Vector Quantile Regression on Manifolds

Marco Pegoraro [1 2] **Sanketh Vedula** [2 3] **Aviv A. Rosenberg** [2 3] **Irene Tallini** [1 2]
Emanuele Rodolà [1] **Alex M. Bronstein** [2 3]

Abstract

Quantile regression (QR) is a statistical tool for distribution-free estimation of conditional quantiles of a target variable given explanatory features. QR is limited by the assumption that the target distribution is univariate and defined on an Euclidean domain. Although the notion of quantiles was recently extended to multi-variate distributions, QR for multi-variate distributions on manifolds remains underexplored, even though many important applications inherently involve data distributed on, e.g., spheres (climate measurements), tori (dihedral angles in proteins), or Lie groups (attitude in navigation). By leveraging optimal transport theory and the notion of c-concave functions, we meaningfully define conditional vector quantile functions of high-dimensional variables on manifolds (M-CVQFs). Our approach allows for quantile estimation, regression, and computation of conditional confidence sets. We demonstrate the approach's efficacy and provide insights regarding the meaning of non-Euclidean quantiles through preliminary synthetic data experiments.

1. Introduction

Quantile regression (QR) (Koenker & Bassett, 1978) is a powerful statistical tool that estimates *conditional quantiles* of a target variable Y, given covariates $\mathbf{X}$. QR solves a regression problem that minimizes the pinball loss, the definition of which inherently assumes both scalar and Euclidean data. QR has thus been extensively employed in applications with scalar target variables in the Euclidean domain, where the notion of a quantile is both well-defined and widely understood. However, there exist many real-world applications in which the target data are distributed on a *manifold*, such as a sphere or a cylinder in case of climate measurements (Lagona, 2018; Scher & Messori, 2020), a torus in case of protein dihedral angles (Rosenberg et al., 2022), or more general manifolds for medical images (Pennec et al., 2019). The ability to meaningfully extend QR to manifolds would therefore unlock the power of this statistical tool for a broader range of applications.

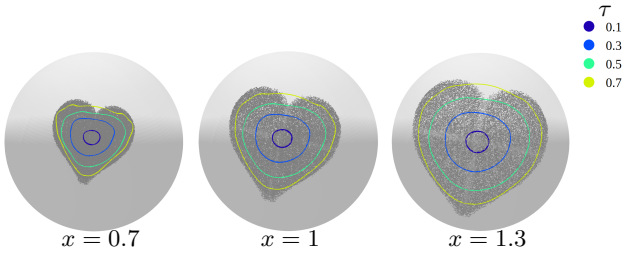


Figure 1. **Sampling and confidence set estimation for a conditional heart-shaped distribution $\mathrm{Y}|\mathbf{X}$ on $\mathcal{S}^2$.** The covariate x controls the scale of the distribution. Confidence sets containing τ probability shown as colored lines.

Recently, new perspectives on quantile functions have emerged, allowing for their extension into high-dimensional data. Notably, Carlier et al. (2016) proposed Vector Quantile Regression (VQR), defining the quantile function as a co-monotonic map between a uniform base distribution and the target distribution, and framing quantile regression as a conditional optimal transport (OT) problem to obtain the map. This approach is limited by a linearity assumption and the use of the primal OT formulation. Rosenberg et al. (2023) extended VQR by proposing a non-linear extension and introducing a more scalable solver which exploits the entropic-regularized dual of the resulting OT problem. Although this work introduced a learnable embedding that can incorporate inductive bias through the structure of $\mathbf{X}$, the approach is unable to exploit any intrinsic structure present in $\mathbf{Y}$. Regarding quantile functions on manifolds, Hallin et al. (2022) proposed directional quantiles for spherical data, by solving OT between the base and target distributions on the n-sphere. However, their approach is defined only for spherical data, and furthermore relies on the primal OT formulation, which involves a large-scale linear program, and thus quickly becomes impractical even for moderately-sized problems. Despite these limitations, Hallin et al. provide a

[1]Sapienza, University of Rome [2]Department of Computer Science, Technion [3]Sibylla, UK. Correspondence to: Marco Pegoraro <pegoraro@di.uniroma1.it>.

Workshop on New Frontiers in Learning, Control, and Dynamical Systems at the International Conference on Machine Learning (ICML), Honolulu, Hawaii, USA, 2023.

motivating example of the potential of applying OT methods to quantify distributions on manifolds, and is, to our knowledge, the only approach proposed thus far for non-Euclidean quantiles.

In this paper, we propose a novel approach for estimating high-dimensional conditional quantile functions on manifolds. Building on the dual formulation of the OT VQR problem, we propose a parametrization of the convex potentials using partially input c-concave neural networks. We thus address the major limitations of both previous works (Rosenberg et al., 2023; Hallin et al., 2022), in a holistic framework that supports multivariate distributions defined on arbitrary manifolds for which the exponential map is known. Furthermore, it is free of any distributional assumptions about $\mathbf{Y}|\mathbf{X}$, and also allows for estimating confidence sets. Through preliminary experiments, using challenging synthetic datasets defined on the 2-sphere and torus, we demonstrate the effectiveness of our approach for quantile regression with confidence sets on manifolds.

We recommend the reader to review Appendix A, which provides a concise background on the notation used in this paper, as well as information on manifolds and c-convexity. This will help enhance the understanding of the material presented in the subsequent sections.

2. Quantile Estimation on Manifolds

We formulate the problem of Manifold Quantile Regression as an optimal transport (OT) problem, similar to (Hallin et al., 2022; Rosenberg et al., 2023) but taking care to account for the manifold geometry. Consider the cost function $c(\boldsymbol{u}, \boldsymbol{y}) = \frac{1}{2} d_{\mathcal{M}}(\boldsymbol{u}, \boldsymbol{y})^2$, where $d_{\mathcal{M}}$ is an intrinsic distance function on the manifold $\mathcal{M}$. Given the set $\mathcal{S}(\mu, \nu)$ of diffeomorphisms $\boldsymbol{s} : \mathcal{M} \to \mathcal{M}$ pushing a base probability measure $\mu \in \mathcal{P}(\mathcal{M})$ to a target measure $\nu \in \mathcal{P}(\mathcal{M})$ ($\boldsymbol{s}_{\#}\mu = \nu$) with $\boldsymbol{u} \sim \mu$, the Monge problem on $\mathcal{M}$ consists in minimizing the transportation cost,

$$\inf_{s} \int_{\mathcal{M}} c(\boldsymbol{u}, \boldsymbol{s}(\boldsymbol{u})) d\mu(\boldsymbol{u}). \tag{1}$$

To avoid the non-convex optimization in Monge's formulation, we consider the celebrated Kantorovich relaxation instead. Given the set of joint distributions $\Gamma(\mu, \nu)$ on $\mathcal{M} \times \mathcal{M}$ marginalizing to μ, ν with $(\boldsymbol{u}, \boldsymbol{y}) \sim \gamma$, the Kantorovich problem minimizes the cost,

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{\mathcal{M} \times \mathcal{M}} c(\boldsymbol{u}, \boldsymbol{y}) d\gamma(\boldsymbol{u}, \boldsymbol{y}). \tag{2}$$

In this case, γ corresponds to a transport plan. The minimum of the Kantorovich problem can also be obtained through its dual formulation,

$$\begin{aligned} \sup_{\varphi,\psi} & \int_{\mathcal{M}} \varphi(\boldsymbol{u}) d\mu + \int_{\mathcal{M}} \psi(\boldsymbol{y}) d\nu \\ \text{s.t.} \quad & \varphi(\boldsymbol{u}) + \psi(\boldsymbol{y}) \leq c(\boldsymbol{u}, \boldsymbol{y}) \end{aligned} \tag{3}$$

where $\varphi, \psi : \mathcal{M} \to \mathbb{R}$ are bounded and continuous functions. The optimal dual potentials φ^*, ψ^* are each other's c-transform and, therefore both c-concave. Given the dual potential φ on a smooth compact manifold $\mathcal{M}$ with no boundary, Theorem 9 of (McCann, 2001) shows that the optimal transport map $Q_{\mathbf{Y}} : \mathcal{M} \to \mathcal{M}$, for a random target variable $\mathbf{Y}$, is a unique minimizer of the Monge problem (Equation 1) and is given by

$$Q_{\mathbf{Y}}(\boldsymbol{u}) = \exp_{\boldsymbol{u}}[-\nabla_{\boldsymbol{u}} \varphi(\boldsymbol{u})] \tag{4}$$

where ∇ is the intrinsic gradient on $\mathcal{M}$, and exp is the exponential map. Note that $Q_{\mathbf{Y}}(\boldsymbol{u})$ upholds important properties of a quantile function: it is c-cyclically monotone (Hallin et al., 2022) by virtue of being the gradient of a c-convex function, and admits strong representation (Rosenberg et al., 2023), i.e. $\mathbf{Y} = Q_{\mathbf{Y}}(\mathbf{U})$ for $\mathbf{U} \sim \mu$. The application of the exponential map here maps the gradient of the c-convex function from the tangent plane back onto the manifold surface. We thus interpret the optimal transport map $Q_{\mathbf{Y}}$ as the *manifold vector quantile function* of $\mathbf{Y}$. Similarly, can define the *manifold distribution function* $R_{\mathbf{Y}} := Q_{\mathbf{Y}}^{-1}$ as:

$$R_{\mathbf{Y}}(\boldsymbol{y}) = \exp_{\boldsymbol{y}}[-\nabla_{\boldsymbol{y}} \psi(\boldsymbol{y})]. \tag{5}$$

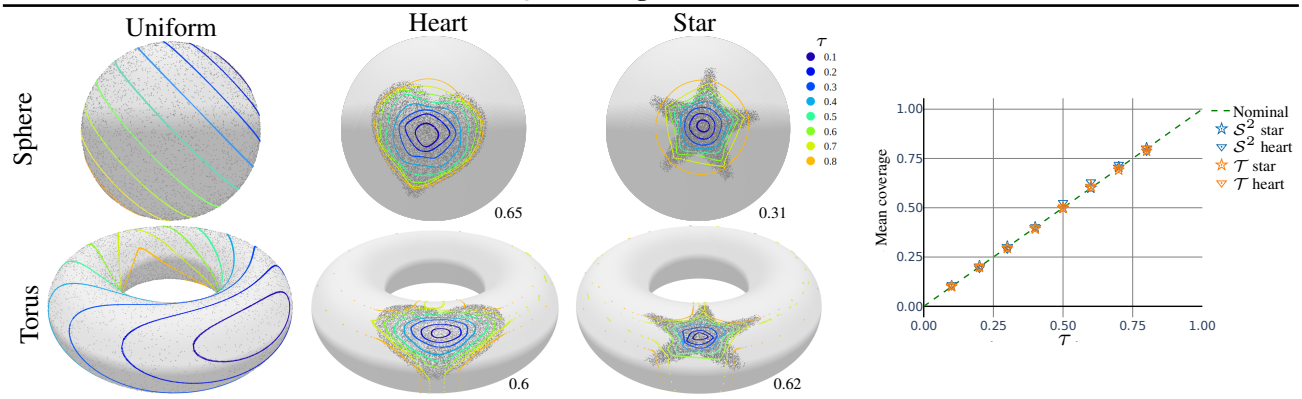
Estimation. Given samples $\{\boldsymbol{u}_i\}_{i=1}^T \sim \mathcal{U}_{\mathcal{M}}$ from a manifold uniform distribution $\mathcal{U}_{\mathcal{M}}$ (see Appendix A.2 for definition) and $\{\boldsymbol{y}_j\}_{j=1}^N \sim \mathbf{Y}$ from a random target variable $\mathbf{Y}$, the dual formulation of the Kantorovich problem (Equation (3)) can be discretized as

$$\begin{aligned} \max_{\varphi,\psi} & \sum_{i=1}^{T} \mu_i \varphi(\boldsymbol{u}_i) + \sum_{j=1}^{N} \nu_j \psi(\boldsymbol{y}_j) \\ \text{s.t.} \, \forall i, j: \quad & \varphi(\boldsymbol{u}_i) + \psi(\boldsymbol{y}_j) \leq c(\boldsymbol{u}_i, \boldsymbol{y}_j). \end{aligned} \tag{6}$$

By writing one of the potentials in terms of the other leveraging the c-transform, we obtain the following optimization problem,

$$\max_{\varphi} \sum_{i=1}^{T} \mu_i \varphi(\boldsymbol{u}_i) + \sum_{j=i}^{N} \nu_j \min_{\boldsymbol{u} \in \mathcal{M}} \left\{ c(\boldsymbol{u}, \boldsymbol{y}_j) - \varphi(\boldsymbol{u}) \right\}, \tag{7}$$

with marginal constraints $\boldsymbol{\mu} = \frac{1}{T}\mathbf{1}_T$, $\boldsymbol{\nu} = \frac{1}{N}\mathbf{1}_N$. We solve the inner problem by sampling on $\boldsymbol{u} \sim \mathcal{U}_{\mathcal{M}}$, and approximate the minimum as a soft-minimum to allow differentiability. Following (Cohen et al., 2021), we parameterize φ with $\{(\boldsymbol{z}_i, \alpha_i\}_{i=1}^M \subset \mathcal{M} \times \mathbb{R}$ where $\alpha_i = \alpha(\boldsymbol{z}_i)$ are the



(a) τ-contours (b) Mean coverages

Figure 2. **τ-contours of unconditional heart- and star-shaped distributions accurately estimated on $\mathcal{S}^2$ and $\mathcal{T}$.** *Left:* Estimated τ-contours, each containing $100 \cdot \tau$ percent of the data plotted over the ground truth data. Shown are the contours in the base uniform distribution (rotated for visual clarity), and the target heart- and star-shaped distributions. For the latter two, we report the KDE-L_1 loss ($\times 10^{-5}$). *Right:* The obtained mean coverage on the Heart and Star distributions closely matches the desired nominal coverage.

values of a function defined on $\mathcal{M}$ that we sample at learned points $\boldsymbol{z}_i$:

$$\varphi(\boldsymbol{u}) = \min_i \{c(\boldsymbol{z}_i, \boldsymbol{u}) + \alpha(\boldsymbol{z}_i)\}_{i=1}^M , \tag{8}$$

and again approximate the minimum using soft-minimum.

Confidence sets. With the estimated quantile function $Q_{\mathbf{Y}}(\boldsymbol{u})$, we can compute confidence sets on the target distribution. To define a set of nested regions with μ-probability contents $\tau \in [0, 1]$ on $\mathcal{M}$, we must first choose a central point or *pole* $\boldsymbol{\omega} \in \mathcal{M}$. This point will play the role of the median for ν, around which the contours will be nested. We opted to compute the pole using the *Fréchet mean* of ν, defined as

$$\boldsymbol{\omega} = \arg\min_{\boldsymbol{y} \in \mathcal{M}} \mathbb{E}_\nu[c(\mathbf{y}, \boldsymbol{y})] \tag{9}$$

Then, the τ-contour containing μ-probability τ centered at $\boldsymbol{\omega}$ is defined as

$$\mathcal{C}_\tau^{\mathbf{U}} = \{\boldsymbol{u} \in \mathcal{M} : C_{\boldsymbol{\omega}}^*(\boldsymbol{u}) = \tau\} \tag{10}$$

where $C_{\boldsymbol{\omega}}^*$ is a function that maps the distance from the pole $\boldsymbol{\omega}$ to the probability τ; see Appendix B for a detailed explanation. Correspondingly, the τ-quantile contour of $\mathbf{Y}$ is the image $\mathcal{C}_\tau^{\mathbf{Y}} := Q_{\mathbf{Y}}(\mathcal{C}_\tau^{\mathbf{U}})$. In Figure 2a, we show $\mathcal{C}_\tau^{\mathbf{Y}}$ on multiple distributions and different values of τ.

3. Quantile Regression on Manifolds

In numerous real-world applications, the target distribution may depend on external variables $\mathbf{X}$ that belong to a potentially different domain than the target $\mathbf{Y}$. In such cases we are interested in the *conditional* quantiles, i.e. the quantiles of $\mathbf{Y}|\mathbf{X}$ which can be estimated using (vector) quantile regression. If $\mathbf{Y}$ resides on a manifold, an extension of VQR for manifolds is required.

We consider a target distribution $P_{(\mathbf{Y},\mathbf{X})}$ from which we sample a set of points $\{\boldsymbol{y}_j, \boldsymbol{x}_j\}_{i=1}^N \sim P_{(\mathbf{Y},\mathbf{X})}$. In the case of quantile regression, the OT discrete dual problem 6 becomes (Rosenberg et al., 2023):

$$\begin{aligned} &\max_{f,\psi} \sum_{i=1}^T \mu_i \sum_{j=1}^N \nu_j f(\boldsymbol{u}_i; \boldsymbol{x}_j) + \sum_{j=1}^N \nu_j \psi(\boldsymbol{y}_j; \boldsymbol{x}_j) \\ &\text{s.t.} \, \forall i, j: \; f(\boldsymbol{u}_i; \boldsymbol{x}_j) + \psi(\boldsymbol{y}_j; \boldsymbol{x}_j) \leq c(\boldsymbol{u}_i, \boldsymbol{y}_j) \end{aligned} \tag{11}$$

where $f(\boldsymbol{u}; \boldsymbol{x}) : \mathcal{M} \times \mathbb{R}^k \to \mathbb{R}$ is a partial c-concave function in $\boldsymbol{u}$.

As in Equation (7), we reformulate to optimize for only one of the potentials,

$$\begin{aligned} \max_f \sum_{i=1}^T \mu_i \sum_{j=1}^N \nu_j f(\boldsymbol{u}_i; \boldsymbol{x}_j) + \\ \sum_{j=i}^N \nu_j \min_{\boldsymbol{u} \in \mathcal{M}} \{c(\boldsymbol{u}, \boldsymbol{y}_j) - f(\boldsymbol{u}; \boldsymbol{x}_j)\} , \end{aligned} \tag{12}$$

and retrieve the other using the c-transform,

$$\psi(\boldsymbol{y}; \boldsymbol{x}) = \min_{\boldsymbol{u} \in \mathcal{M}} \{c(\boldsymbol{u}, \boldsymbol{y}) - f(\boldsymbol{u}; \boldsymbol{x}_j)\} . \tag{13}$$

We thus define the *manifold conditional vector quantile function* (M-CVQF) $Q_{\mathbf{Y}|\mathbf{X}}$ as

$$Q_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{u}; \boldsymbol{x}) = \exp_{\boldsymbol{u}} [-\nabla_{\boldsymbol{u}} (f(\boldsymbol{u}; \boldsymbol{x}))] , \tag{14}$$

and the *manifold conditional distribution function* $R_{\mathbf{Y}|\mathbf{X}} := Q_{\mathbf{Y}|\mathbf{X}}^{-1}$ as

$$R_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{y}; \boldsymbol{x}) = \exp_{\boldsymbol{y}}[-\nabla_{\boldsymbol{y}} \psi(\boldsymbol{y}; \boldsymbol{x})]. \tag{15}$$

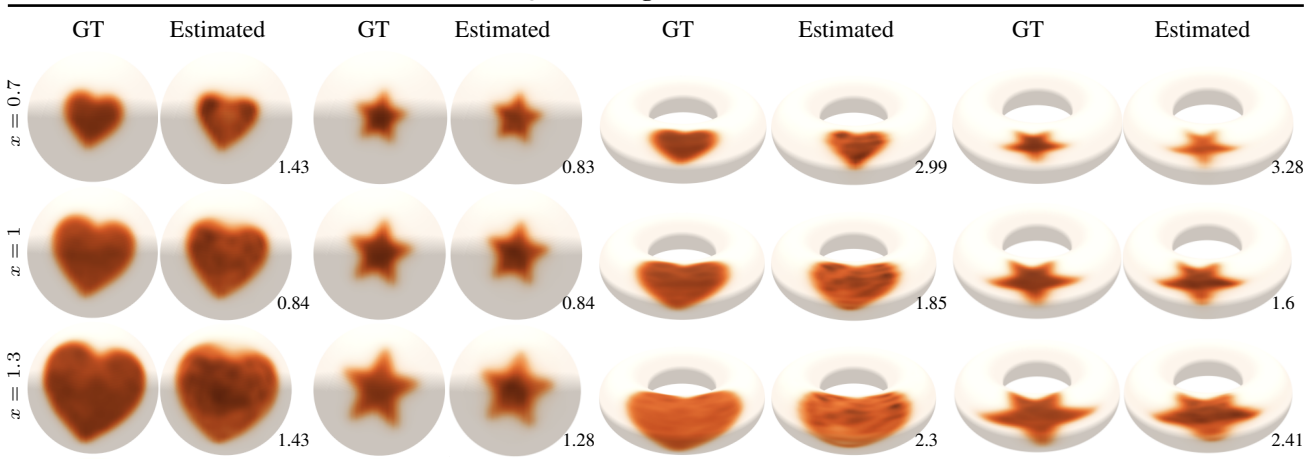


Figure 3. **Reconstruction of distribution $\mathbf{Y}|\mathbf{X}$ using the estimated M-CVQF.** We also report the KDEL-L1 loss for different conditioning. All the values are multiplied by 10^{-5}.

Conditional τ-contours. To compute contours in the conditional case, we employ the M-CVQF. Given the unconditional τ-contours defined in Equation (10), we define the conditional τ-contours $\mathcal{C}_\tau^{\mathbf{Y}|\mathbf{X}}$ as

$$\mathcal{C}_\tau^{\mathbf{Y}|\mathbf{X}} := Q_{\mathbf{Y}|\mathbf{X}}(\mathcal{C}_\tau^{\mathbf{U}}). \tag{16}$$

Implementation and training. We implement $f(\boldsymbol{u};\boldsymbol{x})$ as a non-negative sum of functions which are c-concave in $\boldsymbol{u}$ and non-linear in $\boldsymbol{x}$: $f(\boldsymbol{u};\boldsymbol{x}) = \varphi(\boldsymbol{u}) + \boldsymbol{\beta}(\boldsymbol{u})^\top \boldsymbol{g}(\boldsymbol{x})$ with $\boldsymbol{\beta} : \mathcal{M} \to \mathbb{R}^{\tilde{k}}$ and $\boldsymbol{g} : \mathbb{R}^k \to \mathbb{R}^{\tilde{k}}_{\geq 0}$. Both $\boldsymbol{\beta}, \boldsymbol{g}$ are parametrized using neural networks. In particular, $\boldsymbol{\beta}$ is implemented as a $\tilde{k}$-stack of c-concave functions (8): $\boldsymbol{\beta} = [\beta_1 \dots \beta_{\tilde{k}}]$ where $\beta_i : \mathcal{M} \to \mathbb{R}$ with $i \in [1 \dots \tilde{k}]$. On the other hand, $\boldsymbol{g}$ can be any learned embedding suitable for the domain of $\mathbf{X}$; e.g., if the conditioning variable is an image, $g(\boldsymbol{x})$ may be a CNN to leverage the translation equivariance and hierarchical nature of image features. In our experiments, we implement $\boldsymbol{g}$ as a multilayer perceptron (MLP) with rectified linear unit (ReLU) activation as the last layer to ensure non-negative values.

4. Experiments

We validate the proposed method on synthetic distributions generated on two manifolds: the *2-sphere* $\mathcal{S}^2$ and the *torus* $\mathcal{T} = \mathcal{S}^1 \times \mathcal{S}^1$. Both manifolds have a closed form for the inner distance and exponential map, which we report in the Appendix A.2. As target distributions, we consider two manifold uniform distributions $\mathcal{U}_A$ in a subset $A \subset \mathcal{M}$ with heart and star shape. Full experimental details can be found in Appendix C.

Quantile estimation. We evaluate the unconditional τ-contours computed on the target distribution ν, both qualitatively and quantitatively. Figure 2a shows the τ-contours for uniform, heart- and star-shaped distributions. The results show that all the contours have the desired properties described by (Hallin et al., 2022): closed, connected and nested. As discussed by (Cohen et al., 2021), adopting the soft-minimum in Equation (8) results in estimated potentials that may not be strictly c-concave, as this property is not enforced by the optimization procedure. Our empirical results demonstrate that in these cases the obtained potentials maintain the desired property.

To quantitatively evaluate the probability contained in each τ-contour, we formulate a coverage test. Figure 2b shows the coverage results for the contours shown in Figure 2a. In all considered cases, the coverage error has a mean of 6.5×10^{-3}.

Quantile regression. In Figure 3, we reconstruct the target distribution $P_{(\mathbf{Y},\mathbf{X})}$ using the estimated M-CVQF, $\widehat{Q}_{\mathbf{Y}|\mathbf{X}}$. For each conditioning $\boldsymbol{x}$, we perform Kernel Density Estimation (KDE) on the manifold to approximate both the ground truth and the estimated distributions from their samples. We consider two distributions (Heart and Star), where the conditioning variable $\boldsymbol{x}$ controls the scale. In all the cases the estimated distribution qualitatively resembles the true one, with only minor artifacts in the inner region.

5. Discussion and Conclusions

Our work provides the first formulation of conditional vector quantile functions on manifolds, by extending non-linear VQR (Rosenberg et al., 2023) to non-Euclidean domains. Our key contribution is a novel formulation of nonlinear VQR, which allows us to estimate conditional quantiles and construct confidence sets on manifolds by fitting data sampled directly from the joint distribution.

One potential limitation of this work is that the estimated potential might not be c-concave, as this property is only

promoted but not enforced. In future studies, we hope to analyze it from a theoretical perspective and provide ways to overcome this limitation. Another avenue for future work is to extend the proposed approach to additional domains possessing closed-form formulations for computing c-concave functions, such as Lie groups. Finally, to test its efficacy in challenging applications, we plan to apply our method to real-world data in diverse domains such as pose estimation, weather modeling, and protein structure prediction.

Acknowledgements

This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. 863839). Moreover, it has been supported by ERC grant no.802554 (SPECGEO), PRIN 2020 project no.2020TA3K9N (LEGO.AI), and PNRR MUR project PE0000013-FAIR.

References

Boyd, S. P. and Vandenberghe, L. *Convex optimization*. Cambridge university press, 2004.

Carlier, G., Chernozhukov, V., and Galichon, A. Vector quantile regression: An optimal transport approach. *Annals of Statistics*, 44(3):1165–1192, 2016. ISSN 00905364. doi: 10.1214/15-AOS1401.

Cohen, S., Amos, B., and Lipman, Y. Riemannian convex potential maps. In *International Conference on Machine Learning*, pp. 2028–2038. PMLR, 2021.

Hallin, M., Liu, H., and Verdebout, T. Nonparametric measure-transportation-based methods for directional data. *arXiv preprint arXiv:2212.10345*, 2022.

Koenker, R. and Bassett, G. Regression Quantiles. *Econometrica*, 46(1):33, 1978. ISSN 00129682. doi: 10.2307/1913643.

Lagona, F. Correlated cylindrical data. In *Applied Directional Statistics*, pp. 61–76. Chapman and Hall/CRC, 2018.

McCann, R. J. Polar factorization of maps on riemannian manifolds. *Geometric & Functional Analysis GAFA*, 11 (3):589–608, 2001.

Pennec, X., Sommer, S., and Fletcher, T. *Riemannian geometric statistics in medical image analysis*. Academic Press, 2019.

Rosenberg, A. A., Marx, A., and Bronstein, A. M. Codon-specific ramachandran plots show amino acid backbone conformation depends on identity of the translated codon. *Nature communications*, 13(1):2815, 2022.

Rosenberg, A. A., Vedula, S., Romano, Y., and Bronstein, A. M. Fast nonlinear vector quantile regression. *International Conference on Learning Representations (ICLR)*, 2023.

Scher, S. and Messori, G. Spherical convolution and other forms of informed machine learning for deep neural network based weather forecasts. *arXiv preprint arXiv:2008.13524*, 2020.

A. Background

A.1. Notation

Throughout, Y, $\mathbf{X}$ denote random variables and vectors, respectively; deterministic scalars, vectors and matrices are denoted as y, $\boldsymbol{x}$, and $\boldsymbol{X}$. $P_{(\mathbf{Y},\mathbf{X})}$ denotes the joint distribution of the $\mathbf{X}$ and $\mathbf{Y}$. $\mathbf{1}_N$ denotes an N-dimensional vector of ones, $\odot$ denotes the elementwise product, and $\mathbb{I}\{\cdot\}$ is an indicator. We denote by N the number of samples, T the number of vector quantile levels, k the dimension of the covariates, and $\tilde{k}$ is the embedding dimension. $Q_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{u};\boldsymbol{x})$ is the *manifold conditional vector quantile function* M-CVQF of the variable $\mathbf{Y}|\mathbf{X}$ evaluated at the vector quantile level $\boldsymbol{u}$ for $\mathbf{X}=\boldsymbol{x}$.

A.2. Manifolds

We report the definition of differential geometry used in the paper. Consider a manifold $(\mathcal{M}, g)$ with metric g. The *tangent space* at a point $\boldsymbol{y} \in \mathcal{M}$ is the linear subspace $T_{\boldsymbol{y}}\mathcal{M} = \left\{\boldsymbol{v} \in \mathbb{R}^d : \boldsymbol{v}^\top \boldsymbol{y} = 0\right\}$.

Diffeomorphism. We define a diffeomorphism $s : \mathcal{M} \to \mathcal{M}$ as a differentiable bijective mapping with a differentiable inverse.

Exponential map. On the manifold, we define the exponential map $\exp_{\boldsymbol{y}}$ at point $\boldsymbol{y} \in \mathcal{M}$ as a map from the tangent space $T_{\boldsymbol{y}}\mathcal{M}$ to $\mathcal{M}$.

Definition A.1 (Exponential map)**.** Let $\boldsymbol{y} \in \mathcal{M}$, $\boldsymbol{v} \in T_y\mathcal{M}$ and consider the unique geodesic $\gamma : [0,1] \to \mathcal{M}$ such that $\gamma(0) = \boldsymbol{y}$ and $\gamma'(0) = v$. The exponential map at $\boldsymbol{y}$, $\exp_{\boldsymbol{y}} : T_{\boldsymbol{y}}\mathcal{M} \to \mathcal{M}$, is defined as $\exp_{\boldsymbol{y}}(\boldsymbol{v}) = \gamma(1)$.

Intrinsic distance. On a manifold $\mathcal{M}$, the length of a curve $\gamma : [0,1] \to \mathcal{M}$ is define as $L(\gamma) = \int_0^1 \|\gamma'(t)\|_g \, dt$, where $\|\gamma'(t)\|_g$ means taking the norm of the velocity $\gamma'(t)$ at $T_{\gamma(t)}\mathcal{M}$ with respect to the metric g of the manifold $\mathcal{M}$.

Definition A.2 (Intrinsic distance)**.** The intrinsic distance $d_{\mathcal{M}}$ between $\boldsymbol{y}, \boldsymbol{u} \in \mathcal{M}$ is the infimum over the curves $\gamma : [0,1] \to \mathcal{M} : d_{\mathcal{M}}(\boldsymbol{y}, \boldsymbol{u}) = \inf_\gamma L(\gamma)$, with $\gamma(0) = \boldsymbol{y}$ and $\gamma(1) = \boldsymbol{u}$.

If $\mathcal{M}$ is *complete* (e.g., Hopf-Rinow Theorem), a geodesic realizes the intrinsic distance.

Manifold uniform distribution. Given the volume measure $d\mathcal{M}(\boldsymbol{y})$ representing the infinitesimal volume element at each point of the manifold $\boldsymbol{y} \in \mathcal{M}$, a random variable $\mathbf{Y}$ follows a manifold uniform distribution $\mathcal{U}_A$ on the bounded subset $A \subseteq \mathcal{M}$ if its probability density function (PDF) is constant within A:

$$p_{\mathbf{Y}}(\boldsymbol{y}) = \frac{\mathbb{I}\{A\}(\boldsymbol{y})}{V(A)} \tag{17}$$

where $V(A)$ is the volume of the set A with respect to the volume measure $d\mathcal{M}$. The uniform distribution on the manifold, with respect to the volume measure, assigns probabilities to subsets of the manifold based on their intrinsic volumes.

Sphere. On the n-sphere $\mathcal{S}^n$, the exponential map and the intrinsic distance are provided as closed-form expressions. If $\boldsymbol{y}, \boldsymbol{u} \in \mathcal{S}^n$ and $\boldsymbol{v} \in T_{\boldsymbol{y}}\mathcal{S}^n$,

$$\exp_{\boldsymbol{y}}^{\mathcal{S}^n}(\boldsymbol{v}) = \boldsymbol{y}\cos(\|\boldsymbol{v}\|) + \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}\sin(\|\boldsymbol{v}\|) \tag{18}$$

$$d_{\mathcal{S}^n}(\boldsymbol{y}, \boldsymbol{u}) = \arccos(\boldsymbol{u}^T\boldsymbol{v}), \tag{19}$$

where $\|\cdot\|$ is the standard Euclidean norm.

Torus. The torus $\mathcal{T}$ can be defines as a product manifold between two 1-spehere: $\mathcal{T} = \mathcal{S}^1 \times \mathcal{S}^1$. On general product manifolds of the form $\mathcal{M} = \mathcal{M}_1 \times \ldots \times \mathcal{M}_l$, the squared intrinsic distance is simply

$$d^2_{\mathcal{M}}(\boldsymbol{y}, \boldsymbol{u}) = d^2_{\mathcal{M}_1}(\boldsymbol{y}_1, \boldsymbol{u}_1) + \ldots + d^2_{\mathcal{M}_l}(\boldsymbol{y}_l, \boldsymbol{u}_l). \tag{20}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l)$, and $\boldsymbol{u}_j \in \mathcal{M}_j, \;\; j \in [l]$ (and similarly for y). The exponential map on the product manifold is the cartesian product of exponential maps on the individual manifolds. Therefore, the exponential map and intrinsic distance on

the torus $\mathcal{T}$ is defined as:

$$\exp^{\mathcal{T}}_{\boldsymbol{y}}(\boldsymbol{v}) = \exp^{\mathcal{S}^1}_{\boldsymbol{y}_1}(\boldsymbol{v}_1) \times \exp^{\mathcal{S}^1}_{\boldsymbol{y}_2}(\boldsymbol{v}_2) \tag{21}$$

$$d_{\mathcal{T}}(\boldsymbol{y}, \boldsymbol{u}) = \sqrt{d^2_{\mathcal{S}^1}(\boldsymbol{y}_1, \boldsymbol{u}_1) + d^2_{\mathcal{S}^1}(\boldsymbol{y}_2, \boldsymbol{u}_2)}. \tag{22}$$

c-concave and c-convex functions Given the cost function $c(\boldsymbol{z}, \boldsymbol{u}) = \frac{1}{2} d_{\mathcal{M}}(\boldsymbol{z}, \boldsymbol{u})^2$, we define the c-concave and c-convex functions as :

Definition A.3 (c-concave)**.** A function $\varphi : \mathcal{M} \to \mathbb{R} \cup \{-\infty\}$ is *c-concave* if it is not identically $\{-\infty\}$ and there exists $\alpha : \mathcal{M} \to \mathbb{R} \cup \{\pm\infty\}$ s.t.

$$\varphi_{concave}(\boldsymbol{u}) = \inf_{\boldsymbol{z} \in \mathcal{M}} \{c(\boldsymbol{z}, \boldsymbol{u}) + \alpha(\boldsymbol{z})\} \tag{23}$$

Definition A.4 (c-convex)**.** A function $\varphi : \mathcal{M} \to \mathbb{R} \cup \{+\infty\}$ is *c-convex* if it is not identically $\{+\infty\}$ and there exists $\alpha : \mathcal{M} \to \mathbb{R} \cup \{\pm\infty\}$ s.t.

$$\varphi_{convex}(\boldsymbol{u}) = \sup_{\boldsymbol{z} \in \mathcal{M}} \{-c(\boldsymbol{z}, \boldsymbol{u}) + \alpha(\boldsymbol{z})\} \tag{24}$$

If φ is c-concave, then $-\varphi$ is c-convex. It is important to notice that a non-negative weighted sum of c-convex functions is still c-convex (Boyd & Vandenberghe, 2004). We also define the c-transform of a function $\varphi : \mathcal{M} \to \mathbb{R}$ as:

$$\varphi^c(\boldsymbol{u}) = \inf_{\boldsymbol{z} \in \mathcal{M}} \{c(\boldsymbol{z}, \boldsymbol{u}) - \varphi(\boldsymbol{z})\}. \tag{25}$$

c-cyclically monotone. We define the concept of *c-cyclically monotone* using the definition from (Hallin et al., 2022):

Definition A.5 (c-cyclically monotone)**.** A subset $\Omega \in \mathcal{M} \times \mathcal{M}$ is c-cyclically monotone if, denoting by $\Sigma(k)$ the set of permutations of $\{1, \dots, k\}$, for all $k \in \mathbb{N}$, all $\sigma \in \Sigma(k)$, and all $(\boldsymbol{x}_1, \boldsymbol{y}_1), \dots, (\boldsymbol{x}_k, \boldsymbol{y}_k) \in \Omega$,

$$\sum_{i=1}^{k} c(\boldsymbol{x}_i, \boldsymbol{y}_i) \leq \sum_{i=1}^{k} c(\boldsymbol{x}_{\sigma(i)}, \boldsymbol{y}_i)$$

The solution of the Kantorovich problem (Equation (2)) is always supported on a c-cyclically monotone subset of $\mathcal{M} \times \mathcal{M}$.

B. Creating confidence sets

Given a point $\boldsymbol{\omega} \in \mathcal{M}$, we define a confidence set with a level of confidence $(1 - \tau)$ on a manifold uniform distribution as the set of points contained in a τ-contour with pole $\boldsymbol{\omega}$:

$$\mathbb{C}^{\mathbf{U}}_{(1-\tau)} = \{\boldsymbol{u} \in \mathcal{M} : C^*_{\boldsymbol{\omega}}(\boldsymbol{u}) \leq \tau\} \tag{26}$$

$C^*_{\boldsymbol{\omega}}$ is a function that maps distances $d_{\mathcal{M}}(\boldsymbol{\omega}, \boldsymbol{u})$, with $\boldsymbol{u} \in \mathcal{M}$, to the probabilities $\tau \in [0, 1]$. We build this function empirically for each pole $\boldsymbol{\omega}$. We compute the set of points $\mathcal{C}^*_\kappa = \{\boldsymbol{u} \in \mathcal{M} : d_{\mathcal{M}}(\boldsymbol{\omega}, \boldsymbol{u}) = \kappa\}$ and the amount of probability τ_κ contained in $\mathcal{C}^*_\kappa$ as the percentage of points $\boldsymbol{u} \in \mathcal{M}$ with $d_{\mathcal{M}}(\boldsymbol{\omega}, \boldsymbol{u}) \leq \kappa$. The function $C^*_{\boldsymbol{\omega}}$ is then the interpolation of the pairs $\{(\kappa, \tau_\kappa)\}_{i=1}^{N_\kappa}$. Figure 4 shows the amount of probability contained in the contours as their distance from the pole increases with and without using $C^*_{\boldsymbol{\omega}}$.

C. Experiments details

C.1. Implementation

We train our models for 6×10^3 iterations and use the Adam optimizer with a learning rate of 10^{-3}. We ensure the optimality of the obtained M-CVQF using the c-cyclically monotone test defined in Definition A.5.

Spherical uniform distribution $\mathcal{U}_{\mathcal{S}^n}$ We compute the manifold uniform distribution on the n-sphere sampling each component of the vector $\boldsymbol{v} \in \mathbb{R}^n$ from a random uniform distribution and then normalizing $\boldsymbol{v}$ such that $\| \boldsymbol{v} \| = 1$.

Torus uniform distribution $\mathcal{U}_{\mathcal{T}}$ We compute the manifold uniform distribution on the torus sampling from the spherical uniform distributions $\mathcal{U}_{\mathcal{S}^1}$ on the $\mathcal{S}^1$ and then concatenating the results: $\boldsymbol{u} = [\boldsymbol{u}_1, \boldsymbol{u}_2] \sim \mathcal{U}_{\mathcal{T}}$ with $\boldsymbol{u}_1, \boldsymbol{u}_2 \sim \mathcal{U}_{\mathcal{S}^1}$.

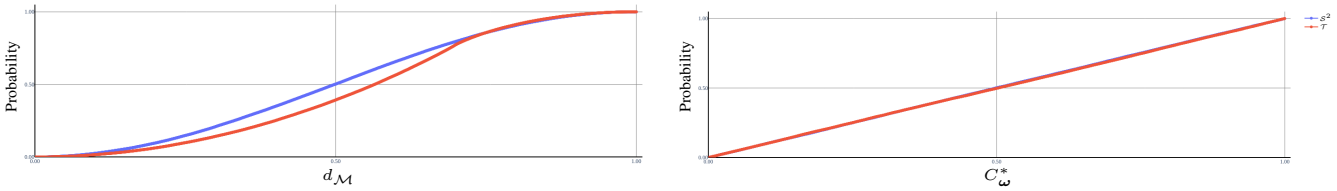


Figure 4. Probability contained in the contours computed using the inner distance $d_{\mathcal{M}}(\omega, \boldsymbol{u})$ (left) and the mapping function $C^*_{\boldsymbol{\omega}}(\boldsymbol{u})$ (right). The distances are normalized to be in the range $[0, 1]$. The mapping function $C^*_{\boldsymbol{\omega}}$ creates a linear dependency between the distance of the points in a contour and the amount of probability contained.

C.2. Distribution distance (KDE-L_1)

To evaluate the learned MQR model $\hat{Q}_{\mathbf{Y}|\boldsymbol{X}}$, we use the L_1-loss between the estimated density functions $h_{\mathbf{Y}|\boldsymbol{X}}$. Given a set of L evaluation covariate vectors $\{\boldsymbol{x}_l\}_{l=1}^{L}$, from the learned $\hat{Q}_{\mathbf{Y}|\mathbf{X}}$, we sample a set of points on the target distribution: $\{\hat{\boldsymbol{y}}_{m,l} = \hat{Q}_{\mathbf{Y}|\mathbf{X}}(u_m, x_l)\}_{m=1}^{M}$. We estimate the density function $\hat{h}_{\nu|\boldsymbol{X}=\boldsymbol{x}_l}$ from $\{\hat{\boldsymbol{y}}_{m,l}\}$ using the Kernel Density Estimation (KDE). We do the same for a set of points $\{\tilde{\boldsymbol{y}}_{m,l}\}$ sampled from the ground truth distribution $\mathbf{Y}|\mathbf{X} = \boldsymbol{x}_l$, obtaining the estimated function $\tilde{h}_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}_l}$. We calculate the KDE-$L_1$ loss as:

$$|\tilde{h}_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}_l} - \hat{h}_{\mathbf{Y}|\mathbf{X}=\boldsymbol{x}_l}|. \tag{27}$$

C.3. τ-contours

τ-contours computation. In the case of the Heart and Star distribution, the τ-contours are computed using the vector quantile function $\hat{Q}_{\mathbf{Y}}$. First of all, we compute the τ-contours $\mathcal{C}^{\mathbf{U}}_{\tau}$ on the base distribution μ according to Equation 10 (first row in Figure 2a). Then we map $\mathcal{C}^{\mathbf{U}}_{\tau}$ to the target distribution using the learned vector quantile function $\hat{Q}_{\mathbf{Y}}$: $\mathcal{C}^{\mathbf{Y}}_{\tau} := \hat{Q}_{\mathbf{Y}}(\mathcal{C}^{\mathbf{U}}_{\tau})$.

Coverage test. To quantitatively evaluate the probability contained in each τ-contour, we formulate a coverage test. We sample points $\{\tilde{\boldsymbol{y}}_m\}$ from the ground truth distribution ν and map them back to the base distribution $\mathcal{U}_{\mathcal{M}}$ using the rank function defined in Equation 5: $\{\hat{\boldsymbol{u}}_m = \hat{R}(\tilde{\boldsymbol{y}}_m)\}$. For each τ-contour $\mathcal{C}^{\mathbf{U}}_{\tau}$ computed on the uniform distribution $\mathcal{U}_{\mathcal{M}}$, we compute the average coverage of $\{\hat{\boldsymbol{u}}_m\}$ as $\frac{1}{M}\sum_{m=1}^{M} \text{cov}_{\mu,\boldsymbol{\omega}}(\hat{\boldsymbol{u}}_m)$ where $\text{cov}_{\mu,\boldsymbol{\omega}}(\boldsymbol{u}, \tau)$ is defined as follows:

$$\text{cov}_{\mu,\boldsymbol{\omega}}(\boldsymbol{u}, \tau) = \begin{cases} 1 & \text{if } C^*_{\boldsymbol{\omega}}(\boldsymbol{u})) \leq \tau \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

where $\boldsymbol{\omega}$ is the pole of the set of contours $\{\mathcal{C}^{\mathbf{U}}_{\tau}\}$.