7.50	0.896
	0.912		0.983		0.984		1		0.999
	0.999		1		1		1		1

completely similar arguments, hence, for the sake of brevity we will only prove the result for the one-sided case. Our approach is based on a *empirical trimming process* that we introduce next:

$$\mathbb{V}_n(h) = \sqrt{n}\big(\mathcal{W}_2^2\big((P_n)_h, Q\big) - \mathcal{W}_2^2(P_h, Q)\big), \quad h \in \mathcal{C}_\alpha. \tag{18}$$

Here we are using the parametrization of Example 2.3. We define also

$$\mathbb{V}(h) = 2\int_0^1 \frac{B(t)}{f(F^{-1}(t))}\big(F^{-1}(t) - G^{-1}\big(h(t)\big)\big)h'(t)\,\mathrm{d}t, \quad h \in \mathcal{C}_\alpha, \tag{19}$$

where $B(t)$ is a Brownian bridge on $(0, 1)$. Note that $\{\mathbb{V}(h)\}_{h\in\mathcal{C}_\alpha}$ is a centered Gaussian process with covariance function

$$K(h_1, h_2) = 4\int_0^1 l_1(t)l_2(t)\,\mathrm{d}t - 4\int_0^1 l_1(t)\,\mathrm{d}t\int_0^1 l_2(t)\,\mathrm{d}t, \quad h_1, h_2 \in \mathcal{C}_\alpha,$$

where

$$l_i(t) = \int_{F^{-1}(1/2)}^{F^{-1}(t)} \big(x - G^{-1}\big(h_i\big(F(x)\big)\big)\big)h_i'\big(F(x)\big)\,\mathrm{d}x, \quad i = 1, 2.$$

This follows from noting that, after integration by parts, $\mathbb{V}(h_i) = -2\int_0^1 l_i(t)\,\mathrm{d}B(t)$. It is an easy exercise to show that l_i is square integrable provided (12) and (13) hold.

The main result regarding $\mathbb{V}_n$ is the following.

Theorem 3.2. *If* (12) *and* (13) *hold then* $\mathbb{V}$ *is a tight, Borel measurable map into* $\ell^\infty(\mathcal{C}_\alpha)$ *and* $\mathbb{V}_n$ *converges weakly to* $\mathbb{V}$ *in* $\ell^\infty(\mathcal{C}_\alpha)$.

Proof. We set $\rho_n(t) = \sqrt{n} f(F^{-1}(t))(F_n^{-1}(t) - F^{-1}(t))$ (the weighted quantile process). A little rewriting shows that

$$\mathbb{V}_n(h) = 2\int_0^1 \frac{\rho_n(h^{-1}(t))}{f(F^{-1}(h^{-1}(t)))}\big(F^{-1}\big(h^{-1}(t)\big) - G^{-1}(t)\big)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\int_0^1 \frac{\rho_n^2(h^{-1}(t))}{f^2(F^{-1}(h^{-1}(t)))}\,\mathrm{d}t.$$

We can assume w.l.o.g. (Theorem 6.2.1 in [10]) that $\{X_n\}_n$ are defined in a sufficiently rich probability space in which there exist Brownian bridges B_n satisfying

$$n^{1/2-\nu} \sup_{1/n\le t\le 1-1/n} \frac{|\rho_n(t) - B_n(t)|}{(t(1-t))^\nu} = \begin{cases} \mathrm{O}_P(\log n), & \text{if } \nu = 0,\\ \mathrm{O}_P(1), & \text{if } 0 < \nu \le 1/2.\end{cases} \tag{20}$$

We define also $N_n(h) = 2\int_{1/n}^{1-1/n} \frac{B_n(t)}{f(F^{-1}(t))}(F^{-1}(t) - G^{-1}(h(t)))h'(t)\,\mathrm{d}t$ and note that

$$\begin{aligned}
\|\mathbb{V}_n - N_n\|_{\mathcal{C}_\alpha} &:= \sup_{h\in\mathcal{C}_\alpha}\big|\mathbb{V}_n(h) - N_n(h)\big| \\
&\le \frac{1}{(1-\alpha)\sqrt{n}}\int_0^1 \frac{\rho_n^2(t)}{f^2(F^{-1}(t))}\,\mathrm{d}t \\
&\quad + \frac{2}{1-\alpha}\bigg(\int_0^{1/n} \frac{\rho_n^2(t)}{f^2(F^{-1}(t))}\,\mathrm{d}t\bigg)^{1/2}\bigg(\int_0^{1/n}\big(F^{-1}(t)\big)^2\,\mathrm{d}t + \int_0^{1/(1-\alpha)n}\big(G^{-1}(t)\big)^2\,\mathrm{d}t\bigg)^{1/2} \\
&\quad + \frac{2}{1-\alpha}\bigg(\int_{1-1/n}^1 \frac{\rho_n^2(t)}{f^2(F^{-1}(t))}\,\mathrm{d}t\bigg)^{1/2}\bigg(\int_{1-1/n}^1\big(F^{-1}(t)\big)^2\,\mathrm{d}t + \int_{1-1/(1-\alpha)n}^1\big(G^{-1}(t)\big)^2\,\mathrm{d}t\bigg)^{1/2} \\
&\quad + 2\sup_{h\in\mathcal{C}_\alpha}\bigg|\int_{1/n}^{1-1/n} \frac{\rho_n(t) - B_n(t)}{f(F^{-1}(t))}\big(F^{-1}(t) - G^{-1}\big(h(t)\big)\big)h'(t)\,\mathrm{d}t\bigg| \\
&=: A_{n,1} + A_{n,2} + A_{n,3} + A_{n,4}.
\end{aligned}$$

Now, (12) and Lemma 3.3 below imply $A_{n,i} \to_{\mathrm{Pr}} 0$, $i = 1, 2, 3$. From (20) we get

$$\sup_{h\in\mathcal{C}_\alpha}\bigg|\int_{1/n}^{1-1/n} \frac{\rho_n(t) - B_n(t)}{f(F^{-1}(t))} F^{-1}(t)h'(t)\,\mathrm{d}t\bigg| \le \frac{\mathrm{O}_P(1)}{n^{1/2-\nu}}\int_{1/n}^{1-1/n} \frac{(t(1-t))^\nu}{f(F^{-1}(t))}\big|F^{-1}(t)\big|\,\mathrm{d}t \tag{21}$$

for $\nu \in (0, 1/2)$. If F has finite moment of order $4 + \delta$ we can take $\nu \in (2/(4+\delta), 1/2)$, ensuring that $\int_0^1 \frac{(t(1-t))^\nu}{f(F^{-1}(t))}|F^{-1}(t)|\,\mathrm{d}t < \infty$ and, consequently that the right-hand side in (21) vanishes in probability. Similarly, using Hölder's inequality we get

$$\sup_{h\in\mathcal{C}_\alpha}\left|\int_{1/n}^{1-1/n}\frac{\rho_n(t)-B_n(t)}{f(F^{-1}(t))}G^{-1}\big(h(t)\big)h'(t)\,\mathrm{d}t\right| \tag{22}$$
$$\le \frac{\mathrm{O}_P(1)}{n^{1/2-\nu}}\left(\int_0^1 \left|G^{-1}(t)\right|^q \mathrm{d}t\right)^{1/q}\left(\int_{1/n}^{1-1/n}\frac{(t(1-t))^{p\nu}}{f^p(F^{-1}(t))}\,\mathrm{d}t\right)^{1/p}$$

for p and q such that $\frac{1}{p}+\frac{1}{q}=1$. We choose $q>4$ such that F and G have finite moment of order q. Then we will have that the right-hand side of (22) vanishes in probability if we show that

$$\frac{1}{n^{p(1/2-\nu)}}\int_{1/n}^{1-1/n}\frac{(t(1-t))^{p\nu}}{f^p(F^{-1}(t))}\,\mathrm{d}t \to 0. \tag{23}$$

But taking $\nu=\frac{1}{q}$ (which implies $\nu<\frac{1}{4}$) we obtain (23) from Lemma 3.3.

Now, combining (21), (22) and (23) we have that $A_{n,4}\to 0$ in probability. Therefore $\|\mathbb{V}_n - N_n\|_{\mathcal{C}_\alpha}\to 0$ in probability. In fact, arguing as above we can show that $\|\mathbb{V}_n - \tilde{N}_n\|_{\mathcal{C}_\alpha}\to 0$ in probability, where

$$\tilde{N}_n(h) = 2\int_0^1 \frac{B_n(t)}{f(F^{-1}(t))}\big(F^{-1}(t)-G^{-1}\big(h(t)\big)\big)h'(t)\,\mathrm{d}t.$$

Hence, to end the proof we have to show that $\mathbb{V}$ is tight, which amounts to showing that $\tilde{N}_n$ is uniformly d-equicontinuous in probability for some metric d for which $\mathcal{C}_\alpha$ is totally bounded (see Theorems 1.5.7 and 1.10.2 in [23]). We take d to be the uniform norm in $\mathcal{C}_\alpha$ (for this choice $\mathcal{C}_\alpha$ is indeed compact) and note that we have to prove that for any given $\varepsilon, \eta > 0$ there exists $\delta > 0$ such that

$$P\Big(\sup_{\|h_1-h_2\|_\infty<\delta}\big|\mathbb{V}(h_1)-\mathbb{V}(h_2)\big|>\varepsilon\Big)<\eta.$$

From Markov's inequality and compactness we see that it is enough to show that the map

$$h\mapsto E\big|\mathbb{V}(h)\big|$$

is $\|\cdot\|_\infty$-continuous. We take $h_n, h_0\in\mathcal{C}_\alpha$ such that $\|h_n-h_0\|_\infty\to 0$. After a change of variable we can write

$$\mathbb{V}(h_n) = 2\int_0^1 \frac{B(h_n^{-1}(y))}{f(F^{-1}(h_n^{-1}(y)))}\big(F^{-1}\big(h_n^{-1}(y)\big)-G^{-1}(y)\big)\,\mathrm{d}y =: \int_0^1 z_n(y)\,\mathrm{d}y.$$

From (13) we have continuity of $F^{-1}(x)/f(F^{-1}(x))$, which together with continuity of the trajectories of the Brownian bridge and the fact that $h_n^{-1}(y)\to h_0^{-1}(y)$ for almost every y shows that $z_n(y)\to z_0(y)$ for almost every $y\in(0,1)$. Continuity and the dominated convergence theorem yield that $\int_{h_n(\delta)<y\le h_n(1-\delta)} z_n(y)\,\mathrm{d}y \to \int_{h_0(\delta)<y\le h_0(1-\delta)} z_0(y)\,\mathrm{d}y$ almost surely for $\delta\in(0,\frac{1}{2})$. Note that

$$\int_{y\le h_n(\delta)}\frac{|B(h_n^{-1}(y))|}{f(F^{-1}(h_n^{-1}(y)))}\big|F^{-1}\big(h_n^{-1}(y)\big)\big|\,\mathrm{d}y \le \frac{1}{1-\alpha}\int_0^\delta \frac{|B(x)|}{f(F^{-1}(x))}\big|F^{-1}(x)\big|\,\mathrm{d}x$$

and the last upper bound can be made arbitrarily small for small δ since the fourth moment assumption on F ensures that $\frac{|B(x)|}{f(F^{-1}(x))}|F^{-1}(x)|$ is integrable. Similarly, taking $q>4$ such that both F and G have finite moment of order q and p such that $\frac{1}{p}+\frac{1}{q}=1$ we have

$$\int_{y\le h_n(\delta)}\frac{|B(h_n^{-1}(y))|}{f(F^{-1}(h_n^{-1}(y)))}\big|G^{-1}(y)\big|\,\mathrm{d}y \le \left(\int_0^\delta \frac{|B(x)|^p}{f^p(F^{-1}(x))}\,\mathrm{d}x\right)^{1/p}\left(\frac{1}{(1-\alpha)^{q-1}}\int_0^1 \left|G^{-1}(y)\right|^q \mathrm{d}y\right)^{1/q}.$$

Lemma 3.3 shows that $\frac{|B(x)|^p}{f^p(F^{-1}(x))}|F^{-1}(x)|$ is integrable and we can also control the last upper bound. With the same argument we can deal with the integral at 1 and this shows that $\mathbb{V}(h_n) \to \mathbb{V}(h_0)$ almost surely if $\|h_n - h_0\|_\infty \to 0$. $\mathbb{V}(h_n)$ being Gaussian random variables, this implies that $E|\mathbb{V}(h_n)| \to E|\mathbb{V}(h_0)|$ and completes the proof. □

The following technical result has been used in the proof of Theorem 3.2. Its proof is elementary and we omit it.

Lemma 3.3. *If F has finite moment of order 4 then*:

(i) $\sqrt{n}\int_0^{1/n}(F^{-1}(t))^2\,\mathrm{d}t \to 0$; $\sqrt{n}\int_{1-1/n}^1(F^{-1}(t))^2\,\mathrm{d}t \to 0$.

(ii) $\sqrt{n}\int_0^{1/n}(F_n^{-1}(t))^2\,\mathrm{d}t \to 0$; $\sqrt{n}\int_{1-1/n}^1(F_n^{-1}(t))^2\,\mathrm{d}t \to 0$ *in probability.*

Further, if F satisfies (13), *then*:

(iii) $\sqrt{x}\int_x^{1-x}\frac{t(1-t)}{f^2(F^{-1}(t))}\,\mathrm{d}t \to 0$ *as* $x \to 0$.

(iv) *If F has finite moment of order $q > 4$ and $\frac{1}{p} + \frac{1}{q} = 1$ then*

$$x^{p/2-p/q}\int_x^{1-x}\frac{(t(1-t))^{p/q}}{f^p(F^{-1}(t))}\,\mathrm{d}t \to 0$$

as $x \to 0$. We also have

$$\int_0^1\frac{(t(1-t))^{p/2}}{f^p(F^{-1}(t))}\,\mathrm{d}t < \infty$$

(v) $\sqrt{n}\int_0^1(F_n^{-1}(t) - F^{-1}(t))^2\,\mathrm{d}t \to 0$ *in probability.*

The proof of Theorem 3.1 will now follow easily.

Proof of Theorem 3.1. Let $P_{n,\alpha}, P_\alpha$ be the best α trimmings of P_n and P, respectively, for Q. Then $P_{n,\alpha} = (P_n)_{h_{n,\alpha}}$ and $P_\alpha = P_{h_\alpha}$ for some $h_{n,\alpha}, h_\alpha \in \mathcal{C}_\alpha$. Uniqueness of P_α and (13) ensure that h_α is also unique. From Theorem 2.18 we have $\mathcal{W}_2(P_{n,\alpha}, P_\alpha) \to 0$ which implies $\|h_{n,\alpha}\circ F_n - h_\alpha \circ F\|_\infty \to 0$ and, since $\|h_{n,\alpha} - h_\alpha\|_\infty \le \|h_{n,\alpha}\circ F_n - h_\alpha \circ F\|_\infty + \frac{1}{1-\alpha}\|F_n - F\|_\infty$ we also have $\|h_{n,\alpha} - h_\alpha\|_\infty \to 0$.

Observe next that the variance of $\mathbb{V}(h_\alpha)$ is $4(\int_0^1 l^2(t)\,\mathrm{d}t - (\int_0^1 l(t)\,\mathrm{d}t)^2)$, where, for some constants C_i

$$\begin{aligned}2l\big(F(x)\big) &= 2\int_{C_1}^x\big(y - G^{-1}\big(h_\alpha\big(F(y)\big)\big)\big)h_\alpha'\big(F(y)\big)\,\mathrm{d}y = 2\int_{C_1}^x\big(y - G^{-1}\big(F_\alpha(y)\big)\big)\frac{\mathrm{d}P_\alpha}{\mathrm{d}P}(y)\,\mathrm{d}y + C_2\\ &= \frac{2}{1-\alpha}\int_{C_1}^x\big(y - G^{-1}\big(F_\alpha(y)\big)\big)\,\mathrm{d}y + C_2,\end{aligned}$$

where the last equality follows from the fact that, if $(1-\alpha)\frac{\mathrm{d}P_\alpha}{\mathrm{d}P}(y) \in (0,1)$ then $G^{-1}(F_\alpha(y)) = y$ a.s. (this can be shown with the same proof of Theorem 2.15). As a consequence, $\mathbb{V}(h_\alpha)$ is centered normal with variance $\sigma_{1,\alpha}^2(P,Q)$.

Now,

$$\sqrt{n}\big(\mathcal{W}_2^2(P_{n,\alpha}, Q) - \mathcal{W}_2^2(P_\alpha, Q)\big) = \sqrt{n}\big(\mathcal{W}_2^2\big((P_n)_{h_{n,\alpha}}, Q\big) - \mathcal{W}_2^2\big((P_n)_{h_\alpha}, Q\big)\big) + \mathbb{V}_n(h_\alpha).$$

Hence, by Theorem 3.2 it suffices to show that $\sqrt{n}(\mathcal{W}_2^2((P_n)_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2((P_n)_{h_\alpha}, Q))$ vanishes in probability. To check this, observe that, by optimality

$$\sqrt{n}\big(\mathcal{W}_2^2\big((P_n)_{h_{n,\alpha}}, Q\big) - \mathcal{W}_2^2\big((P_n)_{h_\alpha}, Q\big)\big) \le 0. \tag{24}$$

On the other hand

$$\begin{aligned}&\sqrt{n}\big(\mathcal{W}_2^2\big((P_n)_{h_{n,\alpha}}, Q\big) - \mathcal{W}_2^2\big((P_n)_{h_\alpha}, Q\big)\big) - \sqrt{n}\big(\mathcal{W}_2^2(P_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2(P_{h_\alpha}, Q)\big)\\ &\quad = \mathbb{V}_n(h_{n,\alpha}) - \mathbb{V}_n(h_\alpha).\end{aligned}$$

Since $\|h_{n,\alpha} - h_\alpha\|_\infty \to 0$ a.s., asymptotic equicontinuity yields that $\mathbb{V}_n(h_{n,\alpha}) - \mathbb{V}_n(h_\alpha) \to 0$ in probability, while

$$\sqrt{n}\big(\mathcal{W}_2^2(P_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2(P_{h_\alpha}, Q)\big) \geq 0, \tag{25}$$

again by optimality. Thus, combining (24) and (25) we conclude that $\sqrt{n}(\mathcal{W}_2^2((P_n)_{h_{n,\alpha}}, Q) - \mathcal{W}_2^2((P_n)_{h_\alpha}, Q)) \to 0$ in probability and complete the proof. □

Acknowledgment

The authors want to thank an anonymous referee who helped to improve a previous statement of Theorem 3.1.

References

[1] P. C. Álvarez-Esteban, E. del Barrio, J. A. Cuesta-Albertos and C. Matrán. Trimmed comparison of distributions. *J. Amer. Statist. Assoc.* **103** (2008) 697–704. MR2435470
[2] L. Ambrosio. *Lecture Notes on Optimal Transport Problems, Mathematical Aspects of Evolving Interfaces. Lecture Notes in Math.* **1812**. Springer, Berlin/New York, 2003. MR2011032
[3] P. J. Bickel and D. A. Freedman. Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** (1981) 1196–1217. MR0630103
[4] Y. Brenier. Polar decomposition and increasing rearrangement of vector fields. *C. R. Acad. Sci. Paris Ser. I Math.* **305** (1987) 805–808. MR0923203
[5] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** (1991) 375–417. MR1100809
[6] L. A. Caffarelli, M. Feldman and R. J. McCann. Constructing optimal maps for Monge's transport problem as a limit of strictly convex costs. *J. Amer. Math. Soc.* **15** (2002) 1–26. MR1862796
[7] L. A. Caffarelli and R. J. McCann. Free boundaries in optimal transport and Monge–Ampére obstacle problems. *Ann. of Math.* (2006), to appear.
[8] I. J. Cascos and M. López-Díaz. Integral trimmed regions. *J. Multivariate Anal.* **96** (2005) 404–424. MR2204986
[9] I. J. Cascos and M. López-Díaz. Consistency of the α-trimming of a probability. Applications to central regions. *Bernoulli* **14** (2008) 580–592. MR2544103
[10] M. Csörgő and L. Horváth. *Weighted Approximations in Probability and Statistics.* Wiley, New York, 1993. MR1215046
[11] J. A. Cuesta-Albertos and C. Matrán. Notes on the Wasserstein metric in Hilbert spaces. *Ann. Probab.* **17** (1989) 1264–1276. MR1009457
[12] J. A. Cuesta-Albertos, C. Matrán and A. J. Tuero. Optimal transportation plans and convergence in distribution. *J. Multivariate Anal.* **60** (1997) 72–83. MR1441460
[13] J. A. Cuesta-Albertos, C. Matrán and A. J. Tuero. On the monotonicity of optimal transportation plans. *J. Math. Anal. Appl.* **215** (1997) 86–94. MR1478852
[14] L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions. Studies in Advanced Mathematics*. CRC Press, Boca Raton, FL, 1992. MR1158660
[15] M. Feldman and R. J. McCann. Uniqueness and transport density in Monge's mass transportation problem. *Calc. Var.* **15** (2002) 81–113. MR1920716
[16] A. J. Figalli. The optimal partial transport problem. *Arch. Rational Mech. Anal.* **195** (2010), 533–560. MR2592287
[17] W. Gangbo and R. J. McCann. Shape recognition via Wasserstein distance. *Quart. Appl. Math.* **58** (2000) 705–737. MR1788425
[18] A. Gordaliza. Best approximations to random variables based on trimming procedures. *J. Approx. Theory* **64** (1991) 162–180. MR1091467
[19] R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.* **80** (1995) 309–323. MR1369395
[20] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems* **2**. Springer, New York, 1998.
[21] L. Rüschendorf and S. T. Rachev. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.* **32** (1990) 48–54. MR1035606
[22] A. Tuero. On the stochastic convergence of representations based on Wasserstein metrics. *Ann. Probab.* **21** (1993) 72–85. MR1207216
[23] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, New York, 1996. MR1385671
[24] C. Villani. *Topics in Optimal Transportation. Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI, 2003. MR1964483