Vector quantile regression beyond the specified case

G. Carlier[*], V. Chernozhukov [†], A. Galichon [‡]

March 18, 2017

Abstract

This paper studies vector quantile regression (VQR), which models the dependence of a random vector with respect to a vector of explanatory variables with enough flexibility to capture the whole conditional distribution, and not only the conditional mean. The problem of vector quantile regression is formulated as an optimal transport problem subject to an additional mean-independence condition. This paper provides results on VQR beyond the specified case which had been the focus of previous work. We show that even beyond the specified case, the VQR problem still has a solution which provides a general representation of the conditional dependence between random vectors.

Keywords: vector quantile regression, optimal transport, duality.

1 Introduction

Vector quantile regression was recently introduced in [4] in order to generalize the technique of quantile regression when the dependent random variable is multivariate. Quantile regression, pioneered by Koenker and Bassett [10], provides a powerful way to study dependence between random variables assuming a linear form for the quantile of the endogenous variable Y given the explanatory variables X. It has therefore become a very popular tool in many areas of economics, program evaluation, biometrics, etc. However, a well-known limitation of the approach is that Y should be scalar so that its quantile map is defined. When Y is multivariate, there is no canonical notion of quantile, and the picture is less clear than in the univariate case[1].

[*]CEREMADE, UMR CNRS 7534, Université Paris IX Dauphine, Pl. de Lattre de Tassigny, 75775 Paris Cedex 16, FRANCE, and MOKAPLAN Inria Paris, carlier@ceremade.dauphine.fr

[†]Department of Economics, MIT, 50 Memorial Drive, E52-361B, Cambridge, MA 02142, USA, vchern@mit.edu

[‡]Economics Department and Courant Institute of Mathematical Sciences, NYU, 70 Washington Square South, New York, NY 10013, USA ag133@nyu.edu.

[1]There is actually an important literature that aims at generalizing the notion of quantile to a multidimensional setting and various different approaches have been proposed; see in particular [1], [9], [12] and the references therein.

The approach proposed in [4] is based on optimal transport ideas and can be described as follows. For a random vector vector Y taking values in $\mathbb{R}^d$, we look for a random vector U uniformly distributed on the unit cube $[0,1]^d$ and which is maximally correlated to Y, finding such a U is an optimal transport problem. A celebrated result of Brenier [2] implies that such an optimal U is characterized by the existence of a convex function φ such that $Y = \nabla\varphi(U)$. When $d = 1$, of course, the optimal transport map of Brenier $\nabla\varphi = Q$ is the quantile map of Y. In higher dimensions it still retains one of the main properties of univariate quantiles, namely monotonicity. Thus Brenier's map $\nabla\varphi$ is a natural candidate to be considered as the vector quantile of Y, and one advantage of such an approach is the pointwise relation $Y = \nabla\varphi(U)$ where U is a uniformly distributed random vector which best approximates Y in L^2.

If, in addition, we are given another random vector X capturing a set of observable explanatory variables, we may wish to have a tractable method to estimate the conditional quantile of Y given $X = x$, i.e. the map $u \in [0,1]^d \mapsto Q(x,u) \in \mathbb{R}^d$. In the univariate case $d = 1$, and if the conditional quantile is affine in x i.e. $Q(x,u) = \alpha(u) + \beta(u)x$, the quantile regression method of Koenker and Bassett gives a constructive and powerful linear programming approach to compute the coefficients $\alpha(t)$ and $\beta(t)$ for any fixed $t \in [0,1]$. When quantile regression is specified, i.e. when the true conditional quantile is affine in x, this variational approach estimates the true coefficients $\alpha(t)$ and $\beta(t)$. In [4], we have shown that in the multivariate case as well, when the true vector quantile is affine in x, one may estimate it by a variational problem which consists in finding the uniformly distributed random variable U such that $\mathbb{E}(X|U) = \mathbb{E}(X)$ (mean independence) and maximally correlated with Y.

The purpose of the present paper is to understand what these variational approaches tell about the dependence between Y and X in the general case i.e. without assuming any particular form for the conditional quantile. We will characterize the solution of the optimal transport problem with a mean-independence constraint from [4] and relate it to a relaxed form of specified quantile regression. To be more precise, our theorem 3.3 below will provide the following general representation of the distribution of (X,Y):

$$\begin{aligned} &Y \in \partial\Phi_X^{**}(U) \text{ with } X \mapsto \Phi_X(U) \text{ affine},\\ &\Phi_X(U) = \Phi_X^{**}(U) \text{ almost surely},\\ &U \text{ uniformly distributed on } [0,1]^d,\ \mathbb{E}(X|U) = \mathbb{E}(X), \end{aligned}$$

where Φ_x^{**} denotes the convex envelope of $u \to \Phi_x(u)$ for a fixed x, and ∂ denotes the subdifferential. The main ingredients are convex duality and an existence theorem for optimal dual variables. The latter is a non-trivial extension of Kantorovich duality: indeed, the existence of a Lagrange multiplier associated to the mean-independence constraint is not straightforward and we shall prove it thanks to Komlos' theorem (theorem 3.2). Vector quantile regression is *specified* if $\Phi_x(u)$ is convex for all x in the support, in which case one can write

$$Y = \nabla\Phi_X(U) \text{ with } \Phi_X(.) \text{ convex, } X \mapsto \Phi_X(U) \text{ affine,}$$
$$U \text{ uniformly distributed on } [0,1]^d,\ \mathbb{E}(X|U) = \mathbb{E}(X).$$

While our previous paper [4] focused on the specified case, the results we obtain in the present paper are general.

2 Vector quantiles and optimal transport

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be some nonatomic probability space, and let (X, Y) be a random vector, where the vector of explanatory variables X is valued in $\mathbb{R}^N$ and the vector of dependent variables Y is valued in $\mathbb{R}^d$.

The notion of vector quantile was recently introduced by Ekeland, Galichon and Henry [6], Galichon and Henry [8] and was used in the framework of quantile regression in our companion paper [4]. The starting point for this approach is the correlation maximization problem

$$\max\{\mathbb{E}(V \cdot Y),\ \mathrm{Law}(V) = \mu\} \tag{2.1}$$

where $\mu := \mathrm{uniform}([0,1]^d)$ is the uniform measure on the unit d-dimensional cube $[0,1]^d$. This problem is equivalent to the optimal transport problem which consists in minimizing $\mathbb{E}(|Y - V|^2)$ among uniformly distributed random vectors V. This quadratic optimal transport problem has received a lot of attention since the 80's, see [16], [13], [5]. An important result in this field is the polar factorization theorem of Brenier [2]. Indeed, as shown in [2], (2.1) has a solution U which is characterized by the condition

$$Y = \nabla\varphi(U)$$

for some (essentially uniquely defined) convex function φ which is obtained by solving a dual formulation of (2.1). Arguing that gradients of convex functions are the natural multivariate extension of monotone nondrecreasing functions, the authors of [6] and [8] considered the function $Q := \nabla\varphi$ as the vector quantile of Y. We therefore shall define the quantile of Y as the optimal transport map (for the quadratic cost) $Q = \nabla\varphi$ between the uniform measure on $[0,1]^d$ and $\mathrm{Law}(Y)$. We refer to the textbooks [17], [14] and [15] for a presentation of optimal transport theory, and to [7] for a survey of applications to economics.

Let us now assume that in addition, there is an N-dimensional random vector X of regressors, $\nu := \mathrm{Law}(X, Y)$, $m := \mathrm{Law}(X)$, $\nu = \nu^x \otimes m$ where m is the law of X and ν^x is the law of Y given $X = x$. One can consider $Q(x, u) = \nabla\varphi(x, u)$ as the optimal transport between μ and ν^x

$$Y = Q(X, U) = \nabla_u \varphi(X, U), \; U \text{ uniform on } [0, 1]^d.$$

By definition, $Q(X, .)$ is the conditional vector quantile of Y given X.

Note that in the specified case i.e. when the conditional quantile function is affine in X and $Y = Q(X, U) = \alpha(U) + \beta(U)X$ where U is uniform and independent from X, the function $u \mapsto \alpha(u) + \beta(u)x$ should be the gradient of some function of u which requires

$$\alpha = \nabla\varphi, \; \beta = Db^T$$

for some potential φ and some vector-valued function b in which case, $Q(x, .)$ is the gradient of $u \mapsto \varphi(u) + b(u) \cdot x$. Moreover, since quantiles are gradients of convex potentials one should also have that $u \in [0, 1]^d \mapsto \varphi(u) + b(u) \cdot x$ is convex.

3 Vector quantile regression

3.1 Correlation maximization

Without loss of generality we normalize X so that it is centered i.e. $\mathbb{E}(X) = 0$. Our approach to vector quantile regression is based on the following correlation maximization problem, subject to a mean-independence constraint:

$$\max\{\mathbb{E}(V \cdot Y), \; \mathrm{Law}(V) = \mu, \; \mathbb{E}(X|V) = 0\}. \tag{3.1}$$

where $\mu = \mathrm{uniform}([0, 1]^d)$ is the uniform measure on the unit d-dimensional cube. The connection with the specification of vector quantile regression (i.e. the validity of an affine in x form for the conditional quantile) is given by the following result from [4]:

Proposition 3.1. *If* $Y = \nabla\varphi(U) + Db(U)^T X$ *with*

- $u \mapsto \varphi(u) + b(u) \cdot x$ *convex and smooth for* m*-a.e* x,
- $\mathrm{Law}(U) = \mu$, $\mathbb{E}(X|U) = 0$,

then U *solves (3.1).*

3.2 Duality

We now aim at emphasizing the kind of information provided by relation (3.1) regarding the dependence of X and Y. A good starting point is convex duality. As explained in [4], the dual of (3.1) takes the form

$$\inf_{(\psi,\varphi,b)} \mathbb{E}(\psi(X, Y) + \varphi(U)) \; : \; \psi(x, y) + \varphi(t) + b(t) \cdot x \geq t \cdot y. \tag{3.2}$$

where U is any uniformly distributed random vector on $[0,1]^d$ i.e. $\mathrm{Law}(U) = \mu = \mathrm{uniform}([0,1]^d)$ and the infimum is taken over continuous functions $\psi \in C(\mathrm{spt}(\nu),\mathbb{R})$, $\varphi \in C([0,1]^d,\mathbb{R})$ and $b \in C([0,1]^d,\mathbb{R}^N)$ satisfying the pointwise constraint

$$\psi(x,y) + \varphi(t) + b(t)\cdot x \geq t\cdot y, \ \ \forall (x,y,t) \in \mathrm{spt}(\nu)\times[0,1]^d. \tag{3.3}$$

Since for fixed (φ, b), the smallest ψ that satisfies the pointwise constraint in (3.2) is given by the convex function

$$\psi(x,y) := \max_{t\in[0,1]^d}\{t\cdot y - \varphi(t) - b(t)\cdot x\}$$

one may equivalently rewrite (3.2) as the minimization over continuous functions φ and b of

$$\int \max_{t\in[0,1]^d}\{t\cdot y - \varphi(t) - b(t)\cdot x\}\nu(dx,dy) + \int_{[0,1]^d}\varphi(t)\mu(dt). \tag{3.4}$$

By standard approximation techniques, one can show that the infimum in (3.4) over continuous functions (φ, b) coincides with the infimum over smooth or simply integrable functions.

The existence of optimal (L^1) functions ψ, φ and b is our first main result:

Theorem 3.2. *Assume that*

- *the support of ν, is of the form $\mathrm{spt}(\nu) := \overline{\Omega}$ where Ω is an open bounded convex subset of $\mathbb{R}^N\times\mathbb{R}^d$,*
- *$\nu \in L^\infty(\Omega)$, and ν is bounded away from zero on compact subsets of Ω that is for every K compact, included in Ω there exists $\alpha_K > 0$ such that $\nu \geq \alpha_K$ a.e. on K,*

then, the dual problem (3.2) admits at least a solution.

Proof. Let us denote by $(0,\overline{y})$ the barycenter of ν:

$$\int_\Omega x\ \nu(dx,dy) = 0,\ \int_\Omega y\ \nu(dx,dy) =: \overline{y}$$

and observe that $(0,\overline{y}) \in \Omega$ (otherwise, by convexity, ν would be supported on $\partial\Omega$ which would contradict our assumption that $\nu \in L^\infty(\Omega)$).

We wish to prove the existence of optimal potentials for the problem

$$\inf_{\psi,\varphi,b}\int_\Omega \psi(x,y)d\nu(x,y) + \int_{[0,1]^d}\varphi(u)d\mu(u) \tag{3.5}$$

subject to the pointwise constraint that

$$\psi(x,y)+\varphi(u)\geq u\cdot y-b(u)\cdot x,\ (x,y)\in\overline{\Omega},\ u\in[0,1]^d. \tag{3.6}$$

Of course, we can take ψ that satisfies

$$\psi(x,y):=\sup_{u\in[0,1]^d}\{u\cdot y-b(u)\cdot x-\varphi(u)\}$$

so that ψ can be chosen convex and 1 Lipschitz with respect to y. In particular, we have

$$\psi(x,\overline{y})-|y-\overline{y}|\leq\psi(x,y)\leq\psi(x,\overline{y})+|y-\overline{y}|. \tag{3.7}$$

The problem being invariant by the transform $(\psi,\varphi)\to(\psi+C,\psi-C)$ (C being an arbitrary constant), we can add as a normalization the condition that

$$\psi(0,\overline{y})=0. \tag{3.8}$$

This normalization and the constraint (3.6) imply that

$$\varphi(t)\geq t\cdot\overline{y}-\psi(0,\overline{y})\geq-|\overline{y}|. \tag{3.9}$$

We note that there is one extra invariance of the problem: if one adds an affine term $q\cdot x$ to ψ, this does not change the cost and neither does it affect the constraint, provided one modifies b accordingly by subtracting the constant vector q from it. Take then q in the subdifferential of $x\mapsto\psi(x,\overline{y})$ at 0 and change ψ into $\psi-q\cdot x$, we obtain a new potential with the same properties as above and with the additional property that $\psi(.,\overline{y})$ is minimal at $x=0$, and thus $\psi(x,\overline{y})\geq 0$, together with (3.7) this gives the lower bound

$$\psi(x,y)\geq-|y-\overline{y}|\geq-C \tag{3.10}$$

where the bound comes from the boundedness of Ω (from now on, C will denote a generic constant which may change from one line to another).

Now take a minimizing sequence $(\psi_n,\varphi_n,b_n)\in C(\overline{\Omega},\mathbb{R})\times C([0,1]^d,\mathbb{R})\times C([0,1]^d,\mathbb{R}^N)$ where for each n, ψ_n has been chosen with the same properties as above. Since φ_n and ψ_n are bounded from below ($\varphi_n\geq-|\overline{y}|$ and $\psi_n\geq C$) and since the sequence is minimizing, we deduce immediately that ψ_n and φ_n are bounded sequences in L^1. Let $z=(x,y)\in\Omega$ and $r>0$ be such that the distance between z and the complement of Ω is at least $2r$, (so that $B_r(z)$ is in the set of points that are at least at distance r from $\partial\Omega$), by assumption there is an $\alpha_r>0$ such that $\nu\geq\alpha_r$ on $B_r(z)$. We then deduce from the convexity of ψ_n:

$$C\leq\psi_n(z)\leq\frac{1}{|B_r(z)|}\int_{B_r(z)}\psi_n\leq\frac{1}{|B_r(z)|\alpha_r}\int_{B_r(z)}|\psi_n|\nu\leq\frac{1}{|B_r(z)|\alpha_r}\|\psi_n\|_{L^1(\nu)}$$

so that ψ_n is actually bounded in L^∞_{loc} and by convexity, we also have

$$\|\nabla\psi_n\|_{L^\infty(B_r(z))} \leq \frac{2}{R-r}\|\psi_n\|_{L^\infty(B_R(z))}$$

whenever $R > r$ and $B_R(z) \subset \Omega$ (see for instance Lemma 5.1 in [3] for a proof of such bounds). We can thus conclude that ψ_n is also locally uniformly Lipschitz. Therefore, thanks to Ascoli's theorem, we can assume, taking a subsequence if necessary, that ψ_n converges locally uniformly to some potential ψ.

Let us now prove that b_n is bounded in L^1, for this take $r > 0$ such that $B_{2r}(0, \overline{y})$ is included in Ω. For every $x \in B_r(0)$, any $t \in [0,1]^d$ and any n we then have

$$-b_n(t) \cdot x \leq \varphi_n(t) - t \cdot \overline{y} + \|\psi_n\|_{L^\infty(B_r(0,\overline{y}))} \leq C + \varphi_n(t),$$

and maximizing in $x \in B_r(0)$ immediately gives

$$|b_n(t)|r \leq C + \varphi_n(t),$$

from which we deduce that b_n is bounded in L^1 since φ_n is.

From Komlos' theorem (see [11]), we may find a subsequence such that the Cesaro means

$$\frac{1}{n}\sum_{k=1}^{n}\varphi_k,\ \frac{1}{n}\sum_{k=1}^{n}b_k$$

converge a.e. respectively to some φ and b. Clearly ψ, φ and b satisfy the linear constraint (3.6), and since the sequence of Cesaro means $(\psi_n', \varphi_n', b_n') := n^{-1}\sum_{k=1}^n(\psi_k, \varphi_k, b_k)$ is also minimizing, we deduce from Fatous' Lemma

$$\begin{aligned}&\int_\Omega \psi(x,y)d\nu(x,y) + \int_{[0,1]^d}\varphi(u)d\mu(u)\\ &\leq \liminf_n \int_\Omega \psi_n'(x,y)d\nu(x,y) + \int_{[0,1]^d}\varphi_n'(u)d\mu(u) = \inf(3.5).\end{aligned}$$

□

3.3 Vector quantile regression as optimality conditions

Let U solve (3.1) and (ψ, φ, b) solve its dual (3.2). Recall that, without loss of generality, we can take ψ convex and given by

$$\psi(x,y) = \sup_{t\in[0,1]^d}\{t \cdot y - \varphi(t) - b(t) \cdot x\}. \tag{3.11}$$

The constraint of the dual is

$$\psi(x,y)+\varphi(t)+b(t)\cdot x \geq t\cdot y,\ \forall (x,y,t)\in \Omega\times[0,1]^d, \tag{3.12}$$

and the primal-dual relations give that, almost-surely

$$\psi(X,Y)+\varphi(U)+b(U)\cdot X = U\cdot Y. \tag{3.13}$$

Which, since ψ given by (3.11) is convex, yields

$$(-b(U),U)\in\partial\psi(X,Y),\ \text{or, equivalently}\ (X,Y)\in\partial\psi^*(-b(U),U).$$

Problems (3.1) and (3.2) have thus enabled us to find:

- U uniformly distributed with X mean-independent from U,
- $\varphi: [0,1]^d\to\mathbb{R}$, $b: [0,1]^d\to\mathbb{R}^N$ and $\psi: \Omega\to\mathbb{R}$ convex,

such that $(X,Y)\in\partial\psi^*(-b(U),U)$. Specification of vector quantile regression rather asks whether one can write $Y=\nabla\varphi(U)+Db(U)^TX := \nabla\Phi_X(U)$ with $u\mapsto\Phi_x(u) := \varphi(u)+b(u)x$ convex in u for fixed x. The smoothness of φ and b is actually related to this specification issue. Indeed, if φ and b were smooth then (by the envelope theorem) we would have

$$Y=\nabla\varphi(U)+Db(U)^TX=\nabla\Phi_X(U)$$

But the smoothness of φ and b are not enough to guarantee that the conditional quantile is affine in x, which would also require $u\mapsto\Phi_x(u)$ to be convex. Note also that if ψ was smooth, we would then have

$$U=\nabla_y\psi(X,Y),\ -b(U)=\nabla_x\psi(X,Y).$$

In general (without assuming any smoothness), define

$$\psi_x(y)=\psi(x,y).$$

We then have, thanks to (3.12)-(3.13)

$$U\in\partial\psi_X(Y)\ \text{i.e.}\ Y\in\partial\psi_X^*(U).$$

The constraint of (3.2) also gives

$$\psi_x(y)+\Phi_x(t)\geq t\cdot y$$

since Legendre Transform is order-reversing, this implies

$$\psi_x\geq\Phi_x^* \tag{3.14}$$

hence

$$\psi_x^*\leq(\Phi_x)^{**}\leq\Phi_x$$

(where Φ_x^{**} denotes the convex envelope of Φ_x). Duality between (3.1) and (3.2) thus gives:

Theorem 3.3. *Let U solve (3.1), (ψ, φ, b) solve its dual (3.2) and set $\Phi_x(t) := \varphi(t) + b(t) \cdot x$ for every $(t, x) \in [0,1]^d \times \mathrm{spt}(m)$ then*

$$\Phi_X(U) = \Phi_X^{**}(U) \text{ and } U \in \partial\Phi_X^*(Y) \text{ i.e. } Y \in \partial\Phi_X^{**}(U) \text{ a.s..} \tag{3.15}$$

Proof. From the duality relation (3.13) and (3.14), we have

$$U \cdot Y = \psi_X(Y) + \Phi_X(U) \geq \Phi_X^*(Y) + \Phi_X(U)$$

so that $U \cdot Y = \Phi_X^*(Y) + \Phi_X(U)$ and then

$$\Phi_X^{**}(U) \geq U \cdot Y - \Phi_X^*(Y) = \Phi_X(U).$$

Hence, $\Phi_X(U) = \Phi_X^{**}(U)$ and $U \cdot Y = \Phi_X^*(Y) + \Phi_X^{**}(U)$ i.e. $U \in \partial\Phi_X^*(Y)$ almost surely, and the latter is equivalent to the requirement that $Y \in \partial\Phi_X^{**}(U)$.

□

The previous theorem thus gives the following interpretation of the correlation maximization with a mean independence constraint (3.1) and its dual (3.2). These two variational problems in duality lead to the pointwise relations (3.15) which can be seen as best approximations of a specification assumption:

$$Y = \nabla\Phi_X(U), \ (X, U) \mapsto \Phi_X(U) \text{ affine in } X, \text{ convex in } U.$$

Indeed in (3.15), Φ_X is replaced by its convex envelope, the uniform random variable U solving (3.1) is shown to lie a.s. in the contact set $\Phi_X = \Phi_X^{**}$ and differentiability is replaced by a subdifferential condition.

References

[1] A. Belloni, R.L. Winkler, On Multivariate Quantiles Under Partial Orders, *The Annals of Statistics*, **39** (2), 1125-1179 (2011).

[2] Y. Brenier, Polar factorization and monotone rearrangement of vector-valued functions, *Comm. Pure Appl. Math.* 44 **4**, 375–417 (1991).

[3] G. Carlier, A. Galichon, Exponential convergence for a convexifying equation, *ESAIM, Control, Optimisation and Calculus of Variations*, **18** (3), 611-620 (2012).

[4] G. Carlier, V. Chernozhukov, A. Galichon, Vector quantile regression: an optimal transport approach, *The Annals of Statistics*, **44** (3), 1165-1192 (2016)

[5] J.A. Cuesta-Albertos, C. Matrán, Notes on the Wasserstein metric in Hilbert spaces, *The Annals of Probability*, 17(3), 1264-1276 (1989).

[6] I. Ekeland, A. Galichon, M. Henry, *Comonotonic measures of multivariate risks, Math. Finance*, **22** (1), 109-132 (2012).

[7] A. Galichon, *Optimal Transport Methods in Economics*, Princeton University Press (2016).

[8] A. Galichon, M. Henry, Dual theory of choice with multivariate risks, *Journal of Economic Theory*, **47** (4), 1501-516 (2012).

[9] M. Hallin, D. Paindaveine, M. Siman, Multivariate quantiles and multiple-output regression quantiles: From L^1 optimization to halfspace depth, *The Annals of Statistics*, **38** (2), 635-669 (2010).

[10] R. Koenker, G. Bassett, Regression Quantiles, *Econometrica*, **46**, 33-50 (1978).

[11] J. Komlos, A generalization of a problem of Steinhaus, *Acta Mathematica Academiae Scientiarum Hungaricae*, **18** (1–2), 1967, pp. 217–229.

[12] G. Puccetti, M. Scarsini, Multivariate comonotonicity, *Journal of Multivariate Analysis* 101, 291–304 (2010).

[13] L. Rüschendorf, S.T. Rachev, A characterization of random variables with minimum L^2-distance, *Journal of Multivariate Analysis*, 32(1), 48-54 (1990).

[14] L. Rüschendorf, S.T. Rachev, *Mass transportation problems. Vol. I. Theory, Vol. II. Applications*, Probability and its Applications, Springer-Verlag, New York, (1998).

[15] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Progress in Nonlinear Differential Equations and Their Applications 87, Birkhäuser Basel, (2015).

[16] C. S. SmithM. Knott, Note on the optimal transportation of distributions, *J. Optim. Theory Appl.*, 52 (2), 323–329 (1987).

[17] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, 58, American Mathematical Society, Providence, RI (2003).