Statistics and Probability Letters 100 (2015) 115–123

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

A note on quadratic transportation and divergence inequality

Ying Ding

Department of Applied Mathematics, Zhejiang University of Technology, Zhejiang, 310023, PR China

ARTICLE INFO

Article history:
Received 21 May 2014
Received in revised form 6 February 2015
Accepted 10 February 2015
Available online 18 February 2015

MSC:
60E15
60E05

Keywords:
Quadratic transportation and divergence inequality
Tensorization
Poincaré inequality

ABSTRACT

In this paper, we study the quadratic transportation and divergence inequality, which is the generalization of Talagrand's transportation inequality. We obtain its tensorization property, prove it is implied by Poincaré inequality, and give a criterion for it.

1. Introduction

Let (E, d) be a Polish space. Suppose μ and ν are two probability measures on E with finite moments of order $p \geq 1$. The Kantorovich distance of order p between μ and ν is defined as

$$W_p(\nu, \mu) = \left(\inf_{\pi} \iint d(x, y)^p d\pi(x, y) \right)^{\frac{1}{p}}.$$

Here infimum takes over all joint probability measures π on $E \times E$ with μ and ν as marginals. The relative entropy of ν with respect to μ is defined as $H(\nu\|\mu) = \begin{cases} \int_E f \ln f d\mu, & \text{if } \nu \ll \mu, \\ +\infty, & \text{else,} \end{cases}$ where $f = \frac{d\nu}{d\mu}$. We say a probability measure μ satisfies the transportation and entropy inequality (W_pH), if there exists some constant $C > 0$, such that for any probability measure ν,

$$W_p(\nu, \mu) \leq \sqrt{CH(\nu\|\mu)}.$$

This kind of inequalities plays an important role in Gaussian concentration. When $p = 1$, it is equivalent to say there exists some constant $\varepsilon > 0$ and $x_0 \in E$, such that $\int e^{\varepsilon d(x,x_0)^2} d\mu(x) < +\infty$. When $p = 2$, W_2H has the dimensional free tensorization property, which leads to the Gaussian dimensional free concentration. One may refer to Djellout et al. (2004), Gozlan (2009), Talagrand (1996) and Villani (2003), etc.

Transportation and entropy inequality are only concerned with probability measures with finite exponential moments, while $W_p(\nu, \mu)$ only needs finite pth moments. So it is natural to replace $H(\nu\|\mu)$ by some weaker quantity to enlarge the

E-mail address: yding@zjut.edu.cn.

http://dx.doi.org/10.1016/j.spl.2015.02.009
0167-7152/

extents of measures. Given a real number $\alpha > 1$. The Rényi divergence power of order α between μ and ν is defined as

$$D_\alpha(\nu\|\mu) = \begin{cases} \dfrac{1}{\alpha-1}\left(\displaystyle\int f^\alpha d\mu - 1\right), & \text{if } \nu \ll \mu, \\ +\infty, & \text{else,} \end{cases}$$

where $f = \frac{d\nu}{d\mu}$. $D_\alpha(\nu\|\mu)$ is nonnegative and equivalent to zero if and only if $\nu = \mu$. It is an increasing function on α. Especially, when α goes to 1 decreasingly, it reduces to $H(\nu\|\mu)$. So it is a nice replacement of the relative entropy. For more details, one may refer to van Erven and Harremoës (2010, 2014).

Definition 1.1. Given a probability measure μ on E. If there exists some constant $C > 0$ such that for any probability measure ν,

$$W_p(\nu, \mu) \le \sqrt{CD_\alpha(\nu\|\mu)},$$

then we say μ satisfies the transportation and divergence inequality (W_pD_α). Especially, when α goes decreasingly to 1, it reduces to W_pH.

Obviously, when α is fixed, W_1D_α is the weakest one due to Jensen's inequality. In this case, we can derive some polynomial-type concentration inequality by Marton's argument. In fact, suppose μ satisfies the transportation and divergence inequality W_1D_α with constant C. Let $\mu_A = \frac{1_A}{\mu(A)}\mu$ and $\mu_B = \frac{1_B}{\mu(B)}\mu$ be two probability measures, where A and B are two Borel sets on E. Then we have by the triangle inequality,

$$\begin{aligned} W_1(\mu_A, \mu_B) &\le W_1(\mu_A, \mu) + W_1(\mu_B, \mu) \\ &\le \sqrt{\frac{C}{\alpha-1}\left[\mu(A)^{1-\alpha} - 1\right]} + \sqrt{\frac{C}{\alpha-1}\left[\mu(B)^{1-\alpha} - 1\right]}. \end{aligned}$$

If we set $\mu(A) \ge 1/2$, and choose $B = A_r^c$, the complement of $A_r = \{x \in E : \inf_{y\in A} d(x, y) \le r\}$, we have

$$r \le \sqrt{\frac{C}{\alpha-1}(2^{\alpha-1} - 1)} + \sqrt{\frac{C}{\alpha-1}\left[\mu(A_r^c)^{1-\alpha} - 1\right]},$$

that is for any $r \ge r_0 = \sqrt{\frac{C}{\alpha-1}(2^{\alpha-1} - 1)}$,

$$\mu(A_r^c) \le \left[1 + \frac{\alpha-1}{C}(r - r_0)^2\right]^{-\frac{1}{\alpha-1}}.$$

Ding (2014) has obtained some conclusions on W_1D_α.

In this paper, we focus on the quadratic transportation and divergence inequality W_2D_α, $\alpha > 1$, the generalization of W_2H. The first purpose of this paper is the tensorization property of W_2D_α, which will leads to the polynomial concentration results for the product probability measure.

Theorem 1.1. *Let $p \in [1, 2]$ and $\alpha > 1$. Suppose μ_i are probability measures on Polish spaces (E_i, d), $i = 1, 2, \ldots, n$. If there exists some constant C such that for any ν_i on E_i,*

$$W_p(\nu_i, \mu_i) \le \sqrt{CD_\alpha(\nu_i\|\mu_i)},$$

then the product measure $\mu^{\otimes n}$ on product space $(E^{\otimes n}, \left(\sum_{i=1}^n d(x_i, y_i)^p\right)^{\frac{1}{p}})$ satisfies

$$W_p(\nu, \mu^{\otimes n}) \le \sqrt{Cn^{\frac{2}{p}} D_\alpha(\nu\|\mu^{\otimes n})}$$

for any ν on $E^{\otimes n}$. Especially, when $p = 2$, we have

$$W_2(\nu, \mu^{\otimes n})^2 \le CnD_\alpha(\nu\|\mu^{\otimes n}), \quad \forall \nu.$$

We say a probability measure μ satisfies the Poincaré inequality with constant C_P, if for any Lipschitz function $f : E \mapsto \mathbb{R}$,

$$\int f^2 d\mu - \left(\int f d\mu\right)^2 \le C_P \int |\nabla f|^2 d\mu.$$

Here $|\nabla f|(x) = \limsup_{y\to x} \frac{|f(y)-f(x)|}{d(x,y)}$. Since both the quadratic transportation and divergence inequality and the Poincaré inequality are implied by Talagrand's transportation inequality (see Bobkov et al., 2001, Cattiaux and Guillin, 2006, and Otto and Villani, 2000), it is natural to ask which one is stronger? This is the second purpose of this paper. We obtain following conclusion inspired by Cattiaux and Guillin (2006), whose results can also be recovered here.