Multiple-Output Quantile Regression through Optimal Quantization

Isabelle Charlier[(1)], Davy Paindaveine[(1)] and Jérôme Saracco[(2)]

[(1)] *Université Libre de Bruxelles, ECARES and Département de Mathématique*

[(2)] *Université de Bordeaux and INRIA Bordeaux Sud-Ouest, CQFD team*

Abstract: A new nonparametric quantile regression method based on the concept of *optimal quantization* was developed recently and was showed to provide estimators that often dominate their classical, kernel-type, competitors. The construction, however, remains limited to single-output quantile regression. In the present work, we therefore extend the quantization-based quantile regression method to the multiple-output context. We show how quantization allows to approximate population multiple-output regression quantiles based on halfspace depth. We prove that this approximation becomes arbitrarily accurate as the size of the quantization grid goes to infinity. We also consider a sample version of the proposed regression quantiles and derive a weak consistency result. Through simulations, we compare the performances of the proposed estimators with (local constant and local bilinear) kernel competitors. We also compare the corresponding sample quantile regions. The results reveal that the proposed quantization-based estimators, which are local constant in nature, outperform their kernel counterparts and even often dominate their local bilinear kernel competitors.

Key words and phrases: multivariate quantilem, nonparametric estimation, optimal quantization, quantile regression.

1 Introduction

Since its introduction in the seminal paper Koenker and Bassett (1978), quantile regression has met a tremendous success. Unlike standard least squares regression that focuses on the mean of a scalar response Y conditional on a d-dimensional covariate $\boldsymbol{X}$, quantile regression is after the corresponding conditional quantile of any order $\alpha \in (0,1)$, hence provides a thorough description of the conditional distribution of the response. It is well-known that quantile regression, as an L_1-type method, dominates least squares regression methods in terms of robustness, while remaining very light on the computational side since it relies on linear programming methods. Quantile regression was first defined for linear regression but was later extended to nonlinear/nonparametric regression. For a modern account of quantile regression, we refer to the monograph Koenker (2005).

Quantile regression for long has been restricted to the single-output (that is, univariate response) context. The reason is of course that, due to the lack of a canonical ordering in $\mathbb{R}^m$ ($m > 1$), there is no universally accepted definition of multivariate quantile. In the last two decades, the problem of defining a suitable concept of multivariate quantile has been an active research topic. Chaudhuri (1996) proposed a concept of geometric quantiles whose corresponding quantile regions coincide with the *spatial depth* regions; see, e.g., Serfling (2010). These quantiles are suitable in large dimensions and even in general Hilbert spaces, which makes them applicable to high-dimensional or functional data. They have been used in a quantile regression framework, even in the functional case; see, e.g., Chakraborty (2003), Cheng and De Gooijer (2007), Chaouch and Laïb (2013, 2015) and Chowdhury and Chaudhuri (2016). Although they can be made affine-equivariant through a transformation-retransformation approach (see Chakraborty, 2001, 2003), geometric (regression) quantiles intrinsically are orthogonal-equivariant only. An alternative definition of multivariate quantile, that is affine-equivariant and still enjoys all properties usually expected from a quantile, was proposed in Hallin et al. (2010). The corresponding quantiles, unlike geometric ones, are related to the (Tukey, 1975) *halfspace depth*. Very recently, multivariate quantiles that enjoy even stronger equivariance properties and are still connected with appropriate depth functions, but that are much harder to compute in practice, were proposed and used for multiple-output quantile regression; see, e.g., Carlier et al. (2016, 2017) and Chernozhukov et al. (2017).

In this paper, we focus on the affine-equivariant quantiles from Hallin et al. (2010)

and on their use in a nonparametric regression framework. These quantiles were already used in this framework in Hallin et al. (2015) — hereafter, HLPS15 — where a (local constant and local bilinear) kernel approach was adopted. The resulting regression quantiles thus extend to the multiple-output setup the single-output (local constant and local linear) kernel regression quantiles from Yu and Jones (1998). Kernel methods, however, are not the only smoothing techniques that can be used to perform nonparametric quantile regression. In the single-output context, Charlier et al. (2015a) indeed recently showed that nonparametric quantile regression can alternatively be based on *optimal quantization*, which is a method that provides a suitable discretization $\widetilde{\boldsymbol{X}}^N$ of size N of the d-dimensional, typically continuous, covariate $\boldsymbol{X}$. As demonstrated in Charlier et al. (2015b) through simulations, this quantization approach provides sample regression quantiles that often dominate their kernel competitors in terms of integrated square errors.

This dominance of quantization-based regression quantiles over their kernel competitors provides a natural motivation to try and define quantization-based analogs of the kernel multiple-output regression quantiles from HLPS15. This is the objective of the present paper. While this objective, conceptually, can be achieved by applying the quantization-based regression methodology from Charlier et al. (2015a) to the multiple-ouput regression quantiles from HLPS15, establishing theoretical guarantees for the resulting regression quantiles is highly non-trivial (as we explain below) and requires the highly technical proofs to be found in the appendix of the present paper.

The paper is organized as follows. Section 2 describes the multivariate quantiles (Section 2.1) and multiple-output regression quantiles (Section 2.2) that will be considered in this work. Section 3 explains how these can be approximated through optimal quantization and shows that the approximation becomes arbitrarily accurate as the number N of grid points used in the quantization goes to infinity. Section 4 defines the corresponding sample quantization-based regression quantiles and establishes their consistency (for the fixed-N approximation of multiple-output regression quantiles). Section 5 is devoted to numerical results : first, a data-driven method to select the smoothing parameter N is described (Section 5.1). Then, a comparison with the kernel-based competitors from HLPS15 is performed, based on empirical integrated square errors and on visual inspection of the resulting conditional quantile regions (Section 5.2). Finally, Section 6 concludes.

2 The multiple-output regression quantiles considered

As mentioned above, the main objective of this paper is to estimate through optimal quantization the population multiple-output regression quantiles from HLPS15. These regression quantiles are the conditional version of the multivariate quantiles from Hallin et al. (2010). To make the paper self-contained, we start by describing these two types of quantiles.

2.1 The multivariate quantiles considered

The multivariate quantiles from Hallin et al. (2010) are indexed by a vector $\boldsymbol{\alpha}$ ranging over $\mathcal{B}^m := \{\boldsymbol{y} \in \mathbb{R}^m : 0 < |\boldsymbol{y}| < 1\}$, the open unit ball of $\mathbb{R}^m$ deprived of the origin (throughout, $|\cdot|$ denotes the Euclidean norm). This index $\boldsymbol{\alpha}$ factorizes into $\boldsymbol{\alpha} = \alpha \boldsymbol{u}$, with $\alpha = |\boldsymbol{\alpha}| \in (0, 1)$ and $\boldsymbol{u} \in \mathcal{S}^{m-1} := \{\boldsymbol{y} \in \mathbb{R}^m : |\boldsymbol{y}| = 1\}$. Letting $\boldsymbol{\Gamma}_{\boldsymbol{u}}$ be an arbitrary $m \times (m-1)$ matrix whose columns form, jointly with $\boldsymbol{u}$, an orthonormal basis of $\mathbb{R}^m$, the multivariate quantiles we consider in this paper are defined as follows.

Definition 1. Let $\boldsymbol{Y}$ be a random m-vector, with probability distribution $P_{\boldsymbol{Y}}$, say, and fix $\boldsymbol{\alpha} = \alpha \boldsymbol{u} \in \mathcal{B}^m$. Then the *$\boldsymbol{\alpha}$-quantile of $\boldsymbol{Y}$*, or *order-α quantile of $\boldsymbol{Y}$ in direction $\boldsymbol{u}$*, is any element of the collection of hyperplanes

$$\boldsymbol{\pi}_{\boldsymbol{\alpha}} := \boldsymbol{\pi}_{\boldsymbol{\alpha}}(P_{\boldsymbol{Y}}) := \left\{ \boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{u}'\boldsymbol{y} = \boldsymbol{c}'_{\boldsymbol{\alpha}} \boldsymbol{\Gamma}'_{\boldsymbol{u}} \boldsymbol{y} + a_{\boldsymbol{\alpha}} \right\}$$

with

$$\boldsymbol{q}_{\boldsymbol{\alpha}} := \begin{pmatrix} a_{\boldsymbol{\alpha}} \\ \boldsymbol{c}_{\boldsymbol{\alpha}} \end{pmatrix} = \arg\min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} - a)], \tag{1}$$

where $Y_{\boldsymbol{u}} := \boldsymbol{u}'\boldsymbol{Y}$, $\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} := \boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{Y}$ and $z \mapsto \rho_\alpha(z) := z(\alpha - \mathbb{I}_{[z<0]})$ is the usual check function (throughout, $\mathbb{I}_A$ is the indicator function of A).

The multivariate quantile $\boldsymbol{\pi}_{\boldsymbol{\alpha}}$ is nothing but the Koenker and Bassett (1978) order-α regression quantile hyperplane obtained when L_1 vertical deviations are computed in the direction $\boldsymbol{u}$. Two direct consequences are the following. First, unlike many competing concepts of multivariate quantiles, such as, e.g., the geometric quantiles from Chaudhuri (1996), the multivariate quantile considered are *hyperplane-valued* rather than *point-valued* (of course, the difference is relevant for $m > 1$ only). This potentially allows to use multivariate quantiles as critical values, in relation with the fact that a (point-valued) test statistic $\boldsymbol{T}$ would take its value "above" ($\boldsymbol{u}'\boldsymbol{T} > \boldsymbol{c}_{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{T} + a_{\boldsymbol{\alpha}}$) or "below" ($\boldsymbol{u}'\boldsymbol{T} \leq \boldsymbol{c}_{\boldsymbol{\alpha}}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{T} + a_{\boldsymbol{\alpha}}$) the quantile hyperplane $\boldsymbol{\pi}_{\boldsymbol{\alpha}}$. Second, like many competing quantiles, they are *directional* quantiles but are related to the direction in a non-trivial way: the quantile hyperplane in direction $\boldsymbol{u}$ will generally not be orthogonal to $\boldsymbol{u}$ (just like the point-valued geometric quantiles in direction $\boldsymbol{u}$ will generally not belong to the halfline with direction $\boldsymbol{u}$ originating from the corresponding median, namely the spatial median).

For fixed $\alpha \in \left(0, \frac{1}{2}\right)$, the collection of multivariate quantiles $\boldsymbol{\pi}_{\alpha\boldsymbol{u}}$ provides the centrality region

$$R_\alpha := \cap_{\boldsymbol{u} \in \mathcal{S}^{m-1}} \left\{ \boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{u}'\boldsymbol{y} \geq \boldsymbol{c}_{\alpha\boldsymbol{u}}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{y} + a_{\alpha\boldsymbol{u}} \right\}, \tag{2}$$

that, as showed in Hallin et al. (2010), coincides with the order-α *halfspace depth region* of $P_{\boldsymbol{Y}}$ (that is, with the set of $\boldsymbol{y}$'s whose halfspace depth $D(\boldsymbol{y}, P_{\boldsymbol{Y}}) = \inf_{\boldsymbol{u} \in \mathcal{S}^{m-1}} P[\boldsymbol{u}'(\boldsymbol{Y} - \boldsymbol{y}) \geq 0]$ is larger than or equal to α); see Tukey (1975). In the univariate case ($m = 1$), the quantile hyperplane $\boldsymbol{\pi}_{\alpha\boldsymbol{u}}$ reduces, for $\boldsymbol{u} = 1$ (resp., for $\boldsymbol{u} = -1$), to the usual α-quantile (resp., $(1-\alpha)$-quantile) of P_Y, and R_α then coincides with the interval whose end points are these two quantiles. Halfspace depth regions provide a very informative description of the distribution of $\boldsymbol{Y}$ and allow to perform robust nonparametric inference for various problems; see, e.g., Liu et al. (1999). It should be noted that other definitions of multivariate quantiles are linked to different concepts of statistical depth; for instance, as already mentioned, the regions resulting from the geometric quantiles from Chaudhuri (1996) coincide with *spatial depth* regions; see, e.g., Serfling (2010).

2.2 The multiple-output regression quantiles considered

Consider now the regression framework where the random m-vector $\boldsymbol{Y}$ from the previous section is regarded as a vector of response variables and where a d-vector $\boldsymbol{X}$ of random covariates is available. Still with $\boldsymbol{\alpha} = \alpha\boldsymbol{u} \in \mathcal{B}^m$, the regression $\boldsymbol{\alpha}$-quantile of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ we are after, namely the one from HLPS15, is then the cartesian product of $\{\boldsymbol{x}\}$ with the quantile hyperplane $\boldsymbol{\pi}_{\boldsymbol{\alpha}}(P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}})$ associated with the conditional distribution of $\boldsymbol{Y}$ upon $\boldsymbol{X} = \boldsymbol{x}$. More specifically, we adopt the following definition.

Definition 2. Fix $\boldsymbol{\alpha} \in \mathcal{B}^m$. The (regression) *$\boldsymbol{\alpha}$-quantile of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$*, or (regression) *order-α quantile of $\boldsymbol{Y}$ given $\boldsymbol{X} = \boldsymbol{x}$ in direction $\boldsymbol{u}$*, is any element of the collection of hyperplanes

$$\boldsymbol{\pi}_{\boldsymbol{\alpha},\boldsymbol{x}} = \boldsymbol{\pi}_{\alpha\boldsymbol{u},\boldsymbol{x}} := \big\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^d \times \mathbb{R}^m : \boldsymbol{u}'\boldsymbol{y} = \boldsymbol{c}'_{\boldsymbol{\alpha},\boldsymbol{x}}\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{y} + a_{\boldsymbol{\alpha},\boldsymbol{x}}\big\}$$

with

$$\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}} := \begin{pmatrix} a_{\boldsymbol{\alpha},\boldsymbol{x}} \\ \boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}} \end{pmatrix} = \arg\min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}^\perp_{\boldsymbol{u}} - a)|\boldsymbol{X} = \boldsymbol{x}], \tag{3}$$

where we still let $Y_{\boldsymbol{u}} := \boldsymbol{u}'\boldsymbol{Y}$ and $\boldsymbol{Y}^\perp_{\boldsymbol{u}} := \boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}$.

Parallel to what was done in the location case, (regression) *α-quantile regions* can be obtained by considering the envelopes of the regression quantiles $\boldsymbol{\pi}_{\alpha\boldsymbol{u},\boldsymbol{x}}$ for fixed α. More precisely, for any $\alpha \in \big(0, \frac{1}{2}\big)$, the fixed-$\boldsymbol{x}$ order-α regression quantile region is

$$R_{\alpha,\boldsymbol{x}} := \cap_{\boldsymbol{u} \in \mathcal{S}^{m-1}} \big\{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{u}'\boldsymbol{y} \geq \boldsymbol{c}'_{\alpha\boldsymbol{u},\boldsymbol{x}}\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{y} + a_{\alpha\boldsymbol{u},\boldsymbol{x}}\big\}; \tag{4}$$

as in the previous section, this region coincides with the order-α halfspace depth region of $P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$. When considering all values of $\boldsymbol{x}$, these fixed-$\boldsymbol{x}$ regions generate the *nonparametric α-quantile/depth regions*

$$R_\alpha := \cup_{\boldsymbol{x}\in\mathbb{R}^d}\Big(\{\boldsymbol{x}\} \times R_{\alpha,\boldsymbol{x}}\Big),$$

which provide nested regression "tubes" indexed by α. Note that the regression quantile regions $R_{\alpha,\boldsymbol{x}}$ are obtained by considering all directions $\boldsymbol{u}$ in the $\boldsymbol{Y}$-space $\mathbb{R}^m$ and not all directions in the $(\boldsymbol{X},\boldsymbol{Y})$-space $\mathbb{R}^{d+m}$, which is in line with the fact that conditional quantiles of $\boldsymbol{Y}$ remain quantiles of the m-variate quantity $\boldsymbol{Y}$.

In the single-output case $m = 1$, the hypersurface $\cup_{\boldsymbol{x}\in\mathbb{R}^d}\boldsymbol{\pi}_{\boldsymbol{\alpha},\boldsymbol{x}}$ associated with $\boldsymbol{\alpha} = \alpha\boldsymbol{u} = \alpha$ ($\boldsymbol{u} = 1$) is the standard conditional α-quantile surface $\cup_{\boldsymbol{x}\in\mathbb{R}^d}\{(\boldsymbol{x}, a_{\alpha,\boldsymbol{x}})\}$ (here, $a_{\alpha,\boldsymbol{x}}$ is the usual α-quantile of Y given $\boldsymbol{X} = \boldsymbol{x}$), so that the multiple-output regression quantiles from Definition 2 provide an extension of the classical single-output ones. The corresponding fixed-$\boldsymbol{x}$ regression quantile region $R_{\alpha,\boldsymbol{x}}$ is the interval $[a_{\alpha,\boldsymbol{x}}, a_{1-\alpha,\boldsymbol{x}}]$, and, for $\alpha \in (0, \frac{1}{2})$, the resulting nonparametric α-quantile region R_α is then the subset of $\mathbb{R}^{d+1}$ between the quantile conditional hypersurfaces of orders α and $1-\alpha$, that is, between $\cup_{\boldsymbol{x}\in\mathbb{R}^d}\{(\boldsymbol{x}, a_{\alpha,\boldsymbol{x}})\}$ and $\cup_{\boldsymbol{x}\in\mathbb{R}^d}\{(\boldsymbol{x}, a_{1-\alpha,\boldsymbol{x}})\}$.

We close this section with the following technical points. If $P_{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$, with a density that has a connected support and admits finite first-order moments, the minimization problem in (3) admits a unique solution; see Hallin et al. (2010). Moreover, $(a, \boldsymbol{c}')' \mapsto G_{a,\boldsymbol{c}}(\boldsymbol{x}) = \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^\perp - a)|\boldsymbol{X} = \boldsymbol{x}]$ is then convex and continuously differentiable on $\mathbb{R}^m$. Therefore, under these asumptions, $\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}} = (a_{\boldsymbol{\alpha},\boldsymbol{x}}, \boldsymbol{c}'_{\boldsymbol{\alpha},\boldsymbol{x}})'$ is alternatively characterized as the unique solution of the system of equations

$$\partial_a G_{a,\boldsymbol{c}}(\boldsymbol{x}) = P[\boldsymbol{u}'\boldsymbol{Y} < a + \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}] - \alpha = 0 \tag{5}$$

$$\nabla_{\boldsymbol{c}} G_{a,\boldsymbol{c}}(\boldsymbol{x}) = \mathrm{E}\big[\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}\big(\alpha - \mathbb{I}_{[\boldsymbol{u}'\boldsymbol{Y}<a+\boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}]}\big)\big|\boldsymbol{X} = \boldsymbol{x}\big] = 0; \tag{6}$$

see Lemma 4 in Appendix A. As we will see in the sequel, twice differentiability of $(a, \boldsymbol{c}')' \mapsto G_{a,\boldsymbol{c}}(\boldsymbol{x})$ actually requires slightly stronger assumptions.

3 Quantization-based multiple-output regression quantiles

Nonparametric single-output quantile regression is classically performed through kernel smoothing (Yu and Jones, 1998). An alternative approach, that relies on the concept of *optimal quantization*, was recently proposed in Charlier et al. (2015a) and was showed in Charlier et al. (2015b) to dominate kernel methods in finite samples. As explained in Charlier et al. (2015b), the dominance of quantization-based quantile regression methods over kernel smoothing ones can be explained by the fact that the amount of smoothing is usually fixed globally for kernel methods (that is, the bandwidth is constant all over the covariate space) whereas the subtle geometry of optimal quantization grids (see below) de facto leads to smooth more in some part of the covariate space than in others. The efficiency of quantization-based quantile regression methods in single-output situations provides a strong motivation to extend these methods to multiple-output problems, which is the topic of this section.

We start by defining optimal quantization. For any fixed $N \in \mathbb{N}_0 (:= \{1, 2, \dots\})$, quantization replaces the random d-vector $\boldsymbol{X}$ by a discrete version $\widetilde{\boldsymbol{X}}^{\boldsymbol{\gamma}^N} := \mathrm{Proj}_{\boldsymbol{\gamma}^N}(\boldsymbol{X})$ obtained by projecting $\boldsymbol{X}$ onto the N-quantization grid $\boldsymbol{\gamma}^N (\in (\mathbb{R}^d)^N)$. The quantization grid is optimal if it minimizes the quantization error $\|\widetilde{\boldsymbol{X}}^{\boldsymbol{\gamma}^N} - \boldsymbol{X}\|_p$, where $\|\boldsymbol{Z}\|_p := (\mathrm{E}[|\boldsymbol{Z}|^p])^{1/p}$ denotes the L_p-norm of $\boldsymbol{Z}$. Existence (but not unicity) of such an optimal grid is guaranteed if the distribution of $\boldsymbol{X}$ does not charge any hyperplane; see, e.g., Pagès (1998). In the sequel, $\widetilde{\boldsymbol{X}}^N$ will denote the projection of $\boldsymbol{X}$ onto an arbitrary optimal N-grid. This approximation becomes more and more precise as N increases since $\|\widetilde{\boldsymbol{X}}^N - \boldsymbol{X}\|_p = O(N^{-1/d})$ as $N \to \infty$; see, e.g., Graf and Luschgy (2000). More details on optimal quantization can be found in Pagès (1998), Pagès and Printems (2003) or Graf and Luschgy (2000).

Now, let $p \geq 1$ such that $\|\boldsymbol{X}\|_p < \infty$ and let $\boldsymbol{\gamma}^N$ be an optimal quantization grid.

Replacing $\boldsymbol{X}$ in (3) by its projection $\widetilde{\boldsymbol{X}}^N$ onto $\boldsymbol{\gamma}^N$ leads to considering

$$\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} = \begin{pmatrix} \widetilde{a}^N_{\boldsymbol{\alpha},\boldsymbol{x}} \\ \widetilde{\boldsymbol{c}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} \end{pmatrix} = \arg\min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}^\perp_{\boldsymbol{u}} - a)|\widetilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}], \tag{7}$$

where $\tilde{\boldsymbol{x}}$ denotes the projection of $\boldsymbol{x}$ onto $\boldsymbol{\gamma}^N$. A quantization-based approximation of the multiple-output regression quantile from Definition 2 above is thus any hyperplane of the form

$$\tilde{\boldsymbol{\pi}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} := \big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^d \times \mathbb{R}^m : \boldsymbol{u}'\boldsymbol{y} = (\widetilde{\boldsymbol{c}}^N_{\boldsymbol{\alpha},\boldsymbol{x}})'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{y} + \widetilde{a}^N_{\boldsymbol{\alpha},\boldsymbol{x}}\big\}.$$

This quantization-based quantile being entirely characterized by $\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$, we will investigate the quality of this approximation through $\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$. Since $\widetilde{\boldsymbol{X}}^N - \boldsymbol{X}$ goes to zero in L_p-norm as N goes to infinity, we may expect that $\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{X}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{X}}$ also converges to zero in an appropriate sense. To formalize this, the following assumptions are needed.

Assumption (A) (i) The random vector $(\boldsymbol{X},\boldsymbol{Y})$ is generated through $\boldsymbol{Y} = \boldsymbol{M}(\boldsymbol{X},\boldsymbol{\varepsilon})$, where the d-dimensional covariate vector $\boldsymbol{X}$ and the m-dimensional error vector $\boldsymbol{\varepsilon}$ are mutually independent; (ii) the support $S_{\boldsymbol{X}}$ of the distribution $P_{\boldsymbol{X}}$ of $\boldsymbol{X}$ is compact; (iii) denoting by $GL_m(\mathbb{R})$ the set of $m\times m$ invertible real matrices, the link function $\boldsymbol{M} : S_{\boldsymbol{X}} \times \mathbb{R}^m \to \mathbb{R}^m$ is of the form $(\boldsymbol{x},\boldsymbol{z}) \mapsto \boldsymbol{M}(\boldsymbol{x},\boldsymbol{z}) = \boldsymbol{M}_{a,\boldsymbol{x}} + \boldsymbol{M}_{b,\boldsymbol{x}}\boldsymbol{z}$, where the functions $\boldsymbol{M}_{a,\cdot} : S_{\boldsymbol{X}} \to \mathbb{R}^m$ and $\boldsymbol{M}_{b,\cdot} : S_{\boldsymbol{X}} \to GL_m(\mathbb{R})$ are Lipschitz with respect to the Euclidean norm and operator norm, respectively (see below); (iv) the distribution $P_{\boldsymbol{X}}$ of $\boldsymbol{X}$ does not charge any hyperplane; (v) the distribution $P_{\boldsymbol{\varepsilon}}$ of $\boldsymbol{\varepsilon}$ admits finite pth order moments, that is, $\|\boldsymbol{\varepsilon}\|_p^p = \mathrm{E}[|\boldsymbol{\varepsilon}|^p] < \infty$.

For the sake of clarity, we make precise that the Lipschitz properties of $\boldsymbol{M}_{a,\cdot}$ and $\boldsymbol{M}_{b,\cdot}$ in Assumption (A)(iii) mean that there exist constants $C_1, C_2 > 0$ such that

$$\forall \boldsymbol{x}_1,\boldsymbol{x}_2 \in \mathbb{R}^d, \quad |\boldsymbol{M}_{a,\boldsymbol{x}_1} - \boldsymbol{M}_{a,\boldsymbol{x}_2}| \le C_1|\boldsymbol{x}_1 - \boldsymbol{x}_2|, \tag{8}$$

$$\forall \boldsymbol{x}_1,\boldsymbol{x}_2 \in \mathbb{R}^d, \quad \|\boldsymbol{M}_{b,\boldsymbol{x}_1} - \boldsymbol{M}_{b,\boldsymbol{x}_2}\| \le C_2|\boldsymbol{x}_1 - \boldsymbol{x}_2|, \tag{9}$$

where $\|\boldsymbol{A}\| = \sup_{\boldsymbol{u}\in\mathcal{S}^{m-1}} |\boldsymbol{A}\boldsymbol{u}|$ denotes the operator norm of $\boldsymbol{A}$. The smallest constant C_1 (resp., C_2) that satisfies (8) (resp., (9)) will be denoted as $[\boldsymbol{M}_{a,\cdot}]_{\mathrm{Lip}}$ (resp., $[\boldsymbol{M}_{b,\cdot}]_{\mathrm{Lip}}$). We will also need the following assumption, that ensures in particular that the mapping $(a, \boldsymbol{c}')' \mapsto G_{a,\boldsymbol{c}}(\boldsymbol{x})$ is twice continuously differentiable (see Lemma 4).

Assumption (B) The distribution of $\boldsymbol{\varepsilon}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^m$, with a density $f^{\boldsymbol{\varepsilon}} : \mathbb{R}^m \to \mathbb{R}_0^+$ that is bounded, has a connected support, admits finite second-order moments, and satisfies, for some constants $C > 0$, $r > m-1$ and $s > 0$,

$$\left|f^{\boldsymbol{\varepsilon}}(\boldsymbol{z}_1) - f^{\boldsymbol{\varepsilon}}(\boldsymbol{z}_2)\right| \leq C\,|\boldsymbol{z}_1 - \boldsymbol{z}_2|^s \big(1 + \tfrac{1}{2}|\boldsymbol{z}_1 + \boldsymbol{z}_2|^2\big)^{-(3+r+s)/2}, \qquad (10)$$

for all $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^m$.

This assumption is an extremely slight reinforcement of Assumption (A'_n) in Hallin et al. (2010); more precisely, Assumption (B) above is obtained by replacing the condition $r > m-2$ from Hallin et al. (2010) into $r > m-1$, which we had to do for technical reasons. The resulting Lipschitz-type condition (10) remains very mild, though, and in particular it is satisfied (with $s = 1$) as soon as $f^{\boldsymbol{\varepsilon}}$ is continuously differentiable and that there exists a positive constant C and an invertible $m \times m$ matrix $\mathbf{A}$ such that

$$\sup_{|\mathbf{A}\boldsymbol{z}|\geq R} |\nabla f^{\boldsymbol{\varepsilon}}(\boldsymbol{z})| \leq C(1 + R^2)^{-(r+4)/2}$$

for any $R > 0$. This implies that Condition (10) is satisfied with $s = 1$ at the multinormal distributions and at t distributions with $\nu > 2$ degrees of freedom. We insist, however, that Condition (10) actually does not require that $f^{\boldsymbol{\varepsilon}}$ is everywhere continuously differentiable.

We can now state one of the main results of the paper (the proof is deferred to Appendix A).

Theorem 1. *Let Assumptions (A) and (B) hold. Then, for any $\boldsymbol{\alpha} \in \mathcal{B}^m$,*

$$\sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}} |\widetilde{\boldsymbol{q}}^{N}_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}| \to 0,$$

as $N \to \infty$.

This result confirms that, as the size N of the optimal quantization grid goes to infinity, the quantization-based approximation $\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$ of $\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}$ becomes arbitrarily precise. Clearly, the approximation is actually uniform in $\boldsymbol{x}$. This makes it natural to try and define, whenever observations are available, a sample version of $\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$ that will then be an estimator of $\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}$ from which one will be able to obtain in particular sample versions of the regression quantile regions $R_{\alpha,\boldsymbol{x}}$ in (4).

4 Sample quantization-based multiple-output regression quantiles

We now consider the problem of defining, from independent copies $(\boldsymbol{X}_1, \boldsymbol{Y}_1), \ldots, (\boldsymbol{X}_n, \boldsymbol{Y}_n)$ of $(\boldsymbol{X}, \boldsymbol{Y})$, a sample version $\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}$ of the quantization-based regression quantile coefficients $\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$ in (7).

4.1 Definition of the estimator

No closed form is available for an optimal quantization grid, except in some very particular cases. The definition of $\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}$ thus first requires constructing an (approximate) optimal grid. This may be done through a stochastic gradient algorithm, which proceeds as follows to quantize a d-dimensional random vector $\boldsymbol{X}$.

Let $(\boldsymbol{\xi}^t)_{t\in\mathbb{N}_0}$ be a sequence of independent copies of $\boldsymbol{X}$, and let $(\delta_t)_{t\in\mathbb{N}_0}$ be a deterministic sequence in $(0,1)$ satisfying $\sum_{t=1}^{\infty} \delta_t = +\infty$ and $\sum_{t=1}^{\infty} \delta_t^2 < +\infty$. Starting from an initial N-grid $\hat{\boldsymbol{\gamma}}^{N,0}$ with pairwise distinct components, the algorithm recursively defines the grid $\hat{\boldsymbol{\gamma}}^{N,t}$, $t \in \mathbb{N}_0$, as

$$\hat{\boldsymbol{\gamma}}^{N,t} = \hat{\boldsymbol{\gamma}}^{N,t-1} - \frac{\delta_t}{p} \nabla_{\boldsymbol{x}} d^p_N(\hat{\boldsymbol{\gamma}}^{N,t-1}, \boldsymbol{\xi}^t),$$

where $\nabla_{\boldsymbol{x}} d_N^p(\boldsymbol{x}, \boldsymbol{\xi})$ is the gradient with respect to the $\boldsymbol{x}$-component of the so-called *local quantization error* $d_N^p(\boldsymbol{x}, \boldsymbol{\xi}) = \min_{i=1,\ldots,N} |\boldsymbol{x}_i - \boldsymbol{\xi}|^p$, with $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N) \in (\mathbb{R}^d)^N$ and $\boldsymbol{\xi} \in \mathbb{R}^d$. Since $(\nabla_{\boldsymbol{x}} d_N^p(\boldsymbol{x}, \boldsymbol{\xi}))_i = p|\boldsymbol{x}_i - \boldsymbol{\xi}|^{p-2}(\boldsymbol{x}_i - \boldsymbol{\xi})\mathbb{I}_{[\boldsymbol{x}_i = \mathrm{Proj}_{\boldsymbol{x}}(\boldsymbol{\xi})]}$, $i = 1, \ldots, N$, two consecutive grids $\hat{\boldsymbol{\gamma}}^{N,t-1}$ and $\hat{\boldsymbol{\gamma}}^{N,t}$ differ by one point only, namely the point corresponding to the non-zero component of this gradient. The reader can refer to Pagès (1998), Pagès and Printems (2003) or Graf and Luschgy (2000) for more details on this algorithm, which, for $p = 2$, is known as the *Competitive Learning Vector Quantization (CLVQ) algorithm*.

The construction of $\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}$ then proceeds in two steps.

(S1) An "optimal" quantization grid is obtained from the algorithm above. First, an initial grid $\hat{\boldsymbol{\gamma}}^{N,0}$ is selected by sampling randomly without replacement among the $\boldsymbol{X}_i$'s, under the constraint that the same value cannot be picked more than once (a constraint that is relevant only if there are ties in the $\boldsymbol{X}_i$'s). Second, n iterations of the algorithm are performed, based on $\boldsymbol{\xi}^t = \boldsymbol{X}_t$, for $t = 1, \ldots, n$. The resulting optimal grid is denoted as $\hat{\boldsymbol{\gamma}}^{N,n} = (\hat{\boldsymbol{x}}_1^{N,n}, \ldots, \hat{\boldsymbol{x}}_n^{N,n})$.

(S2) The approximation $\widetilde{\boldsymbol{q}}^{N}_{\boldsymbol{\alpha},\boldsymbol{x}} = \arg\min_{(a,\boldsymbol{c}')'} \mathrm{E}[\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{Y} - a)|\widetilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}]$ in (7) is then estimated by

$$\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}} = \begin{pmatrix} \widehat{a}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}} \\ \widehat{\boldsymbol{c}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}} \end{pmatrix} = \arg\min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \sum_{i=1}^{n} \rho_\alpha(\boldsymbol{u}'\boldsymbol{Y}_i - \boldsymbol{c}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{Y}_i - a)\mathbb{I}_{[\hat{\boldsymbol{X}}_i^N = \hat{\boldsymbol{x}}]}, \tag{11}$$

where $\widehat{\boldsymbol{X}}_i^N = \widehat{\boldsymbol{X}}_i^{N,n} = \mathrm{Proj}_{\hat{\boldsymbol{\gamma}}^{N,n}}(\boldsymbol{X}_i)$ and $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^{N,n} = \mathrm{Proj}_{\hat{\boldsymbol{\gamma}}^{N,n}}(\boldsymbol{x})$.

An estimator of the multiple-output regression quantiles from Definition 2 is then any hyperplane of the form

$$\hat{\boldsymbol{\pi}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}} := \big\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^d \times \mathbb{R}^m : \boldsymbol{u}'\boldsymbol{y} = (\widehat{\boldsymbol{c}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}})'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{y} + \widehat{a}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}\big\}.$$

Since this estimator is entirely characterized by $\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}$, we may focus on $\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}$ when investigating the properties of these sample quantiles. We will show that, for fixed $N(\in$

$\mathbb{N}_0$) and $\boldsymbol{x}(\in S_{\boldsymbol{X}})$, $\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}$ is a weakly consistent estimator for $\widetilde{\boldsymbol{q}}^{N}_{\boldsymbol{\alpha},\boldsymbol{x}}$. The result requires restricting to $p = 2$ and reinforcing Assumption (A) into the following assumption.

Assumption (A)′ Same as Assumption (A), but with Assumption (A)(iv) replaced by the following : $P_{\boldsymbol{X}}$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$.

We then have the following result (see Appendix B for the proof).

Theorem 2. *Let Assumption (A)′ hold. Then, for any $\boldsymbol{\alpha} \in \mathcal{B}^m$, $\boldsymbol{x} \in S_{\boldsymbol{X}}$ and $N \in \mathbb{N}_0$,*

$$\left|\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}} - \widetilde{\boldsymbol{q}}^{N}_{\boldsymbol{\alpha},\boldsymbol{x}}\right| \to 0 \qquad \text{as } n \to \infty$$

in probability, provided that quantization is based on $p = 2$.

At first glance, Theorems 1-2 may appear as simple extensions, to the multiple-output case, of the corresponding single-output results in Charlier et al. (2015a). We would like to stress, however, that this extension is by no means trivial and requires different proof techniques. The main reason for this is that the concept of multiple-output regression quantiles considered is actually associated with a single-output quantile regression not only on the covariate vector $\boldsymbol{X}$ but (as soon as $m > 1$) also on the response-based quantity $\boldsymbol{Y}^{\perp}_{\boldsymbol{u}}$; this makes the problem of a different nature for the single-output case ($m = 1$) and for the multiple-output one ($m > 1$). Another reason is that, in the multiple-output case, $\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}$ is not a scalar but a vector, which makes the proof more complex as it requires, e.g., to use Hessian matrices and eigenvalues theory where, in single-output problems, classical optimization theory could be based on second derivatives.

4.2 A bootstrap modification

For small sample sizes, the stochastic gradient algorithm above is likely to provide a grid that is far from being optimal, which may have a negative impact on the proposed sample quantiles. To improve on this, we propose the same bootstrap approach as the one adopted in the single-output context by Charlier et al. (2015a,b) :

(S1) For some integer B, we first generate B samples of size n with replacement from the initial sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, that we denote as $\{\boldsymbol{\xi}_b^t,\, t = 1, \ldots, n\}$, $b = 1, \ldots, B$. We also generate initial grids $\hat{\boldsymbol{\gamma}}_b^{N,0}$ as above, by sampling randomly among the corresponding $\{\boldsymbol{\xi}_b^t,\, t = 1, \ldots, n\}$ under the constraints that the N values are pairwise distinct. We then perform B times the CLVQ algorithm with iterations based on $\{\boldsymbol{\xi}_b^t,\, t = 1, \ldots, n\}$ and with initial grid $\hat{\boldsymbol{\gamma}}_b^{N,0}$. This provides B optimal grids $\hat{\boldsymbol{\gamma}}_b^{N,n}$, $b = 1, \ldots, B$ (each of size N).

(S2) Each of these grids is then used to estimate multiple-output regression quantiles. Working again with the original sample $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$, $i = 1, \ldots, n$, we project the $\boldsymbol{X}$-part onto the grids $\hat{\boldsymbol{\gamma}}_b^{N,n}$, $b = 1, \ldots, B$. Therefore, (11) provides B estimates of $\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}$, denoted as $\hat{\boldsymbol{q}}_{\boldsymbol{\alpha},\boldsymbol{x}}^{(b),N,n}$, $b = 1, \ldots, B$. This leads to the bootstrap estimator

$$\bar{\boldsymbol{q}}_{\boldsymbol{\alpha},\boldsymbol{x}}^{N,n} = \frac{1}{B} \sum_{b=1}^{B} \hat{\boldsymbol{q}}_{\boldsymbol{\alpha},\boldsymbol{x}}^{(b),N,n}, \tag{12}$$

obtained by averaging these B estimates.

Denoting by $\hat{R}_{\alpha,\boldsymbol{x}}$ the resulting sample quantile regions (see Section 5.2.3 for more details), the parameter B should be chosen large enough to smooth the mappings $\boldsymbol{x} \mapsto \hat{R}_{\alpha,\boldsymbol{x}}$, but not too large to keep the computational burden under control. We use $B = 50$ or $B = 100$ in the sequel. The choice of N, that plays the role of the smoothing parameter in the nonparametric regression method considered, has an important impact on the proposed estimators and is discussed in the next section.

5 Numerical results

In this section, we explore the numerical performances of the proposed estimators. We first introduce in Section 5.1 a data-driven method for selecting the size N of the quantization grid. In Section 5.2, we then compare the proposed (bootstrap) quantization-based estimators with their kernel-type competitors from HLPS15.

5.1 Data-driven selection of N

In this section, we extend the N-selection criterion developed in Charlier et al. (2015b) to the present multiple-output context. This criterion is based on the minimization of an empirical integrated square error (ISE) quantity that is essentially convex in N, which allows to identify an optimal value N_{opt} of N.

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J$ be values of interest in $S_{\boldsymbol{X}}$ and $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K$ be directions of interest in $\mathcal{S}^{m-1}$, with J, K finite. The procedure to select N works as follows. For any combination of $\boldsymbol{x}_j$ and $\boldsymbol{u}_k$, we first compute $\bar{\boldsymbol{q}}^{N,n}_{\alpha\boldsymbol{u}_k,\boldsymbol{x}_j} = \frac{1}{B}\sum_{b=1}^{B} \hat{\boldsymbol{q}}^{(b),N,n}_{\alpha\boldsymbol{u}_k,\boldsymbol{x}_j}$ from B bootstrap samples as above. We then generate $\widetilde{B}$ further samples of size n with replacement from the initial sample $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, and we perform $\widetilde{B}$ times the CLVQ algorithm with iterations based on these samples. This provides $\widetilde{B}$ optimal quantization grids. Working again with the original sample $(\boldsymbol{X}_i, \boldsymbol{Y}_i)$, $i = 1, \ldots, n$ and using the $\tilde{b}$th grid, (11) provides $\widetilde{B}$ new estimations, denoted $\hat{\boldsymbol{q}}^{(B+\tilde{b}),N,n}_{\alpha\boldsymbol{u}_k,\boldsymbol{x}_j}$, $\tilde{b} = 1, \ldots, \widetilde{B}$. We then consider

$$\widehat{\text{ISE}}_{\alpha,B,\widetilde{B},J,K}(N) = \frac{1}{J}\sum_{j=1}^{J}\left(\frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{\widetilde{B}}\sum_{\tilde{b}=1}^{\widetilde{B}}\left|\bar{\boldsymbol{q}}^{N,n}_{\alpha\boldsymbol{u}_k,\boldsymbol{x}_j} - \hat{\boldsymbol{q}}^{(B+\tilde{b}),N,n}_{\alpha\boldsymbol{u}_k,\boldsymbol{x}_j}\right|^2\right)\right).$$

To make the notation lighter, we simply denote these integrated square errors as $\widehat{\text{ISE}}_\alpha(N)$ (throughout, our numerical results will be based on $m = 2$, $\widetilde{B} = 30$ and K equispaced directions in $\mathcal{S}^1$; the values of B, K and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_J$ will be made precise in each case).

These sample ISEs are to be minimized in N. Since not all values of N can be considered in practice, we rather consider

$$\hat{N}_{\alpha;\text{opt}} = \arg\min_{N\in\mathcal{N}} \widehat{\text{ISE}}_\alpha(N), \tag{13}$$

where the cardinality of $\mathcal{N}(\subset \mathbb{N}_0)$ is finite and may be chosen as a function of n.

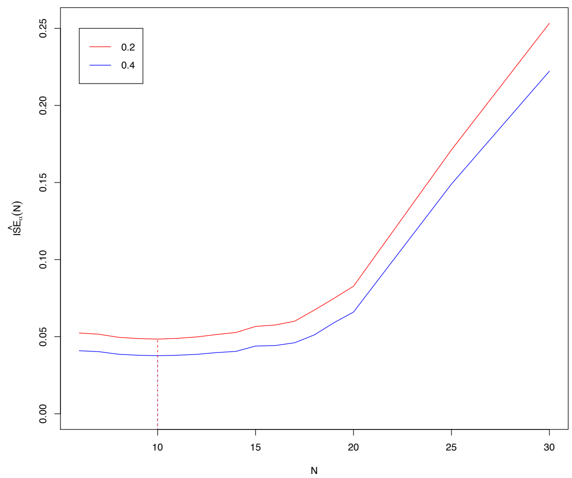


Figure 1: Plot of the mappings $N \mapsto \widehat{\mathrm{ISE}}_\alpha(N)$ ($\alpha = 0.2, 0.4$) with $B = 50$, $\widetilde{B} = 30$ and $K = 40$, averaged over 100 mutually independent replications of Model $(\mathcal{M}1)$ with sample size $n = 999$.

For illustration purposes, we simulated random samples of size $n = 999$ according to the model

$$(\mathcal{M}1) \;\; (Y_1, Y_2) = (X, X^2) + (1 + X^2)\boldsymbol{\varepsilon},$$

where $X \sim U([-2, 2])$, $\boldsymbol{\varepsilon}$ has independent $\mathcal{N}(0, 1/4)$ marginals, and X and $\boldsymbol{\varepsilon}$ are independent. It is easy to check that this model satisfies Assumptions (A)$'$ and (B). Figure 1 plots, for $\alpha = 0.2$ and $\alpha = 0.4$, the graphs of $N \mapsto \widehat{\mathrm{ISE}}_\alpha(N)$, where the ISEs are based on $B = 50$, $K = 40$ and $x_1 = -1.89, x_2 = -1.83, x_3 = -1.77, \ldots, x_J = 1.89$ (more precisely, the figure shows the average of the corresponding graphs, computed from 100 mutually independent replications). It is seen that ISE curves are indeed essentially convex in N and allow to select N equal to 10 for both values of α.

5.2 Comparison with competitors

In this section, we investigate the numerical performances of our estimator $\bar{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}$. In Section 5.2.1, we first define the competitors that will be considered. Then we compare the respective ISEs through simulations (Section 5.2.2) and show how the estimated quantile regions compare on a given sample (Section 5.2.3).

5.2.1 The competitors considered

The main competitors are the local constant and local bilinear estimators from HLPS15, that extend to the multiple-output setting the local constant and local linear estimators of Yu and Jones (1998), respectively. To describe these estimators, fix a kernel function $K : \mathbb{R}^d \to \mathbb{R}^+$ and a bandwidth h. Writing $Y_{i\boldsymbol{u}} := \boldsymbol{u}'\boldsymbol{Y}_i$ and $\mathbf{Y}^{\perp}_{i\boldsymbol{u}} := \boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}_i$, the local constant estimator is then the minimizer $\widehat{\boldsymbol{q}}^c_{\boldsymbol{\alpha},\boldsymbol{x}} = (\widehat{a}^c_{\boldsymbol{\alpha},\boldsymbol{x}}, (\widehat{\boldsymbol{c}}^c_{\boldsymbol{\alpha},\boldsymbol{x}})')'$ of

$$\boldsymbol{q} \mapsto \sum_{i=1}^{n} K\Big(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h}\Big)\rho_\alpha\big(Y_{i\boldsymbol{u}} - \boldsymbol{q}'\mathcal{X}^c_{i\boldsymbol{u}}\big), \quad \text{with } \mathcal{X}^c_{i\boldsymbol{u}} := \begin{pmatrix} 1 \\ \mathbf{Y}^{\perp}_{i\boldsymbol{u}} \end{pmatrix}. \tag{14}$$

As for the local (bi)linear estimator $\widehat{\boldsymbol{q}}^\ell_{\boldsymbol{\alpha},\boldsymbol{x}} = (\widehat{a}^\ell_{\boldsymbol{\alpha},\boldsymbol{x}}, (\widehat{\boldsymbol{c}}^\ell_{\boldsymbol{\alpha},\boldsymbol{x}})')'$, its transpose vector $(\widehat{\boldsymbol{q}}^\ell_{\boldsymbol{\alpha},\boldsymbol{x}})'$ is given by the first row of the $(d+1) \times m$ matrix $\hat{\boldsymbol{Q}}$ that minimizes

$$\boldsymbol{Q} \mapsto \sum_{i=1}^{n} K\Big(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h}\Big)\rho_\alpha\big(Y_{i\boldsymbol{u}} - (\text{vec}\,\boldsymbol{Q})'\mathcal{X}^\ell_{i\boldsymbol{u}}\big), \quad \text{with } \mathcal{X}^\ell_{i\boldsymbol{u}} := \begin{pmatrix} 1 \\ \mathbf{Y}^{\perp}_{i\boldsymbol{u}} \end{pmatrix} \otimes \begin{pmatrix} 1 \\ \boldsymbol{X}_i - \boldsymbol{x} \end{pmatrix}. \tag{15}$$

As explained in HLPS15, the local bilinear approach is more informative than the local constant one and should be more reliable close to the boundary of the covariate support. However, the price to pay is an increase of the covariate space dimension ($\mathcal{X}^c_{i\boldsymbol{u}}$ is of dimension m, whereas $\mathcal{X}^\ell_{i\boldsymbol{u}}$ is of dimension $m(d+1)$). We refer to HLPS15 for more details on these approaches.

In the sequel, we consider $d = 1$ and $m = 2$ in order to provide graphical representations of the corresponding quantile regions. The kernel K will be the density of the bivariate standard Gaussian distribution and we choose, as in most applications in HLPS15,

$$h = \frac{3s_x}{n^{1/5}}, \tag{16}$$

where s_x stands for the empirical standard deviation of $X_1, \ldots, X_n$.

5.2.2 Comparison of ISEs

We now compare our bootstrap estimators with the competitors above in terms of ISEs. To do so, we generated 500 independent samples of size $n = 999$ from

($\mathcal{M}1$) $(Y_1, Y_2) = (X, X^2) + (1 + X^2)\boldsymbol{\varepsilon}_1$,

($\mathcal{M}2$) $(Y_1, Y_2) = (X, X^2) + \boldsymbol{\varepsilon}_1$,

($\mathcal{M}3$) $(Y_1, Y_2) = (X, X^2) + \big(1 + \frac{3}{2}\big(\sin\big(\frac{\pi}{2}X\big)\big)^2\big)\boldsymbol{\varepsilon}_2$,

where $X \sim U([-2, 2])$, $\boldsymbol{\varepsilon}_1$ has independent $\mathcal{N}(0, 1/4)$ marginals, $\boldsymbol{\varepsilon}_2$ has independent $\mathcal{N}(0, 1)$ marginals, and X is independent of $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$. These models, that were already considered in HLPS15, are easily checked to satisfy Assumptions (A)$'$ and (B).

Both the proposed quantization-based quantiles and their competitors are indexed by a scalar order $\alpha \in (0, 1)$ and a direction $\boldsymbol{u} \in \mathcal{S}^1$. In this section, we compare efficiencies when estimating a given conditional quantile $\boldsymbol{q}_{\alpha\boldsymbol{u},x}$. In the sequel, we still work with $\alpha = 0.2, 0.4$ and we fix $\boldsymbol{u} = (0, 1)'$.

For each replication in each model, the various quantile estimators were computed, based on the bandwidth h in (16) for the HLPS15 estimators and based on $B = 100$ and the N-selection procedure described in Section 5.1 (with $x_1 = -1.89, x_2 = -1.83, \ldots, x_J = 1.89$, $\mathcal{N} = \{10, 15, 20\}$ and $K = 1$ direction, namely the direction $\boldsymbol{u} = (0, 1)'$ above) for the quantization-based estimators. For each estimator, we then evaluated

$$\mathrm{ISE}^a_\alpha = \sum_{j=1}^{J} \big(\widehat{a}_{\alpha\boldsymbol{u},x_j} - a_{\alpha\boldsymbol{u},x_j}\big)^2 \quad \text{and} \quad \mathrm{ISE}^c_\alpha = \sum_{j=1}^{J} \big(\widehat{c}_{\alpha\boldsymbol{u},x_j} - c_{\alpha\boldsymbol{u},x_j}\big)^2,$$

still for $x_1 = -1.89, x_2 = -1.83, \ldots, x_J = 1.89$; here, $\widehat{a}_{\alpha\boldsymbol{u},x_j}$ stands for $\bar{a}^{N,n}_{\alpha\boldsymbol{u},x_j}$, $\widehat{a}^c_{\alpha\boldsymbol{u},x_j}$ or $\widehat{a}^\ell_{\alpha\boldsymbol{u},x_j}$ and $\widehat{c}_{\alpha\boldsymbol{u},x_j}$ for $\bar{c}^{N,n}_{\alpha\boldsymbol{u},x_j}$, $\widehat{c}^c_{\alpha\boldsymbol{u},x_j}$ or $\widehat{c}^\ell_{\alpha\boldsymbol{u},x_j}$. Figure 2 reports, for each model and each estimator, the boxplots of ISE^a_α and ISE^c_α obtained from the 500 replications considered.

Results reveal that the proposed estimator $\bar{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},x}$ and the local bilinear estimator $\widehat{\boldsymbol{q}}^\ell_{\boldsymbol{\alpha},x}$ perform significantly better than the local constant estimator $\widehat{\boldsymbol{q}}^c_{\boldsymbol{\alpha},x}$, particularly for the estimation of the first component $a_{\boldsymbol{\alpha},x}$ of $\boldsymbol{q}_{\boldsymbol{\alpha},x}$. In most cases, the proposed estimator $\bar{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},x}$ actually also dominates the local bilinear one $\widehat{\boldsymbol{q}}^{\ell}_{\boldsymbol{\alpha},x}$ (the only cases where the opposite holds relate to the estimation of $c_{\boldsymbol{\alpha},x}$ and the difference of performance is then really small). It should be noted that the quantization-based estimator $\bar{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},x}$ is local constant in nature, which makes it remarkable that it behaves well in terms of ISE compared to its local bilinear kernel competitor.

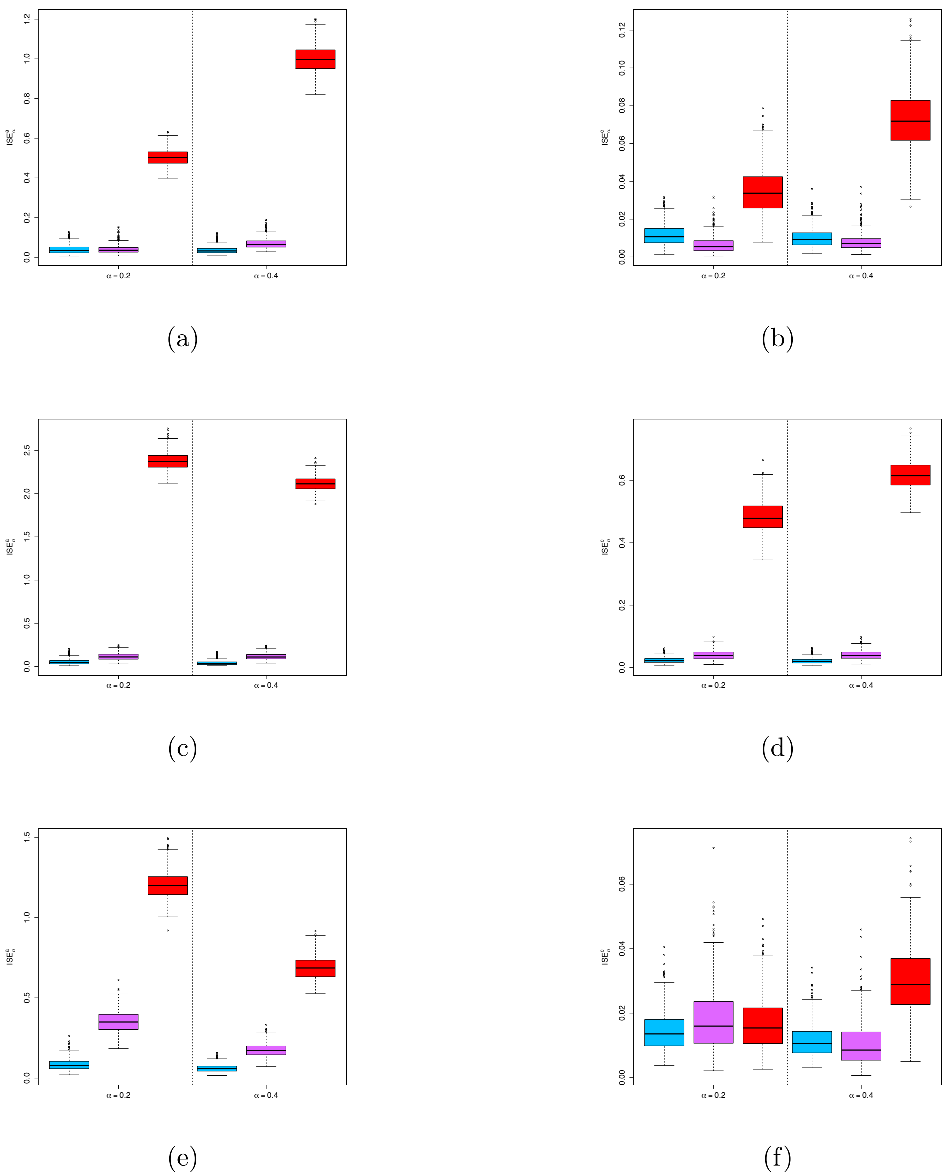
5.2.3 Comparison of sample quantile regions

As explained in HLPS15, the regression quantile regions $R_{\alpha,\boldsymbol{x}}$ in (4) are extremely informative about the conditional distribution of the response, which makes it desirable to obtain well-behaved estimations of these regions. That is why we now compare the sample regions obtained from the proposed quantization-based quantile estimators with the kernel ones from HLPS15. Irrespective of the quantile coefficient estimators $\hat{\boldsymbol{q}}_{\alpha\boldsymbol{u},\boldsymbol{x}} = (\hat{a}_{\alpha\boldsymbol{u},\boldsymbol{x}}, \hat{\boldsymbol{c}}'_{\alpha\boldsymbol{u},\boldsymbol{x}})'$ used, the corresponding sample regions are obtained as

$$\hat{R}_{\alpha,\boldsymbol{x}} := \cap_{\boldsymbol{u}\in\mathcal{S}^{m-1}_F}\big\{\boldsymbol{y}\in\mathbb{R}^m : \boldsymbol{u}'\boldsymbol{y} \geq \hat{\boldsymbol{c}}'_{\alpha\boldsymbol{u},\boldsymbol{x}}\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{y} + \hat{a}_{\alpha\boldsymbol{u},\boldsymbol{x}}\big\},$$

where $\mathcal{S}^{m-1}_F$ is a finite subset of $\mathcal{S}^{m-1}$; compare with (4).

We considered a random sample of size $n = 999$ from Model $(\mathcal{M}1)$ and computed, for the various estimation methods, $\hat{R}_{\alpha,x}$ for $\alpha = 0.2, 0.4$ and for $x = -1.89, -1.83, -1.77, \ldots, 1.89$; in each case, $\mathcal{S}^{m-1}_F = \mathcal{S}^1_F$ is made of 360 equispaced directions in $\mathcal{S}^1$. For the kernel-based estimators,, we did not select h following the data-driven procedure mentioned in Section 5.2.1, but chose it equal to 0.37, as proposed in HLPS15, Figure 3. For the quantization-based estimators, N was selected according to the data-driven method from Section 5.1 (with $B = 100$, $K = 360$, $\mathcal{N} = \{5, 10, 15, 20\}$, and still $x_1 = -1.89$, $x_2 = -1.83$, $\ldots$, $x_J = 1.89$), which led to the optimal value $N = 10$. The resulting sample quantile regions, obtained from the quantization-based method and from the local constant and local bilinear kernel ones, are plotted in Figure 3. For comparison purposes, the population quantile regions $R_{\alpha,x}$ are also reported there. We observe that



(a) (b) (c) (d) (e) (f)

Figure 2: Boxplots, for $\alpha = 0.2, 0.4$ and $\boldsymbol{u} = (0,1)'$, of ISE^a_α (left) and of ISE^c_α (right) for various conditional quantile estimators obtained from 500 independent random samples according to Models ($\mathcal{M}1$) (top), ($\mathcal{M}2$) (middle) and ($\mathcal{M}3$) (bottom), with size $n = 999$. The estimators considered are the quantization-based estimator $\bar{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},x}$ (in blue), the local bilinear estimator $\widehat{\boldsymbol{q}}^{\ell}_{\boldsymbol{\alpha},x}$ (in purple) and the local constant estimator $\widehat{\boldsymbol{q}}^{c}_{\boldsymbol{\alpha},x}$ (in red).

the quantization-based and local bilinear methods provide quantile regions that are nice and close to the population ones. They succeed in particular in catching the underlying heteroscedasticity. Clearly, they perform better than the local constant methods close to the boundary of the covariate range. While the local (bi)linear methods, as already mentioned, are known to exhibit good boundary behaviour, it is surprising that the quantization-based method also behaves well in this respect, since this method is of a local constant nature. Finally, it should be noted that, unlike the smoothing parameter of the local constant/bilinear methods (namely, h), that of the quantization-based method (namely, N) was chosen in a fully data-driven way.

6 Summary and perspectives for future research

In this paper, we defined new nonparametric estimators of the multiple-output regression quantiles proposed in HLPS15. The main idea is to perform localization through the concept of optimal quantization rather than through standard kernel methods. We derived consistency results that generalize to the multiple-output context those obtained in Charlier et al. (2015a). Moreover, the good empirical efficiency properties of quantization-based quantiles showed in Charlier et al. (2015b) extend to the multiple-output context. In particular, the proposed quantization-based sample quantiles, that are local constant in nature, outperform their kernel-based counterparts, both in terms of integrated square errors and in terms of visual inspection of the corresponding quantile regions. The proposed quantiles actually perform as well as (and sometimes even strictly dominate) the local bilinear kernel estimators from HLPS15. The data-driven selection procedure we proposed for the smoothing parameter N involved in the quantization-based method allows to make the estimation fully automatic. Our estimation procedure was actually implemented in R and the code is available from the authors on simple request.

We conclude by stating a few open problems that are left for future research. In

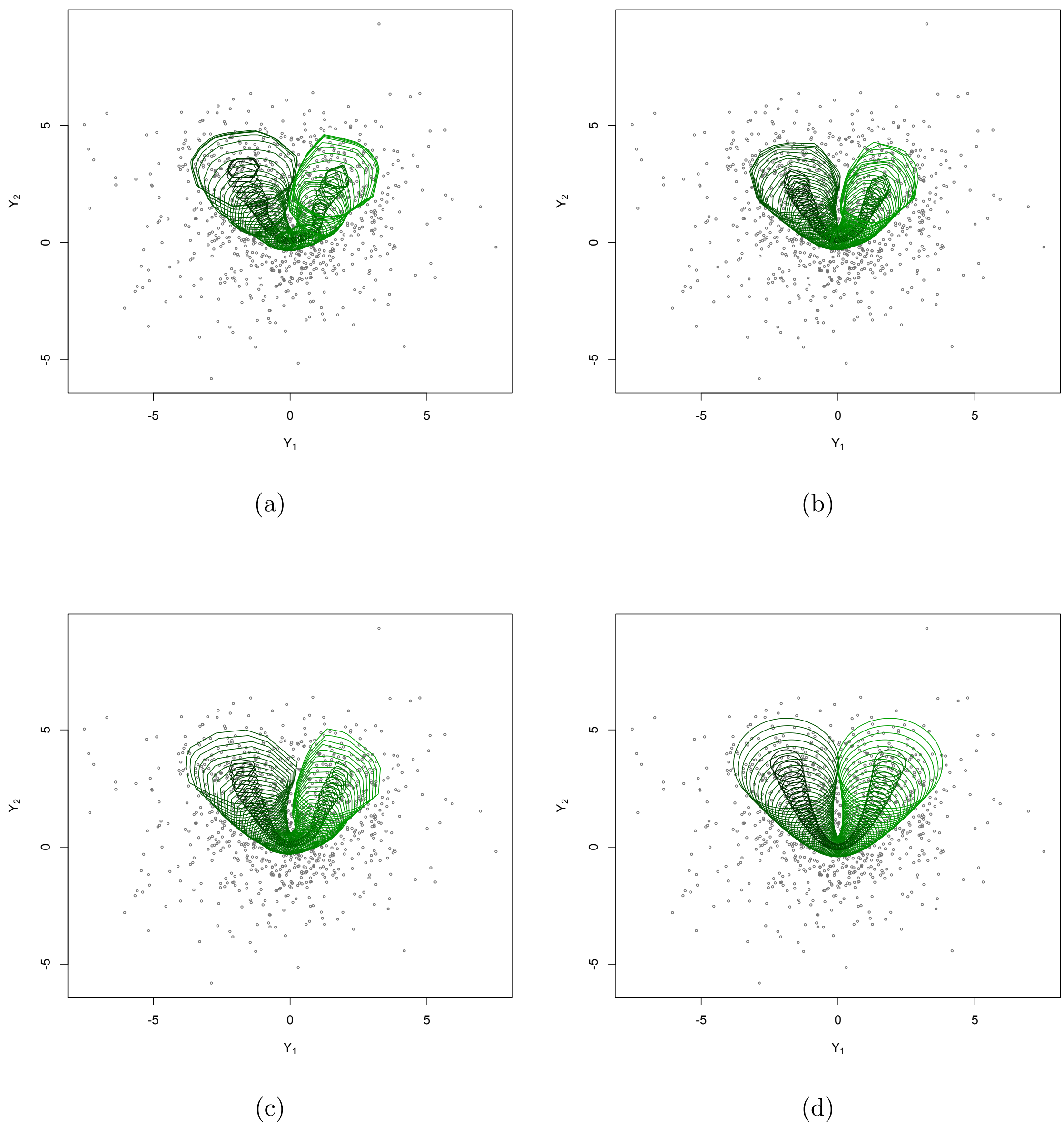


Figure 3: (a)-(c) The sample quantile regions $\hat{R}_{\alpha,x}$, for $\alpha = 0.2, 0.4$ and $x = -1.89, -1.83, -1.77, \ldots, 1.89$, computed from a random sample of size $n = 999$ from Model ($\mathcal{M}1$) by using (a) the quantization-based method, (b) the local constant kernel method, and (c) the local bilinear kernel one. (d) The corresponding population quantile regions $R_{\alpha,x}$.

principle, Theorem 1 and Theorem 2 could be combined to provide an asymptotic result stating that $|\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}| \to 0$ as $n \to \infty$ in probability, with $N = N_n$ going to infinity at an appropriate rate. However, obtaining such a result is extremely delicate, since all convergence results available for the CLVQ algorithm are as $n \to \infty$ with N fixed. Obviously, once such a weak consistency is proved, another challenging task would be to derive the asymptotic distribution of $\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}$ and to design confidence zones for the population quantity $\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}$.

Acknowledgements

The authors would like to thank the Editor-in-chief, the Associate Editor and two anonymous referees for their insightful comments and suggestions, that led to a substantial improvement of a previous version of this work.

Corresponding author : Isabelle Charlier

Université Libre de Bruxelles, ECARES and Département de Mathématique,

50, Avenue F.D. Roosevelt, CP114/04, B-1050, Brussels, Belgium.

E-mail: ischarli@ulb.ac.be

A Proof of Theorem 1

The proof requires several lemmas. First, recall that $G_{a,\boldsymbol{c}}(\boldsymbol{x}) = \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}^{\perp}_{\boldsymbol{u}} - a)|\boldsymbol{X} = \boldsymbol{x}]$ and consider the corresponding quantized quantity $\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) = \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}^{\perp}_{\boldsymbol{u}} - a)|\widetilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}]$. Since $\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}} = (a_{\boldsymbol{\alpha},\boldsymbol{x}}, \boldsymbol{c}'_{\boldsymbol{\alpha},\boldsymbol{x}})'$ and $\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} = (\widetilde{a}^N_{\boldsymbol{\alpha},\boldsymbol{x}}, (\widetilde{\boldsymbol{c}}^N_{\boldsymbol{\alpha},\boldsymbol{x}})')'$ are defined as the vectors achieving the minimum of $G_{a,\boldsymbol{c}}(\boldsymbol{x})$ and $\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})$ respectively, we naturally start controlling the distance between $\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})$ and $G_{a,\boldsymbol{c}}(\boldsymbol{x})$. This is achieved below in Lemma 5, whose proof requires the following preliminary lemmas. Throughout this appendix, C is a constant that may vary from line to line.

Lemma 1. *Let Assumption (A) hold and fix* $\boldsymbol{\alpha} = \alpha \boldsymbol{u} \in \mathcal{B}^m$, $a \in \mathbb{R}$ *and* $\boldsymbol{c} \in \mathbb{R}^{m-1}$. *Then for* $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$, $|G_{a,\boldsymbol{c}}(\boldsymbol{x}_1) - G_{a,\boldsymbol{c}}(\boldsymbol{x}_2)| \leq \max(\alpha, 1-\alpha)\sqrt{1+|\boldsymbol{c}|^2}\,([\boldsymbol{M}_{a,\cdot}]_{\mathrm{Lip}} + [\boldsymbol{M}_{b,\cdot}]_{\mathrm{Lip}}\|\boldsymbol{\varepsilon}\|_1)|\boldsymbol{x}_1 - \boldsymbol{x}_2|$.

Proof. For $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$, we have

$$
\begin{aligned}
&|G_{a,\boldsymbol{c}}(\boldsymbol{x}_1) - G_{a,\boldsymbol{c}}(\boldsymbol{x}_2)| \\
&\quad = \left|\mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^\perp - a)|\boldsymbol{X} = \boldsymbol{x}_1] - \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^\perp - a)|\boldsymbol{X} = \boldsymbol{x}_2]\right| \\
&\quad = \left|\mathrm{E}[\rho_\alpha((\boldsymbol{u} - \boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{c})'\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{\varepsilon}) - a)|\boldsymbol{X} = \boldsymbol{x}_1] - \mathrm{E}[\rho_\alpha((\boldsymbol{u} - \boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{c})'\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{\varepsilon}) - a)|\boldsymbol{X} = \boldsymbol{x}_2]\right| \\
&\quad = \left|\mathrm{E}[\rho_\alpha((\boldsymbol{u} - \boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{c})'\boldsymbol{M}(\boldsymbol{x}_1, \boldsymbol{\varepsilon}) - a) - \rho_\alpha((\boldsymbol{u} - \boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{c})'\boldsymbol{M}(\boldsymbol{x}_2, \boldsymbol{\varepsilon}) - a)]\right|,
\end{aligned}
$$

where we used the independence of $\boldsymbol{X}$ and $\boldsymbol{\varepsilon}$. Using the fact that ρ_α is a Lipschitz function with Lipschitz constant $[\rho_\alpha]_{\mathrm{Lip}} := \max(\alpha, 1-\alpha)$, then the Cauchy-Schwarz inequality, this yields

$$
\begin{aligned}
|G_{a,\boldsymbol{c}}(\boldsymbol{x}_1) - G_{a,\boldsymbol{c}}(\boldsymbol{x}_2)| &\leq [\rho_\alpha]_{\mathrm{Lip}}\mathrm{E}[|(\boldsymbol{u} - \boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{c})'(\boldsymbol{M}(\boldsymbol{x}_1, \boldsymbol{\varepsilon}) - \boldsymbol{M}(\boldsymbol{x}_2, \boldsymbol{\varepsilon}))|], \\
&\leq [\rho_\alpha]_{\mathrm{Lip}}|\boldsymbol{u} - \boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{c}|\,\mathrm{E}[|\boldsymbol{M}(\boldsymbol{x}_1, \boldsymbol{\varepsilon}) - \boldsymbol{M}(\boldsymbol{x}_2, \boldsymbol{\varepsilon})|] \\
&\leq [\rho_\alpha]_{\mathrm{Lip}}|(\boldsymbol{u}, \boldsymbol{\Gamma}_{\boldsymbol{u}})(1, -\boldsymbol{c}')'|\,\mathrm{E}[|\boldsymbol{M}_{a,\boldsymbol{x}_1} - \boldsymbol{M}_{a,\boldsymbol{x}_2} + (\boldsymbol{M}_{b,\boldsymbol{x}_1} - \boldsymbol{M}_{b,\boldsymbol{x}_2})\boldsymbol{\varepsilon}|] \\
&\leq [\rho_\alpha]_{\mathrm{Lip}}\sqrt{1+|\boldsymbol{c}|^2}\,([\boldsymbol{M}_{a,\cdot}]_{\mathrm{Lip}} + [\boldsymbol{M}_{b,\cdot}]_{\mathrm{Lip}}\|\boldsymbol{\varepsilon}\|_1)|\boldsymbol{x}_1 - \boldsymbol{x}_2|,
\end{aligned}
$$

where we used Assumptions (A)(iii)-(v). □

The following lemma shows that, under the assumptions considered, the regularity property (10) extends from the error density $f^{\boldsymbol{\varepsilon}}(\cdot)$ to the conditional density $f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\cdot)$.

Lemma 2. *Let Assumptions (A) and (B) hold and fix* $\boldsymbol{x} \in S_{\boldsymbol{X}}$. *Then, for some constants* $C > 0$, $r > m-1$ *and* $s > 0$, *we have*

$$
\left|f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}_1) - f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}_2)\right| \leq C|\boldsymbol{y}_1 - \boldsymbol{y}_2|^s\left(1 + \tfrac{1}{2}|\boldsymbol{y}_1 + \boldsymbol{y}_2|^2\right)^{-(3+r+s)/2}, \qquad (17)
$$

for all $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{R}^m$.

Proof. Assumption (A) allows to rewrite the conditional density of $\boldsymbol{Y}$ given $\boldsymbol{X}=\boldsymbol{x}$ as

$$f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}) = \frac{1}{|\det(\boldsymbol{M}_{b,\boldsymbol{x}})|} f^{\boldsymbol{\varepsilon}}\big(\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}-\boldsymbol{M}_{a,\boldsymbol{x}})\big),$$

for all $\boldsymbol{y}\in\mathbb{R}^m$. Hence, we have

$$\begin{aligned}
&|f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}_1) - f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}_2)| \\
&\quad = \frac{1}{|\det(\boldsymbol{M}_{b,\boldsymbol{x}})|}\Big| f^{\boldsymbol{\varepsilon}}\big(\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_1-\boldsymbol{M}_{a,\boldsymbol{x}})\big) - f^{\boldsymbol{\varepsilon}}\big(\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_2-\boldsymbol{M}_{a,\boldsymbol{x}})\big)\Big|.
\end{aligned}$$

Now, Assumption (B) entails

$$\begin{aligned}
|f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}_1) - f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(\boldsymbol{y}_2)| &\leq \frac{C\big|\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_1-\boldsymbol{y}_2)\big|^s}{|\det(\boldsymbol{M}_{b,\boldsymbol{x}})|\big(1+\frac{1}{2}\big|\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_1+\boldsymbol{y}_2-2\boldsymbol{M}_{a,\boldsymbol{x}})\big|^2\big)^{(3+r+s)/2}} \\
&\leq C|(\boldsymbol{y}_1-\boldsymbol{y}_2)|^s\big(1+\tfrac{1}{2}\big|\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_1+\boldsymbol{y}_2-2\boldsymbol{M}_{a,\boldsymbol{x}})\big|^2\big)^{-(3+r+s)/2},
\end{aligned}$$

where the second inequality comes from the compactness of $S_{\boldsymbol{X}}$ and the continuity of the mapping $\boldsymbol{x}\mapsto\boldsymbol{M}_{b,\boldsymbol{x}}$. The result then follows from the fact that

$$\begin{aligned}
&\frac{1+\frac{1}{2}\big|\boldsymbol{y}_1+\boldsymbol{y}_2\big|^2}{1+\frac{1}{2}\big|\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_1+\boldsymbol{y}_2-2\boldsymbol{M}_{a,\boldsymbol{x}})\big|^2} \\
&\qquad = \frac{1+\frac{1}{2}\big|\boldsymbol{M}_{b,\boldsymbol{x}}\{\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_1+\boldsymbol{y}_2-2\boldsymbol{M}_{a,\boldsymbol{x}})\}+2\boldsymbol{M}_{a,\boldsymbol{x}}\big|^2}{1+\frac{1}{2}\big|\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_1+\boldsymbol{y}_2-2\boldsymbol{M}_{a,\boldsymbol{x}})\big|^2} \\
&\qquad \leq \frac{C+C\big|\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_1+\boldsymbol{y}_2-2\boldsymbol{M}_{a,\boldsymbol{x}})\big|^2}{1+\frac{1}{2}\big|\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}(\boldsymbol{y}_1+\boldsymbol{y}_2-2\boldsymbol{M}_{a,\boldsymbol{x}})\big|^2} \leq C,
\end{aligned}$$

where we used again the continuity of $\boldsymbol{x}\mapsto\boldsymbol{M}_{a,\boldsymbol{x}}$ and $\boldsymbol{x}\mapsto\boldsymbol{M}_{b,\boldsymbol{x}}$, and the compactness of $S_{\boldsymbol{X}}$. □

We will also need the following result belonging to linear algebra.

Lemma 3. *For $p>q\geq 1$, let $\boldsymbol{V}=(\boldsymbol{v}_1\ldots\boldsymbol{v}_q)$ be a $p\times q$ full-rank matrix and H be a q-dimensional vector subspace of $\mathbb{R}^p$. Then, there exists a $p\times q$ matrix $\boldsymbol{U}=(\boldsymbol{u}_1\ldots\boldsymbol{u}_q)$ whose columns form an orthonormal basis of H and such that $\mathbf{I}_q+\boldsymbol{U}'\boldsymbol{V}$ is invertible.*

Proof. We fix $p \geq 2$ and we prove the result by induction on q between $q = 1$ and $q = p-1$. We start with the case $q = 1$ and take $\boldsymbol{U} = (\boldsymbol{u}_1)$, where $\boldsymbol{u}_1$ is an arbitrary unit p-vector in H. If $\det(1 + \boldsymbol{U}'\boldsymbol{V}) = 1 + \boldsymbol{u}_1'\boldsymbol{v}_1 = 0$, then we may alternatively take $\boldsymbol{U}_* = (-\boldsymbol{u}_1)$, which provides $\det(1 + \boldsymbol{U}_*'\boldsymbol{V}) = 1 - \boldsymbol{u}_1'\boldsymbol{v}_1 = 2 \neq 0$. Assume then that the result holds for q (with $q < p-1$) and let us prove it for $q+1$. Pick an arbitrary $p \times (q+1)$ matrix $\boldsymbol{U} = (\boldsymbol{u}_1 \ldots \boldsymbol{u}_{q+1})$ whose columns form an orthonormal basis of the $(q+1)$-dimensional vector subspace H of $\mathbb{R}^p$. Assume that $\det(\boldsymbol{I}_{q+1} + \boldsymbol{U}'\boldsymbol{V}) = 0$, where $\boldsymbol{V} = (\boldsymbol{v}_1 \ldots \boldsymbol{v}_{q+1})$ is the given $p \times (q+1)$ full-rank matrix. Letting $\boldsymbol{U}_{-1} = (\boldsymbol{u}_2 \ldots \boldsymbol{u}_{q+1})$ and $\boldsymbol{V}_{-1} = (\boldsymbol{v}_2 \ldots \boldsymbol{v}_{q+1})$, an expansion of the determinant along the first row provides

$$0 = \det(\mathbf{I}_{q+1} + \boldsymbol{U}'\boldsymbol{V}) = (\boldsymbol{u}_1'\boldsymbol{v}_1 + 1)\det(\mathbf{I}_q + \boldsymbol{U}_{-1}'\boldsymbol{V}_{-1}) + \textstyle\sum_{i=2}^{q+1}(-1)^{i+1}\boldsymbol{u}_1'\boldsymbol{v}_i \det(\boldsymbol{W}_i),$$

for some $q \times q$ matrices $\boldsymbol{W}_2, \ldots, \boldsymbol{W}_m$. With $\boldsymbol{U}_* = (-\boldsymbol{u}_1, \boldsymbol{u}_2 \ldots \boldsymbol{u}_{q+1})$, we then have

$$\begin{aligned}\det(\mathbf{I}_{q+1} + \boldsymbol{U}_*'\boldsymbol{V}) &= (-\boldsymbol{u}_1'\boldsymbol{v}_1 + 1)\det(\mathbf{I}_q + \boldsymbol{U}_{-1}'\boldsymbol{V}_{-1}) - \textstyle\sum_{i=2}^{q+1}(-1)^{i+1}\boldsymbol{u}_1'\boldsymbol{v}_i \det(\boldsymbol{W}_i)\\ &= 2\det(\mathbf{I}_q + \boldsymbol{U}_{-1}'\boldsymbol{V}_{-1}) - \det(\mathbf{I}_{q+1} + \boldsymbol{U}'\boldsymbol{V})\\ &= 2\det(\mathbf{I}_q + \boldsymbol{U}_{-1}'\boldsymbol{V}_{-1}).\end{aligned}$$

The induction hypothesis guarantees that $\boldsymbol{U}_{-1}$ can be chosen such that $\det(\mathbf{I}_q + \boldsymbol{U}_{-1}'\boldsymbol{V}_{-1})$ is non-zero, which establishes the result. □

We can now calculate explicitly the gradient and the Hessian matrix of the function $(a, \boldsymbol{c}) \mapsto G_{a,\boldsymbol{c}}(\boldsymbol{x})$ for any $\boldsymbol{x}$ in the support $S_{\boldsymbol{X}}$ of $\boldsymbol{X}$, and derive some important properties of this Hessian matrix.

Lemma 4. *Let Assumptions (A) and (B) hold. Then (i) $(a, \boldsymbol{c}) \mapsto G_{a,\boldsymbol{c}}(\boldsymbol{x})$ is twice differentiable at any $\boldsymbol{x} \in S_{\boldsymbol{X}}$, with gradient vector*

$$\boldsymbol{\nabla} G_{a,\boldsymbol{c}}(\boldsymbol{x}) = \begin{pmatrix} \nabla_a G_{a,\boldsymbol{c}}(\boldsymbol{x}) \\ \boldsymbol{\nabla}_{\boldsymbol{c}} G_{a,\boldsymbol{c}}(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} P[\mathbf{u}'\boldsymbol{Y} < a + \boldsymbol{c}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{Y} | \boldsymbol{X} = \boldsymbol{x}] - \alpha \\ \mathrm{E}\big[\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{Y}(\mathbb{I}_{[\boldsymbol{u}'\boldsymbol{Y} < a + \boldsymbol{c}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{Y}]} - \alpha) | \boldsymbol{X} = \boldsymbol{x}\big] \end{pmatrix} \qquad (18)$$

and Hessian matrix

$$\boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x}) = \int_{\mathbb{R}^{m-1}} \begin{pmatrix} 1 & \boldsymbol{t}' \\ \boldsymbol{t} & \boldsymbol{t}\boldsymbol{t}' \end{pmatrix} f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}\big((a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u} + \boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{t}\big)\, d\boldsymbol{t};$$

(ii) for any $(a, \boldsymbol{c}, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}^{m-1} \times S_{\boldsymbol{X}}$*,* $\boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x})$ *is positive definite; (iii)* $(a, \boldsymbol{c}, \boldsymbol{x}) \mapsto \boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x})$ *is continuous over* $\mathbb{R} \times \mathbb{R}^{m-1} \times S_{\boldsymbol{X}}$*.*

Proof. (i) Let

$$\eta_{\boldsymbol{\alpha}}(a, \boldsymbol{c}) = \big(\mathbb{I}_{[\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a < 0]} - \alpha\big) \begin{pmatrix} 1 \\ \boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} \end{pmatrix}.$$

For any $(a, \boldsymbol{c}')', (a_0, \boldsymbol{c}'_0)' \in \mathbb{R}^m$, we then have

$$\begin{aligned} &\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a) - \rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'_0\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a_0) - (a - a_0, \boldsymbol{c}' - \boldsymbol{c}'_0)\eta_{\boldsymbol{\alpha}}(a_0, \boldsymbol{c}_0) \\ &\qquad = (\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a)\Big\{\mathbb{I}_{[\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'_0\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a_0 < 0]} - \mathbb{I}_{[\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a < 0]}\Big\} \geq 0, \end{aligned} \tag{19}$$

so that $\eta_{\boldsymbol{\alpha}}(a, \boldsymbol{c})$ is a subgradient for $(a, \boldsymbol{c}) \mapsto \rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a)$. Hence,

$$\boldsymbol{\nabla} G_{a,\boldsymbol{c}}(\boldsymbol{x}) = \boldsymbol{\nabla}_{a,\boldsymbol{c}} \mathrm{E}[\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a)|\boldsymbol{X} = \boldsymbol{x}] = \mathrm{E}[\eta_{\boldsymbol{\alpha}}(a, \boldsymbol{c})|\boldsymbol{X} = \boldsymbol{x}], \tag{20}$$

which readily provides (18). Let us now show that

$$|\boldsymbol{\nabla} G_{a+\Delta_a, \boldsymbol{c}+\boldsymbol{\Delta}_{\boldsymbol{c}}}(\boldsymbol{x}) - \boldsymbol{\nabla} G_{a,\boldsymbol{c}}(\boldsymbol{x}) - \boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x})(\Delta_a, \boldsymbol{\Delta}'_{\boldsymbol{c}})'| = o\big(|(\Delta_a, \boldsymbol{\Delta}'_{\boldsymbol{c}})'|\big)$$

as $(\Delta_a, \boldsymbol{\Delta}'_{\boldsymbol{c}})' \to 0$. From (20) and the identity

$$\int_{a+\boldsymbol{c}'\boldsymbol{t}}^{(a+\Delta_a)+(\boldsymbol{c}+\boldsymbol{\Delta}_{\boldsymbol{c}})'\boldsymbol{t}} \begin{pmatrix} 1 \\ \boldsymbol{t} \end{pmatrix} dz = \begin{pmatrix} 1 & \boldsymbol{t}' \\ \boldsymbol{t} & \boldsymbol{t}\boldsymbol{t}' \end{pmatrix} \begin{pmatrix} \Delta_a \\ \boldsymbol{\Delta}_{\boldsymbol{c}} \end{pmatrix},$$

we obtain

$$\begin{aligned}
&\boldsymbol{\nabla} G_{a+\Delta_a,\boldsymbol{c}+\boldsymbol{\Delta_c}}(\boldsymbol{x}) - \boldsymbol{\nabla} G_{a,\boldsymbol{c}}(\boldsymbol{x}) - \boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x})(\Delta_a, \boldsymbol{\Delta}'_{\boldsymbol{c}})' \\
&\quad = \mathrm{E}[\eta_{\boldsymbol{\alpha}}(a+\Delta_a, \boldsymbol{c}+\boldsymbol{\Delta_c}) - \eta_{\boldsymbol{\alpha}}(a,\boldsymbol{c}) | \boldsymbol{X}=\boldsymbol{x}] \\
&\qquad\quad - \int_{\mathbb{R}^{m-1}} \begin{pmatrix} 1 & \boldsymbol{t}' \\ \boldsymbol{t} & \boldsymbol{t}\boldsymbol{t}' \end{pmatrix} \begin{pmatrix} \Delta_a \\ \boldsymbol{\Delta_c} \end{pmatrix} f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}\big((a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u} + \boldsymbol{\Gamma_u t}\big)\, d\boldsymbol{t} \\
&\quad = \int_{\mathbb{R}^{m-1}} \int_{\mathbb{R}} \big(\mathbb{I}_{[z-(\boldsymbol{c}+\boldsymbol{\Delta_c})'\boldsymbol{t}-(a+\Delta_a)<0]} - \mathbb{I}_{[z-\boldsymbol{c}'\boldsymbol{t}-a<0]}\big) \begin{pmatrix} 1 \\ \boldsymbol{t} \end{pmatrix} f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(z\boldsymbol{u} + \boldsymbol{\Gamma_u t})\, dz d\boldsymbol{t} \\
&\qquad\quad - \int_{\mathbb{R}^{m-1}} \int_{a+\boldsymbol{c}'\boldsymbol{t}}^{(a+\Delta_a)+(\boldsymbol{c}+\boldsymbol{\Delta_c})'\boldsymbol{t}} \begin{pmatrix} 1 \\ \boldsymbol{t} \end{pmatrix} f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}\big((a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u} + \boldsymbol{\Gamma_u t}\big)\, dz d\boldsymbol{t} \\
&\quad = \int_{\mathbb{R}^{m-1}} \int_{a+\boldsymbol{c}'\boldsymbol{t}}^{(a+\Delta_a)+(\boldsymbol{c}+\boldsymbol{\Delta_c})'\boldsymbol{t}} \begin{pmatrix} 1 \\ \boldsymbol{t} \end{pmatrix} \big\{ f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(z\boldsymbol{u} + \boldsymbol{\Gamma_u t}) - f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}\big((a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u} + \boldsymbol{\Gamma_u t}\big) \big\}\, dz d\boldsymbol{t}.
\end{aligned}$$

Now, by Lemma 2, one has, for any z between $a+\boldsymbol{c}'\boldsymbol{t}$ and $(a+\Delta_a)+(\boldsymbol{c}+\boldsymbol{\Delta_c})'\boldsymbol{t}$,

$$\begin{aligned}
&|f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}(z\boldsymbol{u} + \boldsymbol{\Gamma_u t}) - f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}\big((a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u} + \boldsymbol{\Gamma_u t}\big)| \\
&\qquad\qquad \le \frac{C|z-a-\boldsymbol{c}'\boldsymbol{t}|^s}{\big(1+\frac{1}{2}|(z+a+\boldsymbol{c}'\boldsymbol{t})u + 2\boldsymbol{\Gamma_u t}|^2\big)^{(3+r+s)/2}} \le \frac{C|\Delta_a + \boldsymbol{\Delta}'_{\boldsymbol{c}}\boldsymbol{t}|^s}{|(1,\boldsymbol{t}')'|^{3+r+s}}.
\end{aligned}$$

This entails

$$\begin{aligned}
|\boldsymbol{\nabla} G_{a+\Delta_a,\boldsymbol{c}+\boldsymbol{\Delta_c}}(\boldsymbol{x}) - \boldsymbol{\nabla} G_{a,\boldsymbol{c}}(\boldsymbol{x}) - \boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x})(\Delta_a, \boldsymbol{\Delta}'_{\boldsymbol{c}})'| &\le C \int_{\mathbb{R}^{m-1}} \frac{|\Delta_a + \boldsymbol{\Delta}'_{\boldsymbol{c}}\boldsymbol{t}|^{1+s}}{|(1,\boldsymbol{t}')'|^{2+r+s}}\, d\boldsymbol{t} \\
&\le C|(\Delta_a, \boldsymbol{\Delta}'_{\boldsymbol{c}})'|^{1+s} \int_{\mathbb{R}^{m-1}} \frac{1}{|(1,\boldsymbol{t}')'|^{1+r}}\, d\boldsymbol{t} = o(|(\Delta_a, \boldsymbol{\Delta}'_{\boldsymbol{c}})'|),
\end{aligned}$$

as $(\Delta_a, \boldsymbol{\Delta}'_{\boldsymbol{c}})' \to 0$. Therefore, $(a,\boldsymbol{c}) \mapsto G_{a,\boldsymbol{c}}(\boldsymbol{x})$ is twice continuously differentiable at any $\boldsymbol{x} \in S_{\boldsymbol{X}}$, with Hessian matrix $\boldsymbol{H}\big(G_{a,\boldsymbol{c}}(\boldsymbol{x})\big)$. Eventually, Assumption (A)(iii) implies that

$$\boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x}) = \frac{1}{|\det(\boldsymbol{M}_{b,\boldsymbol{x}})|} \int_{\mathbb{R}^{m-1}} \begin{pmatrix} 1 & \boldsymbol{t}' \\ \boldsymbol{t} & \boldsymbol{t}\boldsymbol{t}' \end{pmatrix} f^{\varepsilon}\big(\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}\big((a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u} + \boldsymbol{\Gamma_u t} - \boldsymbol{M}_{a,\boldsymbol{x}}\big)\big)\, d\boldsymbol{t}. \quad (21)$$

(ii) Positive definiteness then readily follows from (21) and Assumption (B).

(iii) Every entry of $\boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x})$ is an integral involving an integrand of the form (see (21))

$$g_{i,j}(a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}) = \frac{t_i^{\delta_i} t_j^{\delta_j}}{|\det(\boldsymbol{M}_{b,\boldsymbol{x}})|}\Big(f^{\boldsymbol{\varepsilon}}\big(\boldsymbol{M}_{b,\boldsymbol{x}}^{-1}((a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u}+\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{t}-\boldsymbol{M}_{a,\boldsymbol{x}})\big) - f^{\boldsymbol{\varepsilon}}((1,\boldsymbol{t}')')\Big)$$
$$+ \frac{t_i^{\delta_i} t_j^{\delta_j}}{|\det(\boldsymbol{M}_{b,\boldsymbol{x}})|} f^{\boldsymbol{\varepsilon}}((1,\boldsymbol{t}')') =: g_{i,j}^{I}(a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}) + g_{i,j}^{II}(a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}),$$

where $\delta_i,\delta_j \in \{0,1\}$. Clearly, for any $\boldsymbol{t}$, $(a,\boldsymbol{c},\boldsymbol{x}) \mapsto g_{i,j}^{I}(a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t})$ and $(a,\boldsymbol{c},\boldsymbol{x}) \mapsto g_{i,j}^{II}(a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t})$ are continuous. Therefore, in view of Theorem 8.5 in Briane and Pagès (2012), it is sufficient to prove that there exist integrable functions $h_{i,j}^{I}, h_{i,j}^{II} : \mathbb{R}^{m-1} \to \mathbb{R}^{+}$ such that

$$|g_{i,j}^{I}(a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t})| \leq h_{i,j}^{I}(\boldsymbol{t}) \ \text{ and } \ |g_{i,j}^{II}(a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t})| \leq h_{i,j}^{II}(\boldsymbol{t}) \quad \text{for any } (a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t}).$$

Since Assumptions (A)(ii)-(iii) ensure that $\det(\boldsymbol{M}_{b,\boldsymbol{x}})$ stays away from 0 for any $\boldsymbol{x} \in S_{\boldsymbol{X}}$, we can take $\boldsymbol{t} \mapsto h_{i,j}^{II}(\boldsymbol{t}) := t_i^{\delta_i} t_j^{\delta_j} f^{\boldsymbol{\varepsilon}}((1,\boldsymbol{t}')')/(\inf_{\boldsymbol{x}\in S_{\boldsymbol{X}}} |\det(\boldsymbol{M}_{b,\boldsymbol{x}})|)$, whose integrability follows from the fact that $f^{\boldsymbol{\varepsilon}}(\cdot)$ is bounded and $\boldsymbol{\varepsilon}$ has finite second-order moments. Now, Lemma 2 and Assumptions (A)(ii)-(iii) readily entail that there exist $r > m-1$ and $s > 0$ such that

$$\begin{aligned}
|g_{i,j}^{I}(a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t})| &= |t_i^{\delta_i} t_j^{\delta_j}| \times \big|f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}\big((a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u}+\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{t}\big) - f^{\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}}\big(\boldsymbol{M}_{b,\boldsymbol{x}}(1,\boldsymbol{t}')'+\boldsymbol{M}_{a,\boldsymbol{x}}\big)\big| \\
&\leq |t_i^{\delta_i} t_j^{\delta_j}| \frac{C|(a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u}+\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{t}-\boldsymbol{M}_{b,\boldsymbol{x}}(1,\boldsymbol{t}')'-\boldsymbol{M}_{a,\boldsymbol{x}}|^{s}}{\big(1+\frac{1}{2}\big|(a+\boldsymbol{c}'\boldsymbol{t})\boldsymbol{u}+\boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{t}+\boldsymbol{M}_{b,\boldsymbol{x}}(1,\boldsymbol{t}')'+\boldsymbol{M}_{a,\boldsymbol{x}}\big|^{2}\big)^{(3+r+s)/2}} \\
&\leq C|\boldsymbol{t}|^{\delta_i+\delta_j}(1+|\boldsymbol{t}|^{s})\big(1+\tfrac{1}{2}\big|\boldsymbol{t}+\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{M}_{b,\boldsymbol{x}}(1,\boldsymbol{t}')'+\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{M}_{a,\boldsymbol{x}}\big|^{2}\big)^{-(3+r+s)/2} \\
&\leq C|\boldsymbol{t}|^{\delta_i+\delta_j}(1+|\boldsymbol{t}|^{s})\big(1+\tfrac{1}{2}\big|(\mathbf{I}_{m-1}+\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{A}_{\boldsymbol{x}})\boldsymbol{t}+\boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{B}_{\boldsymbol{x}})\big|^{2}\big)^{-(3+r+s)/2},
\end{aligned} \tag{22}$$

where the matrices $\boldsymbol{A}_{\boldsymbol{x}} := (\boldsymbol{M}_{b,\boldsymbol{x}})_{.2}$ and $\boldsymbol{B}_{\boldsymbol{x}} := (\boldsymbol{M}_{b,\boldsymbol{x}})_{.1} - \boldsymbol{M}_{a,\boldsymbol{x}}$ are based on the partition $\boldsymbol{M}_{b,\boldsymbol{x}} = ((\boldsymbol{M}_{b,\boldsymbol{x}})_{.1}\,(\boldsymbol{M}_{b,\boldsymbol{x}})_{.2})$ into an $m \times 1$ matrix $(\boldsymbol{M}_{b,\boldsymbol{x}})_{.1}$ and an $m \times (m-1)$ matrix $(\boldsymbol{M}_{b,\boldsymbol{x}})_{.2}$. Lemma 3 implies that it is always possible to choose $\boldsymbol{\Gamma}_{\boldsymbol{u}}$ in such a way that $\mathbf{I}_{m-1} + \boldsymbol{\Gamma}_{\boldsymbol{u}}'\boldsymbol{A}_{\boldsymbol{x}}$ is invertible. Consequently, one may proceed as in the proof of

Lemma 2 and write

$$\begin{aligned}
&\frac{1+|\boldsymbol{t}|^2}{1+\frac{1}{2}\big|(\mathbf{I}_{m-1}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{x}})\boldsymbol{t}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{B}_{\boldsymbol{x}}\big|^2} \\
&\qquad = \frac{1+\big|(\mathbf{I}_{m-1}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{x}})^{-1}[(\mathbf{I}_{m-1}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{x}})\boldsymbol{t}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{B}_{\boldsymbol{x}}]-(\mathbf{I}_{m-1}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{x}})^{-1}\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{B}_{\boldsymbol{x}}\big|^2}{1+\frac{1}{2}\big|(\mathbf{I}_{m-1}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{x}})\boldsymbol{t}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{B}_{\boldsymbol{x}})\big|^2} \\
&\qquad \leq \frac{C+C\big|(\mathbf{I}_{m-1}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{x}})\boldsymbol{t}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{B}_{\boldsymbol{x}})\big|^2}{1+\frac{1}{2}\big|(\mathbf{I}_{m-1}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{A}_{\boldsymbol{x}})\boldsymbol{t}+\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{B}_{\boldsymbol{x}})\big|^2} \leq C,
\end{aligned}$$

where we used the fact that $\boldsymbol{x}\mapsto\boldsymbol{A}_{\boldsymbol{x}}$ and $\boldsymbol{x}\mapsto\boldsymbol{B}_{\boldsymbol{x}}$ are continuous functions defined over the compact set $S_{\boldsymbol{X}}$. Therefore, (22) provides $|g^I_{i,j}(a,\boldsymbol{c},\boldsymbol{x},\boldsymbol{t})|\leq C|\boldsymbol{t}|^{\delta_i+\delta_j}(1+|\boldsymbol{t}|^s)\big(1+|\boldsymbol{t}|^2\big)^{-(3+r+s)/2}=: h^I_{i,j}(\boldsymbol{t})$, where $h^I_{i,j}(\cdot)$ is integrable over $\mathbb{R}^{m-1}$ (since $r>m-1$). □

The proof of Theorem 1 still requires the following lemma.

Lemma 5. *Let Assumptions (A) and (B) hold, fix* $\boldsymbol{\alpha}\in\mathcal{B}^m$*, and write* $\tilde{\boldsymbol{x}}=\tilde{\boldsymbol{x}}^N=\mathrm{Proj}_{\boldsymbol{\gamma}^N}(\boldsymbol{x})$ *for any* $\boldsymbol{x}$*. Then,*

(i) for any compact set $\boldsymbol{K}(\subset\mathbb{R}^{m-1})$, $\sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}}\sup_{a\in\mathbb{R},\boldsymbol{c}\in\boldsymbol{K}}|\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})-G_{a,\boldsymbol{c}}(\boldsymbol{x})|\to 0$ *as* $N\to\infty$*;*

(ii) $\sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}}|\min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m}\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})-\min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m}G_{a,\boldsymbol{c}}(\boldsymbol{x})|\to 0$ *as* $N\to\infty$*.*

Proof. (i) Fix $a\in\mathbb{R}$ and $\boldsymbol{c}\in\boldsymbol{K}$. First note that $[\widetilde{\boldsymbol{X}}^N=\tilde{\boldsymbol{x}}]$ is equivalent to $[\boldsymbol{X}\in\boldsymbol{C}_{\boldsymbol{x}}]$, where we let $\boldsymbol{C}_{\boldsymbol{x}}=\boldsymbol{C}^N_{\boldsymbol{x}}=\{\boldsymbol{z}\in S_{\boldsymbol{X}}:\mathrm{Proj}_{\boldsymbol{\gamma}^N}(\boldsymbol{z})=\tilde{\boldsymbol{x}}\}$. Hence, one has

$$\begin{aligned}
&\big|\mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}}-\boldsymbol{c}'\boldsymbol{Y}^\perp_{\boldsymbol{u}}-a)|\widetilde{\boldsymbol{X}}^N=\tilde{\boldsymbol{x}}]-\mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}}-\boldsymbol{c}'\boldsymbol{Y}^\perp_{\boldsymbol{u}}-a)|\boldsymbol{X}=\tilde{\boldsymbol{x}}]\big| \\
&\qquad\leq \sup_{\boldsymbol{z}\in\boldsymbol{C}_{\boldsymbol{x}}}\big|\mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}}-\boldsymbol{c}'\boldsymbol{Y}^\perp_{\boldsymbol{u}}-a)|\boldsymbol{X}=\boldsymbol{z}]-\mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}}-\boldsymbol{c}'\boldsymbol{Y}^\perp_{\boldsymbol{u}}-a)|\boldsymbol{X}=\tilde{\boldsymbol{x}}]\big|,
\end{aligned}$$

which provides

$$
\begin{aligned}
&|\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - G_{a,\boldsymbol{c}}(\boldsymbol{x})| \\
&\quad \le |\mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} - a)|\widetilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}] - \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} - a)|\boldsymbol{X} = \tilde{\boldsymbol{x}}]| \\
&\qquad + |\mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} - a)|\boldsymbol{X} = \tilde{\boldsymbol{x}}] - \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} - a)|\boldsymbol{X} = \boldsymbol{x}]| \\
&\quad \le 2 \sup_{\boldsymbol{z} \in \boldsymbol{C}_{\boldsymbol{x}}} |\mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} - a)|\boldsymbol{X} = \boldsymbol{z}] - \mathrm{E}[\rho_\alpha(Y_{\boldsymbol{u}} - \boldsymbol{c}'\boldsymbol{Y}_{\boldsymbol{u}}^{\perp} - a)|\boldsymbol{X} = \tilde{\boldsymbol{x}}]| \\
&\quad \le 2 \sup_{\boldsymbol{z} \in \boldsymbol{C}_{\boldsymbol{x}}} \big|G_{a,\boldsymbol{c}}(\boldsymbol{z}) - G_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})\big| \\
&\quad \le 2 \max(\alpha, 1-\alpha)\sqrt{1 + |\boldsymbol{c}|^2}\,([\boldsymbol{M}_{a,\cdot}]_{\mathrm{Lip}} + [\boldsymbol{M}_{b,\cdot}]_{\mathrm{Lip}}\|\boldsymbol{\varepsilon}\|_1) \sup_{\boldsymbol{z} \in \boldsymbol{C}_{\boldsymbol{x}}} |\boldsymbol{z} - \tilde{\boldsymbol{x}}|,
\end{aligned}
$$

where we used Lemma 1. It directly follows that, for some C that does not depend on N,

$$
\sup_{\boldsymbol{x} \in S_{\boldsymbol{X}}} \sup_{a \in \mathbb{R}, \boldsymbol{c} \in \boldsymbol{K}} |\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - G_{a,\boldsymbol{c}}(\boldsymbol{x})| \le C \sup_{\boldsymbol{x} \in S_{\boldsymbol{X}}} \sup_{\boldsymbol{z} \in \boldsymbol{C}_{\boldsymbol{x}}} |\boldsymbol{z} - \tilde{\boldsymbol{x}}| =: C \sup_{\boldsymbol{x} \in S_{\boldsymbol{X}}} R(\boldsymbol{C}_{\boldsymbol{x}});
$$

the quantity $R(\boldsymbol{C}_{\boldsymbol{x}})$ is the "radius" of the cell $\boldsymbol{C}_{\boldsymbol{x}}$. The result then follows from the fact that $\sup_{\boldsymbol{x} \in S_{\boldsymbol{X}}} R(\boldsymbol{C}_{\boldsymbol{x}}) \to 0$ as $N \to \infty$; see Lemma A.2(ii) in Charlier et al. (2015a).

(ii) For simplicity of notations, we write $\tilde{a} = \widetilde{a}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$ and $\tilde{\boldsymbol{c}} = \widetilde{\boldsymbol{c}}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$. From Lemma 4(ii),

$$
\boldsymbol{v}'\boldsymbol{H}_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\boldsymbol{x})\boldsymbol{v} > 0
$$

for any $\boldsymbol{x} \in S_{\boldsymbol{X}}$ and any $\boldsymbol{v} \in \mathcal{S}^{m-1} = \{\boldsymbol{x} \in \mathbb{R}^m : |\boldsymbol{x}| = 1\}$. The compactness assumption on $S_{\boldsymbol{X}}$ and the continuity of $\boldsymbol{x} \mapsto \boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x})$ (Lemma 4(iii)) yield that

$$
\inf_{\boldsymbol{x} \in S_{\boldsymbol{X}}} \inf_{\boldsymbol{v} \in \mathcal{S}^{m-1}} \boldsymbol{v}'\boldsymbol{H}_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\boldsymbol{x})\boldsymbol{v} > 0.
$$

This, jointly with Part (i) of the result, implies that there exists a positive integer N_0 and a compact set, $K_{\boldsymbol{\alpha}}(\subset \mathbb{R}^m)$ say, such that, for all $N \ge N_0$ and for all $\boldsymbol{x} \in S_{\boldsymbol{X}}$, $\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$ and $\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}$ belong to $K_{\boldsymbol{\alpha}}$. In particular, for all $N \ge N_0$ and for all $\boldsymbol{x} \in S_{\boldsymbol{X}}$, $\widetilde{\boldsymbol{c}}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$ and $\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}$ belong to a compact set $K^{\boldsymbol{c}}_{\boldsymbol{\alpha}}(\subset \mathbb{R}^{m-1})$). Then, with $\mathbb{I}_+ = \mathbb{I}_{[\min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) \geq \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} G_{a,\boldsymbol{c}}(\boldsymbol{x})]}$, we have

$$\Big| \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} G_{a,\boldsymbol{c}}(\boldsymbol{x}) \Big| \mathbb{I}_+ = (\widetilde{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{x}}) - G_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\boldsymbol{x}))\mathbb{I}_+$$

$$\leq \big(\widetilde{G}_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\tilde{\boldsymbol{x}}) - G_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\boldsymbol{x})\big)\mathbb{I}_+ \leq \sup_{a\in\mathbb{R},\boldsymbol{c}\in K^{\boldsymbol{c}}_{\boldsymbol{\alpha}}} |\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - G_{a,\boldsymbol{c}}(\boldsymbol{x})|\mathbb{I}_+. \tag{23}$$

Similarly, with $\mathbb{I}_- := 1 - \mathbb{I}_+$, we have

$$\Big| \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} G_{a,\boldsymbol{c}}(\boldsymbol{x}) \Big| \mathbb{I}_- = \big(G_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\boldsymbol{x}) - \widetilde{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{x}})\big)\mathbb{I}_-$$

$$\leq \big(G_{\tilde{a},\tilde{\boldsymbol{c}}}(\boldsymbol{x}) - \widetilde{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{x}})\big)\mathbb{I}_- \leq \sup_{a\in\mathbb{R},\boldsymbol{c}\in K^{\boldsymbol{c}}_{\boldsymbol{\alpha}}} |\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - G_{a,\boldsymbol{c}}(\boldsymbol{x})|\mathbb{I}_-. \tag{24}$$

From (23)-(24), we readily obtain

$$\Big| \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} G_{a,\boldsymbol{c}}(\boldsymbol{x}) \Big| \leq \sup_{a\in\mathbb{R},\boldsymbol{c}\in K^{\boldsymbol{c}}_{\boldsymbol{\alpha}}} |\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - G_{a,\boldsymbol{c}}(\boldsymbol{x})|.$$

The result then directly follows from Part (i) of the result. □

We can now prove Theorem 1.

Proof of Theorem 1. Write again $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}^N = \mathrm{Proj}_{\boldsymbol{\gamma}^N}(\boldsymbol{x})$ and fix the same integer N_0 and the same compact sets $K_{\boldsymbol{\alpha}}$ and $K^{\boldsymbol{c}}_{\boldsymbol{\alpha}}$ as in the proof of Lemma 5. Then, for $\boldsymbol{x} \in S_{\boldsymbol{X}}$ and $N \geq N_0$, one has

$$\begin{aligned}
&|G_{\tilde{a},\tilde{\boldsymbol{c}}}(\boldsymbol{x}) - G_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\boldsymbol{x})| \\
&\quad\leq |G_{\tilde{a},\tilde{\boldsymbol{c}}}(\boldsymbol{x}) - \widetilde{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{x}})| + |\widetilde{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{x}}) - G_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\boldsymbol{x})| \\
&\quad\leq \sup_{a\in\mathbb{R},\boldsymbol{c}\in K^{\boldsymbol{c}}_{\boldsymbol{\alpha}}} |G_{a,\boldsymbol{c}}(\boldsymbol{x}) - \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})| + \big| \min_{a,\boldsymbol{c}} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - \min_{a,\boldsymbol{c}} G_{a,\boldsymbol{c}}(\boldsymbol{x}) \big| \\
&\quad\leq \sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}} \sup_{a\in\mathbb{R},\boldsymbol{c}\in K^{\boldsymbol{c}}_{\boldsymbol{\alpha}}} |G_{a,\boldsymbol{c}}(\boldsymbol{x}) - \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})| + \sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}} \big| \min_{a,\boldsymbol{c}} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - \min_{a,\boldsymbol{c}} G_{a,\boldsymbol{c}}(\boldsymbol{x}) \big|.
\end{aligned}$$

Therefore, Lemma 5 implies that, as $N \to \infty$,

$$\sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}} |G_{\tilde{a},\tilde{\boldsymbol{c}}}(\boldsymbol{x}) - G_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\boldsymbol{x})| \to 0. \tag{25}$$

Performing a second-order expansion about $\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}} = (a_{\boldsymbol{\alpha},\boldsymbol{x}}, \boldsymbol{c}'_{\boldsymbol{\alpha},\boldsymbol{x}})'$ provides

$$G_{\tilde{a},\tilde{\boldsymbol{c}}}(\boldsymbol{x}) - G_{a_{\boldsymbol{\alpha},\boldsymbol{x}},\boldsymbol{c}_{\boldsymbol{\alpha},\boldsymbol{x}}} = \frac{1}{2}\big(\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}\big)'\boldsymbol{H}^N_*(\boldsymbol{x})\big(\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}\big),$$

with $\boldsymbol{H}^N_{*\boldsymbol{x}} := \boldsymbol{H}_{a^N_{*\boldsymbol{x}},\boldsymbol{c}^N_{*\boldsymbol{x}}}(\boldsymbol{x})$, where $\boldsymbol{q}^N_{*\boldsymbol{x}} = (a^N_{*\boldsymbol{x}}, (\boldsymbol{c}^N_{*\boldsymbol{x}})')' = \theta\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}} + (1-\theta)\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}}$, for some $\theta \in (0,1)$. Write $\boldsymbol{H}^N_{*\boldsymbol{x}} = \boldsymbol{O}_{\boldsymbol{x}}\boldsymbol{\Lambda}^N_{\boldsymbol{x}}\boldsymbol{O}'_{\boldsymbol{x}}$, where $\boldsymbol{O}_{\boldsymbol{x}}$ is an $m \times m$ orthogonal matrix and where $\boldsymbol{\Lambda}^N_{\boldsymbol{x}} = \mathrm{diag}(\lambda^N_{1,\boldsymbol{x}}, \ldots, \lambda^N_{m,\boldsymbol{x}})$ collects the eigenvalues of $\boldsymbol{H}^N_{*\boldsymbol{x}}$ in decreasing order. We then have

$$\begin{aligned}
G_{\tilde{a},\tilde{\boldsymbol{c}}}(\boldsymbol{x}) - G_{a,\boldsymbol{c}}(\boldsymbol{x}) &= \frac{1}{2}\big(\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}\big)'\boldsymbol{H}^N_{*\boldsymbol{x}}\big(\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}\big)\\
&= \frac{1}{2}\sum_{j=1}^m \lambda^N_{j,\boldsymbol{x}}\Big(\big(\boldsymbol{O}_{\boldsymbol{x}}\big(\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}\big)\big)_j\Big)^2 \geq \frac{\lambda^N_{m,\boldsymbol{x}}}{2}\sum_{i=1}^m \Big(\big(\boldsymbol{O}_{\boldsymbol{x}}\big(\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}\big)\big)_j\Big)^2\\
&= \frac{\lambda^N_{m,\boldsymbol{x}}}{2}\big|\boldsymbol{O}_{\boldsymbol{x}}\big(\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}\big)\big|^2 = \frac{\lambda^N_{m,\boldsymbol{x}}}{2}\big|\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}\big|^2.
\end{aligned}$$

Hence,

$$\sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}}\big|\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} - \boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}\big|^2 \leq 2\Big(\inf_{\bar{N}\geq N_0}\inf_{\boldsymbol{x}\in S_{\boldsymbol{X}}}\lambda^{\bar{N}}_{m,\boldsymbol{x}}\Big)^{-1}\sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}}|G_{\tilde{a},\tilde{\boldsymbol{c}}}(\boldsymbol{x}) - G_{a,\boldsymbol{c}}(\boldsymbol{x})|. \tag{26}$$

The result then follows from (25) and from the fact

$$\begin{aligned}
\inf_{\bar{N}\geq N_0}\inf_{\boldsymbol{x}\in S_{\boldsymbol{X}}}\lambda^{\bar{N}}_{m,\boldsymbol{x}} &= \inf_{\bar{N}\geq N_0}\inf_{\boldsymbol{x}\in S_{\boldsymbol{X}}}\inf_{\boldsymbol{v}\in\mathcal{S}^{m-1}}\boldsymbol{v}'\boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x})\big|_{(a,c)=(a^{\bar{N}}_{*\boldsymbol{x}},\boldsymbol{c}^{\bar{N}}_{*\boldsymbol{x}})}\boldsymbol{v}\\
&\geq \inf_{\boldsymbol{x}\in S_{\boldsymbol{X}}}\inf_{\boldsymbol{v}\in\mathcal{S}^{m-1}}\inf_{(a,\boldsymbol{c}')'\in K_{\boldsymbol{\alpha}}}\boldsymbol{v}'\boldsymbol{H}_{a,\boldsymbol{c}}(\boldsymbol{x})\boldsymbol{v} > 0,
\end{aligned}$$

which results from Lemma 4(ii)-(iii) and the compactness of $S_{\boldsymbol{X}}$, $\mathcal{S}^{m-1}$ and $K_{\boldsymbol{\alpha}}$. □

B Proof of Theorem 2

Let $(\boldsymbol{X}_1, \boldsymbol{Y}_1), \ldots, (\boldsymbol{X}_n, \boldsymbol{Y}_n)$ be independent copies of $(\boldsymbol{X}, \boldsymbol{Y})$. Recall that $\boldsymbol{\gamma}^N$ denotes an optimal quantization grid of size N for the random d-vector $\boldsymbol{X}$ and that $\hat{\boldsymbol{\gamma}}^{N,n}$ stands for the grid provided by the CLVQ algorithm on the basis of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. Below, we will write $(\tilde{\boldsymbol{x}}^N_1, \ldots, \tilde{\boldsymbol{x}}^N_N)$ and $(\hat{\boldsymbol{x}}^{N,n}_1, \ldots, \hat{\boldsymbol{x}}^{N,n}_N)$ for the grid points of $\boldsymbol{\gamma}^N$ and $\hat{\boldsymbol{\gamma}}^{N,n}$, respectively.

Throughout this section, we assume that the empirical quantization of $\boldsymbol{X}$, based on $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ converges almost surely towards the population one, i.e., $\widehat{\boldsymbol{X}}^{N,n} = \mathrm{Proj}_{\hat{\boldsymbol{\gamma}}^{N,n}}(\boldsymbol{X}) \to \widetilde{\boldsymbol{X}}^N = \mathrm{Proj}_{\boldsymbol{\gamma}^N}(\boldsymbol{X})$ almost surely as $n \to \infty$. This is justified by classical results in quantization about the convergence in n of the CLVQ algorithm when N is fixed; see Pagès (1998).

The proof of Theorem 2 requires Lemmas 6-7 below.

Lemma 6. *Let Assumption (A)′ hold. Fix $N \in \mathbb{N}_0$ and $\boldsymbol{x} \in S_{\boldsymbol{X}}$. Write $\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{x}}^N = \mathrm{Proj}_{\boldsymbol{\gamma}^N}(\boldsymbol{x})$ and $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}^{N,n} = \mathrm{Proj}_{\hat{\boldsymbol{\gamma}}^{N,n}}(\boldsymbol{x})$. Then, with $\widehat{\boldsymbol{X}}_i^N = \widehat{\boldsymbol{X}}_i^{N,n} = \mathrm{Proj}_{\hat{\boldsymbol{\gamma}}^{N,n}}(\boldsymbol{X}_i)$, $i = 1, \ldots, n$, we have*

(i) $\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{[\hat{\boldsymbol{X}}_i^N = \hat{\boldsymbol{x}}^N]} \xrightarrow[n\to\infty]{a.s.} P[\widetilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}]$;

(ii) after possibly reordering the $\tilde{\boldsymbol{x}}_i^N$'s, $\hat{\boldsymbol{x}}_i^{N,n} \xrightarrow[n\to\infty]{a.s.} \tilde{\boldsymbol{x}}_i^N$, $i = 1, \ldots, N$ (hence, $\hat{\boldsymbol{\gamma}}^{N,n} \xrightarrow[n\to\infty]{a.s.} \boldsymbol{\gamma}^N$).

A proof is given in Charlier et al. (2015a).

Lemma 7. *Let Assumptions (A) and (B) hold. Fix $\boldsymbol{\alpha} = \alpha\boldsymbol{u} \in \mathcal{B}^m$, $\boldsymbol{x} \in S_{\boldsymbol{X}}$ and $N \in \mathbb{N}_0$. Let K $(\subset \mathbb{R}^m)$ be compact and define*

$$\widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) = \widehat{G}_{a,\boldsymbol{c}}^{N,n}(\hat{\boldsymbol{x}}) := \frac{\frac{1}{n}\sum_{i=1}^n \rho_\alpha(\boldsymbol{u}'\mathbf{Y}_i - \boldsymbol{c}'\boldsymbol{\Gamma}_{\boldsymbol{u}}'\mathbf{Y}_i - a)\,\mathbb{I}_{[\hat{\boldsymbol{X}}_i^N = \hat{\boldsymbol{x}}]}}{\frac{1}{n}\sum_{i=1}^n \mathbb{I}_{[\hat{\boldsymbol{X}}_i^N = \hat{\boldsymbol{x}}]}}.$$

Then

(i) $\sup_{(a,\boldsymbol{c}')'\in K} |\widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})| = o_{\mathrm{P}}(1)$ *as* $n \to \infty$;

(ii) $|\min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})| = o_{\mathrm{P}}(1)$ *as* $n \to \infty$;

(iii) $|\widetilde{G}_{\widehat{a}_{\boldsymbol{\alpha},\boldsymbol{x}}^{N,n}, \widehat{\boldsymbol{c}}_{\boldsymbol{\alpha},\boldsymbol{x}}^{N,n}}(\tilde{\boldsymbol{x}}) - \widetilde{G}_{\tilde{a}_{\boldsymbol{\alpha},\boldsymbol{x}}^N, \tilde{\boldsymbol{c}}_{\boldsymbol{\alpha},\boldsymbol{x}}^N}(\tilde{\boldsymbol{x}})| = o_{\mathrm{P}}(1)$ *as* $n \to \infty$.

Proof. (i) Since

$$\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) = \mathrm{E}[\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a)|\widetilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}] = \frac{\mathrm{E}[\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a)\mathbb{I}_{[\tilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}]}]}{P[\widetilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}]},$$

it is sufficient, in view of Lemma 6(i), to show that

$$\sup_{(a,\boldsymbol{c}')'\in K} \Big|\frac{1}{n}\sum_{i=1}^n \rho_\alpha(\boldsymbol{u}'\boldsymbol{Y}_i - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}_i - a)\,\mathbb{I}_{[\hat{\boldsymbol{X}}_i^N = \hat{\boldsymbol{x}}]} - \mathrm{E}\big[\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a)\mathbb{I}_{[\tilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}]}\big]\Big| = o_{\mathrm{P}}(1),$$

as $n \to \infty$. It is natural to decompose it as

$$\begin{aligned}
&\sup_{(a,\boldsymbol{c}')'\in K} \Big|\frac{1}{n}\sum_{i=1}^n \rho_\alpha(\boldsymbol{u}'\boldsymbol{Y}_i - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}_i - a)\,\mathbb{I}_{[\hat{\boldsymbol{X}}_i^N = \hat{\boldsymbol{x}}]} - \mathrm{E}\big[\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a)\mathbb{I}_{[\tilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}]}\big]\Big| \\
&\qquad \leq \sup_{(a,\boldsymbol{c}')'\in K} |T_{a\boldsymbol{c}1}| + \sup_{(a,\boldsymbol{c}')'\in K} |T_{a\boldsymbol{c}2}|,
\end{aligned}$$

with

$$T_{a\boldsymbol{c}1} := \frac{1}{n}\sum_{i=1}^n \rho_\alpha(\boldsymbol{u}'\boldsymbol{Y}_i - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}_i - a)\big(\mathbb{I}_{[\hat{\boldsymbol{X}}_i^N = \hat{\boldsymbol{x}}]} - \mathbb{I}_{[\tilde{\boldsymbol{X}}_i^N = \tilde{\boldsymbol{x}}]}\big),$$

and

$$T_{a\boldsymbol{c}2} := \frac{1}{n}\sum_{i=1}^n \rho_\alpha(\boldsymbol{u}'\boldsymbol{Y}_i - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}_i - a)\mathbb{I}_{[\tilde{\boldsymbol{X}}_i^N = \tilde{\boldsymbol{x}}]} - \mathrm{E}\big[\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a)\mathbb{I}_{[\tilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}]}\big],$$

with $\widetilde{\boldsymbol{X}}_i^N = \mathrm{Proj}_{\boldsymbol{\gamma}^N}(\boldsymbol{X}_i)$, $i = 1, \ldots, n$.

We start by considering $T_{a\boldsymbol{c}2}$. Since $\boldsymbol{x} \mapsto \boldsymbol{M}_{a,\boldsymbol{x}}$ and $\boldsymbol{x} \mapsto \boldsymbol{M}_{b,\boldsymbol{x}}$ are continuous functions defined over the compact set $S_{\boldsymbol{X}}$, one has that, for all $(a, \boldsymbol{c}')' \in K$,

$$\begin{aligned}
&\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a)\mathbb{I}_{[\tilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}]} \\
&\qquad \leq \max(\alpha, 1-\alpha)|\boldsymbol{u}'\boldsymbol{Y} - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y} - a| \leq \max(\alpha, 1-\alpha)|(\boldsymbol{u} - \boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{c})'\boldsymbol{Y} - a| \\
&\qquad \leq \max(\alpha, 1-\alpha)\left[|\boldsymbol{u} - \boldsymbol{\Gamma}_{\boldsymbol{u}}\boldsymbol{c}|\Big(\sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}} |\boldsymbol{M}_{a,\boldsymbol{x}}| + |\boldsymbol{\varepsilon}| \sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}} \|\boldsymbol{M}_{b,\boldsymbol{x}}\|\Big) + |a|\right] \\
&\qquad \leq C_1|\boldsymbol{\varepsilon}| + C_2,
\end{aligned} \tag{27}$$

for some constants C_1, C_2 that do not depend on $(a, \boldsymbol{c}')'$. Since Assumption (A)(v) ensures that $\mathrm{E}[|\boldsymbol{\varepsilon}|] < +\infty$ (recall that $p = 2$ here), the uniform law of large numbers (see, e.g., Theorem 16(a) in Ferguson, 1996) then implies that

$$\sup_{(a,\boldsymbol{c}')'\in K} |T_{a\boldsymbol{c}2}| = o_{\mathrm{P}}(1), \qquad \text{as } n \to \infty. \tag{28}$$

It remains to treat $T_{a\boldsymbol{c}1}$. Let $\ell_n := \{i = 1, \ldots, n : \mathbb{I}_{[\hat{\boldsymbol{X}}_i^N = \hat{\boldsymbol{x}}]} \neq \mathbb{I}_{[\tilde{\boldsymbol{X}}_i^N = \tilde{\boldsymbol{x}}]}\}$ be the set collecting the indices of observations that are projected on the same point as $\boldsymbol{x}$ for $\boldsymbol{\gamma}^N$ but not for $\hat{\boldsymbol{\gamma}}^{N,n}$, or on the same point as $\boldsymbol{x}$ for $\hat{\boldsymbol{\gamma}}^{N,n}$ but not for $\boldsymbol{\gamma}^N$. Proceeding as in (27) then shows that, for any $(a, \boldsymbol{c}')' \in K$,

$$|T_{a\boldsymbol{c}1}| \leq \frac{1}{n} \sum_{i\in\ell_n} |\rho_\alpha(\boldsymbol{u}'\boldsymbol{Y}_i - \boldsymbol{c}'\boldsymbol{\Gamma}'_{\boldsymbol{u}}\boldsymbol{Y}_i - a)| \leq \frac{\#\ell_n}{n} \times \frac{1}{\#\ell_n} \sum_{i\in\ell_n} (C_1 + C_2|\boldsymbol{\varepsilon}_i|) =: S_1 \times S_2,$$

say. Lemma 6(ii) implies that $S_1 = o_{\mathrm{P}}(1)$ as $n \to \infty$. Regarding S_2, the independence between $\#\ell_n$ and the $\boldsymbol{\varepsilon}_i$'s (which follows from the fact that $\#\ell_n$ is measurable with respect to the $\boldsymbol{X}_i$'s) entails that $\mathrm{E}[S_2] = O(1)$ as $n \to \infty$, hence that $S_2 = O_{\mathrm{P}}(1)$ as $n \to \infty$. Therefore,

$$\sup_{(a,\boldsymbol{c}')'\in K} |T_{a\boldsymbol{c}1}| \leq S_1 S_2 = o_{\mathrm{P}}(1) \qquad \text{as } n \to \infty,$$

which, jointly with (28), establishes the result.

(ii) For simplicity, we write $\tilde{\boldsymbol{q}} = (\tilde{a}, \tilde{\boldsymbol{c}}')'$ and $\hat{\boldsymbol{q}} = (\hat{a}, \hat{\boldsymbol{c}}')'$ instead of $\widetilde{\boldsymbol{q}}^N_{\boldsymbol{\alpha},\boldsymbol{x}} = (\widetilde{a}^N_{\boldsymbol{\alpha},\boldsymbol{x}}, (\widetilde{\boldsymbol{c}}^N_{\boldsymbol{\alpha},\boldsymbol{x}})')'$ and $\widehat{\boldsymbol{q}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}} = (\widehat{a}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}, (\widehat{\boldsymbol{c}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}})')'$, respectively. First fix $\delta > 0$ and $\eta > 0$, and choose n_1 and R large enough to have $|\tilde{\boldsymbol{q}}| \leq R$ and $P[|\widehat{\boldsymbol{q}}| > R] < \eta/2$ for any $n \geq n_1$. This is possible since $\widehat{\boldsymbol{q}}$ is nothing but the sample Hallin et al. (2010) quantile of a number of $\boldsymbol{Y}_i$'s that increases to infinity (so that, with arbitrary large probability for n large, $|\widehat{\boldsymbol{q}}|$ cannot exceed $2\sup_{\boldsymbol{x}\in S_{\boldsymbol{X}}} |\boldsymbol{q}_{\boldsymbol{\alpha},\boldsymbol{x}}|$). Define $K_R := \{\boldsymbol{y} \in \mathbb{R}^m : |\boldsymbol{y}| \leq R\}$. Then, with $\mathbb{I}_+ := \mathbb{I}_{[\min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) \geq \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})]}$, we have

$$\begin{aligned}\Big| \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \min_{(a,\boldsymbol{c}')'\in\mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) \Big| \mathbb{I}_+ &= (\widehat{G}_{\hat{a},\hat{\boldsymbol{c}}}(\hat{\boldsymbol{x}}) - \widetilde{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{x}}))\mathbb{I}_+ \\ &\leq \big(\widehat{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\hat{\boldsymbol{x}}) - \widetilde{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{x}})\big)\mathbb{I}_+ \leq \sup_{(a,\boldsymbol{c}')'\in K_R} |\widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})|\mathbb{I}_+, \end{aligned} \tag{29}$$

for $n \geq n_1$. Similarly, with $\mathbb{I}_- := 1 - \mathbb{I}_+$, we have that, under $|\widehat{\boldsymbol{q}}| \leq R$,

$$\Big| \min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) \Big| \mathbb{I}_- = \big(\widetilde{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{x}}) - \widehat{G}_{\hat{a},\hat{\boldsymbol{c}}}(\hat{\boldsymbol{x}}) \big) \mathbb{I}_-$$

$$\leq \big(\widetilde{G}_{\hat{a},\hat{\boldsymbol{c}}}(\tilde{\boldsymbol{x}}) - \widehat{G}_{\hat{a},\hat{\boldsymbol{c}}}(\hat{\boldsymbol{x}}) \big) \mathbb{I}_- \leq \sup_{(a,\boldsymbol{c}')' \in K_R} |\widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})| \mathbb{I}_-, \tag{30}$$

still for $n \geq n_1$. By combining (29) and (30), we obtain that, under $|\widehat{\boldsymbol{q}}| \leq R$,

$$\Big| \min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) \Big| \leq \sup_{(a,\boldsymbol{c}')' \in K_R} |\widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})|,$$

for $n \geq n_1$. Therefore, for any such n, we get

$$P\Big[\Big| \min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) \Big| > \delta \Big]$$

$$\leq P\Big[\Big| \min_{a,\boldsymbol{c}} \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \min_{a,\boldsymbol{c}} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) \Big| > \delta, |\widehat{\boldsymbol{q}}| \leq R \Big] + P\big[|\widehat{\boldsymbol{q}}| > R \big]$$

$$\leq P\Big[\sup_{(a,\boldsymbol{c}')' \in K_R} |\widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})| > \delta \Big] + \frac{\eta}{2}.$$

From Part (i) of the lemma, we conclude that, for n large enough,

$$P\Big[\Big| \min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \min_{(a,\boldsymbol{c}')' \in \mathbb{R}^m} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) \Big| > \delta \Big] < \eta,$$

as was to be shown.

(iii) This proof proceeds in the same way as for (ii). We start with picking N_1 and R large enough so that $P[|\widehat{\boldsymbol{q}}| > R] < \eta/2$ for any $N \geq N_1$, with η fixed. This yields

$$P\Big[|\widetilde{G}_{\hat{a},\hat{\boldsymbol{c}}}(\tilde{\boldsymbol{x}}) - \widetilde{G}_{\tilde{a},\tilde{c}}(\tilde{\boldsymbol{x}})| > \delta \Big] \leq P\Big[|\widetilde{G}_{\hat{a},\hat{\boldsymbol{c}}}(\tilde{\boldsymbol{x}}) - \widetilde{G}_{\tilde{a},\tilde{c}}(\tilde{\boldsymbol{x}})| > \delta, |\widehat{\boldsymbol{q}}| \leq M \Big] + \frac{\eta}{2}. \tag{31}$$

Note then that

$$P\Big[|\widetilde{G}_{\hat{a},\hat{\boldsymbol{c}}}(\tilde{\boldsymbol{x}}) - \widetilde{G}_{\tilde{a},\tilde{\boldsymbol{c}}}(\tilde{\boldsymbol{x}})| > \delta, |\widehat{\boldsymbol{q}}| \leq M \Big]$$

$$\leq P\Big[|\widetilde{G}_{\hat{a},\hat{\boldsymbol{c}}}(\tilde{\boldsymbol{x}}) - \widehat{G}_{\hat{a},\hat{\boldsymbol{c}}}(\hat{\boldsymbol{x}})| > \delta/2, |\widehat{\boldsymbol{q}}| \leq M \Big] + P\Big[|\widehat{G}_{\hat{a},\hat{\boldsymbol{c}}}(\hat{\boldsymbol{x}}) - \widetilde{G}_{\tilde{a},\tilde{c}}(\tilde{\boldsymbol{x}})| > \delta/2, |\widehat{\boldsymbol{q}}| \leq M \Big]$$

$$\leq P\Big[\sup_{(a,\boldsymbol{c}')' \in K_R} |\widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}}) - \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}})| > \delta/2 \Big] + P\Big[|\min_{a,\boldsymbol{c}} \widehat{G}_{a,\boldsymbol{c}}(\hat{\boldsymbol{x}}) - \min_{a,\boldsymbol{c}} \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})| > \delta/2 \Big]$$

$$=: p_1^{(n)} + p_2^{(n)},$$

say. Parts (i) and (ii) of the lemma imply that $p_1^{(n)}$ and $p_2^{(n)}$ can be made arbitrarily small for n large enough. Combining this with (31) yields the result. $\square$

We can now prove Theorem 2.

Proof of Theorem 2. Under the assumptions considered, the function $(a, \boldsymbol{c}')' \mapsto \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})$ has a unique minimizer (that is the Hallin et al. (2010) $\boldsymbol{\alpha}$-quantile of the distribution of $\boldsymbol{Y}$ conditional on $\widetilde{\boldsymbol{X}}^N = \tilde{\boldsymbol{x}}$). Therefore, the convergence in probability of $\widetilde{G}_{\widehat{a}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}},\widehat{\boldsymbol{c}}^{N,n}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\tilde{\boldsymbol{x}})$ towards $\widetilde{G}_{\tilde{a}^{N}_{\boldsymbol{\alpha},\boldsymbol{x}},\tilde{\boldsymbol{c}}^{N}_{\boldsymbol{\alpha},\boldsymbol{x}}}(\tilde{\boldsymbol{x}})$ (Lemma 7(iii)) implies the convergence in probability of the corresponding arguments (note indeed that the function $(a, \boldsymbol{c}')' \mapsto \widetilde{G}_{a,\boldsymbol{c}}(\tilde{\boldsymbol{x}})$ does not depend on n). $\square$

References

Briane, M. and G. Pagès (2012). *Analyse - Théorie de l'intégration: Convolution et transformée de Fourier*. Vuibert.

Carlier, G., V. Chernozhukov, and A. Galichon (2016). Vector quantile regression: an optimal transport approach. *Ann. Statist. 44*, 1165–1192.

Carlier, G., V. Chernozhukov, and A. Galichon (2017). Vector quantile regression beyond the specified case. *J. Multivariate Anal. 161*, 96–102.

Chakraborty, B. (2001). On affine equivariant multivariate quantiles. *Ann. Inst. Statist. Math. 53*(2), 380–403.

Chakraborty, B. (2003). On multivariate quantile regression. *J. Statist. Plann. Inference 110*(1-2), 109–132.

Chaouch, M. and N. Laïb (2013). Nonparametric multivariate l_1-median regression estimation with functional covariates. *Electron. J. Stat. 7*, 1553–1586.

Chaouch, M. and N. Laïb (2015). Vector-on-function quantile regression for stationary ergodic processes. *J. Korean Statist. Soc. 44*, 161–178.

Charlier, I., D. Paindaveine, and J. Saracco (2015a). Conditional quantile estimators through optimal quantization. *J. Statist. Plann. Inference 156*, 14–30.

Charlier, I., D. Paindaveine, and J. Saracco (2015b). Conditional quantile estimation based on optimal quantization: from theory to practice. *Comput. Statist. Data Anal. 91*, 20–39.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *J. Amer. Statist. Assoc. 91*(434), 862–872.

Cheng, Y. and J. De Gooijer (2007). On the uth geometric conditional quantile. *J. Statist. Plann. Inference 137*, 1914–1930.

Chernozhukov, V., A. Galichon, M. Hallin, and M. Henry (2017). Monge–kantorovich depth, quantiles, ranks and signs. *Ann. Statist. 45*, 223–256.

Chowdhury, J. and P. Chaudhuri (2016). Nonparametric depth and quantile regression for functional data. *arXiv preprint, arXiv:1607.03752*.

Ferguson, T. S. (1996). *A Course in Large Sample Theory.* Chapman & Hall/CRC.

Graf, S. and H. Luschgy (2000). *Foundations of quantization for probability distributions*, Volume 1730 of *Lecture Notes in Mathematics*. Berlin: Springer-Verlag.

Hallin, M., Z. Lu, D. Paindaveine, and M. Šiman (2015). Local constant and local bilinear multiple-output quantile regression. *Bernoulli 3*(21), 1435–1466.

Hallin, M., D. Paindaveine, and M. Šiman (2010). Multivariate quantiles and multiple-output regression quantiles: from L_1 optimization to halfspace depth. *Ann. Statist. 38*(2), 635–669.

Koenker, R. (2005). *Quantile Regression.* Number 38 in Econometric Society Monographs. Cambridge: Cambridge Univ. Press.

Koenker, R. and G. Bassett, Jr. (1978). Regression quantiles. *Econometrica 46*(1), 33–50.

Liu, R. Y., J. M. Parelius, and K. Singh (1999). Multivariate analysis by data depth:

descriptive statistics, graphics and inference. *Ann. Statist. 27*(3), 783–858. With discussion and a rejoinder by Liu and Singh.

Pagès, G. (1998). A space quantization method for numerical integration. *J. Comput. Appl. Math. 89*(1), 1–38.

Pagès, G. and J. Printems (2003). Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods Appl. 9*(2), 135–165.

Serfling, R. (2010). Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation. *J. Nonparametr. Stat. 22*(7), 915–936.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*, pp. 523–531. Canad. Math. Congress, Montreal, Que.

Yu, K. and M. C. Jones (1998). Local linear quantile regression. *J. Amer. Statist. Assoc. 93*(441), 228–237.