Central Limit Theorem and bootstrap procedure for Wasserstein's variations with application to structural relationships between distributions

Eustasio del Barrio, Hélène Lescornel, Jean-Michel Loubes

Central Limit Theorem and bootstrap procedure for Wasserstein's variations with application to structural relationships between distributions

Eustasio del Barrio[a], Hélène Lescornel[b] and Jean-Michel Loubes[b*]

[a] *IMUVA, Universidad de Valladolid* and [b] *Institut de Mathématiques de Toulouse*

November 10, 2016

Abstract

Wasserstein barycenters and variance-like criterion using Wasserstein distance are used in many problems to analyze the homogeneity of collections of distributions and structural relationships between the observations. We propose the estimation of the quantiles of the empirical process of the Wasserstein's variation using a bootstrap procedure. Then we use these results for statistical inference on a distribution registration model for general deformation functions. The tests are based on the variance of the distributions with respect to their Wasserstein's barycenters for which we prove central limit theorems, including bootstrap versions.

AMS subject classifications: Primary, 62H10; secondary,62E20
Keywords: Wasserstein distance, Wasserstein variation, Central Limit Theorem, goodness of fit.

1 Introduction

Analyzing the variability of large data sets is a difficult task when the information conveyed by the observations possesses an inner geometry far from the Euclidean one. Indeed, deformations on the data such as translations, scale location models for instance or more general warping procedures prevent the use of the usual methods in statistics. Looking for a way to measure structural relationships between data is of high importance. This kind of issues arises when considering the estimation of probability measures observed with deformations. This situation occurs often in biology, for example when considering gene expression. There has been over the last decade a large amount of work to deal with registrations issues. We refer for instance to Amit et al. (1991), Allasonnière, Amit and Trouvé (2007) or Ramsay and Silverman (2005) and references therein. However, when dealing with the registration of warped distributions, the literature is scarce. We mention here the method provided for biological computational issues known as quantile normalization in Bolstad et al. (2003), Gallón, Loubes and Maza (2013) and references therein. Recently, using optimal transport methodologies, comparisons of distributions have been studied using a notion of Fréchet mean for distributions, see for instance in Agueh and Carlier (2011) or a notion of depth as in Chernozhukov et al. (2014).

*Corresponding author: loubes@math.univ-toulouse.fr

A natural frame for applications is given by observations drawn from a deformation model in the sense that we observe J independent samples of random variables in $\mathbb{R}^d$, with sample j following distribution μ_j, such that

$$X_{i,j} = g_j(\varepsilon_{i,j}), \quad j = 1, \ldots, J, \quad i = 1 \ldots, n,$$

where $(\varepsilon_{i,j})$ are i.i.d. random variables with unknown distribution μ. The functions g_j belong to a class $\mathcal{G}$ of deformation functions, which models how the distributions μ_j's can be warped one to another by functions in the chosen class. This model is the natural extension of the functional deformation models studied in the statistical literature for which estimation procedures are provided in Gamboa, Loubes and Maza (2007) while testing issues are tackled in Collier and Dalalyan (2015). In the setup of warped distributions a main goal is the estimation of the warping functions, possibly as a first step towards registration or alignment of the (estimated) distributions. Of course, without some constraints on the class $\mathcal{G}$ the deformation model is meaningless (we can, for instance, obtain any distribution on $\mathbb{R}^d$ as a warped version of a fixed probability having a density if we take the optimal transportation map as the warping function; see Villani (2009)) and one has to consider smaller classes of deformation functions to perform a reasonable registration. In the case of parametric classes estimation of the warping functions is studied in Agulló-Antolín et al. (2015). However, estimation/registration procedures may lead to inconsistent conclusions if the chosen deformation class $\mathcal{G}$ is too small. It is, therefore, important to be able to assess fit to the deformation model given by a particular choice of $\mathcal{G}$ and this is the main goal of this paper. We note that within this framework, statistical inference on deformation models for distributions has been studied first in Freitag and Munk (2005). Here we provide a different approach which allows to deal with more general deformation classes.

The pioneer works Czado and Munk (1998) and Munk and Czado (1998) study the existence of relationships between distributions F and G by using a discrepancy measure between the distributions, $\Delta(F, G)$, built using the Wasserstein distance. The authors consider the assumption $\Delta(F, G) > \Delta_0$ versus $\Delta(F, G) \leq \Delta_0$ for Δ_0 a chosen threshold. Thus when the test is rejected, this implies that there is a statistical evidence that the two distributions are similar with respect to the chosen criterion. In this direction, we define a notion of variation of distributions using the Wasserstein distance, W_r, in the set of probability measures with finite r-th moments, $\mathcal{W}_r(\mathbb{R}^d)$, $r \geq 1$, which generalizes the notion of variance for random distributions over $\mathbb{R}^d$. This quantity can be defined as

$$V_r(\mu_1, \ldots, \mu_J) = \inf_{\eta \in \mathcal{W}_r(\mathbb{R}^d)} \left(\frac{1}{J} \sum_{j=1}^{J} W_r^r(\mu_j, \eta) \right)^{1/r},$$

which measures the spread of the distributions. Then, to measure closeness to a deformation model we take a look at the minimal variation among warped distributions, a quantity that we could consider as a minimal alignment cost. Under some mild conditions a deformation model holds if and only if this minimal alignment cost is null and we can base our assessment of a deformation model on this quantity. As in Czado and Munk (1998) and Munk and Czado (1998) we provide results (CLT's and bootstrap versions) that enable to reject that the minimal alignment cost exceeds some threshold (hence, to conclude that it is below that threshold). Our results are given in a setup of general, nonparametric classes of warping functions. If, still, one is interested in the more classical goodness-of-fit problem for the deformation model we also provide results in a somewhat more restrictive setup.

The paper is organized as follows. The main facts about Wasserstein variation are presented in Section 2, together with the key idea that fit to a deformation model can be recast in terms of the

minimal Wasserstein variation among warped versions of the distributions. Later, in Section 3 we prove some Lipsichtz bounds for the law of empirical Wasserstein variations as well as of minimal alignment costs on $\mathbb{R}^d$. The implications of these results include that quantiles of the minimal warped variation criterion can be consistently estimated by some suitable bootstrap quantiles, which can be approximated by simulation, yielding some consistent tests of fit to deformation models, provided that the empirical criterion has some regular limiting distribution. This issue, namely, Central Limit Theorems for empirical minimal Wasserstein variation is further explored for univariate distributions in Sections 4, covering non parametric deformation models, and 5, with a sharper analysis for the case of semiparametric deformation models. These sections propose consistent tests for deformation models in the corresponding setups. Finally, proofs are postponed to Section 6.

2 Wasserstein variation and deformation models for distributions

Much recent work has been conducted to measure the spread or the inner structure of a collection of distributions. In this paper we define a notion of variability which relies on the notion of Fréchet mean for the space of probability endowed with the Wasserstein metrics, of which we will recall the definition hereafter. First, for $d \geq 1$, consider the set $\mathcal{W}_r\left(\mathbb{R}^d\right)$ of probabilities with finite r-th moment. For μ and ν in $\mathcal{W}_r\left(\mathbb{R}^d\right)$, we denote by $\Pi(\mu,\nu)$ the set of all probability measures π over the product set $\mathbb{R}^d \times \mathbb{R}^d$ with first (resp. second) marginal μ (resp. ν). The L_r transportation cost between these two measures is defined as

$$\mathcal{W}_r(\mu,\nu)^r = \inf_{\pi\in\Pi(\mu,\nu)} \int \|x-y\|^r \, d\pi(x,y).$$

This transportation cost allows to endow the set $\mathcal{W}_r\left(\mathbb{R}^d\right)$ with the metric $W_r(\mu,\nu)$. More details on Wasserstein distances and their links with optimal transport problems can be found in Rachev (1984) or Villani (2009) for instance.

Within this framework, we can define a global measure of separation of a collection of measures μ_j, $j = 1,\dots,n$, as follows. Given probabilities $\mu_1,\dots,\mu_J \in \mathcal{W}_r(\mathbb{R}^d)$ let

$$V_r\left(\mu_1,\dots,\mu_J\right) = \inf_{\eta\in\mathcal{W}_r(\mathbb{R}^d)} \Big(\frac{1}{J}\sum_{j=1}^{J} W_r^r(\mu_j,\eta)\Big)^{1/r}$$

be the Wasserstein r-variation of $\mu_1,\dots,\mu_J$ or the variance of the μ_j's.

The special case $r = 2$ has been studied in the literature. Existence of a minimizer of the map $\eta \mapsto \frac{1}{J}\sum_{j=1}^{J} W_2^2(\mu_j,\eta)$ is proved in Agueh and Carlier (2011), as well as uniqueness under some smoothness assumptions. Such a minimizer, μ_B, is called a barycenter or Fréchet mean of $\mu_1,\dots,\mu_J$. Hence, $V_2\left(\mu_1,\dots,\mu_J\right) = (\frac{1}{J}\sum_{j=1}^{J} W_2^2(\mu_j,\mu_B))^{1/2}$. Empirical versions of the barycenter are analyzed in Boissard, Le Gouic and Loubes (2015) or Le Gouic, T and Loubes, J-M. (2016). Similar ideas have also been developed in Cuturi, M. and Doucet, A. (2014) or Bigot and Klein (2012).

This quantity, which is an extension of the variance for probability distributions is a good candidate to evaluate the concentration of a collection of measures around its Fréchet mean. In particular, it can be used to measure fit to a distribution deformation model. More precisely, assume as in the Introduction that we observe J independent i.i.d. samples with sample j, $j = 1,\dots,J$ consisting of i.i.d. observations $X_{i,j}$, $i = 1,\dots,n$ with common distribution μ_j. We change for later convenience the notation in the Introduction. We assume that $\mathcal{G}_j$ is a family (parametric or nonparametric) of invertible warping functions and denote $\mathcal{G} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_J$. The deformation model assumes then

that

$$\text{there exists } (\varphi_1^\star, \dots \varphi_J^\star) \in \mathcal{G} \text{ and i.i.d. } (\varepsilon_{i,j})_{\substack{1\le i\le n\\ 1\le j\le J}} \text{ such that}$$
$$X_{i,j} = \left(\varphi_j^\star\right)^{-1}(\varepsilon_{i,j}) \quad \forall 1 \le j \le J \tag{2.1}$$

Equivalently, the deformation model (2.1) means that there exist $(\varphi_1^\star, \dots \varphi_J^\star) \in \mathcal{G}$ such that $\varphi_j^\star(X_{i,j})$, $1 \le j \le J$, $1 \le i \le n$, are all i.i.d. or, if we write $\mu_j(\varphi_j)$ for the distribution of $\varphi_j(X_{i,j})$, that there exists $(\varphi_1^\star, \dots \varphi_J^\star) \in \mathcal{G}$ such that

$$\mu_1(\varphi_1^*) = \cdots = \mu_J(\varphi_J^*). \tag{2.2}$$

We propose to use the Wasserstein variation to measure fit to model ((2.1)), through the minimal alignment cost

$$A_r(\mathcal{G}) := \inf_{(\varphi_1,\dots,\varphi_J)\in\mathcal{G}} V_r^r\left(\mu_1(\varphi_1), \dots, \mu_J(\varphi_J)\right). \tag{2.3}$$

Let us assume that $\mu_1(\varphi_1), \dots, \mu_J(\varphi_J)$, $(\varphi_1, \dots, \varphi_J) \in \mathcal{G}$ are in $\mathcal{W}_r(\mathcal{R}^d)$. If the deformation model (2.1) holds then $A_r(\mathcal{G}) = 0$. Under the additional mild assumption that the minimum in (2.3) is attained we have that the deformation model can be equivalently formulated as

$$A_r(\mathcal{G}) = 0 \tag{2.4}$$

and a goodness-of-fit test to the deformation model becomes, formally, a test of

$$H_0 : A_r(\mathcal{G}) = 0 \quad \text{vs.} \quad H_a : A_r(\mathcal{G}) > 0. \tag{2.5}$$

A testing procedure can be based on the empirical version of $A_r(\mathcal{G})$, namely,

$$A_{n,r}(\mathcal{G}) := \inf_{(\varphi_1,\dots,\varphi_J)\in\mathcal{G}} V_r^r\left(\mu_{n,1}(\varphi_1), \dots, \mu_{n,J}(\varphi_J)\right), \tag{2.6}$$

where $\mu_{n,j}(\varphi_j)$ denotes the empirical measure on $\varphi_j(X_{1,j}), \dots, \varphi_j(X_{n,j})$. We would reject the deformation model (2.1) for large values of $A_{n,r}(\mathcal{G})$.

As noted in Czado and Munk (1998) or Munk and Czado (1998) the testing problem (2.5) can be considered as a mere sanity check for the deformation model, since lack of rejection of the null does not provide statistical evidence that the deformation model holds. Consequently, as in the cited references, we will also consider the alternative testing problem

$$H_0 : A_r(\mathcal{G}) \ge \Delta_0 \quad \text{vs.} \quad H_a : A_r(\mathcal{G}) < \Delta_0, \tag{2.7}$$

where $\Delta_0 > 0$ is a fixed threshold. With this formulation the test decision of rejecting the null hypothesis implies that there is statistical evidence that the deformation model is approximately true. In this case rejection would correspond to small observed values of $A_{n,r}(\mathcal{G})$. In later sections we provide theoretical results that allow the computation of approximate critical values and p-values for the testing problems (2.5) and (2.7) under suitable assumptions.

3 Bootstraping Wasserstein's variations

We present now some general results on Wasserstein distances that will be applied to estimate the asymptotic distribution of the minimal alignment cost statistic, $A_{n,r}(\mathcal{G})$, defined in (2.6). In this section, we write $\mathcal{L}(Z)$ for the law of any random variable Z. We note the abuse of notation in the following, in which W_r is used both for Wasserstein distance on $\mathbb{R}$ and on $\mathbb{R}^d$, but this should not cause much confusion.

Our first result shows that the laws of empirical transportation costs are continuous (and even Lipschitz) functions of the underlying distributions.

Theorem 3.1. *Set ν, ν', η probability measures in $\mathcal{W}_r\left(\mathbb{R}^d\right)$, $Y_1, \ldots, Y_n$ i.i.d. random vectors with common law ν, $Y_1', \ldots, Y_n'$, i.i.d. with law ν' and write ν_n, ν_n' for the corresponding empirical measures. Then*

$$W_r(\mathcal{L}(W_r(\nu_n, \eta)), \mathcal{L}(W_r(\nu_n', \eta))) \leq W_r(\nu, \nu').$$

The deformation assessment criterion introduced in section 2 is basd on the Wasserstein r-variation of distributions, V_r. It is convenient to note that $V_r^r(\nu_1, \ldots, \nu_J)$ can also be expressed as

$$V_r^r(\nu_1, \ldots, \nu_J) = \inf_{\pi \in \Pi(\nu_1, \ldots, \nu_J)} \int T(y_1, \ldots, y_J) d\pi(y_1, \ldots, y_J), \tag{3.1}$$

where $\Pi(\nu_1, \ldots, \nu_J)$ denotes the set of probability measures on $\mathbb{R}^d$ with marginals $\nu_1, \ldots, \nu_J$ and $T(y_1, \ldots, y_J) = \min_{z \in \mathbb{R}^d} \frac{1}{J} \sum_{j=1}^J \|y_j - z\|^r$.

Here we are interested in empirical Wasserstein r-variations, namely, the r-variations computed from the empirical measures $\nu_{n_j,j}$ coming from independent samples $Y_{1,j}, \ldots, Y_{n_j,j}$ of i.i.d. random variables with distribution ν_j. Note that in this case problem (3.1) is a linear optimization problem for which a minimizer always exists.

As before, we consider the continuity of the law of empirical Wasserstein r-variations with respect to the underlying probabilities. This is covered in the next result.

Theorem 3.2. *With the above notation*

$$W_r^r(\mathcal{L}(V_r(\nu_{n_1,1}, \ldots, \nu_{n_J,J})), \mathcal{L}(V_r(\nu_{n_1,1}', \ldots, \nu_{n_J,J}'))) \leq \frac{1}{J} \sum_{j=1}^J W_r^r(\nu_j, \nu_j').$$

A useful consequence of the above results is that empirical Wasserstein distances or r-variations can be bootstrapped under rather general conditions. To be more precise, we take in Theorem 3.1 $\nu' = \nu_n$, the empirical measure on $Y_1, \ldots, Y_n$ and consider a bootstrap sample $Y_1^*, \ldots, Y_{m_n}^*$ of i.i.d. (conditionally given $Y_1, \ldots, Y_n$) observations with common law ν_n. We will assume that the resampling size m_n satisfies $m_n \to \infty$, $m_n = o(n)$ and write $\nu_{m_n}^*$ for the empirical measure on $Y_1^*, \ldots, Y_{m_n}^*$ and $\mathcal{L}^*(Z)$ for the conditional law of Z given $Y_1, \ldots, Y_n$. Theorem 3.1 now reads

$$W_r(\mathcal{L}^*(W_r(\nu_{m_n}^*, \nu)), \mathcal{L}(W_r(\nu_{m_n}, \nu))) \leq W_r(\nu_n, \nu).$$

Hence, if $W_r(\nu_n, \nu) = O_{\mathbb{P}}(1/r_n)$ for some sequence $r_n > 0$ such that $r_{m_n}/r_n \to 0$ as $n \to \infty$, then, using that $W_r(\mathcal{L}(aX), \mathcal{L}(aY)) = aW_r(\mathcal{L}(X), \mathcal{L}(Y))$ for $a > 0$, we see that

$$W_r(\mathcal{L}^*(r_{m_n} W_r(\nu_{m_n}^*, \nu)), \mathcal{L}(r_{m_n} W_r(\nu_{m_n}, \nu))) \leq \frac{r_{m_n}}{r_n} r_n W_r(\nu_n, \nu) \to 0 \tag{3.2}$$

in probability.

Asume that, in addition, $r_n W_r(\nu_n, \nu) \rightharpoonup \gamma(\nu)$ for a smooth distribution $\gamma(\nu)$. Then (see, e.g., Lemma 1 in Janssen and Pauls (2003)) if $\hat{c}_n(\alpha)$ denotes the α quantile of the conditional distribution $\mathcal{L}^*(r_{m_n} W_r(\nu_{m_n}^*, \nu))$

$$\mathbb{P}\left(r_n W_r(\nu_n, \nu) \leq \hat{c}_n(\alpha)\right) \to \alpha \quad \text{as } n \to \infty. \tag{3.3}$$

We conclude in this case that the quantiles of $r_n W_r(\nu_n, \nu)$ can be consistently estimated by the bootstrap quantiles, $\hat{c}_n(\alpha)$, which, in turn, can be approximated through Monte-Carlo simulation.

As an example, if $d = 1$ and $r = 2$, under integrability and smoothness assumptions on ν we have $\sqrt{n} W_2(\nu_n, \nu) \rightharpoonup \left(\int_0^1 \frac{B^2(t)}{f^2(F^{-1}(t))} dt\right)^{1/2}$, where f and F^{-1} are the density and the quantile function of ν, see del Barrio, Giné and Utzet (2005), and (3.3) holds.

For the deformation model (2.1), statistical inference is based on $A_{n,r}(\mathcal{G})$, introduced in (2.6). Now consider $A'_{n,r}(\mathcal{G})$, the corresponding version obtained from samples with underlying distributions μ'_j. Then, a version of Theorem 3.2 is valid for these minimal alignment costs, provided the deformation classes are uniformly Lipschitz, namely, under the assumption that

$$L_j := \sup_{x \neq y, \varphi_j \in \mathcal{G}_j} \frac{\|\varphi_j(x) - \varphi_j(x)\|}{\|x - y\|}, \quad j = 1, \ldots, J \tag{3.4}$$

are finite.

Theorem 3.3. *If $L = \max(L_1, \ldots, L_j) < \infty$, with L_j as in (3.4), then*

$$W_r^r(\mathcal{L}((A_{n,r}(\mathcal{G}))^{1/r}), \mathcal{L}((A'_{n,r}(\mathcal{G}))^{1/r})) \leq L^r \frac{1}{J} \sum_{j=1}^{J} W_r^r(\mu_j, \mu'_j).$$

Hence, the Wasserstein distance of the variance of two collections of distributions can be controlled using the distance between the distributions. The main consequence of this fact is that the minimal alignment cost can be also bootstrapped as soon as a distributional limit theorem exists for $A_{n,r}(\mathcal{G})$, as in the discussion above. In sections 4 and 5 below we present distributional results of this type in the one dimensional case. We note that, while general central limit theorems for the empirical transportation cost are not available in dimension $d > 1$, some recent progress has been made in this line, see, e.g., Rippl, Munk and Sturm (2015) for Gaussian distributions and Sommerfeld and Munk (2016), which gives such type of results for distributions on $\mathbb{R}^d$ with finite support. Further advances in this line would enable to extend the results in the following section to higher dimension.

4 Assessing fit to non-parametric deformation models

We focus in this and the next sections on the case $d = 1$ and $r = 2$ and will simply write $A(\mathcal{G})$ and $A_n(\mathcal{G})$ (instead of $A_2(\mathcal{G})$ and $A_{2,n}(\mathcal{G})$) for the minimal alignment cost and its empirical version, defined in (2.3) and (2.6). Otherwise we keep the notation in section 2, with $X_{1,j}, \ldots, X_{n,j}$ i.i.d. r.v.s with law μ_j being one of the J independent samples. Now $\mathcal{G}_j$ is a class of invertible warping functions from $\mathbb{R}$ to $\mathbb{R}$ which we assume to be increasing. We note that in this case the barycenter of a set of probabilities $\mu_1, \ldots, \mu_J$ with distribution functions $F_1, \ldots, F_J$ is the probability having quantile function $F_B^{-1} := \frac{1}{J} \sum_{j=1}^{J} F_j^{-1}$, see, e.g., Agueh and Carlier (2011). We observe further that $\mu_j(\varphi_j)$ is determined by the quantile function $\varphi_j \circ F_j^{-1}$. We will write

$$F_B^{-1}(\varphi) = \frac{1}{J} \sum_{j=1}^{J} \varphi_j \circ F_j^{-1} \tag{4.1}$$

for the quantile function of the barycenter of $\mu_1(\varphi_1), \ldots, \mu_J(\varphi_J)$, while $\rightharpoonup$ will denote convergence in distribution.

In order to prove a CLT for $A_n(\mathcal{G})$ we need to make assumptions on the integrability and regularity of the distributions μ_j as well as on the smoothness of the warping functions. We consider first the assumptions on the distributions. For each μ_j, $j = 1, \ldots, J$, we denote its distribution function by F_j. We will assume that μ_j is supported on an (possibly unbounded) interval in the interior of which F_j is C^2 and $F'_j = f_j > 0$ and satisfies

$$\sup_x \frac{F_j(x)(1 - F_j(x)) f'_j(x)}{f_j(x)^2} < \infty, \tag{4.2}$$

and, further, that for some $q > 1$

$$\int_0^1 \frac{(t(1-t))^{\frac{q}{2}}}{\left(f_j\left(F_j^{-1}(t)\right)\right)^q} dt < \infty \tag{4.3}$$

and for some $r > 4$

$$\mathbb{E}\left[|X_j|^r\right] < \infty. \tag{4.4}$$

Assumption (4.2) is a classical regularity requirement for the use of strong approximations for the quantile process, as in Csörgő and Horváth (1993) or del Barrio, Giné and Utzet (2005). Our proof relies on the use of these techniques. Then (4.3) and (4.4) are mild integrability conditions. If F_j has regularly varying tails of order $-r$ (as, for instance, Pareto tails) then both conditions hold (and also (4.2)) as long as $r > 4$ and $1 < q < 2r/(r+2)$. Of course the conditions are fulfilled by distributions with lighter tails such as exponential or Gaussian laws (for any $q \in (1, 2)$).

Turning to the assumptions on the classes of warping functions, we recall that a uniform Lipsichtz condition was needed for the approximation bound in Theorem 3.3. For the CLT in this section we need some refinement of that condition, the extent of which will depend on the integrability exponent q in (4.3), as follows. We set $p_0 = \max\left(\frac{q}{q-1}, 2\right)$ and define on $\mathcal{H}_j = C^1(\mathbb{R}) \cap L^{p_0}(X_j)$ the norm $\|h_j\|_{\mathcal{H}_j} = \sup |h_j'(x)| + \mathbb{E}\left[|h_j(X_j)|^{p_0}\right]^{\frac{1}{p_0}}$, and on the product space $\mathcal{H}_1 \times \cdots \times \mathcal{H}_J$, $\|h\|_{\mathcal{H}} = \sum_{j=1}^J \|h_j\|_{\mathcal{H}_j}$ and assume that

$$\mathcal{G}_j \subset \mathcal{H}_j \text{ is compact for } \|\cdot\|_{\mathcal{H}_j} \text{ and } \sup_{h \in \mathcal{G}_j} \left|h'(x_n^h) - h'(x)\right| \underset{\sup_{h\in\mathcal{G}_j} |x_n^h - x| \to 0}{\to} 0, \tag{4.5}$$

and, finally, that for some $r > \max(4, p_0)$,

$$\mathbb{E} \sup_{h \in \mathcal{G}_j} |h(X_j)|^r < \infty. \tag{4.6}$$

We note that (4.6) is a slight strengthening of the uniform moment bound already contained in (4.5) (we could take $p_0 > \max(\frac{q}{q-1}, 4)$ in (4.5) and (4.6) would follow). Our next result gives a CLT for $A_n(\mathcal{G})$ under the assumptions on the distributions and deformation classes described above. The limit can be simply described in terms of a centered Gaussian process indexed by the set of minimizers of the variation functional, namely,

$$U(\varphi) = V_2^2(\mu_1(\varphi_1), \ldots, \mu_J(\varphi_J)).$$

An elementary computation shows that $(U^{1/2}(\varphi) - U^{1/2}(\tilde{\varphi}))^2 \leq \frac{1}{J}\sum_{j=1}^J \mathbb{E}(\varphi_j(X_j) - \tilde{\varphi}_j(X_j))^2$, from which we conclude continuity of U with respect to $\|\cdot\|_{\mathcal{H}}$. In particular, the set

$$\Gamma = \left\{\varphi \in \mathcal{G} : U(\varphi) = \inf_{\phi \in \mathcal{G}} U(\phi)\right\} \tag{4.7}$$

is a nonempty compact subset of $\mathcal{G}$.

Theorem 4.1. *Assume that $(B_j)_{1\leq j\leq J}$ are independent Brownian bridges. Set*

$$c_j(\varphi) = 2\int_0^1 \varphi_j' \circ F_j^{-1}(\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi))\frac{B_j}{f_j \circ F_j^{-1}}$$

and $C(\varphi) = \frac{1}{J}\sum_{j=1}^J c_j(\varphi)$, $\varphi \in \mathcal{G}$. Then, under assumptions (4.2) to (4.6), C is a centered Gaussian process on $\mathcal{G}$ with trajectories a.s. continuous with respect to $\|\cdot\|_{\mathcal{H}}$. Furthermore,

$$\sqrt{n}(A_n(\mathcal{G}) - A(\mathcal{G})) \rightharpoonup \min_{\varphi \in \Gamma} C(\varphi).$$

A proof of Theorem 4.1 is given in the Appendix below. The random variables $\int_0^1 \varphi_j' \circ F_j^{-1} \frac{B_j}{f_j \circ F_j^{-1}}(\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi))$ are centered Gaussian, with variance

$$\int_{[0,1]^2} (\min(s,t) - st) \frac{\varphi_j'(F_j^{-1}(t))}{f_j\big(F_j^{-1}(t)\big)} (\varphi_j(F_j^{-1}(t)) - F_B^{-1}(\varphi)(t)) \\ \times \quad \frac{\varphi_j'(F_j^{-1}(s))}{f_j\big(F_j^{-1}(s)\big)} (\varphi_j(F_j^{-1}(s)) - F_B^{-1}(\varphi)(s)) ds dt.$$

In particular, if U has a unique minimizer the limiting distribution in Theorem 4.1 is normal. However, our result works in more generality, even without uniqueness assumptions.

We remark also that although we have focused for simplicity on the case of samples of equal size, the case of different sample sizes, n_j, $j = 1, \ldots, J$, can also be handled with straightforward changes. More precisely, let us write $A_{n_1,\ldots,n_J}(\mathcal{G})$ for the minimal alignment cost computed from the empirical distribution of the samples and assume that $n_j \to +\infty$ and

$$\frac{n_j}{n_1 + \cdots + n_J} \to (\gamma_j)^2 > 0,$$

then with straightforward changes in our proof we can see that

$$\sqrt{\tfrac{n_1 \ldots n_J}{(n_1+\cdots+n_J)^{J-1}}} (A_{n_1,\ldots,n_J}(\mathcal{G}) - A(\mathcal{G})) \rightharpoonup \min_{\varphi \in \Gamma} \tilde{C}(\varphi), \tag{4.8}$$

where $\tilde{C}(\varphi) = \frac{1}{J} \sum_{j=1}^J \tilde{c}_j(\varphi)$ and $\tilde{c}_j(\varphi) = \big(\Pi_{p \neq j} \gamma_p\big) c_j(\varphi)$.

If we try, as argued in section 2, to base our assessment of fit to the deformation model (2.1) on $A_n(\mathcal{G})$, we should note that the limiting distribution in Theorem 4.1 depends on the unknown distributions μ_j and cannot be used for the computation of approximate critical values or p-values without further adjustments. We show now how this can be done in the case of the testing problem (2.7), namely, the test of

$$H_0 : A_r(\mathcal{G}) \geq \Delta_0 \quad \text{vs.} \quad H_a : A_r(\mathcal{G}) < \Delta_0,$$

for some fixed threshold $\Delta_0 > 0$, through the use of a bootstrap procedure.

Let us consider bootstrap samples $X_{1,j}^*, \ldots, X_{m_n,j}^*$ of i.i.d. observations sampled from $\mu_{n,j}$, the empirical distribution on $X_{1,j}, \ldots, X_{n,j}$. We write $\mu_{m_n,j}^*$ for the empirical measure on $X_{1,j}^*, \ldots, X_{m_n,j}^*$ and introduce

$$A_{m_n}^*(\mathcal{G}) = \inf_{\varphi \in \mathcal{G}} V_2^2(\mu_{m_n,1}^*(\varphi_1), \ldots, \mu_{m_n,J}^*(\varphi_J)).$$

Now, we base our testing procedure on the conditional α-quantiles (given the $X_{i,j}$'s) of $\sqrt{m_n}(A_{m_n}^*(\mathcal{G}) - \Delta_0)$, which we denote $\hat{c}_n(\alpha; \Delta_0)$. Our next result, which follows from Theorems 3.3 and 4.1, shows that the test that rejects H_0 when

$$\sqrt{n}(A_n(\mathcal{G}) - \Delta_0) < \hat{c}_n(\alpha; \Delta_0)$$

is a consistent test of approximate level α for (2.7). We note that the bootstrap quantiles $\hat{c}_n(\alpha; \Delta_0)$ can be computed using Monte-Carlo simulation.

Corollary 4.2. *If $m_n \to \infty$, and $m_n = O(\sqrt{n})$, then under assumptions (4.2) to (4.6)*

$$\mathbb{P}\big(\sqrt{n}(A_n(\mathcal{G}) - \Delta_0) < \hat{c}_n(\alpha; \Delta_0)\big) \to \begin{cases} 0 & \text{if } A(\mathcal{G}) > \Delta_0 \\ \alpha & \text{if } A(\mathcal{G}) = \Delta_0 \\ 1 & \text{if } A(\mathcal{G}) < \Delta_0 \end{cases} \tag{4.9}$$

Rejection in the testing problem (2.7) would result, as noted in section 2, in statistical evidence supporting that the deformation model holds approximately (hence, that related registration methods can be safely applied). If, nevertheless, we were interested in gathering statistical evidence against the deformation model then we should consider the classical goodness-of-fit problem (2.5). Some technical difficulties arise then. Note that if the deformation model holds, that is, if $A(\mathcal{G}) = 0$, then we have $\varphi_j \circ F_j^{-1} = F_B^{-1}(\varphi)$ for each $\varphi \in \Gamma$, which implies that the result of Theorem 4.1 becomes

$$\sqrt{n} A_n(\mathcal{G}) \rightharpoonup 0.$$

Hence, a nondegenerate limit law for $A_n(\mathcal{G})$ in this case requires a more refined analysis, that we handle in the next section.

5 Goodness-of-fit in semiparametric deformation models

In many cases, deformation functions can be made more specific in the sense that they follow a known shape depending on parameters that may differ for sample to sample. In our approach to the classical goodness-of-fit problem (2.5) we consider a parametric model in which $\varphi_j = \varphi_{\theta_j}$ for some finite dimensional parameter θ_j that describes the warping effect within a fixed shape. Now, that the deformation model holds means that there exist $\theta^\star = (\theta_1^\star, \ldots, \theta_J^\star)$ such that for $1 \leq i \leq n, 1 \leq j \leq J$,

$$X_{i,j} = \varphi_{\theta_j^\star}^{-1}(\varepsilon_{i,j}).$$

Hence, from now on, we will consider the following family of deformations, indexed by a parameter $\lambda \in \Lambda \subset \mathbb{R}^p$:

$$\begin{array}{rcl} \varphi : \Lambda \times \mathbb{R} & \to & \mathbb{R} \\ (\lambda, x) & \mapsto & \varphi_\lambda(x) \end{array}$$

The classes $\mathcal{G}_j$ become now $\{\varphi_{\theta_j} : \theta_j \in \Lambda\}$. We denote $\Theta = \Lambda^J$ and write $A_n(\Theta)$ and $A(\Theta)$ instead of $A_n(\mathcal{G})$ and $A(\mathcal{G})$. We also use the simplified notation $\mu_j(\theta_j)$ instead of $\mu_j\left(\varphi_{\theta_j}\right)$, $F_B(\theta)$ for $F_B(\varphi_{\theta_1}, \ldots, \varphi_{\theta_J})$ and similarly for the empirical versions. Our main goal is to prove a weak limit theorem for $A_n(\Theta)$ under the null in (2.5). Therefore, throughout this section *we assume that model (2.1) holds*. This means, in particular, that the quantile functions of the samples satisfy $F_j^{-1} = \varphi_{\theta_j^*}^{-1} \circ G^{-1}$, with G the d.f. of the $\varepsilon_{i,j}$'s. As before, we assume that the warping functions are invertible and increasing, which now means that, for each $\lambda \in \Lambda$, φ_λ is an invertible, increasing function. It is convenient at this point to introduce the notation

$$\psi_j(\lambda, x) = \varphi_\lambda(\varphi_{\theta_j^*}^{-1}(x)), \quad j = 1, \ldots, J \tag{5.1}$$

and ε for a random variable with the same distribution as the $\varepsilon_{i,j}$. Note that $\psi_j(\theta_j^*, x) = x$.

Now, under smoothness assumptions on the functions ψ_j that we present in detail below, if the parameter space is compact then the function

$$U_n(\theta_1, \ldots, \theta_J) = V_2^2(\mu_{n,1}(\theta_1), \ldots, \mu_{n,J}(\theta_J))$$

admits a minimizer, that we will denote by $\hat{\theta}_n$, that is

$$\hat{\theta}_n \in \operatorname*{argmin}_{\theta \in \Theta} U_n(\theta). \tag{5.2}$$

Of course, since we are assuming that the deformation model holds, we know that θ^* is a minimizer of

$$U(\theta_1, \ldots, \theta_J) = V_2^2(\mu_1(\theta_1), \ldots, \mu_J(\theta_J)).$$

For a closer analysis of the asymptotic behavior of $A_n(\Theta)$ under the deformation model we need to make the following identifiability assumption

$$\theta^\star \text{ belongs to the interior of } \Lambda \text{ and is the unique minimizer of } U. \tag{5.3}$$

Note that, equivalently, this means that θ^* is the unique zero of U.

As in the case of nonparametric deformation models, we need to impose some conditions on the class of warping functions and on the distribution of the errors, the $\varepsilon_{i,j}$. For the former, we write D or D_u for derivative operators with respect to parameters (hence, for instance, $D\psi_j(\lambda, x) = (D_1\psi_j(\lambda, x), \ldots, D_p\psi_j(\lambda, x))^T$ is the vector consisting of partial derivatives of ψ_j with respect to its first p arguments evaluated at (λ, x); $D^2\psi_j(\lambda, x) = (D_{u,v}\psi_j(\lambda, x))_{u,v}$ is the hessian matrix for fixed x and so on). $\psi_j'(\lambda, x)$ and similar notation will stand for derivatives with respect to x. Then we will assume that for each $j = 1, \ldots, J$, $u, v = 1, \ldots, p$, and some $r > 4$

$$\psi_j(\cdot, \cdot) \text{ is } C^2, \tag{5.4}$$

$$\mathbb{E}\big[\sup_{\lambda\in\Lambda}\big|\psi_j(\lambda, \varepsilon)\big|^r\big] < \infty, \quad \mathbb{E}\big[\sup_{\lambda\in\Lambda}\big|D_u\psi_j(\lambda, \varepsilon)\big|^r\big] < \infty, \quad \mathbb{E}\big[\sup_{\lambda\in\Lambda}\big|D_{u,v}\psi_j(\lambda, \varepsilon)\big|^r\big] < \infty, \tag{5.5}$$

and

$$\psi_j'(\cdot, \cdot) \text{ is bounded on } \Lambda \times \mathbb{R} \quad \text{and} \quad \sup_{\lambda\in\Lambda}\Big|\psi_j'(\lambda, x_n^\lambda) - \psi_j'(\lambda, x)\Big| \xrightarrow{\sup_{\lambda\in\Lambda}|x_n^\lambda - x|\to 0} 0. \tag{5.6}$$

Turning to the distribution of the errors, we will assume that G is C^2 with $G'(x) = g(x) > 0$ on some interval and

$$\sup_x \frac{G(x)\,(1 - G(x))\,g'(x)}{g(x)^2} < \infty. \tag{5.7}$$

Additionally (but see the comments after Theorem 5.1 below) we make the assumption that

$$\int_0^1 \frac{t(1-t)}{g^2\,(G^{-1}(t))} dt < \infty. \tag{5.8}$$

Finally, before stating the asymptotic result for $A_n(\Theta)$, we introduce the $p \times p$ matrices

$$\Sigma_{i,i} = \frac{2(J-1)}{J^2}\int_0^1 D_i\psi_i(\theta_i^*, G^{-1}(t))\psi_i(\theta_i^*, G^{-1}(t))^T dt,$$

$$\Sigma_{i,j} = -\frac{2}{J^2}\int_0^1 D_i\psi_i(\theta_i^*, G^{-1}(t))\psi_i(\theta_j^*, G^{-1}(t))^T dt, \quad i \neq j$$

and the $(pJ) \times (pJ)$ matrix

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,J} \\ \vdots & & \vdots \\ \Sigma_{J,1} & \cdots & \Sigma_{J,J} \end{bmatrix}. \tag{5.9}$$

Σ is a symmetric, positive semidefinite matrix. To see this, consider $x_1, \ldots, x_J \in \mathbb{R}^p$ and $x^T = [x_1^T, \ldots, x_J^T]$ and note that

$$\begin{aligned} x'\Sigma x &= \frac{2}{J^2}\int_0^1 \Big(\sum_i (J-1)(x_i \cdot D_i\psi_i(\theta_i^*, G^{-1}(t)))^2 \\ &- 2\sum_{i<j}(x_i \cdot D_i\psi_i(\theta_i^*, G^{-1}(t)))(x_j \cdot D_j\psi_j(\theta_j^*, G^{-1}(t)))\Big)dt \\ &= \frac{2}{J^2}\int_0^1 \sum_{i<j}((x_i \cdot D_i\psi_i(\theta_i^*, G^{-1}(t))) - (x_j \cdot D_j\psi_j(\theta_j^*, G^{-1}(t))))^2 dt \geq 0. \end{aligned}$$

In fact, Σ is positive definite, hence invertible, apart from some degenerate cases, For instance, if $p = 1$, Σ is invertible unless all the functions $D_i\psi_i(\theta_i^*, G^{-1}(t))$ are proportional.

We are ready now for the announced distributional limit theorem.

Theorem 5.1. *Assume that the deformation model holds. Under assumptions (5.3) to (5.7)*

$$\hat{\theta}_n \to \theta^*$$

in probability. If, in addition, Φ is invertible, then

$$\sqrt{n}(\hat{\theta}_n - \theta^\star) \rightharpoonup \Sigma^{-1}Y,$$

where $Y = (Y_1^T, \dots, Y_J^T)^T$ with

$$Y_j = \frac{2}{J}\int_0^1 D\psi_j(\theta_j^*, G^{-1}(t))\frac{\tilde{B}_j(t)}{g(G^{-1}(t))}dt,$$

$\tilde{B}_j = B_j - \frac{1}{J}\sum_{k=1}^J B_k$ and $(B_j)_{1\le j\le J}$ independent Brownian bridges. Furthermore, if (5.8) also holds, then

$$nA_n(\Theta) \rightharpoonup \frac{1}{J}\sum_{j=1}^J \int_0^1 \Big(\frac{\tilde{B}_j}{g\circ G^{-1}}\Big)^2 - \frac{1}{2}Y^T\Sigma^{-1}Y.$$

We have to make a number of comments here. First, we note that, while, for simplicity, we have formulated Theorem 5.1 assuming that the deformation model holds, the CLT for $\hat{\theta}_n$ still holds (with some additional assumptions and changes in Φ) in the case when the model is false and θ^* is not the *true* parameter, but the one that gives the best (but imperfect) alignment. Since our focus here is the assessment of the deformation models we refrain from pursuing this issue.

Our second comment is about the indentifiability condition (5.3). At first sight it can seem to be too strong to be realistic. Actually, for some deformation models it could happen that $\varphi_\theta \circ \varphi_\eta = \varphi_{\theta * \eta}$ for some $\theta * \eta \in \Theta$. In this case, if $X_{i,j} = \varphi^{-1}_{\theta_j^*}(\varepsilon_{i,j})$ with $\varepsilon_{i,j}$ i.i.d., then, for any θ, $X_{i,j} = \varphi^{-1}_{\theta * \theta_j^*}(\tilde{\varepsilon}_{i,j})$ with $\tilde{\varepsilon}_{i,j} = \varphi_\theta(\varepsilon_{i,j})$ which are also i.i.d. and, consequently, $(\theta * \theta_1^*, \dots, \theta * \theta_J^*)$ is also a zero of U. This applies, for instance, to location and scale models. A simple fix to this issue is to select one of the signals as the reference, say the J-th signal, and assume that θ_J^* is known (since it can be, in fact, chosen arbitrarily). The criterion function becomes then $\tilde{U}(\theta_1, \dots, \theta_{J-1}) = U(\theta_1, \dots, \theta_{J-1}, \theta_J^*)$. One could then make the (more realistic) assumption that $\tilde{\theta}^* = (\theta_1^*, \dots, \theta_{J-1}^*)$ is the unique zero of $\tilde{U}$ and base the analysis on $\tilde{U}_n(\theta_1, \dots, \theta_{J-1}) = U_n(\theta_1, \dots, \theta_{J-1}, \theta_J^*)$ and $\hat{\tilde{\theta}}_n = \arg\min_{\tilde{\theta}} \tilde{U}_n(\tilde{\theta})$. The results in this section can be adapted almost verbatim to this setup. In particular, $\sqrt{n}(\hat{\tilde{\theta}}_n - \tilde{\theta}^*) \rightharpoonup \tilde{\Sigma}^{-1}\tilde{Y}$, with $\tilde{Y}^T = (Y_1^T, \dots, Y_{J-1}^T)$ and $\tilde{\Sigma} = [\Sigma_{i,j}]_{1\le i,j\le J-1}$. Again, the invertibility of $\tilde{\Sigma}$ is almost granted. In fact, arguing as above, we see that and $\tilde{\Sigma}$ is positive definite if the functions $D\psi_i(\theta_i^*, G^{-1}(t))$, $i = 1, \dots, J-1$, are not null.

Next, we discuss about the smoothness and integrability conditions on the errors. As before, (5.7) is a regularity condition that enables to use strong approximations for the quantile process. One might be surprised that the moment condition (4.4) does not show up here, but in fact it is contained in (5.5) (recall that $\psi_j(\theta_j^*, x) = x$). The integrability condition (5.8) is necessary and sufficient for ensuring $\int_0^1 \frac{B(t)^2}{g^2(G^{-1}(t))}dt < \infty$ (from which we see that the limiting random variable in the last claim in Theorem 5.1 is an a.s. finite random variable) and implies that

$$nW_2^2(G_n, G) \rightharpoonup \int_0^1 \frac{B(t)^2}{g^2(G^{-1}(t))}dt,$$

with G_n the empirical d.f. on a sample of size n and d.f G. We refer to del Barrio, Giné and Utzet (2005) and Samworth and Johnson (2004) for details. Condition (4.4) is a strong assumption on the tails of G and does not include, for instance, normal distributions. On the other hand, under the less stringent condition

$$\int_0^1 \int_0^1 \frac{(s \wedge t - st)^2}{g^2(G^{-1}(s))g^2(G^{-1}(t))} dsdt < \infty, \tag{5.10}$$

which is satisfied for normal laws, it can be shown that the limit as $\delta \to 0$

$$\int_\delta^{1-\delta} \frac{B(t)^2 - t(1-t)}{g^2(G^{-1}(t))} dt,$$

exists in probability and can be expressed as a weighted sum of independent, centered χ_1^2 random variables, see del Barrio, Giné and Utzet (2005) for details. Then, denoting that kind of limits as $\int_0^1 \frac{B(t)^2 - t(1-t)}{g^2(G^{-1}(t))} dt$, under some additional tail conditions (still satisfied by normal laws; these are conditions (2.10) and (2.22) to (2.24) in the cited reference) we have

$$nW_2^2(G_n, G) - c_n \rightharpoonup \int_0^1 \frac{B(t)^2 - t(1-t)}{g^2(G^{-1}(t))} dt,$$

with $c_n = \int_{1/n}^{1-1/n} \frac{EB(t)^2}{g^2(G^{-1}(t))} dt$. A simple look at the proof of Theorem 5.1 shows that under these conditions (instead of (5.8)) we can conclude that

$$nA_n(\Theta) - \tfrac{J-1}{J^2} c_n \rightharpoonup \tfrac{1}{J} \sum_{j=1}^J \int_0^1 \frac{\tilde{B}_j^2(t) - \frac{J-1}{J} t(1-t)}{g^2(G^{-1}(t))} dt - \tfrac{1}{2} Y^T \Sigma^{-1} Y. \tag{5.11}$$

Our last comment about the assumptions for Theorem 5.1 concerns the compactness assumption on the parameter space. This may lead in some examples to artificial constraints on the parameter space. On the other hand, under some conditions (see, e.g., Corollary 3.2.3 in Van der Vaart and Wellner (1996)) it is possible to prove that the global minimizer of the empirical criterion lies in a compact neighborhood of the true minimizer. In such cases the conclusion of Theorem 5.1 would extend for the unconstrained deformation model. As a toy example consider the case of deformations by changes in scale, with $J = 2$. As above we fix the parameters of, say, the first sample, and consider the family of deformations $\varphi_\sigma(x) = \sigma x$. We assume that the deformation model holds, with the first sample having d.f. G and the second $\frac{1}{\sigma^*} G^{-1}$ (hence, σ^* is the unique minimizer of $U(\sigma)$). We obtain that $U_n(\sigma) = \frac{1}{4} \int_0^1 (F_{n,1}^{-1} - \sigma F_{n,2}^{-1})^2$, from which we see that $\hat{\sigma}_n = \big(\int F_{n,1}^{-1} F_{n,2}^{-1} \big) / \big(\int (F_{n,2}^{-1})^2 \big) \to \sigma^*$ a.s. and thus the conclusion of Theorem 5.1 remains valid if we take $\Theta = (0, \infty)$. To avoid further technicalities we prefer to think of this as a different problem that should be handled in an ad hoc way for each particular example.

Turning back to our goal of assessment of the deformation model (2.1) based on the observed value of $A_n(\Theta)$, Theorem 5.1 gives some insight into the threshold levels for rejection of the null in the testing problem (2.5). However, the limiting distribution still depends on unknown objects and designing a tractable test requires to estimate the quantiles of this distribution. This is the goal of our next result.

We consider bootstrap samples $X_{1,j}^*, \ldots, X_{m_n,j}^*$ of i.i.d. observations sampled from μ_j^n, write $\mu_{m_n,j}^*$ for the empirical measure on $X_{1,j}^*, \ldots, X_{m_n,j}^*$ and $A_{m_n}^*(\Theta)$ for the minimal alignment cost computed from the bootstrap samples. We also write $\hat{c}_n(\alpha)$ for the conditional α quantile of $m_n A_{m_n}^*(\Theta)$ given the $X_{i,j}$.

Corollary 5.2. *Assume that the semiparametric deformation models holds. If $m_n \to \infty$, and $m_n/n \to 0$, then under assumptions (5.3) to (5.8) we have that*

$$\mathbb{P}\left(nA_n(\Theta) > \hat{c}_n(1-\alpha)\right) \to \alpha. \tag{5.12}$$

Corollary 5.2 show that the test that rejects $H_0 : A(\Theta) = 0$ (which, as disussed in section 2, is true if and only if the deformation model holds) when $nA_n(\Theta) > \hat{c}_n(1-\alpha)$ is asymptotically of level α. It is easy to check that the test is consistent against alternatives that satisfy regularity and integrability assumptions as in Theorem 5.1.

The key to Corollary 5.2 is that under the assumptions a bootstrap CLT holds for $m_n A^*_{m_n}(\Theta)$. As with Theorem 5.1, the integrability conditions on the errors can be relaxed and still have a bootstrap CLT. That would be the case if we replace (5.12) by (5.10) and the additional conditions mentioned above under which (5.11) holds. Then, the further assumption that the errors have a log-concave distribution and $m_n = O(n^\rho)$ for some $\rho \in (0,1)$ would be enough to prove a bootstrap CLT, see the comments after the proof of Corollary 5.2 in the Appendix. In particular, a bootstrap CLT holds for Gaussian tails.

6 Appendix

6.1 Proofs for section 3

PROOF OF THEOREM 3.1. We set $T_n = W_r(\nu_n, \eta)$ and $T'_n = W_r(\nu'_n, \eta)$ and $\Pi_n(\eta)$ for the set of probabilities on $\{1,\dots,n\} \times \mathbb{R}^d$ with first marginal equal to the discrete uniform distribution on $\{1,\dots,n\}$ and second marginal equal to η and note that we have $T_n = \inf_{\pi \in \Pi_n(\eta)} a(\pi)$ if we denote

$$a(\pi) = \left(\int_{\{1,\dots,n\}\times\mathbb{R}^d} \|Y_i - z\|^r d\pi(i,z) \right)^{1/r}.$$

We define similarly $a'(\pi)$ from the Y'_i sample to get $T'_n = \inf_{\pi \in \Pi_n(\eta)} a'(\pi)$. But then, using the inequality $|\|a\| - \|b\|| \le \|a-b\|$,

$$|a(\pi) - a'(\pi)| \le \left(\int_{\{1,\dots,n\}\times\mathbb{R}^d} \|Y_i - Y'_i\|^r d\pi(i,z) \right)^{1/r} = \left(\frac{1}{n} \sum_{i=1}^n \|Y_i - Y'_i\|^r \right)^{1/r}$$

This implies that

$$|T_n - T'_n|^r \le \frac{1}{n} \sum_{i=1}^n \|Y_i - Y'_i\|^r.$$

If we take now (Y, Y') to be an optimal coupling of ν and ν', so that $E\left[\|Y - Y'\|^r\right] = W_r^r(\nu, \nu')$ and $(Y_1, Y'_1), \dots, (Y_n, Y'_n)$ to be i.i.d. copies of (Y, Y') we see that for the corresponding realizations of T_n and T'_n we have

$$\mathbb{E}\left[|T_n - T'_n|^r\right] \le \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[\|Y_i - Y'_i\|^r\right] = W_r(\nu, \nu')^r.$$

But this shows that $W_r(\mathcal{L}(T_n), \mathcal{L}(T'_n)) \le W_r(\nu, \nu')$, as claimed.

□

PROOF OF THEOREM 3.2. We write $V_{r,n} = V_r(\nu_{n_1,1}, \dots, \nu_{n_J,J})$ and $V'_{r,n} = V_r(\nu'_{n_1,1}, \dots, \nu'_{n_J,J})$. We note that

$$V^r_{r,n} = \inf_{\pi \in \Pi(U_1,\dots,U_J)} \int T(i_1,\dots,i_J) d\pi(i_1,\dots,i_J),$$

where U_j is the discrete uniform distribution on $\{1,\dots,n_j\}$ and $T(i_1,\dots,i_J)=\min_{z\in\mathbb{R}^d}\frac{1}{J}\sum_{j=1}^J\|Y_{i_j,j}-z\|^r$. We write $T'(i_1,\dots,i_J)$ for the equivalent function computed from the $Y'_{i,j}$'s. Hence we have

$$|T(i_1,\dots,i_J)^{1/r}-T'(i_1,\dots,i_J)^{1/r}|^r\le\frac{1}{J}\sum_{j=1}^J\|Y_{i_j,j}-Y'_{i_j,j}\|^r,$$

which implies

$$\begin{aligned}&\left|\left(\int T(i_1,\dots,i_J)d\pi(i_1,\dots,i_J)\right)^{1/r}-\left(\int T(i_1,\dots,i_J)d\pi(i_1,\dots,i_J)\right)^{1/r}\right|^r\\ &\quad\le\int\frac{1}{J}\sum_{j=1}^J\|Y_{i_j,j}-Y'_{i_j,j}\|^r d\pi(i_1,\dots,i_J)\\ &\quad=\frac{1}{J}\sum_{j=1}^J\int\|Y_{i_j,j}-Y'_{i_j,j}\|^r d\pi(i_1,\dots,i_J)=\frac{1}{J}\sum_{j=1}^J\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\|Y_{i,j}-Y'_{i,j}\|^r\right)\end{aligned}$$

So,

$$|V_{r,n}-V'_{r,n}|^r\le\frac{1}{J}\sum_{j=1}^J\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\|Y_{i,j}-Y'_{i,j}\|^r\right).$$

If we take (Y_j,Y'_j) to be an optimal coupling of ν_j and ν'_j and $(Y_{1,j},Y'_{1,j}),\dots,(Y_{n_j,j},Y'_{n_j,j})$ to be i.i.d. copies of (Y_j,Y'_j), for $j=1,\dots,J$, then we obtain

$$\mathbb{E}\left[|V_{r,n}-V'_{r,n}|^r\right]\le\frac{1}{J}\sum_{j=1}^J\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\mathbb{E}\left[\|Y_{i,j}-Y'_{i,j}\|^r\right]\right)=\frac{1}{J}\sum_{j=1}^J W_r^r(\nu_j,\nu'_j).$$

The conclusion follows.

□

Proof of Theorem 3.3. We argue as in the proof of Theorem 3.2 and write

$$A_{n,r}(\mathcal{G})=\inf_{\varphi\in\mathcal{G}}\left[\inf_{\pi\in\Pi(U_1,\dots,U_J)}\int T(\varphi;i_1,\dots,i_J)d\pi(i_1,\dots,i_J)\right],$$

where $T(\varphi;i_1,\dots,i_J)=\min_{y\in\mathbb{R}}\frac{1}{J}\sum_{j=1}^J\|Z_{i_j,j}(\varphi_j)-y\|^r$. We write $T'(\varphi;i_1,\dots,i_J)$ for the same function computed on the $Z'_{i,j}(\varphi_j)$'s. Now, from the fact $\|Z_{i,j}(\varphi_j)-Z'_{i,j}(\varphi_j)\|^r\le L^r\|X_{i,j}-X'_{i,j}\|^r$ we see that

$$|T(\varphi;i_1,\dots,i_J)^{1/r}-T'(\varphi;i_1,\dots,i_J)^{1/r}|^r\le L^r\frac{1}{J}\sum_{j=1}^J\|X_{i_j,j}-X'_{i_j,j}\|^r$$

and, as a consequence, that

$$|V_r(\mu_{n,1}(\varphi_1),\dots,\mu_{n,J}(\varphi_J))-V_r(\mu'_{n,1}(\varphi_1),\dots,\mu'_{n,J}(\varphi_J))|^r\le\frac{L^r}{J}\sum_{j=1}^J\sum_{i_j=1}^{n_j}\frac{1}{n_j}\|X_{i_j,j}-X'_{i_j,j}\|^r$$

which implies

$$|(A_{n,r}(\mathcal{G}))^{1/r}-(A'_{n,r}(\mathcal{G}))^{1/r}|^r\le\frac{L^r}{J}\sum_{j=1}^J\left(\tfrac{1}{n_j}\textstyle\sum_{i=1}^{n_j}\|X_{i,j}-X'_{i,j}\|^r\right).$$

If, as in the proof of Theorem 3.2, we assume that $(X_{i,j}, X'_{i,j})$, $i = 1, \dots, n_j$ are i.i.d. copies of an optimal coupling for μ_j and μ'_j, with different samples independent from each other we obtain that

$$\mathbb{E}\left[|(A_{n,r}(\mathcal{G}))^{1/r} - (A'_{n,r}(\mathcal{G}))^{1/r}|^r\right] \leq \frac{L^r}{J}\sum_{j=1}^{J} W_r^r(\mu_j, \mu'_j).$$

□

6.2 Proofs for sections 4 and 5

We provide here proofs of the main results in sections 4 and 5. Our approach relies on the consideration the processes

$$C_n(\varphi) = \sqrt{n}(U_n(\varphi) - U(\varphi)) \quad \text{and} \quad C(\varphi) = \frac{1}{J}\sum_{j=1}^{J} c_j(\varphi), \quad \varphi \in \mathcal{G}, \tag{6.1}$$

where $U_n(\varphi) = V_2^2(\mu_{n,1}(\varphi_1), \dots, \mu_{n,J}(\varphi_J))$, $U(\varphi) = V_2^2(\mu_1(\varphi_1), \dots, \mu_J(\varphi_J))$,

$$c_j(\varphi) = 2\int_0^1 \varphi'_j \circ F_j^{-1}(\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi))\frac{B_j}{f_j \circ F_j^{-1}}$$

and $(B_j)_{1\leq j\leq J}$ are independent standard Brownian bridges on $(0,1)$. We prove below that the empirical deformation cost process C_n converges weakly to C as random elements in $L^\infty(\mathcal{G})$, the space of bounded, real valued functions on $\mathcal{G}$. Theorem 4.1 will follow as a corollary of this result.

We will make frequent use in this section of the following technical Lemma, which follows easily from the triangle and Holder's inequalities. We omit the proof.

Lemma 6.1. *Under Assumption (4.6)*

i) $\sup_{\varphi_j\in\mathcal{G}_j} \sqrt{n}\int_0^{\frac{1}{n}} (\varphi_j \circ F_j^{-1})^2 \to 0$, $\sup_{\varphi_j\in\mathcal{G}_j} \sqrt{n}\int_{1-\frac{1}{n}}^1 (\varphi_j \circ F_j^{-1})^2 \to 0$.

ii) $\sup_{\varphi_j\in\mathcal{G}_j} \sqrt{n}\int_0^{\frac{1}{n}} (\varphi_j \circ F_{n,j}^{-1})^2 \to 0$, $\sup_{\varphi_j\in\mathcal{G}_j} \sqrt{n}\int_{1-\frac{1}{n}}^1 (\varphi_j \circ F_{n,j}^{-1})^2 \to 0$ *in probability.*

iii) If moreover (4.3) holds then for all $1 \leq j, k \leq J$

$$\int_0^1 \frac{\sqrt{t(1-t)}}{f_k(F_k^{-1}(t))} \sup_{\varphi_j\in\mathcal{G}_j} \left|\varphi_j(F_j^{-1}(t))\right| dt < \infty \tag{6.2}$$

Theorem 6.2. *Under assumptions (4.2) to(4.6)* C_n *and* C *have a.s. trajectories in* $L^\infty(\mathcal{G})$*. Furthermore,* C *is a tight Gaussian random elemnt and* C_n *converges weakly to* C *in* $L^\infty(\mathcal{G})$*.*

Proof. We start noting that $U_n(\varphi) = \frac{1}{J}\sum_{j=1}^J \int_0^1 (\varphi_j \circ F_{n,j}^{-1} - F_{n,B}^{-1}(\varphi))^2$ and $U(\varphi) = \frac{1}{J}\sum_{j=1}^J \int_0^1 (\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi))^2$ with $F_{n,B}^{-1}(\varphi) = \frac{1}{J}\sum_{j=1}^J \varphi_j \circ F_{n,j}^{-1}$, $F_B^{-1}(\varphi) = \frac{1}{J}\sum_{j=1}^J \varphi_j \circ F_j^{-1}$. Now, (4.6) implies that $\sup_{\varphi_j\in\mathcal{G}_j} \int_0^1 (\varphi_j \circ F_j^{-1})^2 < \infty$. Similarly, assumption (4.5) implies $K_j := \sup_{\varphi_j\in\mathcal{G}_j, x\in(c_j,d_j)} |\varphi'_j(x)| < \infty$. Noting that $\int_0^1 (\varphi_j \circ F_{n,j}^{-1})^2 \leq 2\int_0^1 (\varphi_j \circ F_j^{-1})^2 + 2K_j^2 \int_0^1 (F_{n,j}^{-1} - F_j^{-1})^2$, we see that $\sup_{\varphi_j\in\mathcal{G}_j} \int_0^1 (\varphi_j \circ$

$F_{n,j}^{-1})^2 < \infty$ a.s. and, with little additional effort, conclude that C_n has a.s. bounded trajectories. On the other hand, writing $d_{j,k}(\varphi) = \int_0^1 \varphi_j' \circ F_j^{-1} \frac{B_j}{f_j \circ F_j^{-1}} \varphi_k \circ F_k^{-1}$ we see that for $\varphi, \rho \in \mathcal{G}$

$$
\begin{aligned}
|d_{j,k}(\varphi) - d_{j,k}(\rho)| &\leq \|\varphi_j' - \rho_j'\|_\infty \Big| \int_0^1 \frac{B_k}{f_k \circ F_k^{-1}} \varphi_k \circ F_k^{-1} \Big| \\
&+ \Big| \int_0^1 \rho_j' \circ F_j^{-1} \frac{B_k}{f_k \circ F_k^{-1}} (\varphi_k \circ F_k^{-1} - \rho_k \circ F_k^{-1}) \Big| \\
&\leq \|\varphi_j' - \rho_j'\|_\infty \sup_{\varphi_k \in \mathcal{G}_k} \Big| \int_0^1 \frac{B_k}{f_k \circ F_k^{-1}} \varphi_k \circ F_k^{-1} \Big| \\
&+ \sup_{(c_j, d_j)} |\rho_j'| \Big(\int_0^1 \big| \tfrac{B_k}{f_k \circ F_k^{-1}} \big|^q \Big)^{1/q} \Big(\int_0^1 |\varphi_k \circ F_k^{-1} - \rho_k \circ F_k^{-1}|^{p_0} \Big)^{1/p_0}
\end{aligned}
$$

But using iii) of Lemma 6.1

$$
\mathbb{E}\Big[\sup_{\varphi_k \in \mathcal{G}_k} \Big| \int_0^1 \frac{B_k}{f_k \circ F_k^{-1}} \varphi_k \circ F_k^{-1} \Big| \Big] \leq \int_0^1 \frac{\sqrt{t(1-t)}}{f_k(F_k^{-1}(t))} \sup_{\varphi_j \in \mathcal{G}_j} |\varphi_j(F_j^{-1}(t))| dt < \infty.
$$

Hence, almost surely, $\sup_{\varphi \in \mathcal{G}} \Big| \int_0^1 \frac{B_j}{f_j \circ F_j^{-1}} \varphi_j \circ F_j^{-1} \Big| < \infty$. Furthermore, from assumption (4.3), we get that, a.s., $\int_0^1 \big(\frac{B_j}{f_j \circ F_j^{-1}} \big)^q < \infty$ and thus, for some a.s. finite random variable T, $|d_{j,k}(\varphi) - d_{j,k}(\rho)| \leq T \|\varphi - \rho\|_{\mathcal{G}}$ for $\varphi, \rho \in \mathcal{G}$. From this conclude that the trajectories of C are a.s. bounded, uniformly continuous functions on $\mathcal{G}$, endowed with the norm $\|\cdot\|_{\mathcal{G}}$ introduced in (4.5). In particular, C is a tight random element in $L^\infty(\mathcal{G})$, see, e.g., p. 39-41 in Van der Vaart and Wellner (1996).

From this point we pay attention to the quantile processes, namely,

$$
\rho_{n,j}(t) = \sqrt{n} f_j(F_j^{-1}(t))(F_{n,j}^{-1}(t) - F_j^{-1}(t)), \quad 0 < t < 1, \quad j = 1, \ldots, J.
$$

A trivial adaptation of Theorem 2.1, p. 381 in Csörgő and Horváth (1993) shows that, under (4.2), there exist, on a rich enough probability space, independent versions of $\rho_{n,j}$ and independent families of Brownian bridges $\{B_{n,j}\}_{n=1}\infty$, $j = 1, \ldots, J$, satisfying

$$
n^{1/2-\nu} \sup_{1/n \leq t \leq 1-1/n} \frac{|\rho_{n,j}(t) - B_{n,j}(t)|}{(t(1-t))^\nu} = \begin{cases} O_p(\log(n)) \text{ if } \nu = 0 \\ O_p(1) \text{ if } 0 < \nu \leq 1/2 \end{cases} \tag{6.3}
$$

We work, without loss of generality, with these versions of $\rho_{n,j}$ and $B_{n,j}$. We show now that

$$
\sup_{\varphi \in \mathcal{G}} \Big| C_n(\varphi) - \hat{C}_n(\varphi) \Big| \to 0 \text{ in probability} \tag{6.4}
$$

with $\hat{C}_n(\varphi) = \frac{1}{J} \sum_{j=1}^J c_{n,j}(\varphi)$ and $c_{n,j}(\varphi) = 2 \int_0^1 \varphi_j' \circ F_j^{-1} (\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)) \frac{B_{n,j}}{f_j \circ F_j^{-1}}$. To check this we note that some simple algebra yields $C_n(\varphi) = \frac{2}{J} \sum_{j=1}^J \tilde{c}_{n,j} + \frac{1}{J} \sum_{j=1}^J \tilde{r}_{n,j}$ with

$$
\tilde{c}_{n,j} = \sqrt{n} \int_0^1 (\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1})(\varphi_j \circ F_j^{-1} - F_B^{-1}(\varphi)),
$$

$$
\tilde{r}_{n,j} = \sqrt{n} \int_0^1 [(\varphi_j \circ F_{n,j}^{-1} - \varphi_j \circ F_j^{-1}) - (F_{n,B}^{-1}(\varphi) - F_B^{-1}(\varphi))]^2.
$$

From the elementary inequality $(a_1+\cdots+a_J)^2 \le Ja_1^2+\cdots+Ja_J^2$ we get that

$$\frac{1}{J}\sum_{j=1}^{J}\tilde{r}_{n,j} \le \frac{4\sqrt{n}}{J}\sum_{j=1}^{J}\int_0^1(\varphi_j\circ F_{n,j}^{-1}-\varphi_j\circ F_j^{-1})^2 \le \frac{4\sqrt{n}}{J}\sum_{j=1}^{J}K_j\int_0^1(F_{n,j}^{-1}-F_j^{-1})^2,$$

with $K_j := \sup_{\varphi_j\in\mathcal{G}_j, x\in(c_j,d_j)}|\varphi_j'(x)| < \infty$, as above. Now we can use (4.4) and argue as in the proof of Theorem 2 in Álvarez-Esteban et al. (2008) to conclude that $\sqrt{n}\int_0^1(F_{n,j}^{-1}-F_j^{-1})^2 \to 0$ in probability and, as a consequence, that

$$\sup_{\varphi\in\mathcal{G}}\Big|C_n(\varphi)-\frac{1}{J}\sum_{j=1}^{J}\tilde{c}_{n,j}\,(\varphi)\,\Big| \to 0 \text{ in probability.} \tag{6.5}$$

On the other hand, the Cauchy-Schwarz's inequality shows that

$$\begin{aligned} n\Big(\int_0^{\frac{1}{n}}&(\varphi_j\circ F_{n,j}^{-1}-\varphi_j\circ F_j^{-1})(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\Big)^2 \\ &\le \sqrt{n}\int_0^{\frac{1}{n}}(\varphi_j\circ F_{n,j}^{-1}-\varphi_j\circ F_j^{-1})^2\sqrt{n}\int_0^{\frac{1}{n}}(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))^2 \end{aligned}$$

and using i) and ii) of Lemma 6.1, the two factors converge to zero uniformly in φ. A similar argument works for the upper tail and allows to conclude that we can replace in (6.5) $\tilde{c}_{n,j}(\varphi)$ with $\tilde{\tilde{c}}_{n,j}(\varphi) := 2\sqrt{n}\int_{\frac{1}{n}}^{1-\frac{1}{n}}(\varphi_j\circ F_{n,j}^{-1}-\varphi_j\circ F_j^{-1})(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))$. Moreover,

$$\sup_{\varphi\in\mathcal{G}}\Big|\int_0^{\frac{1}{n}}\varphi_j'\circ F_j^{-1}\frac{B_{n,j}}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\Big| \le K_j\int_0^{\frac{1}{n}}\Big|\frac{B_{n,j}}{f_j\circ F_j^{-1}}\Big|\sup_{\varphi\in\mathcal{G}}\big|(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\big|$$

and by iii) of Lemma 6.1 and Cauchy-Schwarz's inequality

$$\mathbb{E}\Big[\int_0^{\frac{1}{n}}\Big|\frac{B_{n,j}}{f_j\circ F_j^{-1}}\Big|\sup_{\varphi\in\mathcal{G}}\big|(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\big|\Big] \le \int_0^{\frac{1}{n}}\frac{\sqrt{t(1-t)}}{f_j(F_j^{-1}(t))}\sup_{\varphi\in\mathcal{G}}\big|\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t)\big|dt\to 0.$$

Hence, $\sup_{\varphi\in\mathcal{G}}\Big|\int_0^{\frac{1}{n}}\varphi_j'\circ F_j^{-1}\frac{B_{n,j}}{f_j\circ F_j^{-1}}(\varphi_j\circ F_j^{-1}-F_B^{-1}(\varphi))\Big|\to 0$ in probability and similarly for the right tail. Now, for every $t\in(0,1)$ we have

$$\varphi_j\circ F_{n,j}^{-1}(t)-\varphi_j\circ F_j^{-1}(t) = \varphi_j'(K_{n,\varphi_j}(t))(F_{n,j}^{-1}(t)-F_j^{-1}(t)) \tag{6.6}$$

for some $K_{n,\varphi_j}(t)$ between $F_{n,j}^{-1}(t)$ and $F^{-1}(t)$. Therefore, (recall (6.6)), to prove (6.4) it suffices to show that

$$\begin{aligned} \sup_{\varphi\in\mathcal{G}}\Big|\int_{\frac{1}{n}}^{1-\frac{1}{n}}&\varphi_j'(F_j^{-1}(t))\frac{B_{n,j}(t)}{f_j(F_j^{-1}(t))}(\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t))dt \\ -\int_{\frac{1}{n}}^{1-\frac{1}{n}}&\varphi_j'(K_{n,\varphi_j}(t))\frac{\rho_{n,j}(t)}{f_j(F_j^{-1}(t))}(\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t))dt\Big| \to 0 \end{aligned} \tag{6.7}$$

in probability. To check it we take $\nu\in(0,1/2)$ in (6.3) to get

$$\begin{aligned} &\int_{\frac{1}{n}}^{1-\frac{1}{n}}\frac{|\rho_{n,j}(t)-B_{n,j}(t)|}{f_j(F_j^{-1}(t))}\sup_{\varphi\in\mathcal{G}}\big|\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t)\big|dt \\ &\quad\le n^{\nu-\frac{1}{2}}O_P(1)\int_{\frac{1}{n}}^{1-\frac{1}{n}}\frac{(t(1-t))^\nu}{f_k(F_k^{-1}(t))}\sup_{\varphi\in\mathcal{G}}\big|\varphi_j(F_j^{-1}(t))-F_B^{-1}(\varphi)(t)\big|dt\to 0 \end{aligned} \tag{6.8}$$

in probability (using dominated convergence and iii) of Lemma 6.1). We observe next that, for each $t \in (0,1)$, $\sup_{\varphi_j \in \mathcal{G}_j} |K_{n,\varphi_j}(t) - F_j^{-1}(t)| \to 0$ a.s., since $K_{n,\varphi_j}(t)$ lies between $F_{n,j}^{-1}(t)$ and $F_j^{-1}(t)$. Therefore, using (4.5) we see that $\sup_{\varphi_j \in \mathcal{G}_j} |\varphi_j'(K_{n,\varphi_j}(t)) - \varphi_j'(F_j^{-1}(t)| \to 0$ a.s. while, on the other hand, $\sup_{\varphi_j \in \mathcal{G}_j} |\varphi_j'(K_{n,\varphi_j}(t)) - \varphi_j'(F_j^{-1}(t))| \leq 2K_j$. But then, by dominated convergence we get that

$$\mathbb{E}\Big[\sup_{\varphi_j \in \mathcal{G}_j} |\varphi_j'(K_{n,\varphi_j}(t)) - \varphi_j'(F_j^{-1}(t))|^2 \Big] \to 0.$$

Since by iii) of Lemma 6.1 we have that $t \mapsto \frac{\sqrt{t(1-t)}}{f_j(F_j^{-1}(t))} \sup_{\varphi \in \mathcal{G}} |\varphi_j(F_j^{-1}(t)) - F_B^{-1}(\varphi)(t)|$ is integrable we conclude that

$$\mathbb{E} \sup_{\varphi \in \mathcal{G}} \int_{\frac{1}{n}}^{1-\frac{1}{n}} |\varphi_j'(K_{n,\varphi_j}(t)) - \varphi_j'(F_j^{-1}(t))| \frac{|B_{n,j}(t)|}{f_j(F_j^{-1}(t))} |\varphi_j(F_j^{-1}(t)) - F_B^{-1}(\varphi)(t)| dt$$

tends to 0 as $n \to \infty$ and, consequently,

$$\sup_{\varphi \in \mathcal{G}} \int_{\frac{1}{n}}^{1-\frac{1}{n}} |\varphi_j'(K_{n,\varphi_j}(t)) - \varphi_j'(F_j^{-1}(t))| \frac{|B_{n,j}(t)|}{f_j(F_j^{-1}(t))} |\varphi_j(F_j^{-1}(t)) - F_B^{-1}(\varphi)(t)| dt$$

vanishes in probability. Combining this fact with (6.8) we prove (6.7) and, as a consequence, (6.4). Finally, observe that for all $n \geq 1$, C has the same law as $\hat{C}_n$. This completes the proof.

□

Proof of Theorem 4.1. From Skohorod Theorem (see, e.g., Theorem 1.10.4 in Van der Vaart and Wellner (1996)) we know that there exists on some probability space versions of C_n and C for which convergence of C_n to C holds almost surely. From now on, we place us on this space and observe that

$$\sqrt{n}(A_n(\mathcal{G}) - A(\mathcal{G})) \leq \sqrt{n} \inf_{\Gamma} U_n - \sqrt{n} \inf_{\Gamma} U = \inf_{\varphi \in \Gamma} C_n(\varphi). \tag{6.9}$$

On the other hand, if we consider the (a.s.) compact set $\Gamma_n = \{\varphi \in \mathcal{G} : U(\varphi) \leq \inf_{\mathcal{G}} U + \frac{2}{\sqrt{n}} \|C_n\|_\infty\}$, then, if $\varphi \notin \Gamma_n$, $U_n(\varphi) \geq \inf_{\mathcal{G}} U + \frac{1}{\sqrt{n}} \|C_n\|_\infty$, while if $\varphi \in \Gamma$, then, $U_n(\varphi) \leq \inf_{\mathcal{G}} U + \frac{1}{\sqrt{n}} \|C_n\|_\infty$. Thus, necessarily, $\inf_{\mathcal{G}} U_n = \inf_{\Gamma_n} U_n = \inf_{\Gamma_n}(U_n - U + U) \geq \inf_{\Gamma_n}(U_n - U) + \inf_{\Gamma_n} U = \inf_{\Gamma_n}(U_n - U) + \inf_{\Gamma} U$. Together with (6.9) this entails

$$\inf_{\varphi \in \Gamma_n} C_n(\varphi) \leq \sqrt{n}(A_n(\mathcal{G}) - A(\mathcal{G})) \leq \inf_{\varphi \in \Gamma} C_n(\varphi) \tag{6.10}$$

Note that for the versions that we are considering $\|C_n - C\|_\infty \to 0$ a.s.. In particular, this implies that $\inf_\Gamma C_n \to \inf_\Gamma C$ a.s.. Hence, the proof will be complete if we show that a.s.

$$\inf_{\Gamma_n} C_n \to \inf_{\Gamma} C. \tag{6.11}$$

To check this last point, consider a sequence $\varphi_n \in \Gamma_n$ such that $C_n(\varphi_n) \leq \inf_{\Gamma_n} C_n + \frac{1}{n}$. By compactness of $\mathcal{G}$, taking subsequences if necessary, $\varphi_n \to \varphi_0$ for some $\varphi_0 \in \mathcal{G}$. Continuity of U yields $U(\varphi_n) \to U(\varphi_0)$ and as a consequence, that $U(\varphi_0) \leq \inf_{\mathcal{G}} U$, that is, $\varphi_0 \in \Gamma$ a.s.. Furthermore,

$$\big|C_n(\varphi_n) - C(\varphi_0)\big| \leq \|C_n - C\|_\infty + |C(\varphi_n) - C(\varphi_0)| \to 0.$$

This shows that

$$\liminf \inf_{\Gamma_n} C_n \geq C(\varphi_0) \geqslant \inf_{\Gamma} C \tag{6.12}$$

and yields (6.11). This completes the proof.

□

PROOF OF COROLLARY 4.2. In Theorem 3.3, take $\mu'_j = \mu_{n,j}$. Then, writing $\mathcal{L}^*$ for the conditional law given the $X_{i,j}$, the result of Theorem 3.3 reads

$$W_2^2(\mathcal{L}((A_{m_n}(\mathcal{G}))^{1/2}), \mathcal{L}^*((A^*_{m_n}(\mathcal{G}))^{1/2})) \leq L^2 \frac{1}{J} \sum_{j=1}^{J} W_2^2(\mu_j, \mu_{n,j}),$$

with $L = \sup_{\varphi \in \mathcal{G}} \left\| \varphi'_j \right\|_\infty < \infty$. Since $W_r(\mathcal{L}(aX+b), \mathcal{L}(aY+b)) = aW_r(\mathcal{L}(X), \mathcal{L}(Y))$ for $a > 0, b \in \mathbb{R}$, the last bound gives

$$\begin{aligned} W_2^2(\mathcal{L}\big(\sqrt{m_n}\big((A_{m_n}(\mathcal{G}))^{1/2} - (A(\mathcal{G}))^{1/2}\big)\big), \mathcal{L}^*\big(\sqrt{m_n}\big((A^*_{m_n}(\mathcal{G}))^{1/2} - (A(\mathcal{G}))^{1/2}\big)\big)) & \\ \leq \quad L^2 \frac{m_n}{\sqrt{n}} \frac{1}{J} \sum_{j=1}^{J} \sqrt{n} W_2^2(\mu_j, \mu_{n,j}). & \end{aligned}$$

As noted in the proof of Theorem 4.1, the assumptions imply that $\sqrt{n} W_2^2(\mu_j, \mu_{n,j})$ vanishes in probability. Also, Theorem 4.1 and the delta method yield that

$$\sqrt{m_n}\big((A_{m_n}(\mathcal{G}))^{1/2} - (A(\mathcal{G}))^{1/2}\big) \rightharpoonup \frac{1}{2(A(\mathcal{G}))^{1/2}} \gamma,$$

with γ the limiting law there, which, combined to the above bound, shows that

$$\sqrt{m_n}\big((A^*_{m_n}(\mathcal{G}))^{1/2} - (A(\mathcal{G}))^{1/2}\big) \rightharpoonup \frac{1}{2(A(\mathcal{G}))^{1/2}} \gamma$$

in probability. A further use of the delta method yields

$$\sqrt{m_n}\big(A^*_{m_n}(\mathcal{G}) - A(\mathcal{G})\big) \rightharpoonup \gamma$$

in probability. The result follows now from Lemma 1 in Janssen and Pauls (2003). □

PROOF OF THEOREM 5.1. We assume for simplicity that $p = 1$. The general case follows with straightforward changes. Let us observe that

$$U_n(\theta) = \frac{1}{J} \sum_{j=1} \int_0^1 (\psi_j(\theta_j, G_{n,j}^{-1}) - \tfrac{1}{J} \textstyle\sum_{k=1}^{J} \psi_k(\theta_k, G_{n,k}^{-1}))^2,$$

with $G_{n,j}$ the empirical d.f. on the $\varepsilon_{i,j}$'s (which are i.i.d. G). A similar expression, replacing $G_{n,j}$ with G is valid for $U(\theta)$. Then (5.6) implies that $\sup_\theta |U_n(\theta) - U(\theta)| \to 0$, from which (recall (5.3) it follows that $\hat{\theta}_n \to \theta^*$ in probability.

From (5.4) we have that U_n is a C^2 function whose derivatives can be computed by differentiation under the integral sign. This implies that

$$\begin{aligned} D_j U_n(\theta) &= \frac{2}{J} \int_0^1 D\psi_j(\theta_j, G_{n,j}^{-1})(\psi_j(\theta_j, G_{n,j}^{-1}) - \tfrac{1}{J} \textstyle\sum_{k=1}^{J} \psi_k(\theta_k, G_{n,k}^{-1})), \\ D_{p,q} U_n(\theta) &= -\frac{2}{J^2} \int_0^1 D\psi_p(\theta_p, G_{n,p}^{-1}) D\psi_q(\theta_q, G_{n,q}^{-1}), \quad p \neq q \end{aligned} \tag{6.13}$$

and

$$D_{p,p}U_n(\theta) = \frac{2}{J}\int_0^1 D^2\psi_p(\theta_p, G_{n,p}^{-1})(\psi_j(\theta_j, G_{n,j}^{-1}) - \tfrac{1}{J}\sum_{k=1}^J \psi_k(\theta_k, G_{n,k}^{-1})) + \frac{2(J-1)}{J^2}\int_0^1 (D\psi_p(\theta_p, G_{n,p}^{-1}))^2.$$

Using also (5.5) we obtain similar expressions for the derivatives of $U(\theta)$, replacing everywhere $G_{n,j}^{-1}$ with G^{-1}. We write $DU_n(\theta) = (D_jU_n(\theta))_{1\le j\le J}$, $DU(\theta) = (D_jU(\theta))_{1\le j\le J}$ for the gradients and $\Sigma_n(\theta) = [D_{p,q}U_n(\theta)]_{1\le p,q\le J}$, $\Sigma(\theta) = [D_{p,q}U(\theta)]_{1\le p,q\le J}$for the Hessians of U_n and U. Note that $\Sigma^* = \Sigma(\theta^*)$ is assumed to be invertible.

We write now $\rho_{n,j}$ for the quantile process based on the $\varepsilon_{i,j}$'s. Observe that (5.7) ensures that we can assume, without loss of generality, that there exist independent Brownian bridges, $B_{n,j}$, satisfying (6.3). Now, recalling that $\psi_j(\theta_j^*, x) = x$ we see that

$$\sqrt{n}D_jU_n(\theta^\star) = \frac{2}{J}\int_0^1 D\psi_j(\theta_j^*, G_{n,j}^{-1}(t))\frac{\rho_{n,j}(t) - \frac{1}{J}\sum_{k=1}^J \rho_{n,k}(t)}{g(G^{-1}(t))}dt. \tag{6.14}$$

Now, using (5.5) and arguing as in the proof of Theorem 4.1 we conclude that

$$\left|\int_0^1 D\psi_j(\theta_j^*, G_{n,j}^{-1}(t))\frac{\rho_{n,k}(t)}{g\,(G^{-1}(t))}dt - \int_0^1 D\psi_j(\theta_j^*, G^{-1}(t))\frac{B_{n,k}(t)}{g\,(G^{-1}(t))}dt\right| \to 0$$

in probability and, consequently,

$$\left|\sqrt{n}D_jU_n(\theta^\star) - \frac{2}{J}\int_0^1 D\psi_j(\theta_j^*, G^{-1}(t))\frac{B_{n,j}(t) - \frac{1}{J}\sum_{k=1}^J B_{n,k}(t)}{g\,(G^{-1}(t))}dt\right| \to 0 \tag{6.15}$$

in probability.

A further Taylor expansion of D_jU_n around θ^* shows that for some $\tilde{\theta}_j^n$ between $\hat{\theta}_n$ and $\theta^\star$ we have

$$D_jU_n(\hat{\theta}_n) = D_jU_n(\theta^\star) + (D_{1j}U_n(\tilde{\theta}_j^n), \ldots, D_{Jj}^2U_n(\tilde{\theta}_j^n))\cdot(\hat{\theta}_n - \theta^\star)$$

and because $\hat{\theta}_n$ is a zero of DU_n, we obtain

$$-D_jU_n(\theta^\star) = (D_{1j}U_n(\tilde{\theta}_j^n), \ldots, D_{Jj}U_n(\tilde{\theta}_j^n))\cdot(\hat{\theta}_n - \theta^\star).$$

Writing $\tilde{\Sigma}_n$ for the $J\times J$ matrix whose J-th row equals $(D_{1j}U_n(\tilde{\theta}_j^n), \ldots, D_{Jj}U_n(\tilde{\theta}_j^n))$, $j = 1, \ldots, J$, we can rewrite the last expansion as

$$-\sqrt{n}DU_n(\theta^*) = \tilde{\Sigma}_n\sqrt{n}(\hat{\theta}_n - \theta^\star). \tag{6.16}$$

Now, recalling (6.13), assumptions (5.4) and (5.5) yield that $\tilde{\Sigma}_n \to \Sigma^* = \Sigma(\theta^*)$ in probability. As a consequence, (6.16) and (6.15) together with Slutsky's Theorem complete the proof of the second claim.

Finally, for the proof of the last claim, since $DU_n(\hat{\theta}_n) = 0$, a Taylor expansion around $\hat{\theta}_n$ shows that

$$nU_n(\theta^*) - nU_n(\hat{\theta}_n) = \frac{1}{2}(\sqrt{n}(\hat{\theta}_n - \theta^*))'\Sigma(\tilde{\theta}_n)(\sqrt{n}(\hat{\theta}_n - \theta^*)) \tag{6.17}$$

for some $\tilde{\theta}_n$ between $\hat{\theta}_n$ and θ^*. Arguing as above we see that $\Sigma(\tilde{\theta}_n) \to \Sigma^*$ in probability. Hence, to complete the proof if suffices to show that

$$nU_n(\theta^*) - \frac{1}{J}\sum_{j=1}^{k}\int_0^1 \frac{\left(B_{n,j}(t) - \frac{1}{J}\sum_{k=1}^{J} B_{n,k}(t)\right)^2}{g(G^{-1}(t))^2}dt \to 0$$

in probability. Since

$$nU_n(\theta^*) = \frac{1}{J}\sum_{j=1}^{k}\int_0^1 \frac{\left(\rho_{n,j}(t) - \frac{1}{J}\sum_{k=1}^{J} \rho_{n,k}(t)\right)^2}{g(G^{-1}(t))^2}dt,$$

this amounts to proving that

$$\int_0^1 \frac{\left(\rho_{n,j}(t) - B_{n,j}(t)\right)^2}{g(G^{-1}(t))^2}dt \to 0$$

in probability. Taking $\nu \in (0, \frac{1}{2})$ in (6.3) we see that

$$\int_{\frac{1}{n}}^{1-\frac{1}{n}} \frac{\left(\rho_{n,j}(t) - B_{n,j}(t))\right)^2}{g(G^{-1}(t))^2}dt \le O_P(1)\frac{1}{n^{1-2\nu}}\int_{\frac{1}{n}}^{1-\frac{1}{n}} \frac{(t(1-t))^{2\nu}}{g(G^{-1}(t))^2} \to 0,$$

using condition (5.8) and dominated convergence. From (5.8) we also see that $\int_{1-\frac{1}{n}}^{1} \frac{B_{n,j}(t)^2}{g(G^{-1}(t))^2}dt \to 0$ in probability. Condition (5.8) implies also that $\int_{1-\frac{1}{n}}^{1} \frac{\rho_{n,j}(t)^2}{g(G^{-1}(t))^2}dt \to 0$ in probability, see Samworth and Johnson (2004). Similar considerations apply to the left tail and complete the proof.

□

Proof of Corollary 5.2. Writing $\mathcal{L}^*$ for the conditional law given the $X_{i,j}$'s, we see from Theorem 3.3 that

$$W_2^2(\mathcal{L}(\sqrt{m_n}(A_{m_n}(\Theta))^{1/2}), \mathcal{L}^*(\sqrt{m_n}(A^*_{m_n}(\Theta))^{1/2}) \le L\frac{m_n}{n}\frac{1}{J}\sum_{j=1}^{J} nW_2^2(\mu, \tilde{\mu}_{n,j}),$$

where $L = \sup_{\lambda,x,j} \psi_j'(\lambda, x)$, μ denotes the law of the errors, $\varepsilon_{i,j}$, and $\tilde{\mu}_{n,j}$ the empirical d.f. on $\varepsilon_{1,j}, \ldots, \varepsilon_{n,j}$. Note that $L < \infty$ by (5.6), while $nW_2^2(\mu, \tilde{\mu}_{n,j}) = O_P(1)$ as in the proof of Theorem 5.1. Hence, we conclude that

$$m_n A^*_{m_n}(\Theta) \rightharpoonup \frac{1}{J}\sum_{j=1}^{J}\int_0^1 \Big(\frac{\tilde{B}_j}{g \circ G^{-1}}\Big)^2 - \frac{1}{2}Y^T\Sigma^{-1}Y$$

in probability. The conclusion now follows from Lemma 1 in Janssen and Pauls (2003).

□

If centering were necessary and we had (5.11) rather than the limit in Theorem 5.1 we could adapt the last argument as follows. If A and B are positive random variables then $E|A - B| \le E(A^{1/2} - B^{1/2})^2 + 2(EAE(A^{1/2} - B^{1/2})^2)^{1/2}$. We can apply this bound to (an optimal coupling of) $m_n A_{m_n}(\Theta)$ and $m_n A^*_{m_n}(\Theta)$. Now if the errors have a log-concave distribution then $nEW_2^2(\mu, \tilde{\mu}_{n,j}) = O(\log n)$, see Corollary 6.12 in Bobkov and Ledoux (2014) and we conclude that

$$W_1(\mathcal{L}(m_n A_{m_n}(\Theta) - c_{m_n}), \mathcal{L}^*(m_n A^*_{m_n}(\Theta) - c_{m_n})) = W_1(\mathcal{L}(m_n A_{m_n}(\Theta)), \mathcal{L}^*(m_n A^*_{m_n}(\Theta)))$$

vanishes in probability if $m_n = O(n^\rho)$ for some $\rho \in (0, 1)$. As a consequence,

$$m_n A^*_{m_n}(\Theta) - c_{m_n} \rightharpoonup \frac{1}{J}\sum_{j=1}^{J}\int_0^1 \frac{\tilde{B}_j^2 - E\tilde{B}_j^2}{(g \circ G^{-1})^2} - \frac{1}{2}Y^T\Sigma^{-1}Y$$

in probability.

References

Agueh, M. and Carlier, G. (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal.*, **43**, 904–924.

Agulló-Antolín, M., Cuesta-Albertos, J.A., Lescornel, H. and Loubes, J.-M. (2015). A parametric registration model for warped distributions with Wasserstein's distance. *J. Multivariate Anal.* **135**, 117–130.

Allassonnière, S. and Amit, Y and Trouvé, A.(2007). Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **69**,1,3–29.

Álvarez-Esteban, P.C., del Barrio, E., Cuesta-Albertos, J.A. and Matrán, C. (2008). Trimmed comparison of distributions, *J. Amer. Statist. Assoc.*, **103**, 697–704.

Amit, Y., Grenander, U. and Piccioni, M.(1991). Structural Image Restoration through deformable template. *J. Amer. Statist.l Assoc.*, **86**, 376–387.

Bobkov, S. and Ledoux, M. (2014). One-dimensional empirical measures, order statistics and Kantorovich transport distances. *Preprint.*

Boissard, E. Le Gouic, T. and Loubes, J.-M. (2015). Distribution's template estimate with Wasserstein metrics. *Bernoulli*, **21**, 740–759.

Bolstad, B. M., Irizarry, R. A., Åstrand, M., and Speed, T. P.(2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19 (2), 185–193.

Chernozhukov, V., Galichon, A., Hallin, M. and Henry, M. (2016). Monge-Kantorovich Depth, Quantiles, Ranks, and Signs. *Ann. Statist.*, to appear.

Collier, O. and Dalalyan, A.S. (2015). Curve registration by nonparametric goodness-of-fit testing. *J. Statist. Plann. Inference*, **162**, 20–42.

Csörgő, M. (1983). *Quantile processes with statistical applications. CBMS-NSF Regional Conference Series in Applied Mathematics, 42.* SIAM.

Csörgő, M. and Horváth, L. (1993). *Weighted approximations in probability and statistics*. Wiley

Cuturi, M. and Doucet, A.(2014) *Fast computation of wasserstein barycenters*, Proceedings of the International Conference on Machine Learning 2014, JMLR W&CP, 32, 1,685–693.

Czado, C. and Munk, A. (1998). Assessing the similarity of distributions–finite sample performance of the empirical Mallows distance. *J. Statist. Comput. Simulation*, **60**, 319–346.

del Barrio, E., Deheuvels, P. and van de Geer, S. (2007). *Lectures on Empirical Processes: Theory and Statistical Applications*. European Mathematical Society.

del Barrio, E., Giné, E. and Utzet, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, **11**, 131–189.

Bigot, J. and Klein, T.(2012) Characterization of barycenters in the Wasserstein space by averaging optimal transport maps *Arxiv preprint, Arxiv 1212.2562*

Freitag, G. and Munk, A. (2005). On Hadamard differentiability in k-sample semiparametric models–with applications to the assessment of structural relationships. *Journal of Multivariate Analysis*, **94**, 123–158.

Gallón, S., Loubes, J.-M. and Maza, E. (2013). Statistical properties of the quantile normalization method for density curve alignment. *Mathematical Biosciences*, **242**, 129–142.

Gamboa, F., Loubes, J.-M. and Maza, E. (2007). Semi-parametric Estimation of Shifts. *Electronic Journal of Statistics*, **1**, 616–640.

Janssen, A. and Pauls, T. (2003). How do bootstrap and permutation tests work? *Ann. Statist.* **31**, 768–806.

Le Gouic, T. and Loubes, J-M.(2016). Existence and consistency of Wasserstein barycenters. *Probability Theory and Related Fields* ,1–17.

Munk, A. and Czado, C. (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **60**, 223–241.

Rachev, S.T. (1984). The Monge-Kantorovich problem on mass transfer and its applications in stochastics. *Teor. Veroyatnost. i Primenen.* **29**, 625–653.

Rajput, B.S. (1972). Gaussian measures on L_p spaces, $1 \leq p < \infty$. *J. Multivariate Anal.*, **2**, 382–403.

Ramsay, J. O. and Silverman, B. W.,(2005). *Functional data analysis*, 2nd Edition. Springer Series in Statistics. Springer, New York.

Rippl, T and Munk, A. and Sturm, A.(2015). Limit laws of the empirical Wasserstein distance: Gaussian distributions, *arXiv preprint arXiv:1507.04090*.

Samworth, R. and Johnson, O. (2004). Convergence of the empirical process in Mallows distance, with an application to bootstrap performance. arXiv:math/0406603

Sommerfeld, M. and Munk, A. (2016). Samworth, R. and Johnson, O. (2004). Inference for Empirical Wasserstein Distances on Finite Spaces. arXiv:1610.03287

van der Vaart, A.W. (2000). *Asymptotic statistics.* Cambridge Univ Press.

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes.* Springer.

Villani, C. (2009). *Optimal transport: old and new.* Springer Verlag.