



HAL
open science

Nonparametric Multiple-Output Center-Outward Quantile Regression

Eustasio del Barrio, Alberto González-Sanz, Marc Hallin

► **To cite this version:**

Eustasio del Barrio, Alberto González-Sanz, Marc Hallin. Nonparametric Multiple-Output Center-Outward Quantile Regression. 2022. <hal-03651586>

HAL Id: hal-03651586

<https://hal.science/hal-03651586v1>

Preprint submitted on 25 Apr 2022

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

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



HAL Authorization

Nonparametric Multiple-Output Center-Outward Quantile Regression*

Eustasio del Barrio¹, Alberto González-Sanz² and Marc Hallin³

¹*IMUVA, Universidad de Valladolid, Spain. e-mail: tasio@eio.uva.es*

²*IMT, Université de Toulouse, France. e-mail: alberto.gonzalezsan@math.univ-toulouse.fr*

³*ECARES and Département de Mathématique Université libre de Bruxelles, Brussels, Belgium. e-mail: mhallin@ulb.ac.be*

Abstract: Based on the novel concept of multivariate center-outward quantiles introduced recently in [Chernozhukov et al. \(2017\)](#) and [Hallin et al. \(2021\)](#), we are considering the problem of nonparametric multiple-output quantile regression. Our approach defines nested *conditional center-outward quantile regression contours* and *regions* with given conditional probability content irrespective of the underlying distribution; their graphs constitute nested *center-outward quantile regression tubes*. Empirical counterparts of these concepts are constructed, yielding interpretable empirical regions and contours which are shown to consistently reconstruct their population versions in the Pompeiu-Hausdorff topology. Our method is entirely non-parametric and performs well in simulations including heteroskedasticity and nonlinear trends; its power as a data-analytic tool is illustrated on some real datasets.

Keywords and phrases: Multiple-output regression, Center-outward quantiles, Optimal transport.

1. Introduction

1.1. Quantile regression, single- and multiple-output

Forty-five years after its introduction by [Koenker and Bassett \(1978\)](#), quantile regression—arguably the most powerful tool in the statistical study of the dependence of a variable of interest Y on covariates $\mathbf{X} = (X_1, \dots, X_m)$ —has become part of statistical daily practice, with countless applications in all domains of scientific research, from economics and social sciences to astronomy, biostatistics, and medicine. Unlike classical regression, which, somewhat narrowly, is focused on conditional means $E[Y|\mathbf{X}]$, quantile regression indeed is dealing with the complete conditional distributions $P_{Y|\mathbf{X}=\mathbf{x}}$ of Y conditional on $\mathbf{X} = \mathbf{x}$. Building on that pioneering contribution, a number of quantile regression methods, parametric, semiparametric, and nonparametric, have been developed for an extremely broad range of statistical topics, including time series, survival analysis, instrumental variables, measurement errors, and functional data—to quote only a few. Sometimes, a simple parametrized regression model allows for a parametric approach, yielding, for instance, linear quantile regression. In most situations, however, parametric models are too rigid and a more agnostic nonparametric approach is in order. We refer to [Koenker \(2005\)](#) for an introductory text and to [Koenker et al. \(2018\)](#) for a comprehensive survey.

*The research of Eustasio del Barrio is partially supported by FEDER, Spanish Ministerio de Economía y Competitividad, grant MTM2017-86061-C2-1-P and Junta de Castilla y León, grants VA005P17 and VA002G18. The research of Alberto González-Sanz is partially supported by the AI Interdisciplinary Institute ANITI, which is funded by the French “Investing for the Future – PIA3” program under the Grant agreement ANR-19-PI3A-0004.

In single-output models (univariate variable of interest Y), this nonparametric approach is well understood and well studied, and the history of non-parametric estimation of conditional quantile functions goes back, at least, to the seminal paper by Stone (1977). The results are much scarcer, however, in the ubiquitous multiple-output case (d -dimensional variable of interest \mathbf{Y} , with $d > 1$), and the few existing ones are less satisfactory—the simple reason for this being the absence of a fully satisfactory concept of multivariate quantiles.

A major difficulty with quantiles in dimension $d > 1$, indeed, is the fact that \mathbb{R}^d , contrary to \mathbb{R} , is not canonically ordered. A number of attempts have been made to overcome that issue, the most remarkable of which is the theory of statistical depth. That theory has generated an abundant literature which we cannot summarize here—we refer to (Serfling and Zuo, 2000) or (Serfling, 2002, 2019a,b) for general expositions and authoritative surveys.

Several depth concepts coexist. The most popular of them is Tukey’s halfspace depth (Tukey, 1975), but all depth concepts (including the most recent ones: see, e.g., Konen and Paidaveine (2022)) are sharing the same basic properties. Tukey’s halfspace depth characterizes, for each distribution P over \mathbb{R}^d (for simplicity, assume P to be Lebesgue-absolutely continuous) *depth regions* $\mathbb{D}_P(\delta)$ (resp., *depth contours* $\mathcal{D}_P(\delta)$) as collections of points with depth (relative to P) larger than or equal to δ (resp., equal to δ), $\delta \in (0, 1/2]$. Depth regions are convex, closed, and nested as δ increases, and have been proposed as notions of quantile regions and contours—an interpretation that is supported by the L_1 nature of Tukey depth (Hallin, Paidaveine and Šiman, 2010).

Among the merits of this interpretation is that it has imposed the idea that quantiles, in dimension $d \geq 2$, should rely on some center-outward ordering with central region of depth $\delta = 1/2$ rather than a southwest-northeast extension of the classical univariate “left-to-right” linear ordering of the real line. Unfortunately, depth regions fail to satisfy the quintessential property of quantile regions: the P -probability content $P[\mathbb{D}_P(\delta)]$ of the quantile region $\mathbb{D}_P(\delta)$ indeed very much depends on P . This is not a minor weakness: the univariate median $Y_{1/2}^P$ of an absolutely continuous distribution P , for instance, is characterized by the fact that $P[(-\infty, Y_{1/2}^P]] = 1/2$ irrespective of P —who would call *median* a quantity Y_{med}^P such that $P_1[(-\infty, Y_{\text{med}}^{P_1}]] = 0.4$ while $P_2[(-\infty, Y_{\text{med}}^{P_2}]] = 0.6$? None of the depth concepts in the literature is satisfying that essential property of quantiles, though, and depth regions, therefore, cannot be considered as *bona fide* quantile regions.

A review of the various existing multiple-output quantile regression models (linear and nonparametric, depth-based and others) can be found in Hallin and Šiman (2018). A nonparametric quantile regression model based on a directional form of Tukey depth is developed in Hallin et al. (2015) but suffers the same lack of control over the probability contents of the quantile regions involved as the depth-based quantile concept itself. So does also the directional concept of M -quantiles proposed by Merlo et al. (2022).

Recently, based on measure transportation ideas, new concepts of quantiles in dimension $d > 1$ have been introduced in (Chernozhukov et al., 2017; Hallin, 2017; Hallin et al., 2021) under the names of *Monge-Kantorovich depth* and *center-outward quantile function*. Center-outward quantile functions define nested closed regions $\mathbb{C}_P(\tau)$ and continuous contours $\mathcal{C}_P(\tau)$ indexed by $\tau \in (0, 1)$ such that, for any absolutely continuous P , $P[\mathbb{C}_P(\tau)] = \tau$ irrespective of P . Unlike the previous concepts, thus, and unlike the depth-based ones (as proposed in Hallin, Paidaveine and Šiman (2010) or Kong and Mizera (2012)), these measure-transportation-based quantiles do satisfy the essential property that the P -probability contents of the resulting quantile regions do not depend on P . Moreover, the corresponding quantile regions are not necessarily convex and, as shown in Figure 1, they are able to cap-

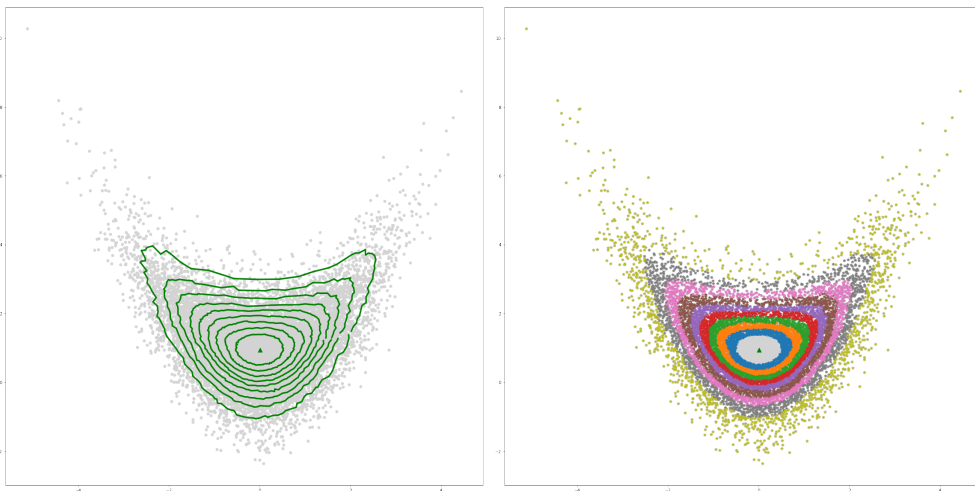


FIG 1. *Quantile contours $C_{\pm}(\tau)$ (left panel) and regions $\mathbb{C}_{\pm}(\tau)$ (right panel) for the banana-shaped Gaussian mixture of Section 4.1.2 and the quantile orders $\tau = 0.1, 0.2, \dots, 0.9$.*

ture the “shape” of the underlying distribution. We refer to Hallin (2022) for a survey of measure-transportation-based center-outward quantiles, the dual concepts of multivariate ranks and signs, and their many applications in inference problems (Ghosal and Sen (2019); Hallin, La Vecchia and Liu (2020); Shi et al. (2021b); Deb and Sen (2021); Shi et al. (2021a); Hallin, Hlubinka and Hudcová (2022); Hallin, La Vecchia and Liu (2022), among others).

Motivated by this long list of successful applications, we are proposing in this paper a novel and meaningful solution, based on the concept of center-outward quantiles, to the problem of nonparametric multiple-output quantile regression. Namely, for a pair of multi-dimensional random variables (\mathbf{X}, \mathbf{Y}) with values in $\mathbb{R}^m \times \mathbb{R}^d$ (\mathbf{Y} the variable of interest, \mathbf{X} the vector of covariates) and joint distribution¹ \mathbb{P} , we define (Section 2.2) the *center-outward quantile map \mathbf{Q}_{\pm} of \mathbf{Y} conditional on $\mathbf{X} = \mathbf{x}$* as

$$\mathbf{u} \in \mathbb{S}_d \mapsto \mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{x}) \in \mathbb{R}^d \quad (1.1)$$

(\mathbb{S}_d the open unit ball in \mathbb{R}^d), with the essential property that, letting

$$\mathbb{C}_{\pm}(\tau | \mathbf{x}) := \mathbf{Q}_{\pm}(\tau \bar{\mathbb{S}}_d | \mathbf{x}) \quad \tau \in (0, 1), \quad \mathbf{x} \in \mathbb{R}^m, \quad (1.2)$$

we have

$$\mathbb{P}[\mathbf{Y} \in \mathbb{C}_{\pm}(\tau | \mathbf{x}) | \mathbf{X} = \mathbf{x}] = \tau \quad \text{for all } \mathbf{x} \in \mathbb{R}^m, \tau \in (0, 1), \text{ and } \mathbb{P}, \quad (1.3)$$

justifying the interpretation of $\mathbf{x} \mapsto \mathbb{C}_{\pm}(\tau | \mathbf{x})$ as the value at \mathbf{x} of a *regression quantile region of order τ* of \mathbf{Y} with respect to \mathbf{X} . For $\tau = 0$,

$$\mathbb{C}_{\pm}(0 | \mathbf{x}) := \bigcap_{\tau \in (0, 1)} \mathbb{C}_{\pm}(\tau | \mathbf{x}) \quad (1.4)$$

¹For simplicity, in this introduction, we tacitly assume all distributions to be Lebesgue-absolutely continuous.

yields the value at $\mathbf{X} = \mathbf{x}$ of the *regression median* $\mathbf{x} \mapsto \mathbb{C}_{\pm}(0 | \mathbf{x})$ of \mathbf{Y} with respect to \mathbf{X} . The same conditional quantile map characterizes nested (no “quantile crossing” phenomenon) “*regression quantile tubes of order τ* ” (in \mathbb{R}^{m+d})

$$\mathbb{T}_{\pm}(\tau) := \{(\mathbf{x}, \mathbf{Q}_{\pm}(\tau \bar{\mathcal{S}}_d | \mathbf{x})) \mid \mathbf{x} \in \mathbb{R}^m\}, \quad \tau \in (0, 1) \quad (1.5)$$

which are such that

$$\mathbb{P}[(\mathbf{X}, \mathbf{Y}) \in \mathbb{T}_{\pm}(\tau)] = \tau \quad \text{irrespective of } \mathbb{P}, \tau \in (0, 1). \quad (1.6)$$

For $\tau = 0$, define

$$\mathbb{T}_{\pm}(0) := \{(\mathbf{x}, \mathbf{y}) \mid \mathbf{x} \in \mathbb{R}^m, \mathbf{y} \in \mathbb{C}_{\pm}(0 | \mathbf{x})\} = \bigcap_{\tau \in (0, 1)} \mathbb{T}_{\pm}(\tau)$$

(the *graph* of $\mathbf{x} \mapsto \mathbb{C}_{\pm}(\tau | \mathbf{x})$); with a slight abuse of language, also call $\mathbb{T}_{\pm}(0)$ the *regression median* of \mathbf{Y} with respect to \mathbf{X} .

None of the earlier attempts to define multiple-output regression quantiles—neither the depth-based definitions in Hallin et al. (2015) or Paindaveine and Siman (2011), the directional concepts of marginal M-quantiles (Breckling and Chambers, 1988) considered by Merlo et al. (2022), nor the measure-transportation-based approach proposed by Carlier, Chernozhukov and Galichon (2016)—is characterizing quantile regions that satisfy requirements (1.3) and (1.6).

Carlier, Chernozhukov and Galichon (2016) deserves special attention, though, as the first attempt to break with directional and depth-based approaches to multiple-output quantile regression by means of innovative measure transportation ideas. Focusing on linear quantile regression, their method is based on a concept of multivariate quantile functions defined over the unit cube $[0, 1]^d$ rather than the unit ball \mathcal{S}_d . While yielding (under the assumption of a linear regression) an asymptotic reconstruction of the distributions of \mathbf{Y} conditional on $\mathbf{X} = \mathbf{x}$, however, their choice of the unit cube does not directly allow for the definition of quantile regions of given order similar to the center-outward regression quantile regions $\mathbb{C}_{\pm}(\tau | \mathbf{x}) \subset \mathbb{R}^d$ or the quantile regression tubes $\mathbb{T}_{\pm}(\tau) \subset \mathbb{R}^{m+d}$.

As for the depth-based quantile regions, moreover, they are necessarily convex, even for distributions with highly non-convex shapes as in Figure 1. To circumvent this convexity problem, Feldman, Bates and Romano (2021) propose a clever transformation of the data turning its distribution into a latent one with convex level sets by fitting a conditional variational auto-encoder (Sohn, Lee and Yan (2015)). The probability contents of the quantile regions resulting from this machine-learning type of “*convexity reparation,*” however, remain out of control (they still depend on \mathbb{P}).

Figure 2 provides, for $m = 1$ and $d = 2$, a visualization of the regression median and quantile tubes of orders $\tau = 0.2, 0.4$, and 0.8 , along with some of the corresponding *conditional regression quantile contours* $\mathcal{C}_{\pm}(\tau | \mathbf{x}) := \{(\mathbf{x}, \mathbf{Q}_{\pm}(\tau \mathcal{S}_{d-1} | \mathbf{x}))\}$ (\mathcal{S}_{d-1} the unit sphere in \mathbb{R}^d), for the model

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \sin(\frac{2\pi}{3}X) + 0.575 e_1 \\ \cos(\frac{2\pi}{3}X) + X^2 + \frac{e_2^3}{2.3} + \frac{1}{4}e_1 + 2.65 X^4 \end{pmatrix} \quad (1.7)$$

where $X \sim U_{[-1, 1]}$, $\mathbf{e} = \sqrt{1 + \frac{3}{2} \sin(\pi X/2)^2} \mathbf{v}$, $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$, and X and \mathbf{v} are mutually independent. Actually, since exact values cannot be obtained for these population concepts,

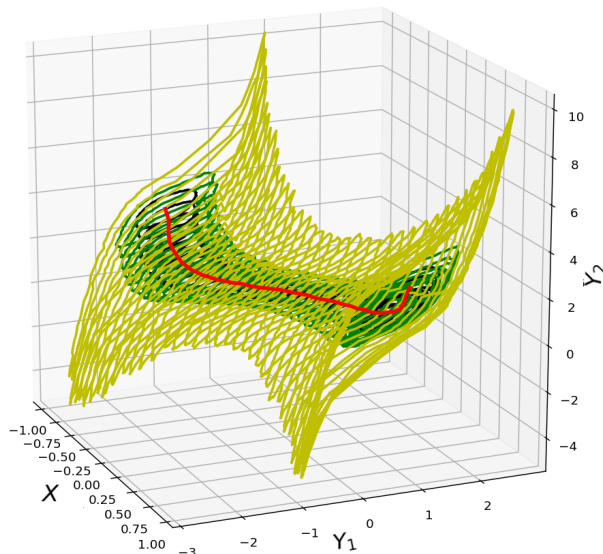


FIG 2. *Quantile regression (two-dimensional variable of interest \mathbf{Y} ; univariate regressor X), showing the conditional center-outward medians (red) and the conditional quantile contours ($\mathbf{x}, \mathbb{C}_{\pm}(\tau | \mathbf{x})$) of order $\tau = 0.2$ (black), $\tau = 0.4$ (green), and $\tau = 0.8$ (yellow).*

a very large sample of $n = 400,040$ observations was generated from (1.7), and the consistent estimation procedure described in Section 3 was performed to obtain the picture. Note the non-convexity of the conditional contours for $\tau = 0.8$, the non-linearity of the regression median, and the marked heteroskedasticity of the regression.

1.2. Outline of the paper

The paper is organized as follows. Section 2 is dealing with the population concept of *conditional center-outward quantile map* and the resulting *center-outward regression quantile contours* and *regions*, Section 3 with their estimation. In Section 3.1, we show how to construct empirical quantile contours and regions, the consistency of which is established in Section 3. Numerical results are provided in Section 4. Monte Carlo experiments in Section 4.1 show the ability of our method to pick heteroskedasticity, nonlinear trends, and the shape of conditional distributions; comparisons also are made with the results of the depth-based method of Hallin et al. (2015). Real datasets are analyzed in Section 4.2, illustrating the power of our method as a data-analytic tool. Section 5 concludes with references to the recent literature on the numerical aspects of optimal transportation and perspectives for future developments.

2. Nonparametric center-outward quantile regression

2.1. Notation

For convenience, we are listing here the main notation to be used throughout the paper. Unless otherwise stated, we denote by $(\Omega, \mathcal{A}, \mathbb{P})$ the triple defining the underlying probability space. Let ℓ_d be the d -dimensional Lebesgue measure, \mathcal{B}_d the Borel σ -field, and $\mathcal{P}(\mathbb{R}^d)$ the space of Borel probability measures on \mathbb{R}^d . The support of a probability $P \in \mathcal{P}(\mathbb{R}^d)$ is denoted as $\text{supp}(P)$; its closure as $\overline{\text{supp}}(P)$. Throughout, (\mathbf{X}, \mathbf{Y}) denotes an \mathbb{R}^{m+d} -valued random vector with probability distribution $\mathbb{P} = P_{\mathbf{X}, \mathbf{Y}} \in \mathcal{P}(\mathbb{R}^{m+d})$, m -dimensional \mathbf{X} -marginal $P_{\mathbf{X}}$ and d -dimensional \mathbf{Y} -marginal $P_{\mathbf{Y}}$. The distribution of \mathbf{Y} conditional on $\mathbf{X} = \mathbf{x}$ is denoted as $P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ ². The open unit ball, the closed unit ball, and the unit hypersphere in \mathbb{R}^d are denoted by \mathbb{S}_d , $\overline{\mathbb{S}}_d$, and \mathcal{S}_{d-1} , respectively. We denote by U_d the spherical uniform over \mathbb{S}_d —that is, the product of a uniform distribution over the unit hypersphere \mathcal{S}_{d-1} (for the directions) and a uniform distribution over $[0, 1]$ (for the distance to the center).

2.2. Conditional center-outward quantiles, regions, and contours.

Let us provide precise definitions for the concepts we briefly presented in the Introduction and properly introduce center-outward quantiles, regions, and contours.

For any $P \in \mathcal{P}(\mathbb{R}^d)$, denote by $\mathbf{Q}_{\pm} = \nabla\varphi$ and call *center-outward quantile map* the (Lebesgue-a.e.) unique gradient of a convex function $\varphi : \mathbb{S}_d \rightarrow \mathbb{R}$ such that $\mathbf{Q}_{\pm}(\mathbf{U}) \sim P$, for any $\mathbf{U} \sim U_d$ —in the measure transportation convenient terminology, \mathbf{Q}_{\pm} is *pushing* P *forward to* U_d , which we denote as $\mathbf{Q}_{\pm} \# P = U_d$. This only defines \mathbf{Q}_{\pm} at φ 's points of differentiability (recall that convex functions are differentiable at almost every point in the interior of their domain: see Theorems 26.1 and 25.5 in Rockafellar (1970)). At φ 's points of non-differentiability \mathbf{u} , let us define $\mathbf{Q}_{\pm}(\mathbf{u})$ as the subdifferential $\partial\varphi(\mathbf{u})$ of φ , namely,

$$\mathbf{Q}_{\pm}(\mathbf{u}) = \partial\varphi(\mathbf{u}) := \{\mathbf{y} \in \mathbb{R}^d \mid \text{for all } \mathbf{z} \in \mathbb{R}^d, \varphi(\mathbf{z}) - \varphi(\mathbf{u}) \geq \langle \mathbf{y}, \mathbf{z} - \mathbf{u} \rangle\}, \quad \mathbf{u} \in \mathbb{S}_d;$$

then, \mathbf{Q}_{\pm} is an everywhere-defined set-valued function. Slightly abusing the notation, we also write $\partial\varphi$ for the set of all points $(\mathbf{u}, \mathbf{y}) \in \mathbb{R}^{m+d}$ such that $\mathbf{y} \in \partial\varphi(\mathbf{u})$. With this notation, we can introduce the concepts of conditional center-outward quantiles, contours, and regions.

Definition 2.1. *Call conditional center-outward quantile map of \mathbf{Y} given $\mathbf{X} = \mathbf{x}$ the center-outward quantile map $\mathbf{u} \mapsto \mathbf{Q}_{\pm}(\mathbf{u}|\mathbf{X} = \mathbf{x})$, $\mathbf{u} \in \mathbb{S}_d$ of $P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$, $\mathbf{x} \in \mathbb{R}^m$. The corresponding conditional center-outward quantile regions and contours of order $\tau \in (0, 1)$ are the sets*

$$\mathbb{C}_{\pm}(\tau|\mathbf{x}) := \mathbf{Q}_{\pm}(\tau\overline{\mathbb{S}}_d|\mathbf{X} = \mathbf{x}) \quad \text{and} \quad \mathcal{C}_{\pm}(\tau|\mathbf{x}) := \mathbf{Q}_{\pm}(\tau\mathcal{S}_{d-1}|\mathbf{X} = \mathbf{x}),$$

respectively. The conditional center-outward quantile maps also characterize (see Definitions (1.4), (1.5), and (1.6)) conditional medians $\mathbb{C}_{\pm}(0|\mathbf{x})$ and regression quantile tubes $\mathbb{T}_{\pm}(\tau)$.

When no confusion is possible, we also write $\mathbf{Q}_{\pm}(\mathbf{u}|\mathbf{x})$ for $\mathbf{Q}_{\pm}(\mathbf{u}|\mathbf{X} = \mathbf{x})$. The terminology center-outward *regression* quantile region, contour, and median is used for the mappings $\mathbf{x} \mapsto \mathbb{C}_{\pm}(\tau|\mathbf{x})$, $\mathbf{x} \mapsto \mathcal{C}_{\pm}(\tau|\mathbf{x})$, and $\mathbf{x} \mapsto \mathbb{C}_{\pm}(0|\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^m$.

²The existence of the regular conditional probability is a direct consequence of the disintegration theorem (see, e.g., Theorem 2.5.1 in Lehmann and Romano (2005)).

Recall, however, that, in the absence of any assumptions on the conditional probabilities $P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$, the mappings $\mathbf{u} \mapsto \mathbf{Q}_{\pm}(\mathbf{u}|\mathbf{X} = \mathbf{x})$ typically are set-valued, see [Rockafellar and Wets \(1998\)](#). Whenever continuous, single-valued functions (typically, on the punctured unit ball $\mathbb{S}_d \setminus \{\mathbf{0}\}$) are needed, we will make the following assumption.

Assumption (R) For $P_{\mathbf{X}}$ -a.e. $\mathbf{x} \in \mathbb{R}^m$, the conditional distribution $P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ admits, with respect to the Lebesgue measure, a density $p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$ with convex support $\text{supp}(P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}})$; moreover, for every $R > 0$, there exist constants $0 < \lambda_R^{\mathbf{x}} \leq \Lambda_R^{\mathbf{x}} < \infty$ such that

$$\lambda_R^{\mathbf{x}} \leq p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) \leq \Lambda_R^{\mathbf{x}} \quad \text{for all } \mathbf{y} \in \text{supp}(P_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}) \cap R\mathbb{S}_d. \quad (2.1)$$

In the classical single-output case ($d = 1$), consistent estimation of conditional quantiles similarly requires the continuity of the conditional quantile maps (see [Stone \(1977\)](#)). In dimension $d > 1$, the continuity of center-outward quantile maps follows from assumptions similar to Assumption (R)—see [Figalli \(2018\)](#) and [del Barrio, González-Sanz and Hallin \(2020\)](#).

3. Empirical center-outward quantile regression

We now proceed with the construction of empirical versions of the conditional center-outward quantile concepts defined in Section 2.2 and their consistency properties.

3.1. Empirical conditional center-outward quantiles

Let $(\mathbf{X}, \mathbf{Y})^{(n)} := ((\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n))$ be a sample of n i.i.d. copies of $(\mathbf{X}, \mathbf{Y}) \sim P_{\mathbf{X}\mathbf{Y}}$. In this section, we develop an estimator of the conditional center-outward quantile maps $\mathbf{u} \mapsto \mathbf{Q}_{\pm}(\mathbf{u}|\mathbf{X} = \mathbf{x})$, $\mathbf{x} \in \mathbb{R}^m$. Our estimator is obtained in two steps: in Step 1, we construct an empirical distribution of \mathbf{Y} conditional on $\mathbf{X} = \mathbf{x}$ and, in Step 2, we compute the corresponding empirical center-outward quantile map.

Step 1. For each value of $\mathbf{x} \in \mathbb{R}^m$, our estimation of the conditional distribution of \mathbf{Y} conditional on $\mathbf{X} = \mathbf{x}$ involves a sequence of *weight functions* $w^{(n)} : \mathbb{R}^{m(n+1)} \rightarrow \mathbb{R}^n$ measurable with respect to \mathbf{x} and the sample $\mathbf{X}^{(n)} := (\mathbf{X}_1, \dots, \mathbf{X}_n)$ of \mathbf{X} observations, of the form

$$(\mathbf{x}, \mathbf{X}^{(n)}) \mapsto w^{(n)}(\mathbf{x}, \mathbf{X}^{(n)}) := \left(w_1(\mathbf{x}; \mathbf{X}^{(n)}), \dots, w_n(\mathbf{x}; \mathbf{X}^{(n)}) \right) \quad (3.1)$$

where $w_j^{(n)} : \mathbb{R}^{m(n+1)} \rightarrow \mathbb{R}$, $j = 1, \dots, n$ are such that

$$w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) \geq 0 \quad \text{and} \quad \sum_{j=1}^n w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) = 1 \quad \text{a.s. for all } n. \quad (3.2)$$

We refer to a function $w^{(n)}$ of the form (3.1) satisfying (3.2) as a *probability weight function* and define the *empirical conditional distribution* of \mathbf{Y} given $\mathbf{X} = \mathbf{x}$ as

$$P_{w(\mathbf{x})}^{(n)} := \sum_{j=1}^n w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) \delta_{\mathbf{Y}_j}, \quad (3.3)$$

where $\delta_{\mathbf{Y}_j}$ is the Dirac function computed at \mathbf{Y}_j . Following Stone (1977), we say that the sequence $w^{(n)}$ is a *consistent* weight function if, whenever $(\mathbf{X}, Y), (\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ are i.i.d., where Y is real-valued and such that $E|Y|^r < \infty$ for $r > 1$,

$$E \left| \sum_{j=1}^n w_j^{(n)}(\mathbf{X}; \mathbf{X}^{(n)}) Y_j - E(Y|\mathbf{X}) \right|^r \longrightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (3.4)$$

Step 2. To estimate the conditional quantiles, consider a *regular grid* $\mathfrak{G}^{(N)}$ of \mathbb{U}_d consisting of N gridpoints denoted as $\mathfrak{g}_1^{(N)}, \dots, \mathfrak{g}_N^{(N)}$. The number N here is arbitrarily chosen as factorizing into a product of integers of the form $N = N_R N_S + N_0$ with $N_0 = 0$ or 1 . That regular grid is created as the intersection between

- the rays generated by an N_S -tuple $\mathbf{u}_1, \dots, \mathbf{u}_{N_S} \in \mathcal{S}_{d-1}$ of unit vectors such that $N_S^{-1} \sum_{j=1}^{N_S} \delta_{\mathbf{u}_j}$ converges weakly to the uniform over \mathcal{S}_{d-1} as $N_S \rightarrow \infty$, and
 - the N_R hyperspheres with center $\mathbf{0}$ and radii $j/(N_R + 1)$, $j = 1, \dots, N_R$,
- along with the origin if $N_0 = 1$. Based on this grid, we define the sequence of discrete uniform measures

$$U_d^{(N)} := \frac{1}{N} \sum_{j=1}^N \delta_{\mathfrak{g}_j^{(N)}} \in \mathcal{P}(\mathbb{R}^d), \quad N \in \mathbb{N}$$

over $\mathfrak{G}^{(N)}$ and require that $N \rightarrow \infty$ with both $N_R \rightarrow \infty$ and $N_S \rightarrow +\infty$. By construction, $U_d^{(N)}$ converges weakly to U_d as $N \rightarrow \infty$. Note that imposing $N_0 = 0$ or 1 is not a problem, since N , unlike n , is chosen by the practitioner. Having $N_0 = 0$ or 1 yields the fundamental advantage that all points of $\mathfrak{G}^{(N)}$ have multiplicity one and that Corollary 3.1 in Hallin et al. (2021), to be used below, applies.

Our estimation of the conditional center-outward quantile maps relies on the optimal transport pushing $U_d^{(N)}$ forward to $P_{w(\mathbf{x})}^{(n)}$ —more precisely, adopting (since typically $N \neq n$) the Kantorovich formulation of the optimal transport problem, on the solution of the linear program (solvable using efficient numerical methods such as the auction or Hungarian algorithms—see Peyré and Cuturi (2019) and references therein)³

$$\begin{aligned} & \min_{\pi := \{\pi_{i,j}\}} \sum_{i=1}^N \sum_{j=1}^n \frac{1}{2} |\mathbf{Y}_j - \mathfrak{g}_i|^2 \pi_{i,j}, \\ \text{s.t. } & \sum_{j=1}^n \pi_{i,j} = \frac{1}{N}, \quad i \in \{1, 2, \dots, N\}, \\ & \sum_{i=1}^N \pi_{i,j} = w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}), \quad j \in \{1, 2, \dots, n\}, \\ & \pi_{i,j} \geq 0, \quad i \in \{1, 2, \dots, N\}, \quad j \in \{1, 2, \dots, n\}. \end{aligned} \quad (3.5)$$

Here, any Nn -tuple $\pi := \{\pi_{i,j} | i = 1, \dots, N \quad j = 1, \dots, n\}$ satisfying the constraints in (3.5) represents a transport plan from $U_d^{(N)}$ to $P_{w(\mathbf{x})}^{(n)}$ —that is, a discrete distribution over $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $U_d^{(N)}$ and $P_{w(\mathbf{x})}^{(n)}$. Let $\pi^*(\mathbf{x}) = \{\pi_{i,j}^*(\mathbf{x}) | i = 1, \dots, N \quad j = 1, \dots, n\}$ be a solution of (3.5) (an optimal *transport plan*). Theorem 2.12(i) in Villani (2003) implies that

³We are dropping the superscripts (N) and (n) when no confusion is possible.

its support $\text{supp}(\pi^*(\mathbf{x})) := \{(\mathfrak{e}_i, \mathbf{Y}_j) \mid \pi_{i,j}^*(\mathbf{x}) > 0\}$ is *cyclically monotone*,⁴ hence is contained in the graph of the *subdifferential* of a convex function. Therefore, the idea is to construct a smooth interpolation of $\pi^*(\mathbf{x})$ that maintains this property.

Note that, for any gridpoint $\mathfrak{e}_i, i \in \{1, \dots, N\}$, the constraints in (3.5) imply that there exists at least one $j \in \{1, \dots, n\}$ such that $(\mathfrak{e}_i, \mathbf{Y}_j) \in \text{supp}(\pi^*(\mathbf{x}))$. Since more than one such j may exist, we choose the one which “gets the highest mass” from \mathfrak{e}_i , and in case of ties we choose the “smallest” one: namely, let

$$\mathbf{T}^*(\mathfrak{e}_i \mid \mathbf{x}) := \arg \inf \left\{ \|\mathbf{y}\| : \mathbf{y} \in \text{conv} \left(\{\mathbf{Y}_J : J \in \arg \max_j \pi_{i,j}^*(\mathbf{x})\} \right) \right\}, \quad (3.6)$$

where $\text{conv}(A)$ denotes the convex hull of a set A . Since $\text{conv}(\{\mathbf{Y}_J : J \in \arg \max_j \pi_{i,j}^*(\mathbf{x})\})$ is closed and convex in \mathbb{R}^d , (3.6) defines a unique $\mathbf{T}^*(\mathfrak{e}_i \mid \mathbf{x})$. Due to the cyclical monotonicity of $\text{supp}(\pi^*(\mathbf{x}))$, there exists a convex function $\varphi^*(\cdot \mid \mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}$ with subdifferential $\partial\varphi^*(\cdot \mid \mathbf{x})$ such that, for all $1 \leq i \leq N$,

$$\emptyset \neq \{\mathbf{Y}_j : (\mathfrak{e}_i, \mathbf{Y}_j) \in \text{supp}(\pi^*(\mathbf{x}))\} \subset \partial\varphi^*(\mathfrak{e}_i \mid \mathbf{x}).$$

Since sub-differentials are convex sets, this entails

$$\text{conv}\{\mathbf{Y}_j : (\mathfrak{e}_i, \mathbf{Y}_j) \in \text{supp}(\pi^*(\mathbf{x}))\} \subset \partial\varphi^*(\mathfrak{e}_i \mid \mathbf{x}).$$

Consequently, $\{(\mathfrak{e}_i, \mathbf{T}^*(\mathfrak{e}_i \mid \mathbf{x})) : i = 1, \dots, N\}$ is cyclically monotone and satisfies the assumptions of Corollary 3.1 in Hallin et al. (2021). This implies the existence, for all \mathbf{x} , of a continuous cyclically monotone map $\mathbf{u} \mapsto \mathbf{Q}_{w,\pm}^{(n)}(\mathbf{u} \mid \mathbf{x})$, say, interpolating the N -tuple $(\mathfrak{e}_1, \mathbf{T}^*(\mathfrak{e}_1 \mid \mathbf{x})), \dots, (\mathfrak{e}_N, \mathbf{T}^*(\mathfrak{e}_N \mid \mathbf{x}))$, i.e., such that $\mathbf{Q}_{w,\pm}^{(n)}(\mathfrak{e}_i \mid \mathbf{x}) = \mathbf{T}^*(\mathfrak{e}_i \mid \mathbf{x})$ for $i = 1, \dots, N$.

In particular, we proceed as in Hallin et al. (2021) by choosing the smooth cyclically monotone interpolation with largest Lipschitz constant. This continuous map $\mathbf{u} \mapsto \mathbf{Q}_{w,\pm}^{(n)}(\mathbf{u} \mid \mathbf{x})$ from \mathbb{S}_d to \mathbb{R}^d will be called the *empirical conditional center-outward quantile function* of \mathbf{Y} given $\mathbf{X} = \mathbf{x}$. It defines the *empirical center-outward regression quantile regions* and *contours*

$$\mathbb{C}_{w,\pm}^{(n)}(\tau \mid \mathbf{x}) := \mathbf{Q}_{w,\pm}^{(n)}(\tau \bar{\mathbb{S}}_d \mid \mathbf{x}) \quad \text{and} \quad \mathcal{C}_{w,\pm}^{(n)}(\tau \mid \mathbf{x}) := \mathbf{Q}_{w,\pm}^{(n)}(\tau \mathcal{S}_{d-1} \mid \mathbf{x}), \quad \tau \in (0, 1) \quad (3.7)$$

which we are proposing as estimators of $\mathbb{C}_{\pm}(\tau \mid \mathbf{x})$ and $\mathcal{C}_{\pm}(\tau \mid \mathbf{x})$, respectively. The intersection $\bigcap_{\tau \in (0,1)} \mathbb{C}_{w,\pm}^{(n)}(\tau \mid \mathbf{x})$ yields the *empirical conditional center-outward regression median region*. The definition of *empirical regression quantile tubes*

$$\mathbb{T}_{w,\pm}^{(n)}(\tau) := \left\{ \left(\mathbf{x}, \mathbf{Q}_{\pm}^{(n)}(\tau \bar{\mathbb{S}}_d \mid \mathbf{x}) \right) \mid \mathbf{x} \in \mathbb{R}^m \right\}, \quad \tau \in (0, 1)$$

naturally follows.

Remark 3.1. Note that the results of this section and the next one still hold for any continuous map with cyclically monotone graph satisfying

$$(\mathbf{u}_i, \mathbf{Q}_{w,\pm}^{(n)}(\mathbf{u}_i \mid \mathbf{x})) \in \text{conv}(\{\mathbf{Y}_j : (\mathbf{u}_i, \mathbf{Y}_j) \in \text{supp}(\pi^*(\mathbf{x}))\}) \quad \text{for all } i = 1, \dots, N.$$

The reason for choosing the “smallest” \mathbf{y} in (3.6) is to have a “universal criterion.”

⁴Recall from Rockafellar (1970) that a set $S \subset \mathbb{R}^d \times \mathbb{R}^d$ is *cyclically monotone* if any finite subset $\{(x_{k_1}, y_{k_1}), \dots, (x_{k_\nu}, y_{k_\nu})\} \subset S$, $\nu \in \mathbb{N}$ satisfies $\sum_{\ell=1}^{\nu-1} \langle y_{k_\ell}, x_{k_{\ell+1}} - x_{k_\ell} \rangle + \langle y_{k_\nu}, x_{k_1} - x_{k_\nu} \rangle \leq 0$, where $\langle \cdot, \cdot \rangle$ stands for the scalar product in \mathbb{R}^d .

3.2. Consistency of empirical conditional center-outward quantiles, regression quantile regions, and regression quantile contours

The objective of this section is to justify the definitions of Section 3.1 by showing the consistency of the empirical quantile regions and contours defined in (3.7). The asymptotic behavior of these regions and contours, quite naturally, depends on the regularity of the conditional distributions involved. In fact, as discussed before, when Assumption (R) does not hold, the population conditional quantile maps are not necessarily defined for every $\mathbf{u} \in \mathbb{S}_d$, but only for a set of U_d -probability one. Consistency results can be obtained despite this a.s. definition, provided that population quantile maps are extended into set-valued maps. The following theorem shows⁵, under mild assumptions, that any possible limit of $\mathbf{Q}_{w,\pm}^{(n)}(\cdot | \mathbf{x})$ asymptotically belongs to the set $\mathbf{Q}_{\pm}(\cdot | \mathbf{X} = \mathbf{x})$.

Theorem 3.2. *Let $(\mathbf{X}, \mathbf{Y}), (\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ be pairs of i.i.d. random vectors with values in $\mathbb{R}^m \times \mathbb{R}^d$ and let $w^{(n)}$ be a consistent sequence of weight functions. Then, for every $\mathbf{u} \in \mathbb{S}_d$ and $\epsilon > 0$,*

$$\mathbb{P} \left(\mathbf{Q}_{w,\pm}^{(n)}(\mathbf{u} | \mathbf{X}) \not\subset \mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X}) + \epsilon \mathbb{S}_d \right) \longrightarrow 0 \quad \text{as } n \text{ and } N \rightarrow \infty,$$

and, for every $\tau \in (0, 1)$,

$$\mathbb{P} \left(\mathbb{C}_{\pm}^{(n)}(\tau | \mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau | \mathbf{X}) + \epsilon \mathbb{S}_d \right) \rightarrow 0 \quad \text{and} \quad \mathbb{P} \left(\mathcal{C}_{\pm}^{(n)}(\tau | \mathbf{X}) \not\subset \mathcal{C}_{\pm}(\tau | \mathbf{X}) + \epsilon \mathbb{S}_d \right) \rightarrow 0$$

as n and $N \rightarrow \infty$.

Neater convergence results—avoiding the notion of set-valued maps—are obtained if it can be assumed that Assumption (R) holds, which implies that, for any $\mathbf{X} = \mathbf{x}$ and any $\mathbf{u} \in \mathbb{S}_d \setminus \{\mathbf{0}\}$, the set $\mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X} = \mathbf{x})$ is a singleton. Then, the map $\mathbf{u} \mapsto \mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X} = \mathbf{x})$ can be seen as continuous on $\mathbb{S}_d \setminus \{\mathbf{0}\}$, see Theorem 25.5 in Rockafellar (1970), hence single-valued on $\mathbb{S}_d \setminus \{\mathbf{0}\}$ since the gradient of a convex function is single-valued at a point if and only if it is continuous at this point. We then can state the following theorem (see Appendix A.1 for the proof), the second part of which describes the convergence of the contours in terms of the Pompeiu-Hausdorff distance d_{∞} . Recall that the Pompeiu-Hausdorff distance between two sets A and B in \mathbb{R}^d is defined as

$$d_{\infty}(A, B) := \inf\{\nu \geq 0 : A \subset B + \nu \mathbb{S}_d \text{ and } B \subset A + \nu \mathbb{S}_d\}$$

(see Rockafellar and Wets (1998)).

Theorem 3.3. *Let $(\mathbf{X}, \mathbf{Y}), (\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_n, \mathbf{Y}_n)$ be pairs of i.i.d. random vectors with values in $\mathbb{R}^m \times \mathbb{R}^d$ and let $w^{(n)}$ be a consistent sequence of weight functions. Suppose moreover that Assumption (R) holds. Then, for every compact $K \subset \mathbb{S}_d \setminus \{\mathbf{0}\}$, as n and $N \rightarrow \infty$,*

$$\sup_{\mathbf{u} \in K} |\mathbf{Q}_{w,\pm}^{(n)}(\mathbf{u} | \mathbf{X}) - \mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X})| \xrightarrow{\mathbb{P}} 0$$

and, for every $\tau \in (0, 1)$ and $\epsilon > 0$,

$$\mathbb{P} \left(d_{\infty} \left(\mathcal{C}_{\pm}^{(n)}(\tau | \mathbf{X}), \mathcal{C}_{\pm}(\tau | \mathbf{X}) \right) > \epsilon \right) \rightarrow 0.$$

⁵Note that, although N does not appear in the notation, $\mathbf{Q}_{w,\pm}^{(n)}$ depends on both N and n .

Under the assumptions of Theorem 3.3, consistency in Pompeiu-Hausdorff distance of the quantile contours holds in case the median is a single point⁶—the continuity of quantile maps then extends to the whole open unit ball. This, however, is not necessarily the case for $d > 3$ (see Figalli (2018)), and Pompeiu-Hausdorff consistency may fail due to the fact that our empirical version is continuous over \mathbb{S}_d while $\mathbf{Q}_\pm(\mathbf{0}|\mathbf{x})$ could be a set rather than a single point: convergence then holds along subsequences of $\mathbf{Q}_{w,\pm}^{(n)}(\cdot|\mathbf{x})$ towards an element of $\mathbf{Q}_\pm(\mathbf{0}|\mathbf{x})$. This, obviously, has an impact on convergence in terms of the Pompeiu-Hausdorff distance—although it does not affect the control over the asymptotic probability contents of quantile regions. More precisely, the following corollary holds (see Appendix A.1 for the proof).

Corollary 3.4. *Under the conditions of Theorem 3.3, as n and $N \rightarrow \infty$,*

$$\mathbb{P}\left(\mathbf{Y} \in \mathbb{C}_\pm^{(n)}(\tau|\mathbf{X})|\mathbf{X}\right) \xrightarrow{\mathbb{P}} \tau \quad \text{for all } \tau \in (0, 1). \quad (3.8)$$

Remark 3.5. *Under the weaker conditions of Theorem 3.2, in view of (A.12), still some asymptotic control of the probability content of the empirical regions can be derived. More precisely, letting $N = N(n)$ be such that $N(n) \rightarrow \infty$ as $n \rightarrow \infty$, for all $\tau \in (0, 1)$ and every subsequence n_k , there exists a further subsequence n_{k_j} such that*

$$\limsup_m \mathbb{P}\left(\mathbb{C}_\pm^{(n_{k_j})}(\tau|\mathbf{x})|\mathbf{X} = \mathbf{x}\right) \leq \tau, \quad \mathbf{x}\text{-a.e. in } \mathbb{R}^m.$$

The above results, as well as the proposed regularization, are valid for any consistent sequence of weight functions. This includes—along with adequate additional assumptions—most of the classic choices of weight functions. Here are three examples.

(i) The kernel weight function, usually defined (see chapter 5 in Györfi et al. (2002)) as

$$w_i^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) := K\left(\frac{\mathbf{X}_i - \mathbf{x}}{h_n}\right) / \sum_{j=1}^n K\left(\frac{\mathbf{X}_j - \mathbf{x}}{h_n}\right), \quad i = 1, \dots, n$$

where h_n is the bandwidth, and $K : \mathbb{R}^m \rightarrow \mathbb{R}$ the kernel. Sufficient conditions for $w_i^{(n)}$ to form a consistent sequence of weight functions are

- (a) $h_n \rightarrow 0$,
- (b) $c_1 \min\left(\mathbb{1}_{\|\mathbf{x}\| \leq r}, H(\|\mathbf{x}\|)\right) \leq K(\mathbf{x}) \leq c_2 H(\|\mathbf{x}\|)$, where c_1, c_2, r are positive constants, and $H : [0, \infty) \rightarrow \mathbb{R}$ is bounded, decreasing, and such that $H(t)t^m \rightarrow 0$ as $t \rightarrow \infty$, and
- (c) $\lim_{n \rightarrow \infty} n^\alpha h_n^m / \log(n) = \infty$ for any $\alpha \in (0, 1)$;

when the kernel K is compactly supported, the assumptions are much simpler, and we only need (a) and $\lim_{n \rightarrow \infty} h_n^m n = \infty$, see Theorem 5.1 in Györfi et al. (2002). The particular case $K(\mathbf{x}) = e^{-|\mathbf{x}|^2}$ is known as the Gaussian kernel.

(ii) The (classical) k -nearest neighbors weight function: the k -nearest neighborhood of $\mathbf{x} \in \mathbb{R}^m$ is obtained (Chapter 6 in Györfi et al. (2002)) by ordering $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$

⁶This is always the case for $d = 2$ and $d = 3$: see Figalli (2018).

according to increasing values of $|\mathbf{X}_j - \mathbf{x}|$. Denoting by $\{\mathbf{X}_{(0,\mathbf{x})}, \dots, \mathbf{X}_{(n,\mathbf{x})}\}$ the re-ordered sequence, the set of k -nearest neighbors of \mathbf{x} is $\mathcal{N}_k^{(n)}(\mathbf{x}) := \{\mathbf{X}_{(j,\mathbf{x})} : j \leq k\}$ and the k -nearest neighbors weight function is defined as

$$w_i^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) := \frac{1}{k} \mathbb{1}_{\mathbf{X}_i \in \mathcal{N}_k^{(n)}(\mathbf{x})}, \quad i = 1, \dots, n.$$

Sufficient conditions for this k -nearest neighbors $w_i^{(n)}$ to form a consistent sequence of weight functions are (see Stone (1977))

$$k \rightarrow +\infty \quad \text{and} \quad k/n \rightarrow 0. \quad (3.9)$$

If a k -nearest neighbors weight function $w^{(n)}$ is satisfying (3.9) and Assumption (R) holds, we thus have the convergence—in probability—of the conditional quantile map described in Theorem 3.3; without this assumption we still obtain the slightly weaker result of Theorem 3.2.

The k -nearest neighbors in (ii) were understood in the classical sense of the Euclidean distance in \mathbb{R}^m , which does not take into account the distribution $\mathbb{P}_{\mathbf{X}}$ of \mathbf{X} . An alternative k -nearest neighbors weight function can be derived from a notion of nearness based on the ordering induced by empirical center-outward distribution functions. This new weight function is obtained as follows.

- (iii) An alternative k -nearest neighbors weight function. Fixing $\mathbf{x} \in \mathbb{R}^m$, first compute, as in Hallin et al. (2021), the empirical center-outward distribution function associated with

$$\frac{1}{n+1} \sum_{j=1}^n \delta_{\mathbf{X}_j} + \frac{1}{n+1} \delta_{\mathbf{x}} \in \mathcal{P}(\mathbb{R}^m).$$

That distribution function is the solution $T_{\mathbf{x}}^*$ of the minimization problem

$$\min_{T \in \Gamma_{n+1}} \sum_{k=0}^n |\mathbf{X}_k - T(\mathbf{X}_k)|^2$$

where $\mathbf{X}_0 = \mathbf{x}$, Γ_{n+1} is the set of all bijections T between $\{\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_n\}$ and a regular grid $\mathfrak{G}^{(n+1)}$ of \mathbb{S}_m , of the form described in Section 3.1, consisting of $(n+1)$ gridpoints denoted as $\mathfrak{g}_0, \mathfrak{g}_1, \dots, \mathfrak{g}_n$, obtained via a factorization of n into a product of non-negative integers of the form $n+1 = n_R n_S + n_0$ with $n_0 < \min(n_R, n_S)$. That regular grid is created as the intersection between

- the rays generated by an n_S -tuple $\mathbf{u}_1, \dots, \mathbf{u}_{n_S} \in \mathcal{S}_{m-1}$ of unit vectors such that $n_S^{-1} \sum_{j=1}^{n_S} \delta_{\mathbf{u}_j}$ converges weakly, as $n_S \rightarrow \infty$, to the uniform over \mathcal{S}_{m-1} and
- the n_R hyperspheres with center $\mathbf{0}$ and radii $j/(n_R+1)$, $j = 1, \dots, n_R$,

along with n_0 copies of the origin whenever $n_0 > 0$. Based on this grid, we define the sequence of discrete uniform measures

$$\mathbb{U}_d^{(n+1)} := \frac{1}{n+1} \sum_{j=1}^{n+1} \delta_{\mathfrak{g}_j} \in \mathcal{P}(\mathbb{R}^m) \quad N \in \mathbb{N}$$

This map is defined only at the $(n + 1)$ points $\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_n$, but, as in the previous section, it can be countinuously extended (see also Hallin et al. (2021)) to the whole space \mathbb{R}^m —call $\mathbf{F}_{\mathbf{x};\pm}^{(n)} : \mathbb{R}^m \rightarrow \mathbb{S}_m$ this extension—with the properties that $\mathbf{F}_{\mathbf{x};\pm}^{(n)}$ coincides, on $\{\mathbf{x}, \mathbf{X}_1, \dots, \mathbf{X}_n\}$, with $T_{\mathbf{x}}^*$, is the gradient of a differentiable convex function with domain \mathbb{R}^m , and satisfies $\mathbf{F}_{\mathbf{x};\pm}^{(n)}(\mathbb{R}^m) \subset \overline{\mathbb{S}}_m$. We then define the *set of k -nearest center-outward neighbors* of \mathbf{x} as

$$\mathcal{K}_k^{(n)}(\mathbf{x}) := \{\mathbf{X}_j : \mathbf{F}_{\pm}^{(n)}(\mathbf{X}_j) \in \mathcal{N}_k(\mathbf{F}_{\mathbf{x};\pm}^{(n)}(\mathbf{x}))\}, \quad (3.10)$$

where, for each $\mathbf{a} \in \mathbf{B}_m$ and $k \in \mathbb{N}$, $\mathcal{N}_k(\mathbf{a})$ denotes the set of k -nearest neighbors (in the sense of Euclidean distance) of \mathbf{a} . Based on this, define the *center-outward nearest neighbor weight function*

$$w_j^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) := \frac{1}{k} \mathbb{1}_{\mathbf{X}_j \in \mathcal{K}_k^{(n)}(\mathbf{x})}, \quad j = 1, \dots, k \quad (3.11)$$

and proceed as in Section 3.1 with the estimation (3.5) of the conditional quantile functions.

The next result shows that, for a suitable choice of $k = k(n)$, center-outward nearest neighbors weight functions form a consistent sequence of weights (see the appendix for a proof).

Lemma 3.6. *If $k = k(n)$ is such that $k(n) \rightarrow \infty$ and $k(n)/n \rightarrow 0$ as $n \rightarrow \infty$, the sequence of weight functions defined in (3.11) is consistent in the sense of (3.4).*

This means, in particular, that Theorem 3.2 applies when the weight function (3.11) is used under the assumptions of Lemma 3.6, and that the resulting estimators are consistent.

Theorems 3.2 and 3.3 provide *weak* (in probability) *consistency* results under minimal assumptions. For sequences of weights satisfying, as $n \rightarrow \infty$,

$$\sum_{j=1}^n w_j^{(n)}(\mathbf{X}; \mathbf{X}^{(n)}) Y_j \longrightarrow \mathbb{E}[Y|\mathbf{X}] \quad \text{a.s.} \quad (3.12)$$

(*strongly consistent* sequences), the conclusion in Theorem 3.2 can be upgraded to *strong* (almost sure) *consistency*. For the particular case of k -nearest neighbors, (3.12) and strong consistency hold if (3.9) is replaced with

$$k/\log(n) \rightarrow \infty \quad \text{and} \quad k/n \rightarrow 0, \quad (3.13)$$

see Devroye et al. (1994) and Devroye (1982).

4. Numerical results

This section is devoted to a numerical assessment of the performance of the estimation procedures described in Section 3. We first analyze (Section 4.1) some artificial datasets—including the motivating example of Hallin et al. (2015)—then turn (Section 4.2) to real-data cases. These examples showcase three important features of our estimators: their ability to capture heteroskedasticity, to deal (non-parametrically) with highly nonlinear regression, and to adapt to non-convex noise.

4.1. Simulated examples

4.1.1. Parabolic trend and periodic heteroskedasticity; spherical conditional densities.

We start with analyzing the motivating example given in Hallin et al. (2015). The model (with $m = 1$, $d = 2$) is

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X \\ X^2 \end{pmatrix} + \left(1 + \frac{3}{2} \sin\left(\frac{\pi}{2}X\right)\right) \mathbf{e}, \quad X \sim U_{[-2,2]} \text{ and } \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{Id}), \quad (4.1)$$

with X and \mathbf{e} mutually independent. In this case, the population conditional (on $X = x$) quantile contours are circles with radii depending on x and can be computed exactly; trend is parabolic and heteroskedasticity periodic.

Figure 3 illustrates the convergence of our estimated contours to the population counterparts. Compared to Figure 1 in Hallin et al. (2015), our method produces less smooth contours, at least for smaller sample sizes. On the other hand, our method is able to capture non-convex contour shapes—something the method in Hallin et al. (2015) cannot, see Section 4.1.2. We also underline that, from a computational point of view, our method is able to handle rather large datasets (in contrast, the R packaged `modQR` cannot handle sample sizes over 10,000, as explained in the documentation).

Model (4.1), as pointed out in Hallin et al. (2015), allows for testing the capacity of a method to estimate the trend while catching potential heteroskedasticity. A comparison with Figure 1 in Hallin et al. (2015) shows that both methods estimate the parabolic trend quite well, but that our method performs much better at capturing heteroskedasticity. Estimations are based on the k -nearest neighbors weights (ii) of Section 3, with $N = k = 401, 1,000$, and $6,401$ and $n = 3,601, 10,000$, and $128,020$, respectively. Note that, in this univariate covariate case, the weight choices (ii) and (iii) coincide.

The performance of the Gaussian kernel (i) (with various bandwidth choices) and classical k -nearest neighbors (ii) (with various choices of k) weight functions are investigated in Figures 4 and 5, still for Model 4.1, with sample size $n = 3,601$. More precisely, estimation in Figure 4 is based on the weight function

$$w_i^{(n)}(\mathbf{x}; \mathbf{X}^{(n)}) := e^{-\left(\frac{|\mathbf{x}_i - \mathbf{x}|}{h_n}\right)^2} / \sum_{j=1}^n e^{-\left(\frac{|\mathbf{x}_j - \mathbf{x}|}{h_n}\right)^2}, \quad i = 1, \dots, n, \quad (4.2)$$

for $h_n = 0.05, 0.1, 0.2$, and 0.3 ; the discretization of the spherical uniform is based on a grid of size $N = n$. Estimation in Figure 5 is based on k -nearest neighbors weights with $k = 101, 256$, and 625 . Here we chose $N = k$, which yields a one-to-one solution in (3.5). In both figures the conditional contours (of order $\tau = 0.2$ (black) and $\tau = 0.4$ (green)) and the estimated conditional center-outward medians (red) are shown for

$$x \in \{-2, -1.6, -1.1, -0.7, -0.2, 0.2, 0.7, 1.1, 1.6, 2\}.$$

For this sample size, the Gaussian kernel weights—due to the fact that they better exploit the information available on the x 's—yield better results than the k -nearest neighbors ones. But the Gaussian kernel has a drawback for large datasets; the optimization problem (3.5) requires the whole dataset and cannot be efficiently computed. For instance, the Gaussian kernel counterpart of Figure 3 (where $n = 128,020$) cannot be computed on a standard desktop computer: large-sample datasets should be handled either with nearest neighbors or compactly supported kernel weights. On the other hand, the bandwidth h_n and the neighborhood size k apparently have little impact on the result.

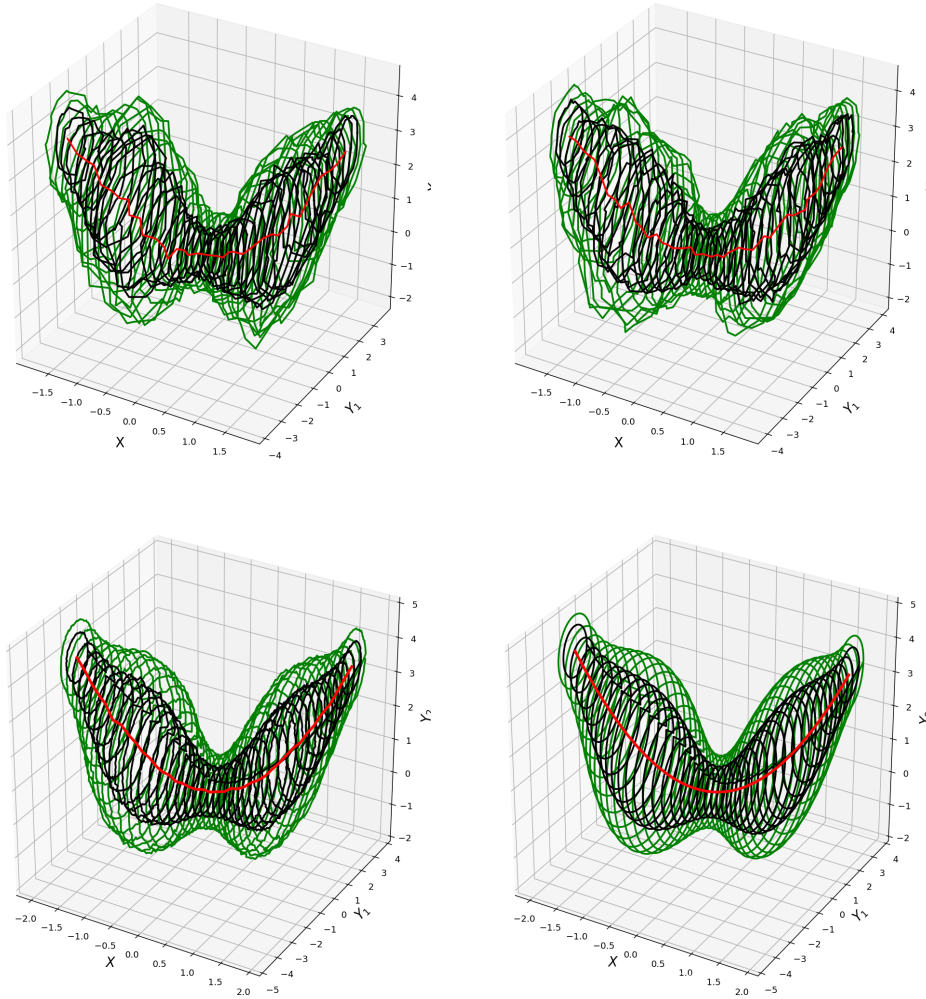


FIG 3. Estimated (sample sizes 3,601 in the upper left panel, 10,000 in the upper right panel, 128,020 in the lower left panel) and population (lower right panel) quantile contours of order $\tau = 0.2$ (black) and 0.4 (green) for Model (4.1); the (estimated) conditional center-outward medians are shown in red. Estimations are based on the k -nearest neighbors weights (ii) with $N = k = 401, 1,000,$ and $6,401$ and $n = 3,601, 10,000,$ and $128,020,$ respectively.

4.1.2. Parabolic trend and periodic heteroskedasticity; banana-shaped conditional densities.

We now consider a model in which the trend and heteroskedasticity are the same as in Model (4.1), but the quantile contours are non-convex (conditional densities are banana-shaped):

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X \\ X^2 \end{pmatrix} + \begin{pmatrix} \left(1 + \frac{3}{2} \sin\left(\frac{\pi}{2}X\right)^2\right) 1.15 e_1 \\ \left(1 + \frac{3}{2} \sin\left(\frac{\pi}{2}X\right)^2\right) \left(\frac{e_2}{1.15} + 0.5(e_1^2 + 1.21)\right) \end{pmatrix}, \quad (4.3)$$

with $X \sim U_{[-2,2]}$ and $\mathbf{e} = (e_1, e_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{Id})$, X and \mathbf{e} mutually independent. The dataset shown in Figure 1 was generated from that model, with $X = 0$ and justifies the terminology “banana-shaped.”

Population conditional quantile contours here cannot be computed analytically. Figure 6 shows their estimations (same method as in Section 4.1.1, with a k -nearest neighbors weight function, $k = 14,401$) for orders $\tau = 0.2$ (black), 0.4 (green), and 0.8 (yellow), along with the conditional medians (red). The sample size $n = 576,040$ is very large, so that, in view of consistency results, these estimations can be considered as close numerical approximations of their theoretical counterparts. This example illustrates further the ability of our method

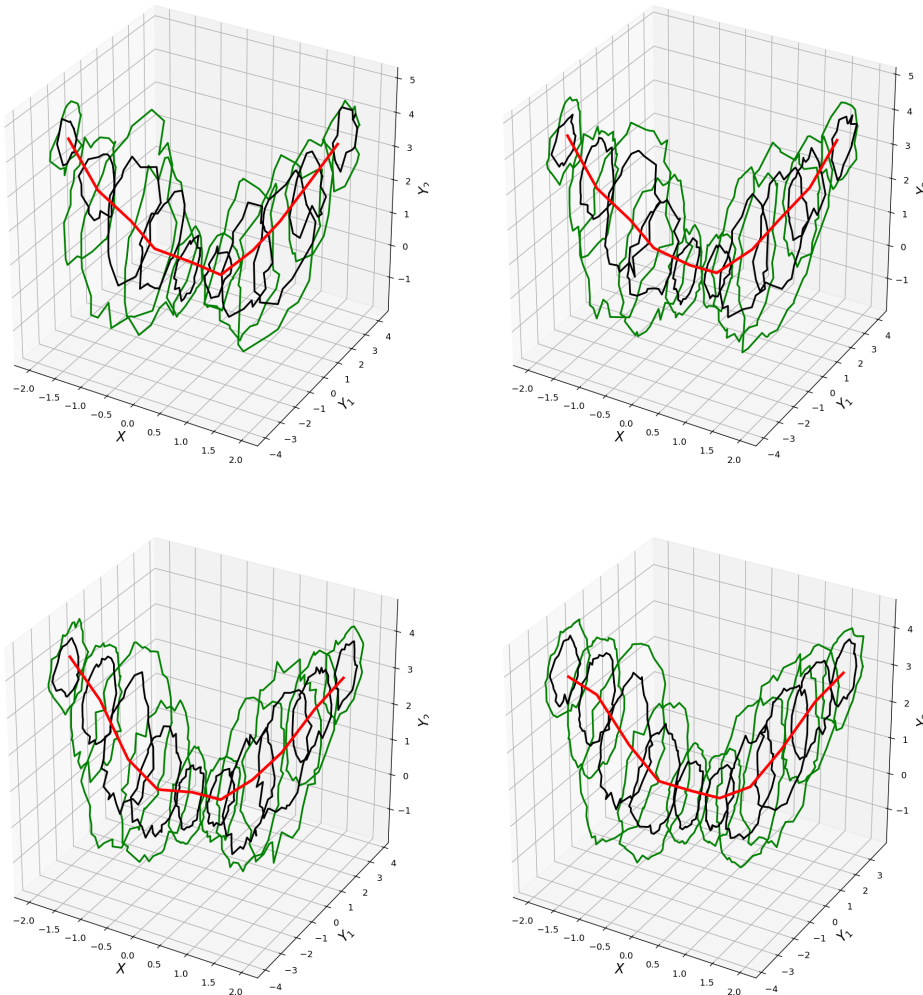


FIG 4. Estimated quantile contours (Model (4.1)) based on Gaussian kernel weight functions for different choices of the bandwidth h . The sample size is $n = 3,601$; bandwidth values are $h = 0.05$ (upper left panel), $h = 0.1$ (upper right panel), $h = 0.2$ (lower left panel) and $h = 0.3$ (lower right panel). The estimated contour orders are $\tau = 0.2$ (black) and 0.4 (green); the estimated conditional center-outward medians are shown in red.

to handle vary large sample sizes. The Hallin et al. (2015) approach produces contours that are convex by construction—hence cannot capture the “banana shape” of the conditional densities. The same comparative analysis of weight functions is performed as for Model (4.1). Figures 7 and 8 show the results for Gaussian kernel weights (various bandwidths) and k -nearest neighbors weights (various values of k), respectively. The sample size is $n = 3601$. In all the examples, for the ease of computation, $N = k$. Note that, for the k -nearest neighbor weights, this choice creates a one-to-one (between the sample and the grid) transport map.

Still for Model (4.3), Figure 9 shows, for sample size $n = 3601$ and various choices of the bandwidth h and the neighborhood size k , the behavior of the empirical conditional center-

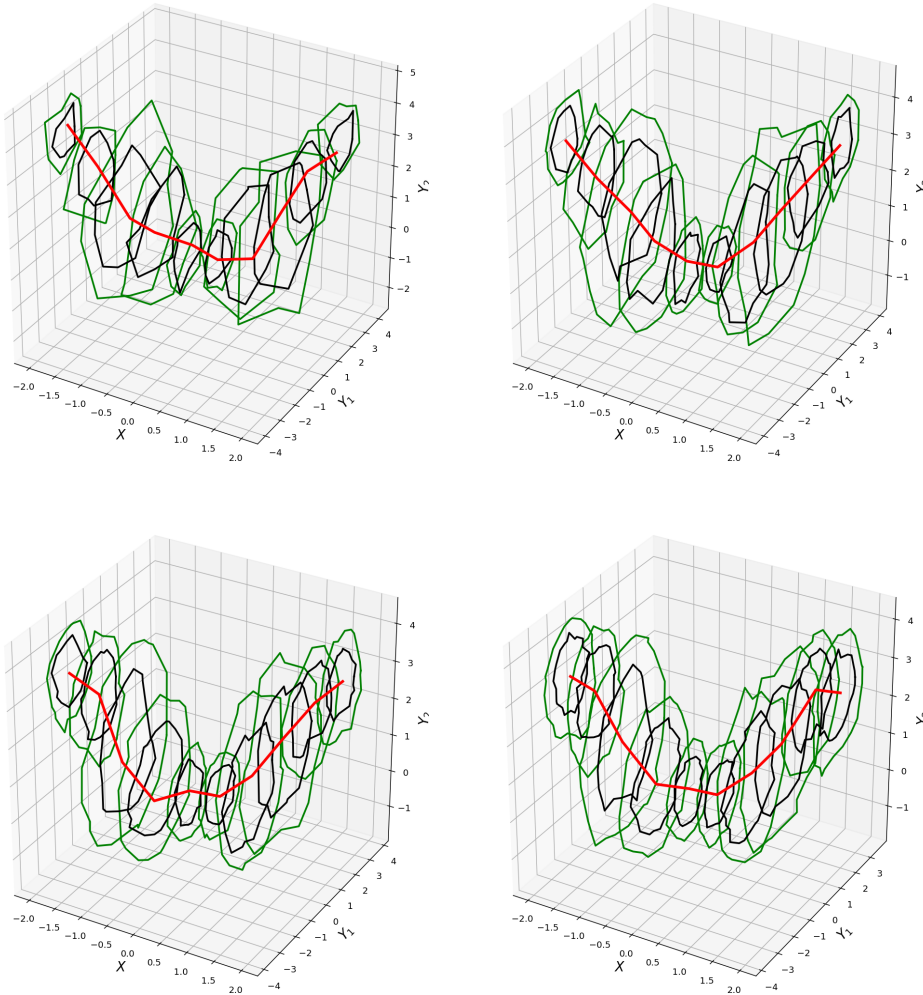


FIG 5. Estimated quantile contours (Model (4.1)) based on k -nearest neighbors weight functions for different choices of k . The sample size is $n = 3601$; $k = 101$ (upper left panel), $k = 256$ (upper right panel), $k = 401$ (lower left panel) and $k = 625$ (lower right panel). The estimated contour orders are $\tau = 0.2$ (black) and 0.4 (green); the estimated conditional center-outward medians are shown in red.

outward contours in $X = 0$ —and compares them with those of Figure 6 (considered as the population contours). The empirical conditional quantiles are computed for $\tau = 0.2, 0.4,$ and 0.8 with Gaussian kernel weights (bandwidths $h = 0.1, 0.2,$ and 0.3) and the k -nearest neighbors weights ($k = 226, 485,$ and 901). The influence of the choice of h and k is clearly seen here: the bigger h (the bigger k), the smoother the estimation of the shape of the contours but also, unfortunately, the worse the estimation of their location.

4.2. Some real-data examples

4.2.1. The CalCOFI oceanographic dataset: depth, temperature, and salinity in the oceans

The dataset “CalCOFI Over 60 Years of Oceanographic Data,” available at <https://www.kaggle.com/sohier/calcofi>, contains the longest (1949-present) and most complete (more than 50,000 sampling stations; sample size $n = 814,247$) time series of oceanographic and larval fish data worldwide. Data collected at depths down to 500 m include temperature, salinity, oxygen, phosphate, silicate, nitrate and nitrite, chlorophyll, transmissometer, PAR, C14 primary productivity, phytoplankton biodiversity, zooplankton biomass, and zooplankton biodiversity. We are focusing here on the influence of $X = \textit{‘depth’}$ (in meters) on the pair $\mathbf{Y} = (\textit{‘temperature,’ ‘salinity’})'$ (in degrees and grams of salt per kilogram of water, respectively).

Figure 10 shows the corresponding 3D observations and the estimated conditional center-outward quantile contours obtained from the same method as in Section 4.1 (nearest neighbors weight function with $k = 6,401$); Figure 11 shows the projections of the same contours

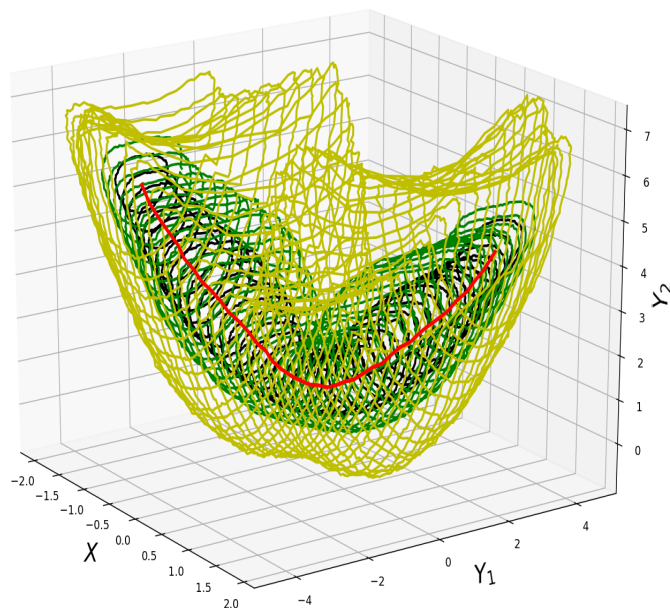


FIG 6. A numerical approximation (estimation based on a simulated sample of size $n = 576,040$) of the quantile contours of order $\tau = 0.2$ (black), 0.4 (green), and 0.8 (yellow) for Model (4.3); the conditional center-outward medians are shown in red.

on the ('depth,' 'salinity') and ('depth,' 'temperature') axes, respectively. Inspection of these figures reveals a nonlinear center-outward conditional median; heteroskedasticity also appears as the area of the conditional quantile regions clearly decreases as a function of depth, while a positive dependence between temperatures and salinity, which is present at the surface, gradually disappears as depth increases. The projection plots of Figure 11 also provide clearer views on marginal dependencies. For example, the decrease of temperature as a function of depth is monotone and almost linear, while the dependence on depth of salinity is more complex, high at shallow depths, lower at medium depths, and higher again at greater

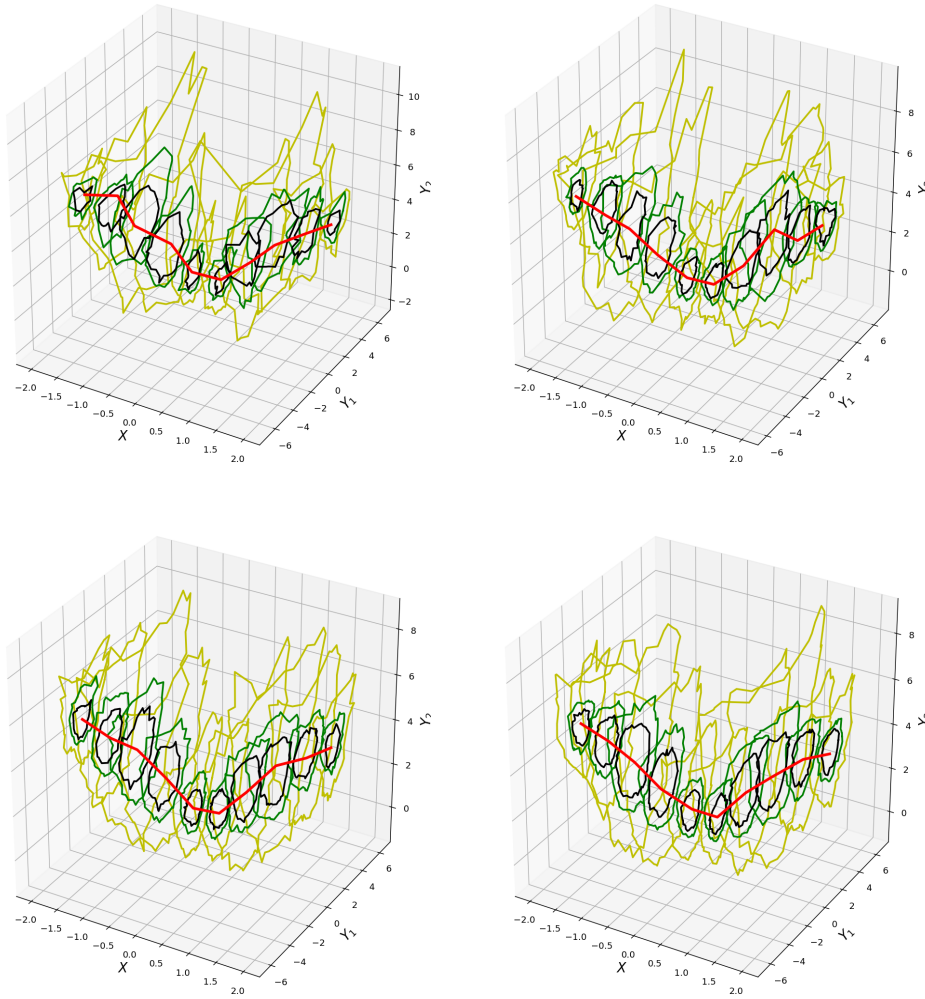


FIG 7. Performance, in Model (4.3), of Gaussian kernel weight functions-based estimation for various choices of the bandwidth. The sample size is $n = 3,601$, and the bandwidths are $h = 0.05$ (upper left panel), $h = 0.1$ (upper right panel), $h = 0.2$ (lower left panel), and $h = 0.3$ (lower right panel). The empirical contour levels are $\tau = 0.2$ (black), 0.4 (green) and 0.8 (yellow); the estimated conditional center-outward medians are shown in red.

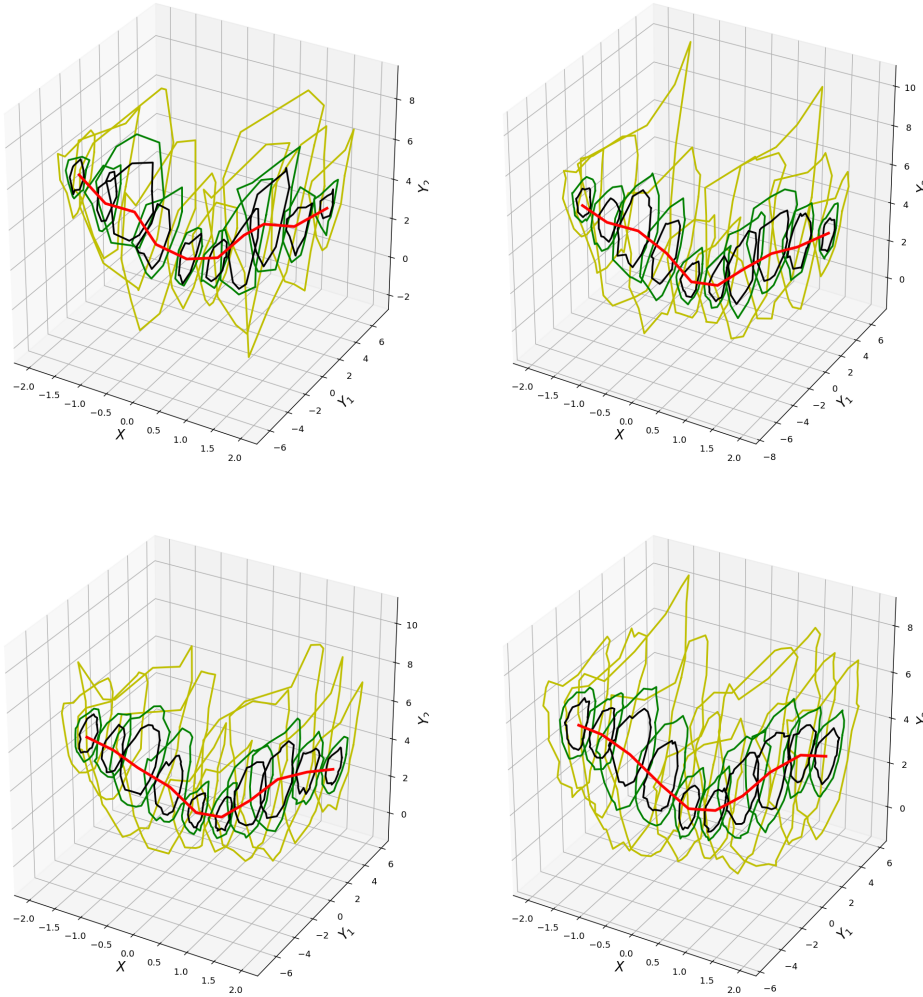


FIG 8. Performance, in Model (4.3), of the k -nearest neighbors weight functions-based estimation for various choices of k . The sample size is $n = 3601$; $k = 101$ (upper left panel), $k = 256$ (upper right panel), $k = 401$ (lower left panel), and $k = 625$ (lower left panel). The empirical contour levels are $\tau = 0.2$ (black), 0.4 (green), and 0.8 (yellow); the estimated conditional center-outward medians are shown in red.

depths. However, these marginal analyses, to some degree, are hiding the heteroskedasticity effects (in particular, the dependence on depth of the relation between salinity and temperature) which are clearly visible in Figure 10. Since the dataset is quite large, we used a nearest neighbors weight function, see the comments about the empirical performance of different weights in Section 4.1.1.

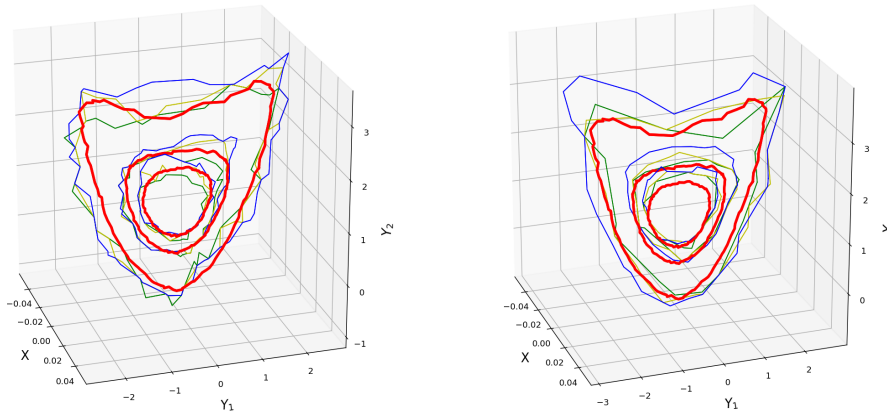


FIG 9. Comparison between the empirical conditional contours $C_{\pm}^{(n)}(\tau|0)$ (levels $\tau = 0.2, 0.4, 0.8$; sample size $n = 3,601$) in Model (4.3) based on Gaussian kernel weight functions (bandwidths $h = 0.1$ (green), $h = 0.2$ (yellow), and $h = 0.3$ (blue)) (left panel) and those in Figure 6, based on k -nearest neighbors weight functions ($k = 226$ (green), $k = 485$ (yellow), and $k = 901$ (blue); sample size $n = 576,040$) (right panel). The center-outward quantiles of Figure 6 (to be considered as population values) are shown in red.

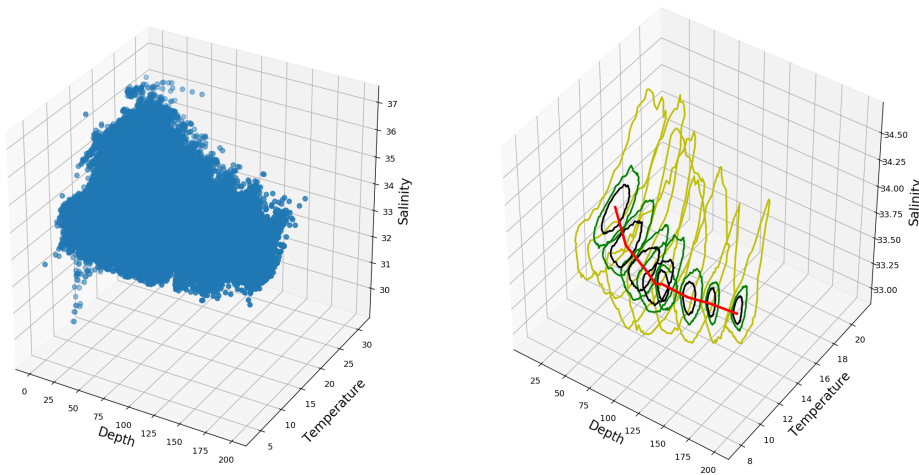


FIG 10. CalCOFI dataset. Left panel: the original dataset ('depth,' 'temperature,' 'salinity') for 'depth' ≤ 200 (sample size for 'depth' ≤ 200 , dropping empty values, is $n = 505,829$). Right panel: the empirical conditional center-outward quantile contours of orders $\tau = 0.2$ (black), $\tau = 0.4$ (green), and $\tau = 0.8$ (yellow) and the empirical conditional center-outward median (red) for the multiple-output regression of $(Y_1, Y_2) = ('temperature,' 'salinity')$ with respect to $X = 'depth.'$ Estimation based on a k -nearest neighbors weight function with $k = 6,401$.

4.2.2. The Female ANSUR 2 dataset: stature, foot length and tibial height of female US Army personnel .

Our second real-data example involves a smaller sample size n . The Female Anthropometric Survey of US Army Personnel (Female ANSUR 2 or Female ANSUR II) featured in this section consists in ninety-three direct measures and 41 derived ones, as well as three-dimensional head, foot, and whole-body scans of $n = 1,986$ women of the United States' army. These measurements were collected between October 4, 2010 and April 5, 2012, in May 2014 and in May 2015; they are available online at <https://www.openlab.psu.edu/ansur2/>.

We want to analyze the influence of the covariate $X = \text{'stature'}$ (in centimeters) on the variable of interest $\mathbf{Y} = (\text{'foot length,' 'tibial height'})$ (both in centimeters). Figure 12 provides a 3D view of the center-outward quantile contours/tubes (levels $\tau = 0.2, 0.4, 0.7$), along with the center-outward regression median (red); Figure 13 shows the projections on the axes of the same contours. Inspecting these two figures reveals the absence of heteroskedasticity, the spherical shape of conditional distributions, and a roughly linear regression. Since the size of the model is not too large, a Gaussian kernel is convenient. The bandwidth was chosen as $h = 15$, which, up to scale changes, corresponds to the choice $h = 0.2$ in Figure 7.

5. Some concluding remarks

5.1. Relation to the recent literature on numerical optimal transportation

The estimation of transport maps beyond the sample points currently is a hot topic, and a fastly developing strand of literature is proposing such estimators. However, the objec-

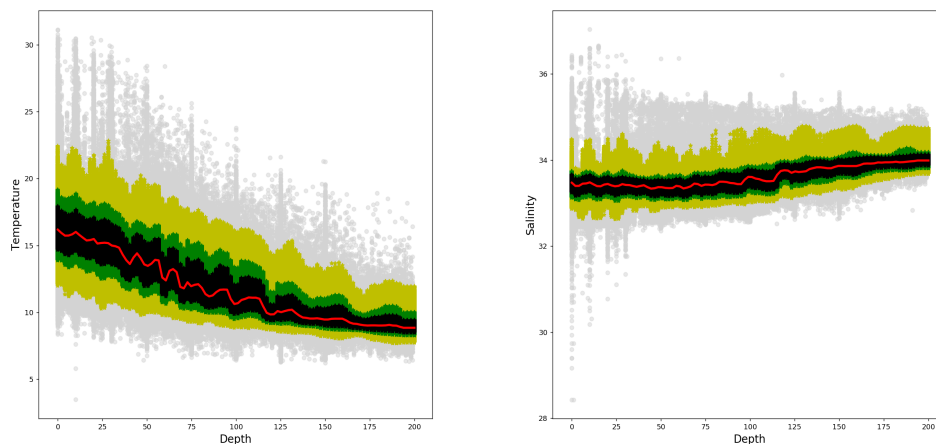


FIG 11. CalCOFI dataset. Left panel: projection of the empirical center-outward quantile regions shown in Figure 10 (orders $\tau = 0.2, 0.4, 0.8$) and median on the axes ('depth,' 'salinity'), for the multiple-output regression of $(Y_1, Y_2) = (\text{'temperature,' 'salinity'})$ with respect to $X = \text{'depth.'}$ Right panel: projection of the same on the axes ('depth,' 'temperature'); see Figure 10 for the color code.

tive of most authors is to reach near-optimal convergence rates, for which they typically impose fairly strong assumptions. Some estimators (Hütter and Rigollet (2021), Manole et al. (2021)) are computationally quite heavy and sometimes numerically almost infeasible; others (Pooladian and Niles-Weed, 2021) use a regularized version (Cuturi (2013)) of the optimal transport problem to provide consistent near-optimal estimators, which needs stringent assumptions on the shape of the underlying distributions—such as being compactly supported, with densities bounded away from 0 and ∞ over their convex supports. This is redhibitory in our case, since the density of U_d (a choice which plays an essential role in the interpretation of transports as quantile functions) is unbounded at $\mathbf{0}$. Other solutions (Makkuva et al. (2020) and González-Sanz et al. (2022)) are based on deep learning and neural network methods; they achieve excellent empirical performance, but the lack of theoretical results for the first one, the Lipschitz constraint on transport maps for the second, preclude their use in this quantile regression context.

We also could estimate the conditional quantiles through the optimal map from the uniform reference measure to $P_n^{w(\mathbf{x})}$. From a numerical point of view, however, this would lead to the computation of a semi-discrete optimal transportation plan, which has complexity $O(n^{d/2})$, hence is unfeasible even for moderate d . While the computational complexity of our procedure does not depend on the dimension, its statistical performance does (see Fournier and Guillin (2015)) and, in that sense, we do not escape the curse of dimensionality—up to the case where P is finitely supported, see del Barrio, González-Sanz and Loubes (2021). Despite the fact that the literature on the computation of such maps is growing quite fastly (see Lévy, Mohayae and von Hausegger (2020); Gallouët and Mérigot

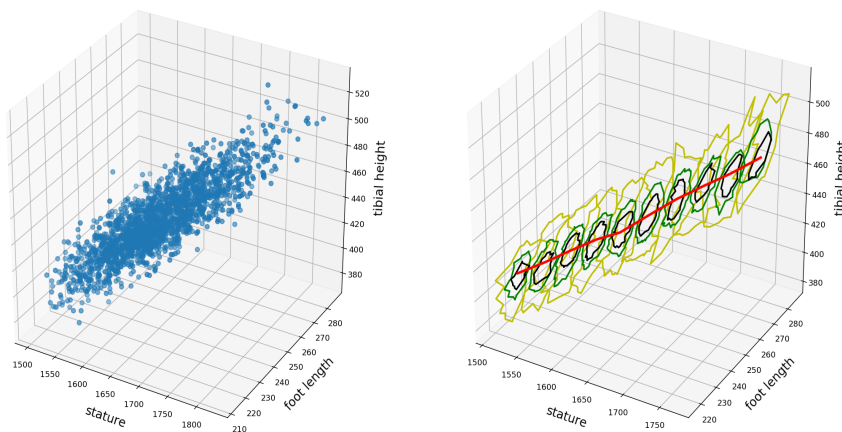


FIG 12. ANSUR 2 dataset (sample size $n = 1,986$). Left panel: the original dataset of $X = \text{'stature'}$ and $(Y_1, Y_2) = (\text{'foot length,' 'tibial height'})$. Right panel: the empirical conditional center-outward quantile contours of orders $\tau = 0.2$ (black), $\tau = 0.4$ (green), $\tau = 0.7$ (yellow) and the empirical conditional center-outward median (red) for the multiple-output regression of (Y_1, Y_2) with respect to X ; estimation based on a Gaussian kernel weight function with bandwidth $h = 15$.

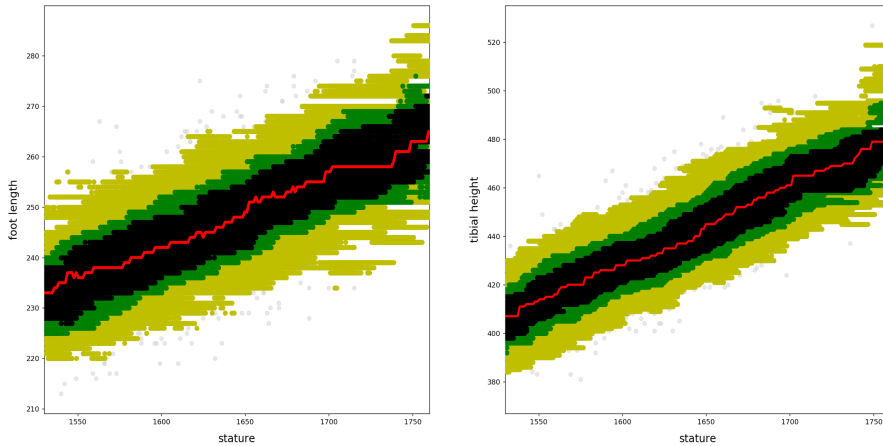


FIG 13. ANSUR 2 dataset (sample size $n = 1,986$). Left panel: projection of the empirical center-outward quantile regions (orders $\tau = 0.2, 0.4, 0.8$) and median on the axes ('stature,' 'foot length'), for the multiple-output regression of $(Y_1, Y_2) = ('foot\ length', 'tibial\ height')$ with respect to $X = 'stature.'$ Right panel: projection of the same on the axes ('stature,' 'tibial height'); see Figure 12 for the color code.

(2018); Meyron (2019); de Goes et al. (2012)), the existing methods are restricted to dimension two, sometimes three. A further issue is that the solution of the semi-discrete problem does not produce quantile contours but creates a Voronoi tessellation of \mathbb{S}_d , each piece of which is mapped to a single sample point.

5.2. Conclusions and perspectives for further developments

Building on the concepts of center-outward quantiles recently developed in Hallin et al. (2021), we are proposing here a fully nonparametric solution to the problem of nonparametric multiple-output quantile regression. Contrary to earlier attempts, our solution is enjoying the quintessential property that the (conditional) probability content of its quantile regions is under control irrespective of the underlying distribution. This is only a first step into the multifarious applications of multiple-output quantile regression, though. Due to the minimality of the assumptions it requires, a completely agnostic nonparametric approach indeed is attractive, but also comes at a cost: linear or polynomial quantile regression remain justified whenever a priori knowledge of the analytical form of the regression is available and should be taken advantage of. A center-outward version of the results of Carlier, Chernozhukov and Galichon (2016), thus, is highly desirable. Single-output quantile regression has been considered in a variety of contexts: survival analysis, longitudinal data, instrumental variable regression, directional, functional, and high-dimensional data, ... Quantile regression versions of time-series models such as the quantile autoregressive model also have been investigated (Koenker and Xiao, 2006). All these applications call for multiple-output extensions with important real-life consequences; they should and can be based on the concept of center-outward quantile, regions, and contours and are the subject of our ongoing research.

References

- BAGH, A. and WETS, R. J. (1996). Convergence of set-valued mappings: equi-outer semi-continuity. *Set-Valued Anal.* **4** 333–360.
- BRECKLING, J. and CHAMBERS, R. (1988). M-Quantiles. *Biometrika* **75** 761–771.
- CARLIER, G., CHERNOZHUKOV, V. and GALICHON, A. (2016). Vector quantile regression: An optimal transport approach. *The Annals of Statistics* **44** 1165–1192.
- CHERNOZHUKOV, V., GALICHON, A., HALLIN, M. and HENRY, M. (2017). Monge-Kantorovich depth, quantiles, ranks and signs. *The Annals of Statistics* **45** 223–256.
- CUTURI, M. (2013). Sinkhorn distances: lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* (C. J. C. BURGESS, L. BOTTOU, M. WELLING, Z. GHAHRAMANI and K. Q. WEINBERGER, eds.) **26**. Curran Associates, Inc.
- DE GOES, F., BREEDEN, K., OSTROMOUKHOV, V. and DESBRUN, M. (2012). Blue noise through optimal transport. *ACM Trans. on Graphics* **31** 1–11.
- DEB, N. and SEN, B. (2021). Multivariate rank-based distribution-free nonparametric testing using measure transportation. *Journal of the American Statistical Association* **0** 1–16.
- DEL BARRIO, E., GONZÁLEZ-SANZ, A. and HALLIN, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *J. Multivar. Anal.* **180** 104671.
- DEL BARRIO, E., GONZÁLEZ-SANZ, A. and LOUBES, J.-M. (2021). A central limit theorem for semidiscrete Wasserstein distances. arXiv:2105.11721.
- DEL BARRIO, E. and LOUBES, J.-M. (2019). Central limit theorems for empirical transportation cost in general dimension. *Ann. Probab.* **47** 926–951.
- DEVROYE, L. (1982). Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **61** 467–481.
- DEVROYE, L., GYORFI, L., KRZYŻAK, A. and LUGOSI, G. (1994). On the strong universal consistency of nearest neighbor regression function estimates. *The Annals of Statistics* **22** 1371–1385.
- FELDMAN, S., BATES, S. and ROMANO, Y. (2021). Calibrated multiple-output quantile regression with representation learning. arXiv:2110.00816.
- FIGALLI, A. (2018). On the continuity of center-outward distribution and quantile functions. *Nonlinear Anal.* **177** 413–21.
- FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probability Theory and Related Fields* **162** 707–738.
- GALLOUËT, T. and MÉRIGOT, Q. (2018). A Lagrangian scheme à la Brenier for the incompressible Euler equations. *Found. Comput. Math.* **18** 835–865.
- GHOSAL, P. and SEN, B. (2019). Multivariate ranks and quantiles using optimal transportation and applications to goodness-of-fit testing. arXiv:1905.05340.
- GONZÁLEZ-SANZ, A., DE LARA, L., BÉTHUNE, L. and LOUBES, J.-M. (2022). GAN estimation of Lipschitz optimal transport maps. arXiv:2202.07965.
- GYÖRFI, L., KOHLER, M., KRZYŻAK, A. and WALK, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer Series in Statistics. Springer.
- HALLIN, M. (2017). On distribution and quantile functions, ranks and signs in \mathbb{R}^d : a measure transportation approach. Available at <https://ideas.repec.org/p/eca/wpaper/2013-258262.html>.
- HALLIN, M. (2022). Measure transportation and statistical decision theory. *Annual Review*

- of *Statistics and its Application* **9** 401–424.
- HALLIN, M., HLUBINKA, D. and HUDECOVÁ, Š. (2022). Efficient Fully Distribution-Free Center-Outward Rank Tests for Multiple-Output Regression and MANOVA. *Journal of the American Statistical Association* **0** 1–43.
- HALLIN, M., LA VECCHIA, D. and LIU, H. (2020). Center-outward R-estimation for semi-parametric VARMA models. *Journal of the American Statistical Association* **0** 1–14.
- HALLIN, M., LA VECCHIA, D. and LIU, H. (2022). Rank-based testing for semiparametric VAR models: a measure transportation approach. *Bernoulli* **0** 1–14.
- HALLIN, M., PAINDAVEINE, D. and ŠIMAN, M. (2010). Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *The Annals of Statistics* **38** 635–703.
- HALLIN, M. and ŠIMAN, M. (2018). Multiple-output quantile regression. In *Handbook of Quantile Regression* (R. KOENKER, V. CHERNOZHUKOV, X. HE and L. PENG, eds.) 185–207. CRC Press.
- HALLIN, M., LU, Z., PAINDAVEINE, D. and ŠIMAN, M. (2015). Local bilinear multiple-output quantile/depth regression. *Bernoulli* **21** 1435–1466.
- HALLIN, M., DEL BARRIO, E., CUESTA-ALBERTOS, J. and MATRÁN, C. (2021). Distribution and quantile functions, ranks and signs in dimension d : a measure transportation approach. *The Annals of Statistics* **49** 1139–1165.
- HÜTTER, J.-C. and RIGOLLET, P. (2021). Minimax estimation of smooth optimal transport maps. *The Annals of Statistics* **49** 1166–1194.
- KOENKER, R. (2005). *Quantile Regression. Econometric Society Monographs*. Cambridge University Press.
- KOENKER, R. and BASSETT, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.
- KOENKER, R. and XIAO, Z. (2006). Quantile Autoregression [with Comments, Rejoinder]. *Journal of the American Statistical Association* **101** 980–1006.
- KOENKER, R., CHERNOZHUKOV, V., HE, X. and PENG, L., eds. (2018). *Handbook of Quantile Regression*. CRC Press.
- KONEN, D. and PAINDAVEINE, D. (2022). Multivariate ρ -quantiles: a spatial approach. *Bernoulli*.
- KONG, L. and MIZERA, I. (2012). Quantile tomography: using quantiles with multivariate data. *Statistica Sinica* **22** 1589–1610.
- LEHMANN, E. L. and ROMANO, J. P. (2005). *Testing Statistical Hypotheses*. Springer.
- LÉVY, B., MOHAYAEI, R. and VON HAUSEGGER, S. (2020). A fast semi-discrete optimal transport algorithm for a unique reconstruction of the early Universe. arXiv:2012.09074.
- MAKKUVA, A., TAGHVAEI, A., OH, S. and LEE, J. (2020). Optimal transport mapping via input convex neural networks. In *International Conference on Machine Learning* 6672–6681. PMLR.
- MANOLE, T., BALAKRISHNAN, S., NILES-WEED, J. and WASSERMAN, L. (2021). Plugin estimation of smooth optimal transport maps. arXiv:2107.12364.
- MCCANN, R. J. (1995). Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal* **80** 309–323.
- MERLO, L., PETRELLA, L., SALVATI, N. and TZAVIDIS, N. (2022). Marginal M-quantile regression for multivariate dependent data. *Computational Statistics & Data Analysis* 107500.
- MEYRON, J. (2019). Initialization procedures for discrete and semi-discrete optimal transport. *Computer-Aided Design* **115** 13–22.
- PAINDAVEINE, D. and ŠIMAN, M. (2011). On directional multiple-output quantile regres-

- sion. *Journal of Multivariate Analysis* **102** 193-212.
- PEYRÉ, G. and CUTURI, M. (2019). Computational Optimal Transport: With Applications to Data Science. *Foundations and Trends® in Machine Learning* **11** 355-607.
- POOLADIAN, A.-A. and NILES-WEED, J. (2021). Entropic estimation of optimal transport maps. arXiv preprint arXiv:2109.12004.
- ROCKAFELLAR, R. T. (1970). *Convex Analysis. Princeton Mathematical Series*. Princeton University Press, Princeton, N. J.
- ROCKAFELLAR, R. and WETS, R. J. B. (1998). *Variational Analysis*. Springer Verlag, Heidelberg, Berlin, New York.
- SERFLING, R. (2002). Quantile functions for multivariate analysis: approaches and applications. *Statistica Neerlandica* **56** 214-232.
- SERFLING, R. (2019a). Depth functions on general data spaces, I. Perspectives, with consideration of “density” and “local” depths”. Available at <https://www.utdallas.edu/~serfling>.
- SERFLING, R. (2019b). Depth functions on general data spaces, II. Formulation and maximality, with consideration of the Tukey, projection, spatial, and “contour” depths. Available at <https://www.utdallas.edu/~serfling>.
- SERFLING, R. and ZUO, Y. (2000). General notions of statistical depth function. *The Annals of Statistics* **28** 461-482.
- SHI, H., DRTON, M., HALLIN, M. and HAN, F. (2021a). Center-outward sign- and rank-based quadrant, Spearman, and Kendall tests for multivariate independence. arXiv:2111.15567.
- SHI, H., DRTON, M., HALLIN, M. and HAN, F. (2021b). On universally consistent and fully distribution-free rank tests of vector independence. *The Annals of Statistics* **0** 1-14.
- SOHN, K., LEE, H. and YAN, X. (2015). Learning Structured Output Representation using Deep Conditional Generative Models. In *Advances in Neural Information Processing Systems* (C. CORTES, N. LAWRENCE, D. LEE, M. SUGIYAMA and R. GARNETT, eds.) **28**. Curran Associates, Inc.
- STONE, C. J. (1977). Consistent Nonparametric Regression. *Ann. Statist.* **5** 595-620.
- TUKEY, J. W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians, ver, 1975* **2** 523-531.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York, NY.
- VILLANI, C. (2003). *Topics in Optimal Transportation*. American Mathematical Society, Providence, Rhode Island.
- VILLANI, C. (2008). *Optimal Transport: Old and New*. Springer-Verlag Berlin Heidelberg.

Appendix

A.1. Proofs for Section 3

The convergence described of Theorem 3.2 is based on the topology of set-valued maps, in particular the Graphical convergence (or Painlevé-Kuratowski convergence of the graphs). Recall from Rockafellar and Wets (1998) that a sequence of set-valued maps $\{T_n\}_n$ converges *graphically* to another set-valued map T if

- the outer limit $\limsup_n T_n$ —which is the set of $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d$ for which there exists a sequence $\{(\mathbf{x}_n, \mathbf{y}_n)\}$ with $(\mathbf{x}_n, \mathbf{y}_n) \in T_n$ containing a subsequence which converges

- to (\mathbf{x}, \mathbf{y}) —exists and coincides with T and
- the inner limit $\liminf_n T_n$ —which is the set of $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d$ for which there exists a sequence $\{(\mathbf{x}_n, \mathbf{y}_n)\}$, with $(\mathbf{x}_n, \mathbf{y}_n) \in T_n$, which converges to (\mathbf{x}, \mathbf{y}) —exists and coincides with T .

We start with some necessary properties of the population and empirical conditional quantiles and some auxiliary results. The following lemma states that the estimated conditional probability converges weakly in probability to its population counterpart. Recall from Theorem 1.12.4 in [van der Vaart and Wellner \(1996\)](#) that weak convergence can be measured in terms of the bounded Lipschitz norm, which is defined for $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ as

$$d_{BL}(\mu, \nu) := \sup_{f \in \mathcal{F}_{BL}(\mathbb{R}^d)} |\mathbb{E}_{\mathbf{Z} \sim \mu}(f(\mathbf{Z})) - \mathbb{E}_{\mathbf{W} \sim \nu}(f(\mathbf{W}))|,$$

where the class \mathcal{F}_{BL} is the class of functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $|f(\mathbf{z}_1) - f(\mathbf{z}_2)| \leq |\mathbf{z}_1 - \mathbf{z}_2|$ and $|f(\mathbf{z}_1)| \leq 1$, for all $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^d$.

Lemma A.1. *For any $\epsilon > 0$, under the assumptions of Theorem 3.2, we have*

$$\mathbb{P}\left(d_{BL}(P_n^{w(\mathbf{X})}, P_{\mathbf{Y}|\mathbf{X}}) > \epsilon\right) \rightarrow 0.$$

See Section A.2 for the proof.

Let $\pi^{(n)}(\mathbf{x})$ be a solution of (3.5) and

$$\pi^*(\mathbf{x}) := (\mathbf{Id} \times \mathbf{Q}_{\pm}(\cdot | \mathbf{x})) \# U_d.$$

Note that, for each value of \mathbf{x} ,

- (i) there exists a sequence of differentiable convex functions $\psi^{(n)}(\cdot | \mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$, $n \in \mathbb{N}$, such that

$$\mathbf{Q}_{w, \pm}^{(n)}(\mathbf{u}_j | \mathbf{x}) = \nabla \psi^{(n)}(\mathbf{u}_j | \mathbf{x}) \text{ for } j = 1, \dots, k, \text{ and } n \in \mathbb{N};$$

- (ii) there exists a convex function $\psi(\cdot | \mathbf{X} = \mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ such that

$$\mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X} = \mathbf{x}) = \nabla \psi(\mathbf{u} | \mathbf{X} = \mathbf{x}) \text{ for } U_d\text{-a.e. } \mathbf{u} \in \mathbb{S}_d.$$

Using obvious notation, it holds that $\pi^*(\mathbf{x})$ and $\pi^{(n)}(\mathbf{x})$, for $n \in \mathbb{N}$, have cyclically monotone supports. Moreover, as a consequence of Corollary 3.1, we have

$$\text{supp}(\pi^{(n)}(\mathbf{x})) \subset \partial \tilde{\psi}^{(n)}(\cdot | \mathbf{x}) \text{ and } \text{supp}(\pi^*(\mathbf{x})) \subset \partial \psi(\cdot | \mathbf{X} = \mathbf{x}), \quad (\text{A.1})$$

possibly for some other sequence $\tilde{\psi}^{(n)}(\cdot | \mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ of convex functions such that

$$\mathbf{Q}_{w, \pm}^{(n)}(\mathbf{u}_j | \mathbf{x}) \in \partial \tilde{\psi}^{(n)}(\mathbf{u}_j | \mathbf{x}) \text{ for } j = 1, \dots, k, \text{ and } n \in \mathbb{N}. \quad (\text{A.2})$$

The following result then follows from Lemma 9 and Corollary 14 in [McCann \(1995\)](#).

Lemma A.2. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be such that $\mu \ll \ell_d$ is supported on a convex set. Let $\{\mu_n\}_{n \in \mathbb{N}}$ and $\{\nu_n\}_{n \in \mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d)$ converge weakly as $n \rightarrow \infty$ to μ and ν , respectively. Suppose, moreover, that there exists a sequence of probability measures $\{\pi_n\}_{n \in \mathbb{N}} \subset \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ with marginals μ_n and ν_n such that, for some sequence of convex functions $\{\phi_n\}_{n \in \mathbb{N}}$, it holds that, for all $n \in \mathbb{N}$, $\text{supp}(\pi_n) \subset \partial \phi_n$. Then,*

- (i) $\{\pi_n\}_n$ converges weakly as $n \rightarrow \infty$ to $\pi^* = (\mathbf{Id} \times \nabla\phi) \# \mu$, where $\nabla\phi$ is the gradient of a convex function ϕ pushing μ forward to ν , and
- (ii) there exists a sequence $\{a_n\}_{n \in \mathbb{N}} \subset \mathbb{R}$ such that, μ -a.e., $\phi_n + a_n \rightarrow \phi$ and $\partial\phi_n \rightarrow \partial\phi$ graphically as $n \rightarrow \infty$.

See Section A.2 for the proof.

Proof of Theorem 3.2. Suppose that there exist $\mathbf{u}_0 \in \mathbb{S}_d$ and $\epsilon_0 > 0$ such that

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left(\mathbf{Q}_{w,\pm}^{(n)}(\mathbf{u}_0 | \mathbf{x}) \not\subset \mathbf{Q}_{\pm}(\mathbf{u}_0 | \mathbf{X}) + \epsilon_0 \mathbb{S}_d \right) = \delta > 0.$$

We can find a subsequence n_k such that

$$\lim_{k \rightarrow \infty} \mathbb{P} \left(\mathbf{Q}_{w,\pm}^{(n_k)}(\mathbf{u}_0 | w(\mathbf{X}; \mathbf{X}^{[n_k]})) \not\subset \mathbf{Q}_{\pm}(\mathbf{u}_0 | \mathbf{X}) + \epsilon_0 \mathbb{S}_d \right) = \delta > 0. \quad (\text{A.3})$$

Note that the space of probability distributions on \mathbb{R}^d , endowed with the bounded Lipschitz metric, is separable and complete, see Theorem 1.12.4 in [van der Vaart and Wellner \(1996\)](#). Therefore, the convergence described in Lemma A.1 implies that, for the subsequence n_k , there exists a further sub sequence n_{k_i} such that the event

$$\Omega_0 = \left(\sup_{f \in \mathcal{F}_{BL}(\mathbb{R}^d)} \left| \sum_{j=1}^{n_{k_i}} w_j(\mathbf{X}; \mathbf{X}^{(n_{k_i})}) f(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y}) | \mathbf{X}) \right| \rightarrow 0 \right) \quad (\text{A.4})$$

has probability one. Let $\mathbf{x} = \mathbf{X}(\omega)$ with $\omega \in \Omega_0$. Then, by Lemma A.2, we have, as $i \rightarrow \infty$,

$$\partial\tilde{\psi}^{(n_{k_i})}(\cdot | \mathbf{x}) \rightarrow \mathbf{Q}_{\pm}(\cdot | \mathbf{X} = \mathbf{x}) \text{ graphically.}$$

This implies, by Theorem 8.3 in [Bagh and Wets \(1996\)](#), that there exists $i_0 \in \mathbb{N}$ such that, for all $i > i_0$,

$$\partial\tilde{\psi}^{(n_{k_i})}(\mathbf{u}_0 | \mathbf{x}) \subset \mathbf{Q}_{\pm}(\mathbf{u}_0 | \mathbf{X} = \mathbf{x}) + \epsilon_0 \mathbb{S}_d. \quad (\text{A.5})$$

From (A.2) and the fact that (A.5) holds with probability one, we deduce that

$$\mathbb{P} \left(\bigcap_{i_0 \in \mathbb{N}} \bigcup_{i \geq i_0} \mathbf{Q}_{\pm}^{(n_{k_i})}(\mathbf{u}_0 | \mathbf{X}) \not\subset \mathbf{Q}_{w,\pm}(\mathbf{u}_0 | \mathbf{X}) + \epsilon_0 \mathbb{S}_d \right) = 0, \quad (\text{A.6})$$

which implies that

$$\limsup_{i \rightarrow \infty} \mathbb{P} \left(\mathbf{Q}_{\pm}^{(n_{k_i})}(\mathbf{u}_0 | \mathbf{X}) \not\subset \mathbf{Q}_{w,\pm}(\mathbf{u}_0 | \mathbf{X}) + \epsilon_0 \mathbb{S}_d \right) \leq 0.$$

This contradicts (A.3).

The rest of the proof follows from compactness arguments and a refined use of Theorem 8.3. in [Bagh and Wets \(1996\)](#). Suppose that, for some $q_0 \in (0, 1)$ and $\epsilon_0 > 0$, and along a subsequence n_k , we have

$$\lim_{k \rightarrow \infty} \mathbb{P} \left(\mathbb{C}_{\pm}^{(n_k)}(q_0 | \mathbf{X}) \not\subset \mathbb{C}_{\pm}(q_0 | \mathbf{X}) + \epsilon_0 \mathbb{S}_d \right) = \delta > 0. \quad (\text{A.7})$$

Theorem 8.3. in [Bagh and Wets \(1996\)](#) and Lemma [A.2](#) jointly imply the existence of a further subsequence n_{k_i} , such that, for every $\bar{\mathbf{u}} \in \mathbb{S}_d$, there exists $I_{\bar{\mathbf{u}}} \in \mathbb{N}$ and $\lambda_{\bar{\mathbf{u}}} > 0$ satisfying

$$\mathbf{Q}_{\pm}^{(n_{k_i})}(\mathbf{u} | \mathbf{x}) \in \mathbf{Q}_{\pm}(\bar{\mathbf{u}} | X = \mathbf{x}) + \epsilon_0 \mathbb{S}_d \text{ for all } \mathbf{u} \in \bar{\mathbf{u}} + \lambda_{\bar{\mathbf{u}}} \mathbb{S}_d \text{ and all } i > I_{\bar{\mathbf{u}}}.$$

Note that $q\mathbb{S}_d \subset \bigcup_{\bar{\mathbf{u}}: |\bar{\mathbf{u}}| \leq q} \bar{\mathbf{u}} + \lambda_{\bar{\mathbf{u}}} \mathbb{S}_d$. Since $q\bar{\mathbb{S}}_d$ is compact, there exists a finite covering $q\mathbb{S}_d \subset \bigcup_{k=1}^{K_\epsilon} \bar{\mathbf{u}}_k + \delta_{\bar{\mathbf{u}}_k} \mathbb{S}_d$. Set $I_0 = \max(I_{\bar{\mathbf{u}}_1}, \dots, I_{\bar{\mathbf{u}}_{K_\epsilon}})$: then for all $i > I_{\bar{\mathbf{u}}}$, we have

$$\mathbb{C}_{\pm}^{(n_{k_i})}(q_0 | \mathbf{x}) \subset \mathbb{C}_{\pm}(q_0 | \mathbf{x}) + \epsilon_0 \mathbb{S}_d,$$

which holds with probability one and, using the same argument as for [\(A.5\)](#), contradicts [\(A.7\)](#). The same reasoning holds also for the contours. \square

Proof of Theorem 3.3. Under Assumption (R), it follows from [del Barrio, González Sanz and Hallin \(2020\)](#) that, for $\omega \in \Omega_0 \subseteq \Omega$ where $\mathbb{P}(\Omega_0) = 1$, the center-outward quantile function $\mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X} = \mathbf{x} := \mathbf{X}(\omega))$ is a singleton for U_d -almost all values of $\mathbf{u} \in \mathbb{S}_d$. Therefore, we adopt here the slight abuse of notation commented before [Theorem 3.3](#) and write $\mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X} = \mathbf{x}) = \{\mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X} = \mathbf{x})\}$. Set $\mathbf{u} \in \mathbb{S}_d \setminus \{\mathbf{0}\}$ and note that, since $\mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X})$ and $\mathbf{Q}_{w,\pm}^{(n)}(\mathbf{u} | \mathbf{x})$ are a.s. singletons, we have, as n and N tend to infinity,

$$\mathbb{P}\left(|\mathbf{Q}_{w,\pm}^{(n)}(\mathbf{u} | \mathbf{x}) - \mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X})| > \epsilon\right) \rightarrow 0.$$

Let K be a compact subset of $\mathbb{S}_d \setminus \{\mathbf{0}\}$. In order to establish uniform convergence in K , suppose that the contrary holds. Since the space of continuous functions from K to \mathbb{R}^d (endowed with the topology of uniform convergence) is complete and separable, there exists a subsequence n_k such that, for some $\delta > 0$, the probability of

$$\Omega' = \left(\sup_{\mathbf{u} \in K} |\mathbf{Q}_{w,\pm}^{(n_k)}(\mathbf{u} | \mathbf{X}) - \mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X})| \rightarrow \delta\right), \quad (\text{A.8})$$

is one. Set $\omega \in \Omega'$ and consider $\mathbf{x} = \mathbf{X}(\omega)$. There exists a sequence $\{\mathbf{u}_{n_k}\} \subset K$ such that

$$|\mathbf{Q}_{w,\pm}^{(n_k)}(\mathbf{u}_{n_k} | \mathbf{x}) - \mathbf{Q}_{\pm}(\mathbf{u}_{n_k} | \mathbf{X} = \mathbf{x})| \rightarrow \delta,$$

for a possibly different $\delta > 0$. Since the sequence $\{\mathbf{u}_{n_k}\}_k$ is in K , it admits at least one point of accumulation $\bar{\mathbf{u}} \in K$. Hence,

$$\begin{aligned} \liminf_{k \rightarrow \infty} (|\mathbf{Q}_{w,\pm}^{(n_k)}(\mathbf{u}_{n_k} | \mathbf{x}) - \mathbf{Q}_{\pm}(\bar{\mathbf{u}} | \mathbf{X} = \mathbf{x})| \\ + |\mathbf{Q}_{\pm}(\bar{\mathbf{u}} | \mathbf{X} = \mathbf{x}) - \mathbf{Q}_{\pm}(\mathbf{u}_{n_k} | \mathbf{X} = \mathbf{x})|) \geq \delta, \end{aligned}$$

where the second term tends to 0 by the continuity of $\mathbf{u} \mapsto \mathbf{Q}_{\pm}(\mathbf{u} | \mathbf{X} = \mathbf{x})$. This implies that

$$\liminf_{k \rightarrow \infty} |\mathbf{Q}_{w,\pm}^{(n_k)}(\mathbf{u}_{n_k} | \mathbf{x}) - \mathbf{Q}_{\pm}(\bar{\mathbf{u}} | \mathbf{X} = \mathbf{x})| \geq \delta. \quad (\text{A.9})$$

But Lemma [A.2](#) entails the Graphical convergence to $\mathbf{Q}_{\pm}(\cdot | \mathbf{X} = \mathbf{x})$ of $\mathbf{Q}_{w,\pm}^{(n_k)}(\cdot | \mathbf{x})$, which contradicts [\(A.9\)](#). The desired uniformity over K follows.

Finally, the convergence of the contours is a consequence of the previous result on the regions. \square

Proof of Corollary 3.4. To prove (3.8), let us show that, for any subsequence n_k , there exists a further subsequence converging a.s. To avoid repetitions, assume that $N = N(n)$ is such that $N \rightarrow \infty$ as $n \rightarrow \infty$.

Leaving aside the singular points, let us consider the empirical and population *quantile rings*

$$\mathbb{C}_{\pm}^{(n)}(\epsilon, \tau | \mathbf{x}) := \mathbb{C}_{\pm}^{(n)}(\tau | \mathbf{x}) \setminus \mathbb{C}_{\pm}^{(n)}(\epsilon | \mathbf{x})$$

and

$$\mathbb{C}_{\pm}(\epsilon, \tau | \mathbf{x}) := \mathbb{C}_{\pm}(\tau | \mathbf{x}) \setminus \mathbb{C}_{\pm}(\epsilon | \mathbf{x}),$$

respectively. Theorem 3.3 yields, for all $0 < \epsilon < \tau$, as $n \rightarrow \infty$,

$$\mathbb{P} \left(\mathbf{Y} \in \mathbb{C}_{\pm}^{(n)}(\epsilon, \tau | \mathbf{X}) \text{ and } \mathbf{Y} \notin \mathbb{C}_{\pm}(\epsilon, \tau | \mathbf{X}) \mid \mathbf{X} \right) \xrightarrow{\mathbb{P}} 0. \quad (\text{A.10})$$

Let $\{\epsilon_j\}_{j \in \mathbb{N}}$ be a monotone sequence tending to 0. For $j = 1$ there exists a further subsequence n_k^1 , say, and a subset Ω_1 of Ω such that $\mathbb{P}(\Omega_1) = 1$ such that, for every $\mathbf{x} = \mathbf{X}(\omega)$ with $\omega \in \Omega_1$,

$$\mathbb{P} \left(\mathbf{Y} \in \mathbb{C}_{\pm}^{(n_k^1)}(\epsilon_1, \tau | \mathbf{x}) \text{ and } \mathbf{Y} \notin \mathbb{C}_{\pm}(\epsilon_1, \tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right) \rightarrow 0. \quad (\text{A.11})$$

By definition, $\mathbb{P} \left(\mathbb{C}_{\pm}(\epsilon_1, \tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right) = \tau - \epsilon_1$; therefore, in view of (A.11),

$$\mathbb{P} \left(\mathbb{C}_{\pm}^{(n_k^1)}(\epsilon_1, \tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right) \rightarrow \tau - \epsilon_1.$$

Repeating the argument for $j = 2$, there exists a set $\Omega_2 \subset \Omega$, with $\mathbb{P}(\Omega_2) = 1$, and a subsequence (call it n_k^2) of n_k^1 such that, for every $\omega \in \Omega_2$ and $\mathbf{x} = \mathbf{X}(\omega)$,

$$\mathbb{P} \left(\mathbb{C}_{\pm}^{(n_k^2)}(\epsilon_2, \tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right) \rightarrow \tau - \epsilon_2.$$

This argument can be repeated for each $j \in \mathbb{N}$ and the set $\Omega_0 = \bigcap_{j \in \mathbb{N}} \Omega_j$ has probability one. Set $\omega \in \Omega_0 \subset \Omega$ and $\mathbf{x} = \mathbf{X}(\omega)$: then there exists k_1 such that

$$\left| \mathbb{P} \left(\mathbb{C}_{\pm}^{(n_{k_1})}(\epsilon_1, \tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right) - \tau \right| < 2\epsilon_1.$$

Analogously, for each m , there exist k_j such that

$$\left| \mathbb{P} \left(\mathbb{C}_{\pm}^{(n_{k_j})}(\epsilon_j, \tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right) - \tau \right| < 2\epsilon_j.$$

Therefore, we obtain the limit

$$\mathbb{P} \left(\mathbb{C}_{\pm}^{(n_{k_j})}(\epsilon_j, \tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right) \rightarrow \tau$$

and, noticing that

$$\mathbb{P} \left(\mathbb{C}_{\pm}^{(n_{k_j})}(\epsilon_j, \tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right) \leq \mathbb{P} \left(\mathbb{C}_{\pm}^{(n_{k_j})}(\tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right),$$

also the asymptotic upper bound

$$\liminf \mathbb{P} \left(\mathbb{C}_{\pm}^{(n_{k_j})}(\tau | \mathbf{x}) \mid \mathbf{X} = \mathbf{x} \right) \geq \tau.$$

Now, Theorem 3.2 implies, for every subsequence (for which we keep the notation n) of n and $j \in \mathbb{N}$, there existence of some $n_j \in \mathbb{N}$ such that

$$\mathbb{P} \left(\mathbb{C}_{\pm}^{(n)}(\tau | \mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau | \mathbf{X}) + \frac{1}{2^j} \mathbb{S}_d \right) \leq \frac{1}{2^j}$$

for all $n \geq n_j$. The sum

$$\sum_{j=1}^{\infty} \mathbb{P} \left(\mathbb{C}_{\pm}^{(n_j)}(\tau | \mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau | \mathbf{X}) + \frac{1}{2^j} \mathbb{S}_d \right)$$

thus is finite, and the Borel-Cantelli lemma yields

$$\mathbb{P} \left(\bigcap_{J \in \mathbb{N}} \bigcup_{j \geq J} \mathbb{C}_{\pm}^{(n_j)}(\tau | \mathbf{X}) \not\subset \mathbb{C}_{\pm}(\tau | \mathbf{X}) + \frac{1}{2^j} \mathbb{S}_d \right) = 0,$$

or, equivalently, $\mathbb{P}(\Omega^*) = 1$, where

$$\Omega^* := \left(\bigcup_{J \in \mathbb{N}} \bigcap_{j \geq J} \mathbb{C}_{\pm}^{(n_j)}(\tau | \mathbf{X}) \subset \mathbb{C}_{\pm}(\tau | \mathbf{X}) + \frac{1}{2^j} \mathbb{S}_d \right).$$

Setting $\omega \in \Omega^* \subset \Omega$ and $\mathbf{x} = \mathbf{X}(\omega)$, there exists $J \in \mathbb{N}$ such that

$$\mathbb{P} \left(\mathbb{C}_{\pm}^{(n_j)}(\tau | \mathbf{x}) | \mathbf{X} = \mathbf{x} \right) \leq \mathbb{P} \left(\mathbb{C}_{\pm}(\tau | \mathbf{x}) + \frac{1}{2^j} \mathbb{S}_d | \mathbf{X} = \mathbf{x} \right),$$

for all $j \geq J$. Since

$$\mathbb{C}_{\pm}(\tau | \mathbf{x}) + \frac{1}{2^{j+1}} \mathbb{S}_d \subset \mathbb{C}_{\pm}(\tau | \mathbf{x}) + \frac{1}{2^j} \mathbb{S}_d$$

and

$$\bigcap_{j \in \mathbb{N}} \mathbb{C}_{\pm}(\tau | \mathbf{x}) + \frac{1}{2^j} \mathbb{S}_d = \mathbb{C}_{\pm}(\tau | \mathbf{x}),$$

we obtain

$$\limsup \mathbb{P} \left(\mathbb{C}_{\pm}^{(n_j)}(\tau | \mathbf{x}) | \mathbf{X} = \mathbf{x} \right) \leq \tau, \tag{A.12}$$

which concludes the proof. \square

Proof of Lemma 3.6 To prove this lemma, we show that the following five conditions of Stone's theorem (Theorem 1 in Stone (1977)) are satisfied (convergence for $n \rightarrow \infty$):

(a) *there exists $C \geq 1$ such that, for any non-negative measurable function f ,*

$$\mathbb{E} \left(\frac{1}{k} \sum_{j=1}^n \mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} f(\mathbf{X}_j) \right) \leq C \mathbb{E}(f(\mathbf{X}));$$

(b) $\mathbb{P} \left(\frac{1}{k} \sum_{j=1}^n \mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} \leq 1 \right) = 1$ for all $n \in \mathbb{N}$;

- (c) $\frac{1}{k} \sum_{j=1}^n \mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} \mathbb{1}_{|\mathbf{X}_j - X| > a} \rightarrow 0$, in probability, for all $a > 0$;
- (d) $\frac{1}{k} \sum_{j=1}^n \mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} \rightarrow 1$ in probability;
- (e) $\max_{i=j, \dots, n} \frac{1}{k} \mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} \rightarrow 0$ in probability.

Proof of (a). Let \mathbf{X} be a random variable independent of $\mathbf{X}^{(n)}$, with the same distribution as \mathbf{X}_i . Set $\mathbf{u}_0 = \mathbf{F}_{\pm}^{(n)}(\mathbf{X})$ and $\mathbf{u}_j = \mathbf{F}_{\pm}^{(n)}(\mathbf{X}_j)$, $j = 1, \dots, n$. Defining

$$K_n^{k,(i)}(\mathbf{X}) := \{\mathbf{X}_j : i \neq j, \mathbb{F}_{\pm}^n(\mathbf{X}_j) \in N_k(\mathbf{F}_{\pm}^{(n)}(\mathbf{X}_i))\},$$

note that $\mathbb{1}_{\mathbf{X}_i \in K_n^k(\mathbf{X})} f(\mathbf{X}_i)$ and $\mathbb{1}_{X \in K_n^{k,(i)}(\mathbf{X}_i)} f(\mathbf{X})$ have the same distribution. It follows that

$$\mathbb{E} \left(\frac{1}{k} \sum_{j=1}^n \mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} f(\mathbf{X}_j) \right) = \frac{1}{k} \sum_{j=1}^n \mathbb{E} \left(\mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} f(\mathbf{X}_j) \right) = \frac{1}{k} \sum_{j=1}^n \mathbb{E} \left(\mathbb{1}_{X \in K_n^{k,(j)}(\mathbf{X}_j)} f(\mathbf{X}) \right),$$

which implies that

$$\mathbb{E} \left(\frac{1}{k} \sum_{j=1}^n \mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} f(\mathbf{X}_j) \right) \leq \mathbb{E} \left(f(\mathbf{X}) \frac{1}{k} \sum_{j=1}^n \mathbb{1}_{X \in K_n^{k,(j)}(\mathbf{X}_j)} \right). \quad (\text{A.13})$$

Since $\sum_{j=1}^n \mathbb{1}_{X \in K_n^k(\mathbf{X}_j)} = \sum_{j=1}^n \mathbb{1}_{\mathbf{u}_0 \in N_k(\mathfrak{G}_j)}$, we can apply Corollary 6.1. in Györfi et al. (2002) and conclude that there exists $\lambda_d \in \mathbb{R}$ such that $\sum_{j=1}^n \mathbb{1}_{X \in K_n^{k,(j)}(\mathbf{X}_j)} \leq k\lambda_d$. This and (A.13) complete the proof. \square

Proof of (b), (d), and (e). Conditions (b) and (d) are direct consequences of the properties of the weight function, see Corollary 1 in Stone (1977). As for (e), it follows from the fact that $k \rightarrow \infty$. \square

Proof of (c). The following lemma (see Section A.2 for a proof) is a corollary of Lemma 6.1 in Györfi et al. (2002).

Lemma A.3. *Let $k/n \rightarrow 0$ as $n \rightarrow \infty$. Then, as $n \rightarrow \infty$*

$$\sup_{\mathfrak{G}_i \in N_k(\mathfrak{G}_0)} |\mathfrak{G}_i - \mathfrak{G}_0| \rightarrow 0, \quad a.s.$$

Since \mathbf{Q}_{\pm} is a singleton with probability one, let us assume, without loss of generality, that $\mathbf{Q}_{\pm}(\mathfrak{G}_0)$ and $\mathbf{Q}_{\pm}^{(n)}(\mathfrak{G}_k)$ are singletons. Actually, the set

$$\bigcap_{k=1}^{\infty} \bigcap_{n=0}^{\infty} \left\{ \mathbf{Q}_{\pm}^{(n)}(\mathfrak{G}_k) \text{ is a singleton} \right\}$$

also has probability one. Within that set,

- (i) $\mathbf{Q}_{\pm}^{(n)} \rightarrow \mathbf{Q}_{\pm}$ graphically as $n \rightarrow \infty$,
- (ii) for every $\mathbf{X}_i \in K_n^k(\mathbf{X})$, there exists some $\mathfrak{G}_i \in N_k(\mathfrak{G}_0)$ such that $\mathbf{Q}_{\pm}^{(n)}(\mathfrak{G}_i) = \{\mathbf{X}_i\}$, and

(iii) Lemma A.3 yields $\mathfrak{G}_i \rightarrow \mathfrak{G}_0$ as $n \rightarrow \infty$.

This, in view of Proposition 5.33 in Rockafellar and Wets (1998), implies that

$$\sup_{\mathbf{X}_i \in K_n^k(\mathbf{X})} |\mathbf{X}_i - \mathbf{X}| \rightarrow 0, \quad a.s. \quad (\text{A.14})$$

We are ready now to prove (c). Set $a > 0$. Since there are k elements in $K_n^k(\mathbf{X})$, we have

$$\begin{aligned} \mathbb{E} \left(\frac{1}{k} \sum_{j=1}^n \mathbb{1}_{\mathbf{X}_j \in K_n^k(\mathbf{X})} \mathbb{1}_{|\mathbf{X}_j - \mathbf{X}| > a} \right) &= \mathbb{E} \left(\frac{1}{k} \sum_{\mathbf{X}_j \in K_n^k(\mathbf{X})} \mathbb{1}_{|\mathbf{X}_j - \mathbf{X}| > a} \right) \\ &= \mathbb{P} \left(\sup_{\mathbf{X}_i \in K_n^k(\mathbf{X})} |\mathbf{X}_j - \mathbf{X}| > a \right) \end{aligned}$$

which, owing to (A.14), tends to 0. The desired result follows as a direct consequence of Markov's inequality. \square

A.2. Proofs of Lemmas A.1, A.2, and A.3

Proof of Lemma A.1. Let $\delta > 0$ and $\epsilon > 0$ be arbitrary. Denote by $K \subset \mathbb{R}^d$ a compact set such that $P(\mathbf{Y} \in K) \geq 1 - \delta\epsilon/18$. Suppose that $\mathbf{0} \in K$ and define \mathcal{F}_K as the class of 1-Lipschitz functions f supported on K such that $f(\mathbf{0}) = 0$. Such a class, by the Arzelà–Ascoli theorem, is relatively compact for $\mathcal{C}(K)$ and the uniform convergence. Then, there exists a sequence $f_1, \dots, f_{N_\epsilon}$, such that $\sup_{f \in \mathcal{F}_K} \inf_{k=1 \dots N_\epsilon} \|f - f_k\|_\infty \leq \epsilon/8$. Therefore, for every $f \in \mathcal{F}_K$, we have

$$\left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y})|\mathbf{X}) \right| \leq \sup_{k=1 \dots N_\epsilon} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f_k(\mathbf{Y}_j) - \mathbb{E}(f_k(\mathbf{Y})|\mathbf{X}) \right| + \frac{\epsilon}{4}.$$

In consequence, there exists n_0 such that, for $n \geq n_0$, we have

$$\begin{aligned} &\mathbb{P} \left(\sup_{f \in \mathcal{F}_K} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y})|\mathbf{X}) \right| > \frac{\epsilon}{2} \right) \\ &\leq \mathbb{P} \left(\sup_{k=1 \dots N_\epsilon} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f_k(\mathbf{Y}_j) - \mathbb{E}(f_k(\mathbf{Y})|\mathbf{X}) \right| > \frac{\epsilon}{4} \right) \quad (\text{A.15}) \\ &\leq \sum_{k=1}^{N_\epsilon} \mathbb{P} \left(\left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f_k(\mathbf{Y}_j) - \mathbb{E}(f_k(\mathbf{Y})|\mathbf{X}) \right| > \frac{\epsilon}{4} \right) \leq \frac{\delta}{3}, \end{aligned}$$

where the last inequality follows from the fact that the weight function is consistent, Note that every $f \in \mathcal{F}_{BL}$ can be approximated by $f\mathbb{1}_K$. This yields

$$\begin{aligned} & \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y})|\mathbf{X}) \right| \\ & \leq \sup_{f \in \mathcal{F}_K} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y})|\mathbf{X}) \right| \\ & \quad + \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}) \right|. \end{aligned} \tag{A.16}$$

Inequality (A.15) provides an upper bound for the first term. The second one, denoted as A_n , is bounded by

$$\left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) \right| + |\mathbb{E}(f(\mathbf{Y}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X})|.$$

Since the weights are positive and $\sup_{\mathbf{x} \in \mathbb{R}^d} |f(\mathbf{x})| \leq 1$,

$$\begin{aligned} A_n & \leq \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) + \mathbb{E}(\mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}) \\ & \leq \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) - \mathbb{E}(\mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}) \right| + 2\mathbb{E}(\mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}). \end{aligned}$$

Note that the bound does not depend on the function f . Consequently,

$$\begin{aligned} & \sup_{f \in \mathcal{F}_{BL}} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}) \right| \\ & \leq \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) - \mathbb{E}(\mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}) \right| + 2\mathbb{E}(\mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}). \end{aligned}$$

Taking expectations on both sides we obtain

$$\begin{aligned} & \mathbb{E} \left(\sup_{f \in \mathcal{F}_{BL}} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}) \right| \right) \\ & \leq \mathbb{E} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) - \mathbb{E}(\mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}) \right| + 2\mathbb{P}(\mathbf{Y} \notin K). \end{aligned}$$

Since the weights are consistent, there exists n_1 such that

$$\mathbb{E} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) - \mathbb{E}(\mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y})|\mathbf{X}) \right| \leq \frac{\delta\epsilon}{12}$$

for all $n > n_1$. Since $\mathbb{P}(\mathbf{Y} \notin K) \leq \delta\epsilon/12$, if $n > n_1$, then

$$\mathbb{E} \left(\sup_{f \in \mathcal{F}_{BL}} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}) | \mathbf{X}) \right| \right) \leq \frac{\delta\epsilon}{3}. \quad (\text{A.17})$$

Using Markov's inequality in (A.17), we obtain

$$\mathbb{P} \left(\sup_{f \in \mathcal{F}_{BL}} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y}) \mathbb{1}_{\mathbb{R}^d \setminus K}(\mathbf{Y}) | \mathbf{X}) \right| > \frac{\epsilon}{2} \right) \leq \frac{2\delta}{3} \quad (\text{A.18})$$

for all $n > n_1$. Finally using (A.15), (A.18) and (A.16), we conclude that

$$\mathbb{P} \left(\sup_{f \in \mathcal{F}_{BL}} \left| \sum_{j=1}^n w_j(X; \mathbf{X}^{(n)}) f(\mathbf{Y}_j) - \mathbb{E}(f(\mathbf{Y}) | \mathbf{X}) \right| > \epsilon \right) \leq \delta \quad (\text{A.19})$$

for all $n > N = \max(n_0, n_1)$. \square

Proof of Lemma A.2. Due to the fact that finite second-order moments are not required in our setting, Theorem 2.8 in del Barrio and Loubes (2019) does not directly apply. Their proof, however, relies on the weak convergence of the couplings (the joint measures solving the Kantorovich problem). In our case, we can prove a similar result using Lemma 9 in McCann (1995). Indeed, since the sequences $\{\mu_n\}_n$ and $\{\nu_n\}_n$ are tight with respect to weak convergence, the same result holds also for the class $\Gamma(\mu_n, \nu_n)$ of probabilities on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals $\{\mu_n\}_n$ and $\{\nu_n\}_n$, see Lemma 4.4 in Villani (2008). Note that all the measures π_n , $n \in \mathbb{N}$, belong to $\Gamma(\mu_n, \nu_n)$. Denote by $\pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ the weak limit of $\{\pi_n\}_n$ along a subsequence; for simplicity, we keep the index n for the subsequence. Lemma 9 (ii) in McCann (1995) implies that the marginals of π are μ and ν . Moreover, the support of π_n is cyclically monotone: indeed, it is contained in the subdifferential of ϕ_n . Therefore, using Lemma 9 (i) in McCann (1995), we know that also π is supported on a cyclically monotone set. Corollary 14 in McCann (1995) yields—since μ is uniformly continuous with respect to the Lebesgue measure—the existence of a unique measure with cyclically monotone support and marginals μ and ν . As a consequence, $\pi = (\mathbf{Id} \times \phi) \# \mu$. Since this holds along all possible subsequence, we have the weak convergence of $(\mathbf{Id} \times \phi_n) \# \mu_n$ to $(\mathbf{Id} \times \phi)$. At this point, to conclude the proof of Lemma A.2, we can repeat verbatim the rest of the proof of Theorem 2.8 in del Barrio and Loubes (2019). \square

Proof of Lemma A.3. Note that, for all $\epsilon > 0$, $\sup_{\mathfrak{G}_i \in N_k(\mathfrak{G}_0)} |\mathfrak{G}_i - \mathfrak{G}_0| > \epsilon$ if and only if $\sum_{i=1}^n \mathbb{1}_{|\mathfrak{G}_i - \mathfrak{G}_0| < \epsilon} < k/n$. Since

$$\inf_{\mathbf{u} \in \mathbb{S}_m} \sum_{i=1}^n \mathbb{1}_{|\mathfrak{G}_i - \mathfrak{G}_0| < \epsilon} \longrightarrow \inf_{\mathbf{u} \in \mathbb{S}_m} U_m(\mathbf{u} + \epsilon \mathbb{S}_m) > 0$$

and $k/n \rightarrow 0$ as $n \rightarrow \infty$,

$$\sup_{\mathfrak{G}_0 \in \mathbb{S}_m} \sup_{\mathfrak{G}_i \in N_k(\mathfrak{G}_0)} |\mathfrak{G}_i - \mathfrak{G}_0| \longrightarrow 0 \quad \text{as } n \rightarrow \infty.$$

\square