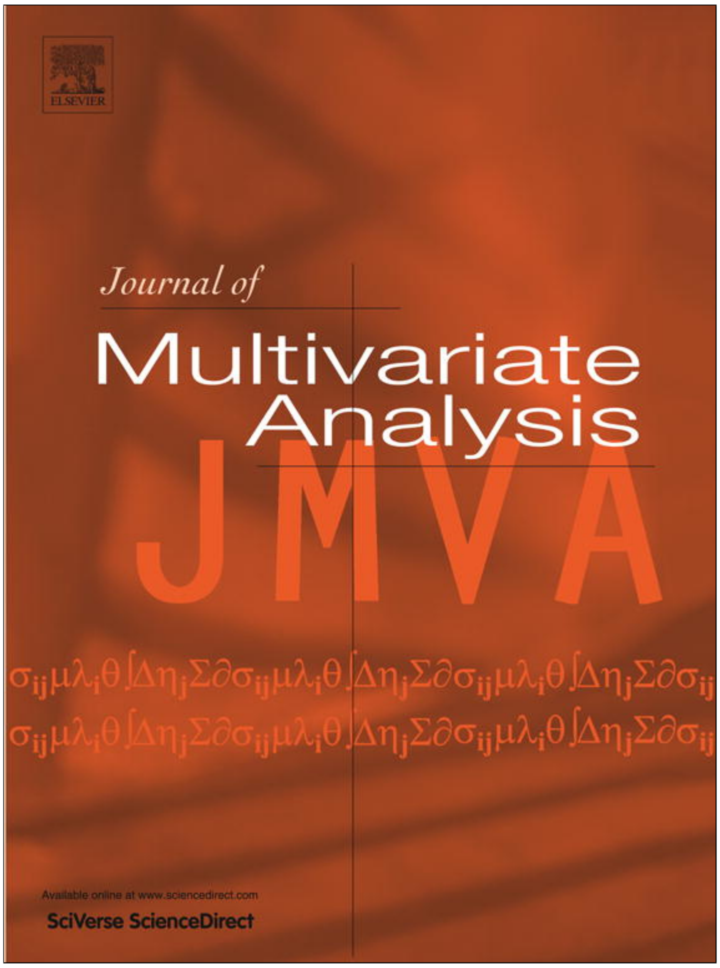


(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Journal of Multivariate Analysis 116 (2013) 163–171

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

# On elliptical quantiles in the quantile regression setup

Daniel Hlubinka [a,*], Miroslav Šiman [b]

[a] Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University in Prague, Sokolovská 83, CZ-186 75 Prague 8, Czech Republic

[b] Institute of Information Theory and Automation of the ASCR, Pod Vodárenskou věží 4, CZ-182 08 Prague 8, Czech Republic

ARTICLE INFO

*Article history:*
Received 3 February 2012
Available online 23 December 2012

*AMS 2000 subject classifications:*
62H05
62J99
62G08

*Keywords:*
Regression quantile
Elliptical quantile
Multivariate quantile
Quantile regression
Nonparametric regression

ABSTRACT

This article defines a meaningful concept of elliptical location quantile with the aid of quantile regression, discusses its basic properties, and suggests its extension to a general regression framework through a locally constant nonparametric approach.



## 1. Introduction

Recently, a directional approach to multivariate quantiles has led to two new and promising concepts that first define multivariate quantiles for each direction in the response space as hyperplanes and then create the quantile regions by intersecting all the induced upper quantile halfspaces at the same quantile level; see [6,11,9] for the theory and [13,12] for the implementation. Not only these approaches extend quantiles to the multivariate setup, but they are also computationally viable in low-dimensional spaces by means of parametric linear programming and provide plenty of statistics for further inference as well as some links to halfspace depth and portfolio optimization. Although these two approaches are so different at first sight, they appear intimately connected to each other and lead to the same multivariate quantile regions.

Nevertheless, these concepts in their present form still suffer from several problems. For example, (a) they do not directly provide any vector/matrix estimator of non-directional location/scatter, (b) they lead to the affine equivariant sample quantile regions only in a computationally very demanding way, (c) they do not have the probability content of the quantile regions under control, (d) they only give rise to polyhedral sample quantile regions, and (e) they cannot employ any apriori information about the underlying probability distribution.

This partly justifies the elliptical quantiles presented here because they need not suffer from the first three drawbacks and partly solve the other two. Although they are tailored to elliptical distributions when the elliptical shape of quantile regions is highly desirable, they can be useful for other symmetric distributions as well and they may be turned into some alternatively shaped quantile regions quite analogously. On the one hand, the semiparametric quantile concept presented here is based on quantile regression and thus computationally feasible, and it directly leads to elliptical quantile regions, vector/matrix estimators of location/scatter, and also to some Lagrange multipliers with high potential for statistical

* Corresponding author.
*E-mail address:* hlubinka@karlin.mff.cuni.cz (D. Hlubinka).

0047-259X/$ – see front matter 
doi:10.1016/j.jmva.2012.11.016

inference. Furthermore, it reflects some symmetry of the underlying distribution, makes it easy to include some a priori information by means of additional convex constraints, and allows for elliptical quantile regions changing shapes with the quantile level. Last but not least, its weighted modification can handle multiple identical observations, and its locally constant regression variant seems to asymptotically provide elliptical quantiles of the conditional distribution with a direct probabilistic interpretation, contrary to the regression extensions of [6,11,9,5]. On the other hand, this concept of location elliptical quantiles lacks robustness in its present form, and always results in sample elliptical contours containing a certain number of observations, which makes it somewhat impractical for very small data sets. The size of its parametric space also grows rather quickly with the dimension of observations. Were it not for the last three drawbacks, the concept would appear an ideal proposal for elliptical quantiles; see Figs. 1–5. Nevertheless, it still seems best suited to medium or large low-dimensional data sets generated without outliers from an axially or spherically symmetric population distribution.

The idea of using ellipsoids for statistical inference is not new at all. Not only are probability distributions with elliptical density contours intensively studied in theory and used in practice, but ellipsoids have also been used for defining data depth, handling measurement errors, finding outliers, specifying copulas, analyzing or discriminating data clusters, and for estimating location or scatter (not to mention other applications). Of course, they have been also considered as quantile or confidence regions, but probably not yet in the quantile regression framework. Nevertheless, our approach is somewhat close to those based on minimum volume sets (see [14]), $L_1$-estimation (see [15]), multivariate S-estimation (see [18]) and its weighted variant (see [16]), and it may lead to the minimum volume covering ellipsoid in a special sample case; see e.g. [17] and references therein.

This article proceeds as follows. Section 2 introduces necessary notation, explains our definition of elliptical quantiles, and discusses its possible modifications. Section 3 studies the elliptical quantiles in the population case, Section 4 comments on the sample elliptical quantiles, Section 5 proposes their weighted and nonparametric regression extensions, and the last Section 6 illustrates all the concepts and ideas with a few examples.

## 2. Definitions and notation

As a rule, we will use boldface letters for vectors and blackboard bold letters for matrices or number sets. Unfortunately, the notation used throughout this article remains considerable despite our effort to simplify it as much as possible.

Let $\boldsymbol{Y} = (Y^{(1)}, \ldots, Y^{(m)})' \in \mathbb{R}^m$ be a random vector with an absolutely continuous distribution having a connected support, finite second-order moments, and a density differentiable almost everywhere, and let $\rho_\tau(x) = x(\tau - \mathrm{I}(x < 0)) = \max\{(\tau-1)x, \tau x\}$, $\tau \in (0, 1)$, stand for the non-negative convex check function with a unique minimum that plays a prominent role in the definition of standard location and regression quantiles; see [8]. With the gradient conditions of Section 3 in mind, we suggest to use this function also for defining (the whole processes of) elliptical quantiles and related parameters of location and scatter, which might open the door to sophisticated inference procedures and extend the applicability of [10].

**Definition 1.** For any $\tau \in (0, 1)$, we define the elliptical $\tau$-quantile $\varepsilon_\tau(\boldsymbol{Y})$ of $\boldsymbol{Y}$ and corresponding lower and upper $\tau$-quantile regions $\mathcal{E}_\tau^-(\boldsymbol{Y})$ and $\mathcal{E}_\tau^+(\boldsymbol{Y})$ as follows:

$$\begin{aligned}
\varepsilon_\tau(\boldsymbol{Y}) &= \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{y}'\mathbb{A}_\tau\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{b}_\tau - c_\tau = 0\},\\
\mathcal{E}_\tau^-(\boldsymbol{Y}) &= \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{y}'\mathbb{A}_\tau\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{b}_\tau - c_\tau < 0\},\\
\mathcal{E}_\tau^+(\boldsymbol{Y}) &= \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{y}'\mathbb{A}_\tau\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{b}_\tau - c_\tau \geq 0\},
\end{aligned}$$

where $\mathbb{A}_\tau$, $\boldsymbol{b}_\tau$ and $c_\tau$ minimize the objective function

$$\Psi_\tau(\mathbb{A}, \boldsymbol{b}, c) := \mathrm{E}\rho_\tau(\boldsymbol{Y}'\mathbb{A}\boldsymbol{Y} + \boldsymbol{Y}'\boldsymbol{b} - c), \tag{$P_1$}$$

subject to some convex constraints on $\mathbb{A}$ and $\boldsymbol{b}$. (We recommend to require only $\mathbb{A} \in \mathrm{PSD}(m)$, i.e. $\mathbb{A} = \mathbb{A}_{m\times m}$ symmetric positive semidefinite, and meeting the constraint ($C_2$) that is stated below along with an explanation and justification of our proposal.)

This definition already leads to a convex optimization problem with a unique minimum. The scalar parameter $c$ ought to remain unrestricted to guarantee a reasonable probability interpretation of $\mathcal{E}_\tau^-(\boldsymbol{Y})$, see Section 3.

The constraints on $\mathbb{A}$ and $\boldsymbol{b}$ should eliminate the trivial zero solution and make the problem clearly identified, solvable, naturally equivariant, and as simple as possible. They should also result in $\mathbb{A}$ symmetric and positive definite if we want the quantiles really elliptical. While positive semidefiniteness can always be imposed on $\mathbb{A}$ by means of semidefinite programming, another restriction on $\mathbb{A}$ seems necessary to keep the matrix regular. It should also preserve good equivariance properties of the elliptical quantiles, which indicates that it should be based on the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_m > 0$ of $\mathbb{A}$. For example, the following constraints ($C_1$) to ($C_4$) keep the problem convex, $\mathbb{A}$ regular, and the computation in the sample case tractable:

$$\det(\mathbb{A}^{-1}) = 1/(\lambda_1 \ldots \lambda_m) \leq 1, \tag{$C_1$}$$

$$(\det(\mathbb{A}))^{1/m} = (\lambda_1 \ldots \lambda_m)^{1/m} \geq 1, \tag{$C_2$}$$

$$\log\det(\mathbb{A}) = \log\lambda_1 + \cdots + \log\lambda_m \geq 0, \tag{$C_3$}$$

$$tr(\mathbb{A}^{-1}) = 1/\lambda_1 + \cdots + 1/\lambda_m \leq m. \tag{$C_4$}$$

All of these rotationally invariant constraints lead to fully affine-equivariant elliptical quantiles. If we wanted to obtain spherical quantiles (corresponding to $\lambda_1 = \cdots = \lambda_m = 1$), we might add yet another convex constraint such as $\lambda_1 - \lambda_m \leq 0$ (see Fig. 2 for an illustration), but we will not discuss this possibility below. See also [1, p. 118] for the convexity or concavity of all the eigenvalue-based functions considered in this paragraph.

As we want our elliptical quantiles to match the multivariate quantiles usually considered for elliptical distributions, we illustrate an application of $(C_4)$ only in Fig. 1 and focus on the determinant-based constraints $(C_1)$–$(C_3)$ hereinafter. These constraints define the same elliptical quantiles. We decided to employ $(C_2)$ that seems best supported by the symmetric primal/dual solvers in the optimization software used.

The objective function $\Psi_\tau(\mathbb{A}, \boldsymbol{b}, c)$ is then positive and scale equivariant. Therefore, the optimal solution will satisfy the inequality $(C_2)$ only weakly, and the corresponding Lagrange multiplier (say $L_\tau$) will be non-zero.

The elliptical $\tau$-quantile $\mathcal{E}_\tau$ can be described more naturally by the equation

$$(\boldsymbol{y} - \boldsymbol{s}_\tau)'\mathbb{A}_\tau(\boldsymbol{y} - \boldsymbol{s}_\tau) - \kappa_\tau = 0$$

where $\kappa_\tau = (1/4)\boldsymbol{b}_\tau{}'\mathbb{A}_\tau{}^{-1}\boldsymbol{b}_\tau + c_\tau$ and $\boldsymbol{s}_\tau = -\mathbb{A}_\tau{}^{-1}\boldsymbol{b}_\tau/2$. In other words, the elliptical $\tau$-quantiles can be uniquely described either by $\mathbb{A}_\tau$, $\boldsymbol{b}_\tau$, and $c_\tau$, or by $\mathbb{A}_\tau$, $\boldsymbol{s}_\tau$, and $\kappa_\tau$. The latter parametrization is employed in the proof of Theorem 2, it seems very intuitive, and it could be used to modify the elliptical concept in several ways, e.g. to extend it to both parametric and local polynomial regression setup. Unfortunately, such generalizations usually lead to non-convex optimization problems requiring different analysis and software implementation, which is why they will be discussed elsewhere. In fact, even the function

$$\tilde{\Psi}_\tau(\mathbb{A}, \boldsymbol{s}, \kappa) := \mathrm{E}\rho_\tau((\boldsymbol{Y} - \boldsymbol{s})'\mathbb{A}(\boldsymbol{Y} - \boldsymbol{s}) - \kappa)$$

is already non-convex in $\boldsymbol{s}$. Consequently, we prefer the original convex formulation of the problem that moreover shares some properties with those investigated in [6] or [11]. Therefore, we usually work with the original definition and possibly use $\mathbb{A}_\tau$, $\boldsymbol{b}_\tau$, and $c_\tau$ to compute $\boldsymbol{s}_\tau$ and $\kappa_\tau$ from the expressions given above.

To sum up the choice of constraints, we require $\mathbb{A}$ symmetric positive semidefinite and meeting $(C_2)$ from now on, which seem to be the only natural constraints without any other apriori knowledge. In principle, one might further add some equality constraints on the parameters to simplify the model and then use the resulting multipliers to judge their relevance. For example, $\boldsymbol{b}_\tau = \boldsymbol{0}$ would correspond to zero location center $\boldsymbol{s}_\tau \in \mathbb{R}^m$. Similarly, $\boldsymbol{u}'\boldsymbol{b}_\tau = \boldsymbol{0}$ would make the line segment linking the probability mass centers of $\mathcal{E}_\tau^-(\boldsymbol{Y})$ and $\mathcal{E}_\tau^+(\boldsymbol{Y})$ (and passing through the overall center of probability mass) parallel to $\boldsymbol{u} \in \mathbb{R}^m$ (if we interpret probability density like the distribution of mass as usual), which would follow from the corresponding gradient conditions similar to those in Section 3. Needless to say that such additional constraints with corresponding Lagrange multipliers might be very useful for statistical inference, but we will not use them hereinafter.

In the following two sections, we will derive further properties of elliptical quantiles with the aid of the Lagrange duality theory and terminology for convex optimization problems as described in [1, Chapter 5]. As the theory is crucial for understanding what follows, every reader should become familiar with it before reading further.

## 3. Population case

The constraint optimization problem defining elliptical quantiles is convex and it trivially satisfies Slater's condition implying strong duality. Furthermore, both the constraint function and the objective function in $(P_1)$ are differentiable under our assumptions. The Karush–Kuhn–Tucker conditions then lead to the following set of necessary and sufficient conditions for the optimal solution $\mathbb{A}_\tau$, $\boldsymbol{b}_\tau$, and $c_\tau$ of $(P_1)$ accompanied with the Lagrange multiplier (or, dual variable) $L_\tau$ corresponding to $(C_2)$:

$$1 = \det(\mathbb{A}_\tau), \tag{1}$$

$$0 = P(\boldsymbol{Y} \in \mathcal{E}_\tau^-) - \tau, \tag{2}$$

$$0 = \frac{1}{1-\tau}\mathrm{E}[\boldsymbol{Y}\mathrm{I}_{[\boldsymbol{Y}\in\mathcal{E}_\tau^+]}] - \frac{1}{\tau}\mathrm{E}\big[\boldsymbol{Y}\mathrm{I}_{[\boldsymbol{Y}\in\mathcal{E}_\tau^-]}\big], \tag{3}$$

$$L_\tau\frac{1}{m\tau(1-\tau)}\det(\mathbb{A}_\tau)^{1/m}\mathbb{A}_\tau^{-1} = \frac{1}{1-\tau}\mathrm{E}[\boldsymbol{Y}\boldsymbol{Y}'\mathrm{I}_{[\boldsymbol{Y}\in\mathcal{E}_\tau^+]}] - \frac{1}{\tau}\mathrm{E}\big[\boldsymbol{Y}\boldsymbol{Y}'\mathrm{I}_{[\boldsymbol{Y}\in\mathcal{E}_\tau^-]}\big], \tag{4}$$

where $\mathbb{A}_\tau$ is assumed symmetric positive semidefinite.

The conditions (2)–(4) already imply $L_\tau > 0$, which is why (1) is stated with the equality sign. That is to say that if we write $\bullet$ for the Hadamard (elementwise) product and realize that $\boldsymbol{1}'(\mathbb{A}_\tau \bullet (\mathbb{A}_\tau^{-1}))\boldsymbol{1} = m\det(\mathbb{A}_\tau)$, then we can multiply (2)–(4) elementwise respectively by $c_\tau$, $\tau(1-\tau)\boldsymbol{b}_\tau$, and $\tau(1-\tau)\mathbb{A}_\tau$, sum up the results, and thus obtain a link between the multiplier $L_\tau$ and the optimum value of the objective function:

$$L_\tau = \Psi_\tau(\mathbb{A}_\tau, \boldsymbol{b}_\tau, c_\tau)/\det(\mathbb{A}_\tau)^{(m+1)/m} \quad (> 0). \tag{5}$$

Consequently, the equality in (1) is justified and

$$L_\tau = \Psi_\tau(\mathbb{A}_\tau, \boldsymbol{b}_\tau, c_\tau). \tag{6}$$

The Lagrange multiplier $L_\tau$ measures the impact of the boundary constraint but here it also equals the optimum value of the Lagrange problem dual to $(P_1)$, interprets $\Psi_\tau(\mathbb{A}_\tau, \boldsymbol{b}_\tau, c_\tau)$ and appears as a useful functional for statistical inference, together with $\mathbb{A}_\tau$, $\boldsymbol{s}_\tau$, and $\kappa_\tau$.

What do the simplified gradient conditions (1)–(4) of the Lagrangian say? The first Eq. (1) scales the problem and is easy to interpret. The second Eq. (2) guarantees that the resulting lower (elliptical) $\tau$-quantile region $\mathcal{E}_\tau^-(\boldsymbol{Y})$ has its coverage probability equal to $\tau$ and is therefore non-empty. The third Eq. (3) says that the probability mass centers of $\mathcal{E}_\tau^-(\boldsymbol{Y})$ and $\mathcal{E}_\tau^+(\boldsymbol{Y})$ must coincide with each other and thus also with the overall center of probability mass, equivalently $E(\boldsymbol{Y}|\boldsymbol{Y} \in \mathcal{E}_\tau^+) = \mathrm{E}(\boldsymbol{Y}|\boldsymbol{Y} \in \mathcal{E}_\tau^-)$. On the one hand, this also indicates nonrobustness of the elliptical quantiles, but on the other hand, it ensures that the elliptical quantiles of elliptical distributions will be naturally centered and nested as expected. The last condition (4) seems rather complicated, but it can be simplified into

$$\frac{L_\tau}{m}\mathbb{A}_\tau^{-1} = \mathrm{var}(\boldsymbol{Y}|\boldsymbol{Y} \in \mathcal{E}_\tau^+) - \mathrm{var}(\boldsymbol{Y}|\boldsymbol{Y} \in \mathcal{E}_\tau^-),$$

which is also satisfied by the standard elliptical quantiles of elliptical distributions and leads to the link between $\mathbb{A}_\tau^{-1}$ and $\mathrm{var}(\boldsymbol{Y})$ for $\tau \to 0$.

Furthermore, it turns out that the lower elliptical quantile region would always reduce to the natural confidence interval for the median of any symmetric distribution in the univariate location case.

For the sake of clarity, we summarize some results of our brief investigation in the following theorem:

**Theorem 2.** *If $\boldsymbol{Y} \in \mathbb{R}^m$ is a random vector with an absolutely continuous distribution having a connected support, finite second-order moments, and a density $f(\boldsymbol{y})$ differentiable almost everywhere, then its elliptical $\tau$-quantiles with centers $\boldsymbol{s}_\tau$ have the following properties for any $\tau \in (0, 1)$:*

[1] *they behave naturally under affine transformations of $\boldsymbol{Y}$*
[2] $P(\boldsymbol{Y} \in \mathcal{E}_\tau^-) = \tau$
[3] $E(\boldsymbol{Y}|\boldsymbol{Y} \in \mathcal{E}_\tau^+) = \mathrm{E}(\boldsymbol{Y}|\boldsymbol{Y} \in \mathcal{E}_\tau^-)$
[4] $\mathbb{A}_\tau^{-1} = \frac{m}{\Psi_\tau(\mathbb{A}_\tau, \boldsymbol{b}_\tau, c_\tau)}\left(\mathrm{var}(\boldsymbol{Y}|\boldsymbol{Y} \in \mathcal{E}_\tau^+) - \mathrm{var}(\boldsymbol{Y}|\boldsymbol{Y} \in \mathcal{E}_\tau^-)\right)$
[5] *if $\boldsymbol{s}_\tau = (s_1, \ldots, s_m)'$, $\mathbb{A}_\tau = (a_{ij})_{i,j=1}^m$, and $f(\boldsymbol{y}) = f(\mathbb{J}\boldsymbol{y})$ for a sign-change matrix $\mathbb{J} = \mathbb{J}' = \mathbb{J}^{-1} = \mathrm{diag}(j_1, \ldots, j_m)$ with diagonal elements $\pm 1$, then $s_i = 0$ whenever $j_i = -1$, $i \in \{1, \ldots, m\}$, and $a_{ij} = 0$ whenever $j_i j_j = -1$, $i, j \in \{1, \ldots, m\}$.*
[6] *if the distribution of $\boldsymbol{Y}$ is symmetric around a hyperplane, then $\boldsymbol{s}_\tau$ lies on the hyperplane*
[7] *if the distribution of $\boldsymbol{Y}$ is symmetric along an axis, then $\boldsymbol{s}_\tau$ lies on the axis*
[8] *if the distribution of $\boldsymbol{Y}$ is centrally symmetric, then $\boldsymbol{s}_\tau$ coincides with the center of symmetry*
[9] *if the distribution of $\boldsymbol{Y}$ is elliptically symmetric, then $\mathbb{A}_\tau^{-1}$ is proportionate to the scatter matrix.*

**Proof.** Clearly, [1] follows directly from the definition, and [2], [3], and [4] are easy consequences of the gradient conditions (1)–(4). Furthermore, [6], [7], and [8] follow directly from [1] and [5]. Therefore, the only points worthy of serious consideration are [5] and [9].

The assumption of the former also implies

$$\tilde{\Psi}_\tau(\mathbb{A}, \boldsymbol{s}, \kappa) = \tilde{\Psi}_\tau(\mathbb{J}\mathbb{A}\mathbb{J}, \mathbb{J}\boldsymbol{s}, \kappa)$$

for any $\mathbb{A}$, $\boldsymbol{s}$, and $\kappa$. Therefore, if $\mathbb{A}_\tau$, $\boldsymbol{s}_\tau$, and $\kappa_\tau$ minimize $\tilde{\Psi}_\tau$, then $\mathbb{J}\mathbb{A}_\tau\mathbb{J}$, $\mathbb{J}\boldsymbol{s}_\tau$, and $\kappa_\tau$ do it as well. This is possible only if both $\mathbb{A}_\tau = \mathbb{J}\mathbb{A}_\tau\mathbb{J}$ and $\boldsymbol{s}_\tau = \mathbb{J}\boldsymbol{s}_\tau$, because the minimizer of $\tilde{\Psi}_\tau$ must be unique. Consequently, $s_i = 0$ whenever $j_i = -1$, and $a_{ij} = 0$ whenever $j_i j_j = -1$, $i, j \in \{1, \ldots, m\}$.

As for [9], we can assume (thanks to [1] and [8]) that $\boldsymbol{Y}$ is elliptically symmetric around $\boldsymbol{0}$ and $\boldsymbol{s}_\tau = \boldsymbol{0}$. Its density is thus of the form $f(\boldsymbol{y}) = kp(\boldsymbol{y}'\mathbb{S}\boldsymbol{y})$ for some non-negative function $p$, constant $k > 0$, and a positive definite matrix $\mathbb{S}$ with $\det(\mathbb{S}) = 1$. If $c$ satisfies $P(\boldsymbol{Y}'\mathbb{S}\boldsymbol{Y} < c) = \tau$, then

$$\begin{aligned}
\mathrm{var}(\boldsymbol{Y}|\boldsymbol{Y}'\mathbb{S}\boldsymbol{Y} \geq c) &= \frac{k}{1-\tau}\int_{\{\boldsymbol{y}\in\mathbb{R}^m:\boldsymbol{y}'\mathbb{S}\boldsymbol{y}\geq c\}} \boldsymbol{y}\boldsymbol{y}'p(\boldsymbol{y}'\mathbb{S}\boldsymbol{y})d\boldsymbol{y} \\
&= \frac{k}{1-\tau}\int_{\{\boldsymbol{x}\in\mathbb{R}^m:\boldsymbol{x}'\boldsymbol{x}\geq c\}} \mathbb{S}^{-1/2}\boldsymbol{x}\boldsymbol{x}'\mathbb{S}^{-1/2}p(\boldsymbol{x}'\boldsymbol{x})\det(\mathbb{S}^{1/2})d\boldsymbol{x} \\
&= \frac{k}{1-\tau}\det(\mathbb{S}^{1/2})\mathbb{S}^{-1/2}\underbrace{\iint_{\{\boldsymbol{x}\in\mathbb{R}^m:\boldsymbol{x}'\boldsymbol{x}\geq c\}} \boldsymbol{x}\boldsymbol{x}'p(\boldsymbol{x}'\boldsymbol{x})d\boldsymbol{x}}_{\propto\mathbb{I}_m}\mathbb{S}^{-1/2} \propto \mathbb{S}^{-1}
\end{aligned}$$

and, quite analogously,

$$\mathrm{var}(\boldsymbol{Y}|\boldsymbol{Y}'\mathbb{S}\boldsymbol{Y} < c) = \frac{k}{\tau}\int_{\{\boldsymbol{y}\in\mathbb{R}^m:\boldsymbol{y}'\mathbb{S}\boldsymbol{y}<c\}} \boldsymbol{y}\boldsymbol{y}'p(\boldsymbol{y}'\mathbb{S}\boldsymbol{y})d\boldsymbol{y} \propto \mathbb{S}^{-1},$$

where we write $\mathbb{C} \propto \mathbb{D}$ for two matrices $\mathbb{C}$ and $\mathbb{D}$ of the same dimension if $\mathbb{D} = \gamma\mathbb{C}$ for a scalar $\gamma > 0$. Consequently, the gradient conditions (2)–(4) are met by $\mathcal{E}_\tau^- = \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{y}'\mathbb{S}\boldsymbol{y} < c\}$ and $\mathcal{E}_\tau^+ = \{\boldsymbol{y} \in \mathbb{R}^m : \boldsymbol{y}'\mathbb{S}\boldsymbol{y} \geq c\}$, and the rest follows from the uniqueness of such a solution. □

In particular, our elliptical quantiles coincide with the standard ones for all elliptical distributions satisfying the assumptions of Theorem 2.

We conclude this section by pointing out that the problem of finding our elliptical quantiles is equivalent to the auxiliary (convex) optimization problem

$$\min_{c,\boldsymbol{b},\mathbb{A}\in \mathrm{PSD}(m)} -(\det(\mathbb{A}))^{1/m}$$

subject to

$$\Psi_\tau(\mathbb{A}, \boldsymbol{b}, c) \leq \Psi_\tau(\mathbb{A}_\tau, \boldsymbol{b}_\tau, c_\tau).$$

It follows from the fact that the Lagrangians associated with the original and auxiliary problems have smooth derivatives with respect to primal variables equal to zero at the same unique feasible point (but corresponding to different positive Lagrange multipliers). Of course, $\min -(\det(\mathbb{A}))^{1/m}$ might be replaced with $\max \det(\mathbb{A})$ without changing the optimal solution. This reveals some remote similarity between our methodology and S-estimation; see [18,16]. Consequently, some ideas already employed in one concept might be found useful in the other, which may be explored carefully in further research.

## 4. Sample case

In the sample case with $n$ observations $\boldsymbol{Y}_i$'s, $i = 1, \ldots, n$, from the population distribution assumed above, we can simply consider the empirical probability distribution and take the expectations with respect to it. After multiplying the objective function by $n$, we may thus rewrite the original sample optimization problem as

$$\min_{\mathbb{A},\boldsymbol{b},c,\boldsymbol{r}^+,\boldsymbol{r}^-} \sum_{i=1}^n \tau r_i^+ + \sum_{i=1}^n (1-\tau) r_i^-$$

subject to the feasibility constraints (FC):

$$\begin{aligned}
&-\det(\mathbb{A})^{1/m} + 1 \leq 0,\\
&-r_i^+ \leq 0, \quad i = 1, \ldots, n,\\
&-r_i^- \leq 0, \quad i = 1, \ldots, n,\\
&\boldsymbol{Y}_i'\mathbb{A}\boldsymbol{Y}_i + \boldsymbol{Y}_i'\boldsymbol{b} - c - r_i^+ + r_i^- = 0, \quad i = 1, \ldots, n, \quad \text{and}\\
&\mathbb{A} \in \mathrm{PSD}(m).
\end{aligned}$$

Here $\boldsymbol{r} = (r_1, \ldots, r_n)' = \boldsymbol{r}^+ - \boldsymbol{r}^- = (r_1^+, \ldots, r_n^+)' - (r_1^-, \ldots, r_n^-)'$ and $r_i = r_i^+ - r_i^-$ stands for the residual corresponding to $\boldsymbol{Y}_i$, $i = 1, \ldots, n$.

This is still a convex optimization problem meeting the refined Slater's constraint qualification. Therefore the strong duality is guaranteed. The Karush–Kuhn–Tucker conditions can be applied as well because all the functions in this representation are differentiable. Therefore $\mathbb{A}_\tau$, $\boldsymbol{b}_\tau$, and $c_\tau$ solve the minimization problem in the sample case if and only if there exist $\boldsymbol{r}^+$ and $\boldsymbol{r}^-$ as defined above and dual variables $L \geq 0$, $\lambda_i^+ \geq 0$, $i = 1, \ldots, n$, $\lambda_i^- \geq 0$, $i = 1, \ldots, n$, and $\nu_i$, $i = 1, \ldots, n$, such that

$$\text{all the conditions (FC) are satisfied,} \tag{7}$$

$$L(-\det(\mathbb{A})^{1/m} + 1) = 0, \tag{8}$$

$$\lambda_i^+ r_i^+ = 0, \quad i = 1, \ldots, n, \tag{9}$$

$$\lambda_i^- r_i^- = 0, \quad i = 1, \ldots, n, \tag{10}$$

$$\tau - \lambda_i^+ - \nu_i = 0, \quad i = 1, \ldots, n, \tag{11}$$

$$1 - \tau - \lambda_i^- + \nu_i = 0, \quad i = 1, \ldots, n, \tag{12}$$

$$\sum_{i=1}^n \nu_i = 0, \tag{13}$$

$$\sum_{i=1}^n \nu_i \boldsymbol{Y}_i = 0, \quad \text{and} \tag{14}$$

$$\sum_{i=1}^n \nu_i \boldsymbol{Y}_i\boldsymbol{Y}_i' = \frac{L}{m} \det(\mathbb{A}_\tau)^{1/m} \mathbb{A}_\tau^{-1}. \tag{15}$$

We tacitly used the fact that the cone of symmetric positive semidefinite matrices is self-dual; see [1, p. 52].

How can these necessary and sufficient conditions be interpreted? If $r_i^+ > 0$, then (9) implies $\lambda_i^+ = 0$ and $\nu_i = \tau$ thanks to (11). Analogously, $r_i^- > 0$ implies $\lambda_i^- = 0$ and $\nu_i = \tau - 1$ thanks to (10) and (12). If $r_i = 0$, we obtain $(\tau - 1) \leq \nu_i \leq \tau$ from (9) to (12). Consequently, (13)–(15) can be roughly interpreted as corresponding population conditions (2)–(4), respectively, where the small deviations are caused only by the data points with zero residuals. In particular, we could derive $N \leq n\tau \leq Z + N$ from (13) where $Z$ and $N$ are the numbers of points in $\mathcal{E}_\tau$ and $\mathcal{E}_\tau^-$, respectively. Under our distributional assumptions, $Z \leq m + 1$ almost surely. On the other hand, (13) and (14) imply $Z \geq m + 1$ with probability one. Therefore, there will be $m + 1$ zero residuals almost surely. Recall that the conditions like (9)–(14) appear in the theory of standard quantile regression as well, which explains why the results obtained here and there are so similar.

We should also remark that the maximum achieved in the dual problem equals $L$, which is why the link (6) between the multiplier and the optimum value of the objective function holds exactly even in the sample case, due to the strong duality theorem for convex optimization problems. Therefore $L > 0$ almost surely, and then (8) implies $\det(\mathbb{A}) = 1$.

Note that all the conclusions obtained in the sample case are valid without any assumption on the underlying continuous distribution of the observations at all, and even the continuity is invoked only sporadically.

Finally, we provide a basic statement on the consistency of the sample elliptical $\tau$-quantiles.

**Theorem 3.** *Assume that the sample elliptical $\tau$-quantiles, $\tau \in (0, 1)$, associated with a random sample of size n from a given distribution with finite second order moments, are parameterized by $\mathbb{A}_{n,\tau}$, $\boldsymbol{b}_{n,\tau}$, and $c_{n,\tau}$. Then it holds for any compact set K and any $\varepsilon > 0$ that*

$$P\Big(\|\mathrm{vec}(\mathbb{A}_{n,\tau} - \mathbb{A}_\tau)\| + \|\boldsymbol{b}_{n,\tau} - \boldsymbol{b}_\tau\| + |c_{n,\tau} - c_\tau| > \varepsilon \;\&\; (\mathbb{A}_{n,\tau}, \boldsymbol{b}_{n,\tau}, c_{n,\tau}) \notin K\Big) \to 0$$

*as $n \to \infty$, i.e. the sample elliptical $\tau$-quantiles then somehow converge to their population counterpart parameterized by $\mathbb{A}_\tau$, $\boldsymbol{b}_\tau$, and $c_\tau$.*

**Proof.** It follows directly from Theorem 5.14 of [19]. □

## 5. Weighted and nonparametric regression modifications

In principle, we could also formulate the problem by considering weights $w_i > 0, i = 1, \ldots, n$, before the terms $\rho_\tau(\boldsymbol{Y}_i'\mathbb{A}\boldsymbol{Y}_i + \boldsymbol{Y}_i'\boldsymbol{b} - c)$ in the sample version of the objective function. In other words, we could look for elliptical quantiles defined by minimizers of

$$\Psi_\tau^w(\mathbb{A}, \boldsymbol{b}, c) := \sum_{i=1}^{n} w_i \rho_\tau(\boldsymbol{Y}_i'\mathbb{A}\boldsymbol{Y}_i + \boldsymbol{Y}_i'\boldsymbol{b} - c) \tag{16}$$

subject to $\mathbb{A} \in \mathrm{PSD}(m)$ and $(\det(\mathbb{A}))^{1/m} \geq 1$. For example, we could employ integer weights to handle multiple identical observations naturally occurring when we use various bootstrap methods or round the data severely; see Fig. 3.

Finally, we provide an intuitive generalization of the sample elliptical quantiles to the nonparametric regression framework when $n$ responses $\boldsymbol{Y}_i$'s, $i = 1, \ldots, n$, are accompanied with corresponding $q$-dimensional covariate vectors $\boldsymbol{Z}_i$'s. Inspired by the local constant version of local polynomial quantile regression, reviewed in [7] and successfully applied in the multivariate setup in [4,5], we suggest, for any $\boldsymbol{z}_0 \in \mathbb{R}^q$, to use the same weights as in [5] even in this elliptical context to obtain local constant modifications of the elliptical quantiles that could hopefully also converge to the population (elliptical) quantiles of the conditional distribution $\mathcal{L}(\boldsymbol{Y}|\boldsymbol{Z} = \boldsymbol{z}_0)$ under a suitable set of additional assumptions. In other words, we suggest to minimize (16) with special weights $w_i$'s of the type

$$w_i(\boldsymbol{z}_0) = \det(\mathbb{H}_0)^{-1} K\big(\mathbb{H}_0^{-1}(\boldsymbol{z}_0 - \boldsymbol{Z}_i)\big)$$

where $\mathbb{H}_0$ is a symmetric positive definite $q \times q$ bandwidth matrix that shrinks towards the zero matrix with $n \to \infty$ at a convenient rate, and $K$ stands for a $q$-variate non-negative kernel (density) function, $\int_{\mathbb{R}^q} K(\boldsymbol{z})d\boldsymbol{z} = 1$, such as the Epanechnikov kernel

$$K(\boldsymbol{z}) = \frac{(q-1)(q+1)\Gamma\left(\frac{q-1}{2}\right)}{4\pi^{(q-1)/2}}(1 - \boldsymbol{z}'\boldsymbol{z})\mathrm{I}(\boldsymbol{z}'\boldsymbol{z} \leq 1)$$

that we also employ in our demo example illustrating this regression generalization (see Fig. 5). Such weights are frequently used in the theory of density estimation and nonparametric regression. Although we use this sort of weights, it does not necessarily mean that we find this kernel or this kernel approach superior to the other possibilities at hand (such as the nearest-neighbor method). In fact, asymmetric kernels or more sophisticated weighting methods might well prove useful for reducing the boundary effect at extreme $\boldsymbol{z}_0$'s.

In this way, we can obtain elliptical regression quantiles for any $\boldsymbol{z}_0 \in \mathbb{R}^q$ that really seem to converge to those of the conditional distribution of the responses under a suitable set of additional assumptions, see Fig. 5. Nevertheless, their asymptotic properties still remain to be explored and proved rigorously.

The optimization involved in the concept of sample elliptical quantiles in the ordinary, weighted, or local constant case can be performed quite easily by using CVX, see [2,3]. The computation is relatively fast and well-suited even for large sets of observations with dimension $m$ greater than two or three, if there is enough RAM memory available.

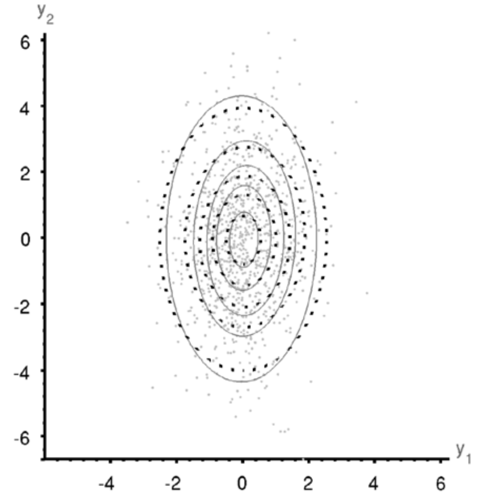


**Fig. 1.** **Elliptical quantiles in** $\mathbb{R}^2$. The plot shows elliptical $\tau$-quantiles based on $(C_2)$ (solid gray curves) and their corresponding $(C_4)$ modifications (dotted black curves) obtained for $\tau = 0.1, 0.3, 0.5, 0.7$, and 0.9 from a random sample of $n = 1000$ (light gray) points from centered bivariate normal distribution $N(0, 1) \times N(0, 4)$.

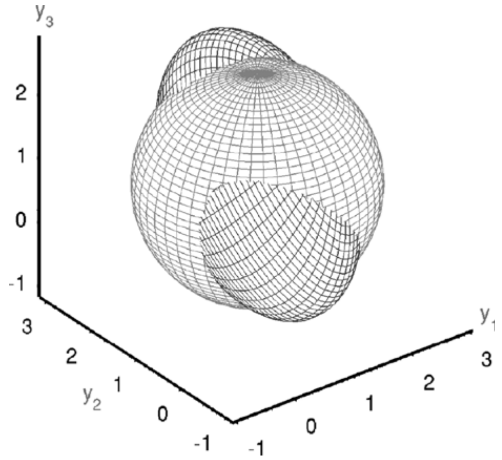


**Fig. 2.** **Elliptical quantiles in** $\mathbb{R}^3$. The plot displays both the elliptical (dark gray) 0.5-quantile and its spherical (light gray) modification computed from $n = 1000$ points drawn from the three-dimensional normal distribution $N(\boldsymbol{\mu}, \mathbb{V})$ with $\boldsymbol{\mu} = (1, 1, 1)'$ and $\text{vech}(\mathbb{V}) = (1, 1/2, 0, 2, 1/2, 1)'$.

## 6. Illustrations

At the very end, we include some pictures to illustrate the elliptical quantiles and their properties.

Fig. 1 presents elliptical $\tau$-quantiles and their modifications using the constraint $(C_4)$ instead of $(C_2)$. They were obtained for $\tau = 0.1, 0.3, 0.5, 0.7$, and 0.9 from a random sample of $n = 1000$ observations generated from the centered bivariate normal distribution $N(0, 1) \times N(0, 4)$. When applied to points coming from elliptical distributions, both types of elliptical quantiles seem to fit the data quite well and to be naturally nested with decreasing $\tau$. Moreover, the default $(C_2)$-based elliptical quantiles closely match the population quantiles commonly used for such distributions.

Fig. 2 shows both the elliptical 0.5-quantile and its spherical modification (enforced by the additional eigenvalue-based constraint $\lambda_1 - \lambda_m \leq 0$) computed from $n = 1000$ points drawn from the three-dimensional normal distribution $N(\boldsymbol{\mu}, \mathbb{V})$ with $\boldsymbol{\mu} = (1, 1, 1)'$ and $\text{vech}(\mathbb{V}) = (1, 1/2, 0, 2, 1/2, 1)'$. This figure is included to illustrate the fact that the elliptical quantiles are naturally affine equivariant and possible to compute even in spaces of dimension higher than two. Furthermore, it reveals how the spherical modification looks like. Last but not least, it demonstrates that the elliptical quantiles are generally compatible with convex constraints that can be used for implementing some a priori information.

The left plot of Fig. 3 displays elliptical $\tau$-quantiles with their center points, obtained for $\tau = 0.1, 0.3, 0.5, 0.7$, and 0.9 from observations $\boldsymbol{Y}_i, i = 1, 2, \ldots, 166$, where $\boldsymbol{Y}_i \sim U(-1, 1) \times U(-2, 2)$ are mutually independent, $i = 1, \ldots, 138$, and $\boldsymbol{Y}_{139} = \boldsymbol{Y}_{140} = \cdots = \boldsymbol{Y}_{166} = (0, 2)'$. This picture shows how the weights can be used to handle multiple identical observations, as the elliptical quantiles were generated from 139 points $\boldsymbol{Y}_i, i = 1, \ldots, 139$, where the last observation was assigned weight $w_{139} = 28$ and the others were left with their default unit weights. This example also demonstrates that the elliptical quantiles might prove useful even for more general distributions with at least one axis of symmetry, and that the process $\{\boldsymbol{s}_\tau\}_\tau$ of their centers might be found useful for testing various symmetry assumptions. After careful analysis, this picture also illustrates the real possibility of quantile crossings (at the very top of the plot) and the fact that each

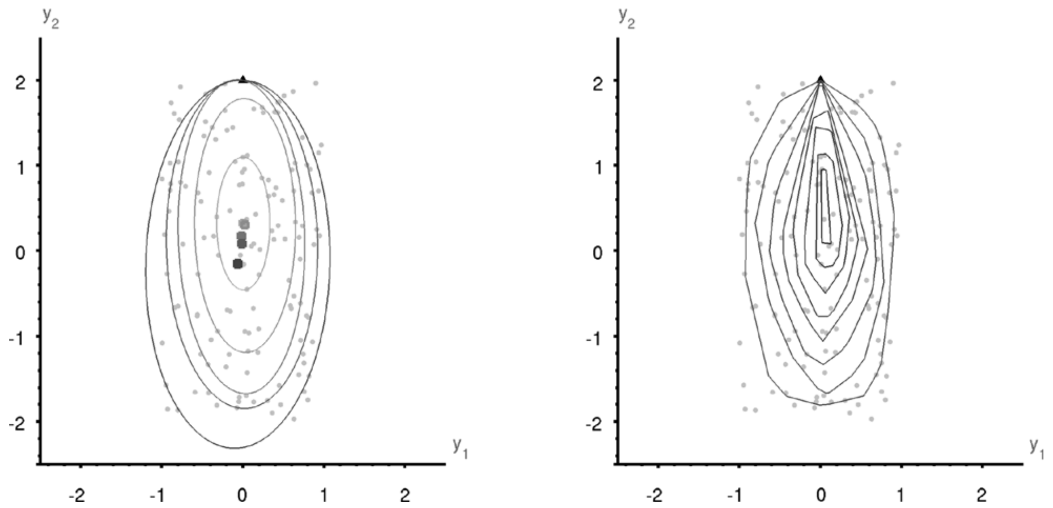


**Fig. 3. Elliptical quantiles with weights.** The left plot shows elliptical $\tau$-quantiles (gray curves) with their center points (small gray squares) where the gray color gets darker with increasing $\tau$, $\tau = 0.1, 0.3, 0.5, 0.7$, and 0.9. The quantiles were computed from a sample of $n = 166$ observations consisting of 138 (light gray) points drawn from the bivariate uniform distribution on $(-1, 1) \times (-2, 2)$ and of the point $(0, 2)$ (marked with a small black triangle) occurring 28 times in the sample. For the sake of comparison, the right plot presents the Tukey depth $\tau$-contours, $\tau = 0.025, 0.075, 0.125, 0.175, 0.225, 0.275, 0.325$, and 0.375, computed from the same data.

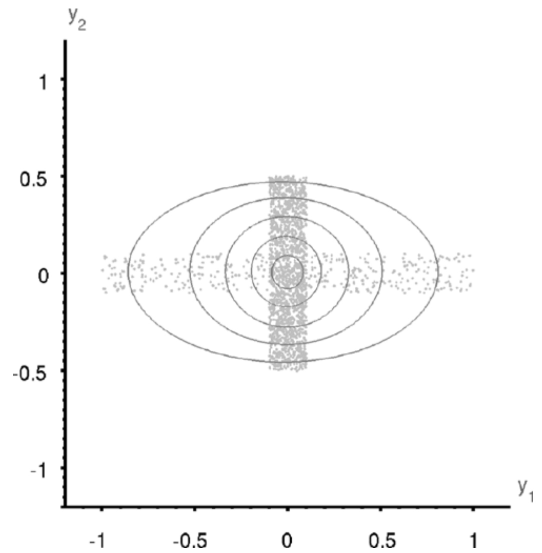


**Fig. 4. Elliptical quantiles and flexibility.** The plot depicts (dark gray) elliptical $\tau$-quantiles, $\tau = 0.1, 0.3, 0.5, 0.7$, and 0.9, computed from $n = 1999$ (light gray) independent observations drawn from the bivariate uniform distribution on either $(-1, 1) \times (-0.1, 0.1)$ (400 data points) or $(-0.1, 0.1) \times (-0.5, 0.5)$ (1599 data points).

elliptical quantile contour passes through a few observations. The Tukey (halfspace) depth $\tau$-contours computed from the same data for $\tau = 0.025, 0.075, 0.125, 0.175, 0.225, 0.275, 0.325$, and 0.375 are included for comparison in the right plot. Nevertheless, the comparison is not straightforward as the probability content of such a $\tau$-contour is not directly related to $\tau$, unfortunately.

Fig. 4 depicts elliptical $\tau$-quantiles, $\tau = 0.1, 0.3, 0.5, 0.7$, and 0.9, computed from $n = 1999$ independent observations drawn from the bivariate uniform distribution on either $(-1, 1) \times (-0.1, 0.1)$ (400 data points) or $(-0.1, 0.1) \times (-0.5, 0.5)$ (1599 data points). The picture makes it evident that elliptical $\tau$-quantiles are very flexible as they need not be only inflated or deflated copies of one another. Furthermore, they could easily incorporate the information about their zero centers and zero non-diagonal elements of $\mathbb{A}_\tau$, $\tau \in (0, 1)$, if it were known in advance.

Finally, Fig. 5 shows elliptical regression 0.5-quantiles and their center points, corresponding to the weights based on the Epanechnikov kernel with bandwidth $\mathbb{H}_0 = 0.15$, that were computed for reference points $z_0 = -1.9, -1.8, \ldots, 1.9$ from $n = 100\,000$ observations following the model $(Y_1, Y_2) = (Z, Z^2) + (1 + 3|\sin(\pi Z/2)|)\boldsymbol{\eta}$ with uniformly distributed $Z \sim U(-2, 2)$ and normally distributed $\boldsymbol{\eta} \sim N(0, 1/4) \times N(0, 1/4)$. While the center points rightly indicate the trend, the volumes and principal semi-axes of the elliptical quantiles may correctly reveal even a complicated form of heteroscedasticity, not detectable by common multivariate regression procedures. The demonstration also shows that the elliptical (regression) quantiles can be computed even for hundred(s) of thousands of observations where the

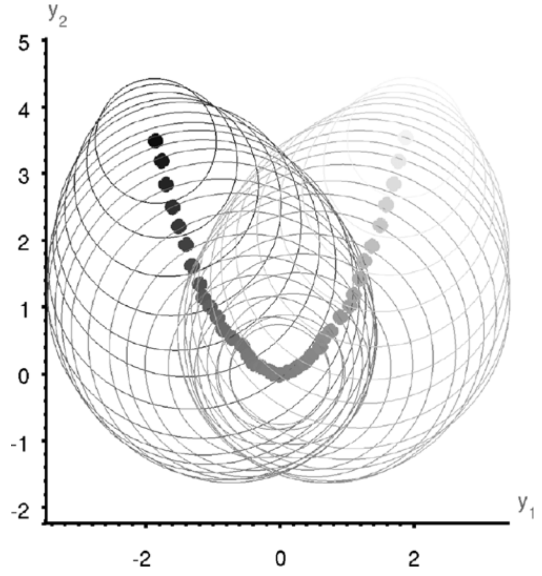


**Fig. 5. Local constant elliptical quantile regression.** The plot displays elliptical regression 0.5-quantiles (gray curves) and their center points (small gray circles), obtained for reference points $\boldsymbol{z}_0 = -1.9, -1.8, \ldots, 1.9$ and lightening with increasing $\boldsymbol{z}_0$. They were computed from $n = 100\,000$ observations following the model $(Y_1, Y_2) = (Z, Z^2) + (1 + 3 \mid \sin(\pi Z/2) \mid)\boldsymbol{\eta}$ where $Z \sim U(-2, 2)$ and $\boldsymbol{\eta} \sim N(0, 1/4) \times N(0, 1/4)$, and the weights were calculated from the Epanechnikov kernel with bandwidth $\mathbb{H}_0 = 0.15$.

implementation of the methods of [6,11] or [5] would be probably not only too time consuming, but also very often failing or numerically unstable at best.

All these examples confirm our belief that the proposed elliptical quantile concepts are indeed worthy of wide attention and further investigation.

## Acknowledgments

The research work of Miroslav Šiman was partly supported by Project 1M06047 of the Ministry of Education, Youth and Sports of the Czech Republic. The author would also like to thank Davy Paindaveine, Marc Hallin, Claude Adan, Nancy de Munck, and Romy Genin for all the good they did for him (and for all the good he could learn from them) during his stay at Université Libre de Bruxelles. The first two mentioned professors also improved our work by many helpful comments, and the latter of them was even playing with the idea of using his expertise in quantile regression for defining elliptical quantiles himself, well before we started to think about that.

## References

[1] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, 2004, http://www.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf.
[2] M. Grant, S. Boyd, Graph implementations for nonsmooth convex programs, in: V. Blondel, S. Boyd, H. Kimura (Eds.), Recent Advances in Learning and Control, in: Lecture Notes in Control and Information Sciences, Springer, 2008, pp. 95–110. http://stanford.edu/~boyd/graph_dcp.html.
[3] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming (web page and software). http://stanford.edu/~boyd/cvx, 2009.
[4] I.W. McKeague, S. López-Pintado, M. Hallin, M. Siman, Analyzing growth trajectories, J. Dev. Orig. Health Dis. 2 (2011) 322–329.
[5] M. Hallin, Z. Lu, D. Paindaveine, M. Šiman, Local bilinear multiple-output quantile/depth regression, 2012 (submitted for publication).
[6] M. Hallin, D. Paindaveine, M. Šiman, Multivariate quantiles and multiple-output regression quantiles: from $L_1$ optimization to halfspace depth, Ann. Statist. 38 (2010) 635–669.
[7] R. Koenker, Quantile Regression, Cambridge University Press, New York, 2005.
[8] R. Koenker, G.J. Bassett, Regression quantiles, Econometrica 46 (1978) 33–50.
[9] L. Kong, I. Mizera, Quantile tomography: using quantiles with multivariate data, Statist. Sinica 22 (2012) 1589–1610.
[10] R.Y Liu, J.M. Parelius, K. Singh, Multivariate analysis by data depth: descriptive statistics, graphics and inference, Ann. Statist. 27 (1999) 783–840.
[11] D. Paindaveine, M. Šiman, On directional multiple-output quantile regression, J. Multivariate Anal. 102 (2011) 193–212.
[12] D. Paindaveine, M. Šiman, Computing multiple-output regression quantile regions from projection quantiles, Comput. Statist. 27 (2012) 29–49.
[13] D. Paindaveine, M. Šiman, Computing multiple-output regression quantile regions, Comput. Statist. Data Anal. 56 (2012) 840–853.
[14] W. Polonik, Minimum volume sets and generalized quantile processes, Stoch. Proc. Appl. 69 (1997) 1–24.
[15] E. Roelant, S. Van Aelst, An L1-type estimator of multivariate location and shape, Stat. Methods Appl. 15 (2007) 381–393.
[16] E. Roelant, S. Van Aelst, G. Willems, The minimum weighted covariance determinant estimator, Metrika 70 (2009) 177–204.
[17] P. Sun, R.M. Freund, Computation of minimum-volume covering ellipsoids, Oper. Res. 52 (2004) 690–706.
[18] S. Van Aelst, G. Willems, Multivariate regression S-estimators for robust estimation and inference, Statist. Sinica 15 (2007) 981–1001.
[19] A.W. van der Vaart, Asymptotic Statistics, Cambridge University Press, Cambridge, 1998.