QUANTIFIED CRAMÉR-WOLD CONTINUITY THEOREM FOR THE KANTOROVICH TRANSPORT DISTANCE

SERGEY G. BOBKOV[1,3] AND FRIEDRICH GÖTZE[2,3]

Abstract. An upper bound for the Kantorovich transport distance between probability measures on multidimensional Euclidean spaces is given in terms of transport distances between one dimensional projections. This quantifies the Cramér-Wold continuity theorem for the weak convergence of probability measures.

1. Introduction

Given a sequence of random vectors $(X_n)_{n\geq 1}$ and a random vector X with values in $\mathbb{R}^d$, the Cramér-Wold continuity theorem indicates that $X_n \Rightarrow X$ weakly in distribution, if and only if this convergences holds true for all one dimensional projections, i.e. if and only if

$$\langle X_n, \theta\rangle \Rightarrow \langle X, \theta\rangle \quad \text{as } n \to \infty$$

on the real line for any $\theta \in \mathbb{R}^d$ (cf. [5], [1]). Of large interest is the problem of how one can quantify this characterization by means of various distances responsible for the weak convergence. Indeed, this could potentially reduce a number of high dimensional questions to dimension one, perhaps under proper moment assumptions. Here we consider the problem with respect to the Kantorovich transport distance.

Let X and Y be random vectors in $\mathbb{R}^d$ with distributions μ and ν having finite first absolute moments. The Kantorovich transport distance, also called the minimal distance between μ and ν, is defined with respect to the Euclidean metric on $\mathbb{R}^d$ by

$$W(X,Y) = W(\mu,\nu) = \inf \mathbb{E}\,|X' - Y'| = \inf \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|\,d\pi(x,y). \tag{1.1}$$

Here the first infimum is taken over all pairs of random vectors X', Y' with distributions μ, ν, and the second one is running over all Borel probability measures π on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. By a well-known characterization of the convergence in W, this metric metrizes the topology of weak convergence in the space of all Borel probability measures μ on $\mathbb{R}^d$ with bounded p-th absolute moments for any fixed $p > 1$ (cf. [16], Theorem 7.12).

2010 *Mathematics Subject Classification.* Primary 60E, 60F.

Key words and phrases. Cramér-Wold continuity theorem, transport inequalities.

1) School of Mathematics, University of Minnesota, Minneapolis, MN, USA, bobkov@math.umn.edu.

2) Faculty of Mathematics, Bielefeld University, Germany, goetze@math-uni.bielefeld.de.

3) Research supported by the NSF grant DMS-2154001, the GRF – SFB 1283/2 2021 – 317210226, and the Hausdorff Research Institute for Mathematics.

According to the Kantorovich duality theorem (cf. [7]),

$$\begin{aligned} W(X,Y) &= \sup_{\|u\|_{\rm Lip} \leq 1} |\mathbb{E} u(X) - \mathbb{E} u(Y)| \\ &= \sup_{\|u\|_{\rm Lip} \leq 1} \Big| \int_{\mathbb{R}^d} u\,d\mu - \int_{\mathbb{R}^d} u\,d\nu \Big|, \end{aligned} \tag{1.2}$$

where the supremum runs over all functions $u : \mathbb{R}^d \to \mathbb{R}$ with Lipschitz semi-norm $\|u\|_{\rm Lip} \leq 1$. Since the functions of the form $u(x) = v(\langle x, \theta\rangle)$ with $|\theta| = 1$ and $v : \mathbb{R} \to \mathbb{R}$ such that $\|v\|_{\rm Lip} \leq 1$ participate in the supremum (1.2), we have

$$W(X,Y) \geq \sup_{|\theta|=1} W(X_\theta, Y_\theta) \tag{1.3}$$

for the linear functionals

$$X_\theta = \langle X, \theta\rangle\,, \ \ Y_\theta = \langle Y, \theta\rangle\,.$$

In the modern literature, the supremum in (1.3) is often called the max-scliced Wasserstein distance. Recall that in dimension one, in big contrast to the multidimensional situation, the Kantorovich distance has a simple description

$$W(X_\theta, Y_\theta) = \int_{-\infty}^\infty |F_\theta(x) - G_\theta(x)|\,dx$$

in terms of the distribution functions $F_\theta(x) = \mathbb{P}\{X_\theta \leq x\}$, $G_\theta(x) = \mathbb{P}\{Y_\theta \leq x\}$.

Here we reverse the inequality (1.3) in a somewhat similar form under a p-th moment assumption.

Theorem 1.1. *Suppose that* $(\mathbb{E}\,|X|^p)^{1/p} \leq b$ *and* $(\mathbb{E}\,|Y|^p)^{1/p} \leq b$ *for some* $p > 1$ *and* $b \geq 0$. *Then*

$$W(X,Y) \leq 15\, b^{1-\alpha} \sup_{|\theta|=1} W(X_\theta, Y_\theta)^\alpha, \tag{1.4}$$

where

$$\alpha = \frac{2}{dp^* + 2}, \quad p^* = \frac{p}{p-1}. \tag{1.5}$$

Note that the distance W is homogeneous with respect to (X,Y), and so is (1.4) when this inequality is written with an optimal value of b.

Letting $p \to \infty$ and assuming that $|X| \leq 1$ and $|Y| \leq 1$ a.s., we obtain a simpler relation

$$W(X,Y) \leq 15 \sup_{|\theta|=1} W(X_\theta, Y_\theta)^{\frac{2}{d+2}}. \tag{1.6}$$

As another interesting case, suppose that $\mathbb{E}\,|X|^2 \leq d$ and $\mathbb{E}\,|Y|^2 \leq d$ (which holds for isotropic distributions). Then, using $d^{\frac{d}{2d+2}} \leq \sqrt{d}$ so that to simplify the factor $b^{1-\alpha}$, (1.4) yields

$$W(X,Y) \leq 15\sqrt{d} \sup_{|\theta|=1} W(X_\theta, Y_\theta)^{\frac{1}{d+1}}.$$

Remark 1.2. One may wonder whether or not it is possible to replace the Kantorovich distance in Theorem 1.1 with other classical distances such as, for example, the Kolmogorov distance

$$\rho_d(X,Y) = \sup |\mathbb{P}\{X \in H\} - \mathbb{P}\{Y \in H\}|,$$

where the supremum is running over all half-spaces H in $\mathbb{R}^d$. In connection with the Cramér-Wold theorem, the computer tomography problems, and continuity properties of the Radon transform (which assignes to every probability distribution on $\mathbb{R}^d$ the collection of all its one-dimensional projections), this question and related inversion issues were discussed in 1990's in a series of works by Zinger, Klebanov and Khalfin, cf. e.g. [9]. However, as was demonstrated by Zaitsev [17], the smallness of $\sup_{|\theta|=1} \rho_1(X_\theta, Y_\theta)$ does not guarantee that $\rho_d(X, Y)$ will be small as well, even if the distributions of X and Y are compactly supported. Moreover, here ρ_1 may be strengthened to the total variation distance. The corresponding counter-example shows that ρ_d is essentially stronger than the Kantorovich metric W.

2. **Transport Distances and Convergence of Empirical Measures**

It is not clear whether or not the exponent $\alpha = \alpha(p, d)$ defined in (1.5) is optimal in the inequality (1.4), even if one can add an additional (p, d)-dependent multiplicative factor. In order to illustrate the strength of the inequality (1.6), we consider the following example involving empirical measures.

Let $X_1, \dots, X_n$ be a sample of size n drawn from μ, that is, independent random vectors in $\mathbb{R}^d$ with distribution μ. They may be treated as independent copies of a random vector X. The associated empirical measures are defined by

$$\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}, \tag{2.1}$$

where δ_x denotes a delta-measure at the point x. Correspondingly, their linear projections represent one dimensional empirical measures

$$\mu_{n,\theta} = \frac{1}{n} \sum_{k=1}^n \delta_{\langle X_k, \theta\rangle}. \tag{2.2}$$

For simplicity, let us restrict ourselves to compactly supported distributions μ supported on the unit ball B_1 in $\mathbb{R}^d$. By concavity of $x \to x^\alpha$ in $x \geq 0$ for $\alpha \in (0, 1]$, it follows from (1.6) that

$$\mathbb{E}\, W(\mu_n, \mu) \leq 15 \left[\mathbb{E} \sup_{|\theta|=1} W(\mu_{n,\theta}, \mu_\theta)\right]^\alpha \tag{2.3}$$

with $\alpha = \frac{2}{d+2}$, In the case $d \geq 3$, it is known (cf. e.g. [6], [3]) that $\mathbb{E}\, W(\mu_n, \mu)$ is of order at most $c_d n^{-1/d}$, and this rate cannot be improved as $n \to \infty$ for the uniform distribution. However, if $d = 1$ and μ is compactly supported, the rate is of the standard order $\mathbb{E}\, W(\mu_n, \mu) \sim \frac{1}{\sqrt{n}}$ up to μ-dependent factors (for a general two-sided bound we refer to [2], Theorem 3.5). Hence, we also have $\mathbb{E}\, W(\mu_{n,\theta}, \mu_\theta) \sim \frac{1}{\sqrt{n}}$ for every fixed θ. In order to determine the upper bound for the transport distance via the bound (2.3), the following result can be used.

Theorem 2.1. *If μ is supported on the unit ball B_1 in $\mathbb{R}^d$, then*

$$\mathbb{E} \sup_{|\theta|=1} W(\mu_{n,\theta}, \mu_\theta) \leq \frac{c_d}{\sqrt{n}} \tag{2.4}$$

with some constants $c_d > 0$ depending on d only.

Applying this bound in (2.3) and using a lower bound with rate $c_d n^{-1/d}$ for the left-hand side with large n, we may conclude that necessarily $\alpha \le 2/d$. Thus, the exponent $\alpha = \frac{2}{d+2}$ is asymptotically optimal for growing dimension d.

For reader's convenience, at the end of the note we include a simple chaining argument leading to (2.4) with $c_d = c\sqrt{d}$, where $c > 0$ is an absolute constant. However, this statement is not new – inequalities for one-dimensional projections of empirical measures such as (2.4) have been the subject of many recent investigations, cf. e.g. [13], [12], [11]. After this paper was submitted, we also learned about the preprint by Boedihardjo [4], where the upper bound (2.4) was derived with a constant independent of d.

3. Reduction to Compactly Supported Lipschitz Functions

We need some preparation for the proof of Theorem 1.1. The argument is based on truncation, smoothing, the Plancherel theorem, together with the Kantorovich duality theorem (1.2).

Let U_r ($r > 0$) denote the collection of all functions $u : \mathbb{R}^d \to \mathbb{R}$ with $\|u\|_{\rm Lip} \le 1$, $u(0) = 0$, which are supported on the Euclidean ball $B_r = \{x \in \mathbb{R}^d : |x| \le r\}$. As a first step, we show that the supremum in (1.2) may be restricted to the set U_r at the expense of a small error for large values of the parameter r under a p-th moment assumption. Define

$$W^{(r)}(X,Y) = \sup_{u \in U_r} |\mathbb{E}u(X) - \mathbb{E}u(Y)| = \sup_{u \in U_r} \Big| \int_{\mathbb{R}^d} u\,d\mu - \int_{\mathbb{R}^d} u\,d\nu \Big|,$$

assuming that the random vectors X and Y have distributions μ and ν.

Lemma 3.1. *Let $(\mathbb{E}\,|X|^p)^{1/p} \le b$ and $(\mathbb{E}\,|Y|^p)^{1/p} \le b$ for some $p > 1$. For any $r > 0$,*

$$W(X,Y) \le 3\,W^{(r)}(X,Y) + 4b\,\Big(\frac{2b}{r}\Big)^{p-1}.$$

Proof. Let u be a Lipschitz function participating in the supremum (1.2) with $u(0) = 0$ (without loss of generality). The latter ensures that $|u(x)| \le |x|$ for all $x \in \mathbb{R}^d$.

Define the function $u_r = u\psi_r$, where using the notation $a^+ = \max(a,0)$,

$$\psi_r(x) = \Big(1 - \frac{2}{r}\,{\rm dist}(B_{r/2}, x)\Big)^+, \quad x \in \mathbb{R}^d.$$

Clearly, $0 \le \psi_r \le 1$ and $\|\psi_r\|_{\rm Lip} \le \frac{2}{r}$, as the distance function

$$x \to {\rm dist}(B,x) = \inf\{|x-y| : y \in B\}$$

has a Lipschitz semi-norm at most one for any non-empty set B in $\mathbb{R}^d$.

By the definition, $|u_r(x)| \le |u(x)|$ for all $x \in \mathbb{R}^d$, and

$$u_r(x) = u(x) \ \text{ for } |x| \le r/2, \qquad u_r(x) = 0 \ \text{ for } |x| \ge r. \tag{3.1}$$

Writing

$$u_r(x) - u_r(y) = (u(x) - u(y))\,\psi_r(x) + u(y)\,(\psi_r(x) - \psi_r(y)),$$

it follows that, for all $x \in \mathbb{R}^d$ and $y \in B_r$,

$$|u_r(x) - u_r(y)| \le |u(x) - u(y)| + |y|\,\frac{2}{r}\,|x-y| \le 3\,|x-y|.$$

A similar final inequality holds true for $x \in B_r$ and $y \in \mathbb{R}^d$. In addition, $u_r(x) - u_r(y) = 0$ in the case $x, y \notin B_r$. Therefore, $\|u_r\|_{\rm Lip} \le 3$.

Next, by (3.1),

$$\begin{aligned}\Big|\int_{\mathbb{R}^d} u\,d\mu - \int_{\mathbb{R}^d} u_r\,d\mu\Big| &\le \int_{|x|>r/2} |u(x)|\,d\mu(x) + \int_{|x|>r/2} |u_r(x)|\,d\mu(x) \\ &\le 2\int_{|x|>r/2} |x|\,d\mu(x) \\ &= 2\,\mathbb{E}\,|X|\,1_{\{|X|>r/2\}} \,\le\, 2\,\frac{\mathbb{E}\,|X|^p}{(r/2)^{p-1}} \,\le\, 2b^p\Big(\frac{2}{r}\Big)^{p-1}.\end{aligned}$$

With a similar inequality for the measure ν, we obtain that

$$\Big|\int_{\mathbb{R}^d} u\,d(\mu-\nu) - \int_{\mathbb{R}^d} u_r\,d(\mu-\nu)\Big| \,\le\, 4b^p\Big(\frac{2}{r}\Big)^{p-1},$$

implying

$$\Big|\int_{\mathbb{R}^d} u\,d(\mu-\nu)\Big| \le \Big|\int_{\mathbb{R}^d} u_r\,d(\mu-\nu)\Big| + 4b^p\Big(\frac{2}{r}\Big)^{p-1}.$$

Since $\|u_r\|_{\rm Lip} \le 3$, we have $\frac{1}{3}\,u_r \in U_r$, so that the last integral does not exceed $3\,W^{(r)}(X,Y)$ in absolute value. Thus,

$$\Big|\int_{\mathbb{R}^d} u\,d(\mu-\nu)\Big| \le 3\,W^{(r)}(X,Y) + 4b^p\Big(\frac{2}{r}\Big)^{p-1}.$$

It remains to take the supremum on the left-hand side over all functions $u:\mathbb{R}^d \rightarrow \mathbb{R}$ such that $\|u\|_{\rm Lip} \le 1$ and $u(0) = 0$. $\square$

4. **Fourier Transforms**

Any integrable compactly supported function u on $\mathbb{R}^d$ has a well-defined Fourier transform

$$\widehat{u}(t) = \int_{\mathbb{R}^d} e^{i\langle t,x\rangle} u(x)\,dx, \quad t \in \mathbb{R}^d, \tag{4.1}$$

which represents a C^∞-smooth function. Towards the proof of Theorem 1.1 let us state now the following integrability property.

Lemma 4.1. *For any function $u:\mathbb{R}^d \rightarrow \mathbb{R}$ which is supported on the ball B_r and has a Lipschitz semi-norm $\|u\|_{\rm Lip} \le 1$,*

$$\int_{\mathbb{R}^d} |\widehat{u}(t)|^2\,|t|^2\,dt \,\le\, \omega_d\,(2\pi r)^d. \tag{4.2}$$

In particuar, this inequality holds true for any function u in U_r. Here and elsewhere ω_d stands for the d-dimensional volume of the unit ball B_1.

Proof. First assume that $\widehat{u}(t)$ decays sufficiently fast at infinity, namely $\widehat{u}(t) = O(1/|t|^p)$ as $|t| \rightarrow \infty$ for any $p > 0$. Then (4.1) may be inverted in the form of the Fourier transform

$$u(x) = \frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} e^{-i\langle t,x\rangle}\,\widehat{u}(t)\,dt.$$

In particular, u is C^∞-smooth on $\mathbb{R}^d$. Moreover, this equality may be differentiated along every coordinate x_k to represent the corresponding partial derivatives as

$$\partial_{x_k} u(x) = -\frac{i}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i\langle t,x\rangle}\, t_k\, \widehat{u}(t)\, dt, \quad k = 1, \dots, d.$$

Hence, by the Plancherel theorem,

$$\int_{\mathbb{R}^d} t_k^2\, |\widehat{u}(t)|^2\, dt = (2\pi)^d \int_{\mathbb{R}^d} (\partial_{x_k} u(x))^2\, dx.$$

Summing over all $k \le d$ and using $|\nabla u(x)| \le 1$ for $x \in B_r$ and $\nabla u(x) = 0$ for $|x| > r$, we get

$$\int_{\mathbb{R}^d} |t|^2\, |\widehat{u}(t)|^2\, dt = (2\pi)^d \int_{\mathbb{R}^d} |\nabla u(x)|^2\, dx \le (2\pi)^d \cdot \omega_d r^d,$$

which is the desired inequality (4.2).

In the general case, a smoothing argument can be used. By the well-known theorem of Ingham [8], there exists a probability density w on $\mathbb{R}^d$ which is supported on the unit ball B_1 and has characteristic function $\widehat{w}(t)$ satisfying $\widehat{w}(t) = O(1/|t|^p)$ as $|t| \to \infty$, for any fixed $p > 0$. Given $\varepsilon > 0$, the probability density $w_\varepsilon(x) = \varepsilon^{-d} w(x/\varepsilon)$ is supported on the ball B_ε and has characteristic function $\widehat{w}_\varepsilon(t) = \widehat{w}(\varepsilon t)$. Consider the convolution

$$u_\varepsilon(x) = (u * w_\varepsilon)(x) = \int_{\mathbb{R}^d} u(x-y)\, w_\varepsilon(y)\, dy, \quad x \in \mathbb{R}^d.$$

This function is supported on $B_{r+\varepsilon}$ and has a Lipschitz semi-norm $\|u_\varepsilon\|_{\rm Lip} \le \|u\|_{\rm Lip} \le 1$. By the Lipschitz property, $|u(x)| \le |u(0)| + r$ for any $x \in B_r$, implying that

$$\sup_{t \in \mathbb{R}^d} |\widehat{u}(t)| \le \int_{B_r} |u(x)|\, dx \le (|u(0)| + r)\, \omega_d r^d < \infty.$$

Hence, the Fourier transform of u_ε satisfies $\widehat{u}_\varepsilon(t) = \widehat{u}(t)\widehat{w}(\varepsilon t) = O(1/|t|^p)$ as $|t| \to \infty$. Thus, one may apply the previous step to the function u_ε which gives

$$\int_{\mathbb{R}^d} |\widehat{u}(t)|^2\, |\widehat{w}(\varepsilon t)|^2\, |t|^2\, dt \ \le\ \omega_d\, (2\pi\, (r+\varepsilon))^d.$$

It remains to send $\varepsilon \to 0$ in this inequality and apply Fatou's lemma together with $\widehat{w}(\varepsilon t) \to 1$ as $\varepsilon \to 0$. $\square$

Next, let us connect the Kantorovich distance with the multivariate characteristic functions

$$f(t) = \mathbb{E}\, e^{i\langle t,X\rangle}, \quad g(t) = \mathbb{E}\, e^{i\langle t,Y\rangle} \qquad (t \in \mathbb{R}^d).$$

Lemma 4.2. *Given random vectors X, Y in $\mathbb{R}^d$ with characteristic functions f, g and finite first absolute moments, we have, for any $t \in \mathbb{R}^d$,*

$$|f(t) - g(t)| \le |t| \sup_{|\theta|=1} W(X_\theta, Y_\theta).$$

Proof. In dimension one, using the property that the function $u_t(x) = \frac{1}{t}\, e^{itx}$ with parameter $t \ne 0$, has a Lipschitz semi-norm at most 1, it follows from (1.2) that

$$|f(t) - g(t)| \le |t|\, W(X,Y).$$

In dimension d, just note that, for any $\theta \in \mathbb{R}^d$, the functions $r \to f(r\theta)$ and $r \to g(r\theta)$ represent the characteristic functions of X_θ of Y_θ. $\square$

5. **Proof of Theorem 1.1**

With $b = \max(\|X\|_p, \|Y\|_p)$, where

$$\|X\|_p = (\mathbb{E}\,|X|^p)^{1/p}, \quad \|Y\|_p = (\mathbb{E}\,|Y|^p)^{1/p},$$

the inequality (1.4) is homogeneous with respect to (X,Y), so one may assume that $b=1$. As a consequence,

$$W(X,Y) \le \mathbb{E}\,|X| + \mathbb{E}\,|Y| \le 2.$$

Let η be a random vector with uniform distribution in the ball B_1, that is, with density $w(x) = \frac{1}{\omega_d}\,1_{B_1}(x)$, and let $h(t)$ denote its characteristic function. Consider the random vectors

$$X(\varepsilon) = X + \varepsilon\eta, \quad Y(\varepsilon) = Y + \varepsilon\eta \quad (\varepsilon > 0),$$

assuming that η is independent of X and Y. Then, by the definition (1.1), or by (1.2),

$$W(X,Y) \le W(X(\varepsilon), Y(\varepsilon)) + 2\varepsilon. \tag{5.1}$$

Indeed, given a function u on $\mathbb{R}^d$ with $\|u\|_{\rm Lip} \le 1$, we have

$$\begin{aligned}\mathbb{E}\,u(X(\varepsilon)) - \mathbb{E}\,u(Y(\varepsilon)) &= \mathbb{E}\,u(X+\varepsilon\eta) - \mathbb{E}\,u(Y+\varepsilon\eta) \\ &\ge \mathbb{E}\,\big(u(X) - \varepsilon|\eta|\big) - \mathbb{E}\,\big(u(Y) + \varepsilon|\eta|\big) \\ &\ge \mathbb{E}\,u(X) - \mathbb{E}\,u(Y) - 2\varepsilon.\end{aligned}$$

Taking the supremum of both sides over all Lipschitz u and applying (1.2), we arrive at (5.1).

On the other hand, using

$$\|X(\varepsilon)\|_p \le 1+\varepsilon, \quad \|Y(\varepsilon)\|_p \le 1+\varepsilon,$$

one may apply Lemma 3.1, which gives that, for any $r>0$,

$$W(X(\varepsilon), Y(\varepsilon)) \le 3\,W^{(r)}(X(\varepsilon), Y(\varepsilon)) + 4(1+\varepsilon)\,\Big(\frac{2(1+\varepsilon)}{r}\Big)^{p-1}.$$

Therefore, by (5.1),

$$W(X,Y) \le 3\,W^{(r)}(X(\varepsilon), Y(\varepsilon)) + 4(1+\varepsilon)\,\Big(\frac{2(1+\varepsilon)}{r}\Big)^{p-1} + 2\varepsilon. \tag{5.2}$$

In order to estimate the first term on the right-hand side, first note that the distributions of $X(\varepsilon)$ and $Y(\varepsilon)$ represent convolutions of the distributions of X and Y with a uniform distribution on the ball B_ε. Hence, these random vectors have densities which we denote by p_ε and q_ε respectively. They have respective characteristic functions

$$f_\varepsilon(t) = f(t)h(\varepsilon t), \quad g_\varepsilon(t) = g(t)h(\varepsilon t) \qquad (t \in \mathbb{R}^d),$$

where f and g denote the characteristic functions of X and Y. As the function $h(t)$ is square integrable, while $|f(t)| \le 1$ and $|g(t)| \le 1$, the functions f_ε and g_ε are square integrable, so that the densities p_ε and q_ε are square integrable as well, by the Plancherel theorem. Thus, given a function u in U_r, one may write

$$\mathbb{E}\,u(X(\varepsilon)) - \mathbb{E}\,u(Y(\varepsilon)) = \int_{\mathbb{R}^d} u(x)\,(p_\varepsilon(x) - q_\varepsilon(x))\,dx.$$

We are in position to apply the Plancherel theorem once more and rewrite the last integral as

$$\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d} \widehat{u}(t)\,(\bar f(t) - \bar g(t))\,h(\varepsilon t)\,dt.$$

Thanks to Lemma 4.2 we then have

$$|\mathbb{E}\, u(X(\varepsilon)) - \mathbb{E}\, u(Y(\varepsilon))| \leq \frac{M}{(2\pi)^d} \int_{\mathbb{R}^d} |\widehat{u}(t)|\, |t|\, |h(\varepsilon t)|\, dt, \tag{5.3}$$

where

$$M = \sup_{|\theta|=1} W(X_\theta, Y_\theta).$$

Moreover, using

$$\begin{aligned} \int_{\mathbb{R}^d} |h(\varepsilon t)|^2\, dt \ &= \ \varepsilon^{-d} \int_{\mathbb{R}^d} |h(t)|^2\, dt \\ &= \ (2\pi)^d\, \varepsilon^{-d} \int_{\mathbb{R}^d} w(x)^2\, dx \ = \ (2\pi)^d\, \varepsilon^{-d} \omega_d^{-1} \end{aligned}$$

and applying Cauchy's inequality in (5.3) together with Lemma 4.1, we obtain that

$$|\mathbb{E}\, u(X(\varepsilon)) - \mathbb{E}\, u(Y(\varepsilon))| \leq M \Big(\frac{r}{\varepsilon}\Big)^{\frac{d}{2}}.$$

Taking the supremum over all $u \in U(r)$ on the left-hand side leads to the similar bound for $W^{(r)}(X(\varepsilon), Y(\varepsilon))$, and using this in (5.2) we are led to

$$W(X,Y) \ \leq \ 3M \left(\frac{r}{\varepsilon}\right)^{\frac{d}{2}} + 4(1+\varepsilon) \left(\frac{2(1+\varepsilon)}{r}\right)^{p-1} + 2\varepsilon.$$

To simplify optimization over free variables $r > 0$ and $\varepsilon > 0$, let us assume that $2(1+\varepsilon) \leq c$ for a constant $c > 2$ (to be chosen later on). Then we have

$$W(X,Y) \ \leq \ 3M \left(\frac{r}{\varepsilon}\right)^{\frac{d}{2}} + 2c \left(\frac{c}{r}\right)^{p-1} + 2\varepsilon.$$

Let us then replace r with cs and ε with $c\delta$ in the above inequality to get

$$W(X,Y) \ \leq \ 3M \left(\frac{s}{\delta}\right)^{\frac{d}{2}} + 2c \left(\frac{1}{s}\right)^{p-1} + 2c\delta.$$

Here, equalizing the terms $M\,(\frac{s}{\delta})^{\frac{d}{2}}$ and $s^{-(p-1)}$, we find the unique value of s for which the above yields

$$W(X,Y) \ \leq \ (3+2c)\, A\delta^{-\beta} + 2c\delta, \tag{5.4}$$

where

$$\beta = \frac{(p-1)\,\frac{d}{2}}{p-1+\frac{d}{2}}, \quad A = M^{\frac{p-1}{p-1+\frac{d}{2}}}.$$

The choice $\delta = A^{\frac{1}{\beta+1}}$ in (5.4) leads to

$$W(X,Y) \ \leq \ (3+4c)\, A^{\frac{1}{\beta+1}},$$

provided that $2(1+c\delta) \leq c$. If we require that $\delta \leq \frac{1}{6}$, the latter condition is satisfied for $c = 3$, and we obtain that

$$W(X,Y) \ \leq \ 15\, A^{\frac{1}{\beta+1}}, \tag{5.5}$$

provided that $A^{\frac{1}{\beta+1}} \leq \frac{1}{6}$. In the other case, the right-hand side in (5.5) is greater than 2, so this inequality is fulfilled automatically due to the property $W(X,Y) \leq 2$.

It remains to note that

$$A^{\frac{1}{\beta+1}} = M^{\frac{2}{dp^*+2}}, \quad p^* = \frac{p}{p-1}.$$

□

6. **Proof of Theorem 2.1**

Let U denote for the space of all functions $u : [-1,1] \to \mathbb{R}$ with $\|u\|_{\rm Lip} \leq 1$, such that $u(0) = 0$. We equip U with the uniform distance

$$\|u - v\|_\infty = \max\{|u(x) - v(x)| : |x|, |y| \leq 1\},$$

which turns this set into a compact space by (Arzelá-Ascoli theorem).

The empirical measures μ_n and $\mu_{n,\theta}$ defined in (2.1)-(2.2) are random with mean μ so that

$$\int_{\mathbb{R}} u\, d\mu_{n,\theta} = \frac{1}{n}\sum_{k=1}^n u(\langle X_k, \theta\rangle), \quad \int_{\mathbb{R}} u\, d\mu_\theta = \mathbb{E}\, u(\langle X, \theta\rangle).$$

According to the Kantorovich duality theorem (1.2),

$$\sup_{|\theta|=1} W(\mu_{n,\theta}, \mu_\theta) \ = \ \frac{1}{n} \sup_{\theta \in S^{d-1}} \sup_{u \in U} \bigg|\sum_{k=1}^n \big(u(\langle X_k, \theta\rangle) - \mathbb{E}\, u(\langle X_k, \theta\rangle)\big)\bigg|, \tag{6.1}$$

where $S^{d-1} = \{\theta \in \mathbb{R}^d : |\theta| = 1\}$ denotes the unit sphere in $\mathbb{R}^d$.

In order to bound the expectation of the right-hand side in (6.1), one may use chaining arguments, in particular, a well-known theorem by Dudley which says the following (for a proof, let us refer to [14], [15]). Given a random variable ξ, define its Orlicz ψ_2-norm

$$\|\xi\|_{\psi_2} = \inf\big\{\lambda > 0 : \mathbb{E}\, e^{\xi^2/\lambda^2} \leq 2\big\}.$$

Suppose that $\xi(t)$ is a mean zero random process defined on some compact metric space (T, ρ), which satisfies the Lipschitz property

$$\|\xi(t) - \xi(s)\|_{\psi_2} \leq \Lambda\rho(t,s), \quad t, s \in T, \tag{6.2}$$

with some $\Lambda > 0$. Then with some absolute constant K we have

$$\mathbb{E} \sup_{t \in T} \xi(t) \ \leq \ K\Lambda \int_0^D \sqrt{\log N(\varepsilon)}\, d\varepsilon, \tag{6.3}$$

where $D = \max\{\rho(t,s) : t, s \in T\}$ is the diameter and $N(\varepsilon) = N(T, \rho, \varepsilon)$ is the minimal number of closed balls in T of radius ε needed to cover the space (recall that $\log N(\varepsilon)$ is called the ε-entropy of (T, ρ)).

In view of (6.1), it is natural to consider the random process

$$\xi(t) = \xi(u,\theta) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \big(u(\langle X_k, \theta\rangle) - \mathbb{E}\, u(\langle X_k, \theta\rangle)\big), \quad t = (u, \theta) \in T = U \times S^{d-1}.$$

We equip T with the metric

$$\rho(t,s) = \|u - v\|_\infty + |\theta - \theta'|, \quad t = (u,\theta), s = (v, \theta') \in T.$$

Endowed with this metric T will be a compact space of diameter $D = 4$.

Now, given two points $t = (u, \theta)$, $s = (v, \theta')$ in T, one may write

$$\xi(t) - \xi(s) = \frac{1}{\sqrt{n}} \sum_{k=1}^n (\eta_k - \mathbb{E}\eta_k),$$

where

$$\eta_k = u(\langle X_k, \theta\rangle) - v(\langle X_k, \theta'\rangle).$$

These random variables are independent and bounded. Indeed, writing

$$\eta_k = \big(u(\langle X_k, \theta\rangle) - v(\langle X_k, \theta\rangle)\big) + \big(v(\langle X_k, \theta\rangle) - v(\langle X_k, \theta'\rangle)\big)$$

and using the Lipschitz property of v together with the assumption $|X_k| \leq 1$ a.s., we have

$$|\eta_k| \leq \rho(t,s) \text{ a.s.}$$

Recall that, by the well-known Hoeffding's lemma, for any random variable η such that $|\eta| \leq r$ a.s.,

$$\mathbb{E}\, e^{z(\eta - \mathbb{E}\eta)} \leq e^{r^2 z^2/2} \quad \text{for all } z \in \mathbb{R}.$$

Hence, this holds for all $\eta = \eta_k$ with $r = \rho(t,s)$, and thus

$$\mathbb{E}\, e^{z(\xi(t)-\xi(s))} \leq e^{\rho^2 z^2/2}, \quad \rho = \rho(t,s).$$

Integrating this inequality with respect to z over the Gaussian measure with mean zero and variance σ^2 $(0 < \sigma < 1/\rho)$, we get

$$\mathbb{E}\, e^{\sigma^2(\xi(t)-\xi(s))^2/2} \leq \frac{1}{\sqrt{1-(\sigma\rho)^2}}$$

which implies

$$\|\xi(t) - \xi(s)\|_{\psi_2} \leq \rho(t,s)\sqrt{8/3}.$$

Thus, the Lipschitz condition (6.2) is fulfilled with an absolute constant Λ, and we may apply the Dudley's bound (6.3). By the definition of the metric ρ in T, any closed ball of radius 2ε in this space contains the product of a closed ball in U and a closed ball in S^{d-1}, both of radius ε. Hence, the corresponding ε-entropies are connected by the relation

$$N(T, \rho, 2\varepsilon) \leq N(U, \varepsilon)\, N(S^{d-1}, \varepsilon). \tag{6.4}$$

It is well-known (cf. [10]) that

$$\log N(U, \varepsilon) \leq \frac{c}{\varepsilon}, \quad N(S^{d-1}, \varepsilon) \leq \Big(\frac{c}{\varepsilon}\Big)^d, \quad 0 < \varepsilon \leq 2,$$

with some absolute constant $c > 0$. Using these bounds in (6.4) and then in (6.3), we conclude that the expectation of both sides in (6.1) does not exceed a multiple of $\sqrt{d/n}$. This gives the desired relation (2.4). □

Acknowledgement. We would like to thank the referee for a careful reading and comments. We also thank Andrey Zaitsev and March Boedihardjo for providing us missing references in the original arxiv version of this paper.

References

[1] Billingsley, P. Probability and Measure. Third ed., John Wiley & Sons. 1995.

[2] Bobkov, S. G.; Ledoux, M. One-dimensional empirical measures, order statistics and Kantorovich transport distances. Memoirs of the AMS, 261 (2019), no. 1259, 126 pp.

[3] Bobkov, S. G.; Ledoux, M. A simple Fourier analytic proof of the AKT optimal matching theorem. Ann. Appl. Probab. 31 (2021), no. 6, 2567–2584.

[4] Boedihardjo, M. T. Sharp bounds for max-scliced Wasserstein distances. arXiv:2403.00666v5 [math,PR], 3 April 2024.

[5] Cramér, H; Wold, H. Some Theorems on Distribution Functions. J. London Math. Soc. 11 (4) (1936), 290–294.

[6] Dudley, R. M. The speed of mean Glivenko–Cantelli convergence. Ann. Math. Stat. 40 (1968), 40–50.

[7] Dudley, R. M. Real analysis and probability. Cambridge Stud. Adv. Math., 74, Cambridge University Press, Cambridge, 2002, x+555 pp.

[8] Ingham, A. E. A Note on Fourier transforms. J. London Math. Soc. 9 (1934), no. 1, 29–32.

[9] Khalfin, L. A.; Klebanov, L. B. A solution of the computer tomography paradox and estimating the distances between the densities of measures with the same marginals. Ann. Probab. 22 (1994), no. 4, 2235–2241.

[10] Kolmogorov, A. N.; Tikhomirov, V. M. ε-entropy and ε-capacity of sets in function spaces. (Russian) Uspekhi Matem. Nauk 14 (1959), no. 2 (86), 3–86.

[11] Manole, T.; Balakrishnan, S.; Wasserman, L. Minimax confidence intervals for the sliced Wasserstein distance. Electronic Journal of Statistics 16 (2022), no. 1, 2252–2345.

[12] Nietert, S.; Goldfeld, Z.; Sadhu, R; Kato, K. Statistical, robustness, and computational guarantees for sliced Wasserstein distances. Advances in Neural Information Processing Systems, 35 (2022), 28179–28193.

[13] Niles-Weed, J.; Rigollet, P. Estimation of Wasserstein distances in the spiked transport model. Bernoulli 28 (2022), 2663–2688.

[14] Talagrand, M. Majorizing measures: the generic chaining. Ann. Probab. 24 (1996), no. 3, 1049–1103.

[15] Talagrand, M. Upper and lower bounds for stochastic processes. Modern methods and classical problems. 60. Springer, Heidelberg, 2014. xvi+626 pp.

[16] Villani, C. Topics in optimal transportation. Grad. Stud. Math., 58 American Mathematical Society, Providence, RI, 2003, xvi+370 pp.

[17] Zaĭtsev, A. Yu. Instability of the inversion of the Radon transform. Zapiski Nauchn. Semin. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 216 (1994), 76–85, 163. J. Math. Sci. (New York) 88 (1998), no. 1, 53–58.