Journal of Fourier Analysis and Applications (2020) 26:60
https://doi.org/10.1007/s00041-020-09766-2

Transport Inequalities on Euclidean Spaces for Non-Euclidean Metrics

Sergey G. Bobkov[1] · **Michel Ledoux**[2,3]

Received: 8 December 2019 / Revised: 26 April 2020 / Published online: 6 July 2020

Abstract
We explore upper bounds on Kantorovich transport distances between probability measures on the Euclidean spaces in terms of their Fourier-Stieltjes transforms, with focus on non-Euclidean metrics. The results are illustrated on empirical measures in the optimal matching problem on the real line.

Keywords Transport distances · Fourier analytic inequalities · non-Euclidean metrics

Mathematics Subject Classification Primary 60E · 60F

1 Introduction

The Kantorovich transport distance between two (Borel) probability measures μ and ν on a separable metric space (E, ρ) is defined as

$$\mathrm{W}(\mu, \nu) = \inf_{\lambda} \iint \rho(x, y)\, d\lambda(x, y), \tag{1.1}$$

Communicated by Massimo Fornasier.

Research was partially supported by the Simons Foundation and NSF Grant DMS-1855575.

✉ Sergey G. Bobkov
bobkov@math.umn.edu

Michel Ledoux
ledoux@math.univ-toulouse.fr

1 School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA

2 Institut de Mathématiques de Toulouse, Université de Toulouse – Paul-Sabatier, 31062 Toulouse, France

3 Institut Universitaire de France, Paris, France

where the infimum is running over all measures λ on the product space $E \times E$ with marginals μ and ν. It describes the minimal cost needed to pay in order to transport one measure to the other one, given that it costs $\rho(x, y)$ to move a particle x to a particle y.

Under mild moment assumptions, W may be used to metrize the topology of weak convergence in the space of probability distributions on E. This metric and related functionals also appear in a natural way in many mathematical areas and concrete problems. It is therefore not surprising that the literature on the optimal transport is rather rich and intensive, reflecting various models and focusing on specific families of probability distributions (see for example [3,14,21] which provide numerous references on the subject). Nevertheless, often it is not easy to compute or even to estimate the distance $\mathrm{W}(\mu, \nu)$. One exceptional case is the real line $E = \mathbb{R}$ with the canonical Euclidean metric $\rho(x, y) = |x - y|$, when (1.1) is reduced to the well-known formula

$$\mathrm{W}(\mu, \nu) = \int_{-\infty}^{\infty} |F_\mu(x) - F_\nu(x)|\, dx \tag{1.2}$$

in terms of the distribution functions F_μ and F_ν associated to μ and ν. The case of the Euclidean space $E = \mathbb{R}^d$ of dimension $d \geq 2$ turns out already to be quite non-trivial.

In analogy with the Esseen's Fourier analytic inequality which serves as a traditional approach to the central limit theorem with respect to the Kolmogorov distance (cf. e.g. [4]), one general upper bound on W has been recently considered in [7] for the class of compactly supported measures on $\mathbb{R}^d$ in terms of their Fourier-Stieltjes transforms (characteristic functions). To describe this bound, in the sequel we denote by $|x| = \sqrt{x \cdot x}$ the Euclidean norm and by $x \cdot y = x_1 y_1 + \cdots + x_d y_d$ the inner product of the vectors $x = (x_1, \dots, x_d)$, $y = (y_1, \dots, y_d) \in \mathbb{R}^d$; m is used to denote a vector with integer components, that is, an element of the lattice $\mathbb{Z}^d$.

Note that any probability measure μ which is supported on the cube (torus) $Q^d = (-\pi, \pi]^d$ is uniquely determined by the multi-indexed sequence

$$f_\mu(m) = \int_{Q^d} e^{im \cdot x}\, d\mu(x), \qquad m \in \mathbb{Z}^d.$$

Assuming that μ and ν are supported on $[0, \pi]^d$, it was shown in [7] that, for any $t > 0$,

$$\mathrm{W}(\mu, \nu) \leq \left(\sum_{m \neq 0} \frac{1}{|m|^2}\, e^{-t|m|^2}\, \big|f_\mu(m) - f_\nu(m)\big|^2 \right)^{1/2} + 2\sqrt{dt}. \tag{1.3}$$

In further applications, the right-hand side of (1.2) should be optimized over the parameter t (or, just a suitable choice of t may work).

In this paper we extend this bound to non-Euclidean metrics of the form

$$\rho(x, y) = \omega(|x - y|), \qquad x, y \in \mathbb{R}^d.$$

This equality does define a metric, once $\omega : [0, \infty) \to \mathbb{R}$ is a modulus of continuity, i.e., a non-decreasing, continuous, subadditive function such that $\omega(0) = 0$ and $\omega(\delta) > 0$ for $\delta > 0$. The subadditivity refers to the property $\omega(\delta_1 + \delta_2) \le \omega(\delta_1) + \omega(\delta_2)$ $(\delta_1, \delta_2 \ge 0)$, ensuring the triangle inequality for ρ.

Thus, in accordance with (1.1) define the transport distance

$$\mathrm{W}_\omega(\mu, \nu) = \inf_\lambda \iint \omega(|x-y|)\, d\lambda(x, y) \tag{1.4}$$

for a modulus of continuity ω, where, as before, the infimum is taken over all probability measures λ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. Here, one interesting choice $\omega(\delta) = \delta^\alpha$ with $0 < \alpha \le 1$ corresponds to the Zolotarev "ideal" metric $\mathrm{W}_\omega(\mu, \nu) = \zeta_\alpha(\mu, \nu)$ of order α (as a consequence of the Kantorovich-Rubinstein theorem). In the class of measures with a fixed compact support, these metrics are getting stronger when the parameter α is decreasing (and, as the limit case, $\zeta_0(\mu, \nu) = \sup_A |\mu(A) - \nu(A)|$ becomes the total variation distance). With this in mind, it is natural to try to strengthen a number of statements known about W with respect to the Euclidean distance ρ by means of the Zolotarev metrics. Here is a main result in this paper whose proof we postpone to Sects. 5–7.

Theorem 1.1 *Given probability measures μ and ν on $[0, \pi]^d$ with Fourier-Stieltjes transforms f_μ and f_ν, for any modulus of continuity ω and any $t > 0$,*

$$\mathrm{W}_\omega(\mu, \nu) \le \sqrt{d} \left(\sum_{m \ne 0} \omega^2\left(\frac{\pi}{|m|}\right) e^{-t|m|^2} \, |f_\mu(m) - f_\nu(m)|^2 \right)^{1/2} + 6\,\omega(\sqrt{dt}). \tag{1.5}$$

Due to a different argument in the proof of (1.5), this generalization of (1.3) contains however an additional factor $\sqrt{d}$ in front of the sum (see, however, Remark 7.2 below for a certain improvement). In both cases, we use smoothing of μ and ν by "Gaussian" measures on the torus. Another choice of the smoothing distribution leads to the following alternative, which may be viewed as a multidimensional variant of the Berry–Esseen bound:

$$\mathrm{W}_\omega(\mu, \nu) \le \sqrt{d} \left(\sum_{1 \le \|m\|_\infty \le T} \omega^2\left(\frac{\pi}{|m|}\right) |f_\mu(m) - f_\nu(m)|^2 \right)^{1/2} + 6\,\omega\left(\frac{\sqrt{12\,d}}{T}\right). \tag{1.6}$$

Here $T > 0$ is arbitrary, and $\|m\|_\infty = \max_{1 \le l \le d} |m_l|$ for $m = (m_1, \dots, m_d)$.

The sharpness of inequalities (1.3) and (1.5)–(1.6) may be illustrated on the example of discrete random measures in the so-called matching problem. Some history of the problem and general consequences from Theorem 1.1 are discussed in Sect. 2. These results are specialized to the class of Zolotarev's distances in Sect. 3, where we explore the way how the asymptotic behaviour of $\zeta_\alpha(\mu_n, \nu_n)$ is influenced by the parameter α (on average). Here, the index $\alpha = 1/2$ turns out to be critical. As shown in Sect. 4, using this parameter, one may study sharper forms in the one-dimensional minimax grid matching (the problem which motivates applications of Zolotarev's distances). In this connection, the role of order statistics is discussed separately in Sect. 8.

2 Empirical Measures

Given two collections $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ of random elements in (E, ρ), the optimal matching problem is concerned with an asymptotic behaviour of the quantities such as

$$\inf_{\sigma} \frac{1}{n} \sum_{k=1}^{n} \rho\left(X_k, Y_{\sigma(k)}\right),$$

where the infimum is taken over all permutations σ of $\{1, \dots, n\}$. This expression can be recognized as the transport distance $\mathrm{W}(\mu_n, \nu_n)$ as in (1.1) between the empirical measures

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

assigning the mass $1/n$ to each point in both collections. Of a large interest is in particular the question on the average value $\mathbb{E}\,(\mathrm{W}(\mu_n, \nu_n))$, assuming that all X_k and Y_l are independent and distributed according to a common law μ.

On the real line $E = \mathbb{R}$ with the Euclidean distance, the question is relatively simple and can be explored directly with the help of the identity (1.2). For compactly supported μ, it easily yields the standard rate (which is best possible), and more generally (cf. [6]), for an arbitrary law μ with distribution function $F(x) = \mu((-\infty, x])$ we have

$$\mathbb{E}\,(\mathrm{W}(\mu_n, \nu_n)) \sim \frac{1}{\sqrt{n}} \iff \int_{-\infty}^{\infty} \sqrt{F(x)(1-F(x))}\,dx < \infty. \tag{2.2}$$

Here and elsewhere, $A \sim B$ means that $c^{-1}B \le A \le cB$ for some constant $c \ge 1$ independent of n.

In dimension $d = 2$, if the two independent samples are drawn from the uniform distribution μ on the square $[0, 1] \times [0, 1]$, the rate at which $\mathbb{E}\left(\mathrm{W}(\mu_n, \nu_n)\right)$ tends to zero with respect to the growing number of observations turns out to be essentially the same as in dimension one, but up to a rather delicate logarithmic correction. More precisely, the famous AKT optimal matching theorem, due to Ajtai et al. [1], asserts that

$$\mathbb{E}\,(\mathrm{W}(\mu_n, \nu_n)) \sim \sqrt{\frac{\log n}{n}}, \tag{2.3}$$

and actually $\mathrm{W}(\mu_n, \nu_n)$ is of order (2.3) for large n with high probability. As for dimensions $d \ge 3$, the asymptotics depends on the dimension and is given by $\mathbb{E}\,(\mathrm{W}(\mu_n, \nu_n)) \sim n^{-1/d}$.

These results, especially the non-trivial two-dimensional relation (2.3), inspired further investigations in this direction, starting with [16,20]. As was shown by Talagrand [18], $\mathbb{E}\,(\mathrm{W}(\mu_n, \nu_n))$ admits an upper bound for any law μ supported on $[0, 1]^d$ with the same rates as in the case of the uniform distribution, i.e.

$$\mathbb{E}\,(\mathrm{W}(\mu_n, \nu_n)) = \begin{cases} O\left(\frac{1}{\sqrt{n}}\right) & \text{if } d = 1, \\ O\left(\sqrt{\frac{\log n}{n}}\right) & \text{if } d = 2, \\ O\left(\frac{1}{n^{1/d}}\right) & \text{if } d \geq 3. \end{cases}$$

In addition, Talagrand [12,19], undertook a deep investigation of optimal matching with the tool of the majorizing measure theory, thus replacing the original combinatorial approach of [1].

Quite surprisingly, the use of (1.3) not only simplifies the proof of the AKT theorem, but also expands the scope of its application, by weakening the independence assumption between X_k's and Y_l's, cf. [7]. Anyhow, the one-dimensional case remains to be exceptional, since the standard rate in it does not follow the general asymptotic rule $n^{-1/d}$ (in which one may include the case of the plane by ignoring a slowly increasing logarithmic factor). This inspires the idea that in some sense the property $\mathbb{E}\,\big(\mathrm{W}(\mu_n, \nu_n)\big) \sim 1/\sqrt{n}$ should be properly strengthened, by modifying the matching problem itself, or by strengthening the Euclidean metric on the line, with involvement of the Zolotarev metrics, for example. Such a strengthening is indeed possible, which has become known only recently due to the work of Dereich et al. [9] and Fournier and Guillin [11] on the asymptotic behaviour of the Kantorovich power distances between empirical measures. Some of the results of [11] are mentioned in the next section.

Now, turning to the Fourier analytic inequality (1.5) about the transport distance (1.4) (associated to an arbitrary modulus of continuity ω), first let us apply it to one general class of empirical measures. For a random variable ξ, define the Orlicz norm

$$\|\xi\|_{\psi_2} = \inf\left\{r > 0 : \mathbb{E}\,\exp\{\xi^2/r^2\} \leq 2\right\},$$

generated by the Young function $\psi_2(t) = e^{t^2} - 1$ ($t \in \mathbb{R}$). In a similar manner, one defines the Orlicz norm for other Young functions including $\psi_s(t) = \exp\{|t|^s\} - 1$ ($s \geq 1$). These norms are commonly used to control large deviations of ξ via Markov's inequality.

Proposition 2.1 *Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be pairwise independent random vectors with values in $[0, \pi]^d \times [0, \pi]^d$ such that X_k and Y_k have equal distributions for each $k \geq 1$. For any $t > 0$, the empirical measures μ_n and ν_n defined in (2.1) satisfy*

$$c\,\mathbb{E}\,(\mathrm{W}_\omega(\mu_n, \nu_n)) \leq \frac{\sqrt{d}}{\sqrt{n}} \left(\sum_{m \neq 0} \omega^2\left(\frac{\pi}{|m|}\right) e^{-t|m|^2}\right)^{1/2} + \omega(\sqrt{dt}) \tag{2.4}$$

with some absolute constant $c > 0$. Moreover, if all X_k, Y_l are independent, a similar inequality also holds for the ψ_2-norm of $\mathrm{W}_\omega(\mu_n, \nu_n)$ in place of the L^1-norm.

As another variant based on the application of the inequality (1.6), we also have the bound

$$c\,\mathbb{E}\,(\mathrm{W}_\omega(\mu_n,\nu_n)) \le \frac{\sqrt{d}}{\sqrt{n}}\Bigg(\sum_{1\le\|m\|_\infty\le T}\omega^2\Big(\frac{\pi}{|m|}\Big)\Bigg)^{1/2} + \omega\Big(\frac{\sqrt{d}}{T}\Big), \tag{2.5}$$

which is valid for any $T > 0$ (and similarly, for the ψ_2-norm).

Proof By the assumption, the expression

$$f_{\mu_n}(m) - f_{\nu_n}(m) = \frac{1}{n}\sum_{k=1}^n \Big(e^{im\cdot X_k} - e^{im\cdot Y_k}\Big) \tag{2.6}$$

represents the normalized sum of non-correlated, complex-valued random variables with mean zero and modulus ≤ 2. Hence, for any $m \in \mathbb{Z}^d$,

$$\mathbb{E}\,\Big(|f_{\mu_n}(m) - f_{\nu_n}(m)|^2\Big) \le \frac{4}{n},$$

and applying Jensen's inequality in (1.5), we then get the desired bound

$$\mathbb{E}\,(\mathrm{W}_\omega(\mu_n,\nu_n)) \le \frac{2\sqrt{d}}{\sqrt{n}}\Bigg(\sum_{m\neq 0}\omega^2\Big(\frac{\pi}{|m|}\Big)\,e^{-t|m|^2}\Bigg)^{1/2} + 6\,\omega(\sqrt{dt}). \tag{2.7}$$

□

In the second assertion of Proposition 2.1, the terms in the sum (2.6) are independent and have mean zero. One can therefore appeal to the following well-known fact.

Lemma 2.2 *Let the complex-valued random variables $\xi_1,\dots,\xi_n$ be independent, with $|\mathrm{Re}(\xi_k)| \le 1$, $|\mathrm{Im}(\xi_k)| \le 1$, and $\mathbb{E}(\xi_k) = 0$ for $k = 1,\dots,n$. Then, for the normalized sum $S_n = \frac{1}{\sqrt{n}}\,(\xi_1+\dots+\xi_n)$, we have*

$$\mathbb{E}\,\Big(e^{|S_n|^2/r^2}\Big) \le 2,$$

where one may take $r = \frac{4}{\sqrt{3}}$. That is, $\|S_n\|_{\psi_2} \le \frac{4}{\sqrt{3}}$.

Let us recall the argument. Any real-valued random variable ξ such that $|\xi| \le 1$ and $\mathbb{E}(\xi) = 0$ has a Laplace transform satisfying $\mathbb{E}\,(e^{t\xi}) \le e^{t^2/2}$ for all $t \in \mathbb{R}$ (the coefficient $1/2$ in the exponent is optimal and is attained for the symmetric Bernoulli distribution on $\{-1, 1\}$). If the random variables in the lemma are real-valued, we therefore obtain a similar bound

$$\mathbb{E}\,(e^{tS_n}) \le e^{t^2/2}, \qquad t \in \mathbb{R}.$$

One can now integrate this inequality over the Gaussian measure on the real line with mean zero and standard deviation $0 < \sigma < 1$, to get

$$\mathbb{E}\left(e^{\sigma^2 S_n^2/2}\right) \le \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t^2/2}\, e^{-t^2/(2\sigma^2)}\, dt = \frac{1}{\sqrt{1-\sigma^2}}.$$

In the complex-valued case as in the lemma, the above inequality may be applied separately to the real and imaginary parts of S_n, which gives, by replacing σ with 2σ,

$$\mathbb{E}\,\exp\left\{2\sigma^2\,|\mathrm{Re}(S_n)|^2\right\} \le \frac{1}{\sqrt{1-4\sigma^2}}, \qquad \mathbb{E}\,\exp\left\{2\sigma^2\,|\mathrm{Im}(S_n)|^2\right\} \le \frac{1}{\sqrt{1-4\sigma^2}}.$$

Hence, by Cauchy's inequality,

$$\begin{aligned}\mathbb{E}\left(e^{\sigma^2|S_n|^2}\right) &= \mathbb{E}\,\exp\left\{\sigma^2\,\left(|\mathrm{Re}(S_n)|^2 + |\mathrm{Im}(S_n)|^2\right)\right\} \\ &\le \left(\mathbb{E}\,\exp\left\{2\sigma^2\,|\mathrm{Re}(S_n)|^2\right\}\right)^{1/2}\left(\mathbb{E}\,\exp\left\{2\sigma^2\,|\mathrm{Im}(S_n)|^2\right\}\right)^{1/2} \le \frac{1}{\sqrt{1-4\sigma^2}}.\end{aligned}$$

Here, the right-hand side is equal to 2 for $\sigma = \frac{1}{4}\sqrt{3}$, proving the lemma.

Thus, returning to (2.6) and applying the lemma, we get $\| f_{\mu_n}(m) - f_{\nu_n}(m)\|_{\psi_2} \le \frac{r}{\sqrt{n}}$ with $r = 4/\sqrt{3}$, or equivalently

$$\big\|\, |f_{\mu_n}(m) - f_{\nu_n}(m)|^2 \,\big\|_{\psi_1} \le \frac{r^2}{n}.$$

Now, squaring (1.5), we have

$$\mathrm{W}_\omega^2(\mu_n, \nu_n) \le 2d \sum_{m \neq 0} \omega^2\left(\frac{\pi}{|m|}\right) e^{-tm^2}\, |f_\mu(m) - f_\nu(m)|^2 + 72\,\omega^2\left(\sqrt{dt}\right),$$

and, by the triangle inequality for the ψ_1-norm,

$$\left\|\mathrm{W}_\omega^2(\mu_n, \nu_n)\right\|_{\psi_1} \le \frac{2dr^2}{n} \sum_{m \neq 0} \omega^2\left(\frac{\pi}{|m|}\right) e^{-t|m|^2} + \frac{72}{\log 2}\,\omega^2\left(\sqrt{dt}\right)$$

(where we used that $\|1\|_{\psi_1} = \frac{1}{\log 2}$). This gives

$$\left\|\mathrm{W}_\omega(\mu_n, \nu_n)\right\|_{\psi_2} \le \frac{\sqrt{2d}\,r}{\sqrt{n}} \left(\sum_{m \neq 0} \omega^2\left(\frac{\pi}{|m|}\right) e^{-t|m|^2}\right)^{1/2} + \sqrt{\frac{72}{\log 2}}\,\omega\left(\sqrt{dt}\right).$$

□

The inequality (2.5) and its ψ_2-version are derived from (1.6) by similar arguments.

3 Kantorovich–Rubinstein Theorem and Zolotarev Distances

In the setting of an abstract separable metric space (E, ρ), the Kantorovich–Rubinstein duality theorem asserts that, for any two (Borel) probability measures μ and ν on E,

$$\mathrm{W}(\mu, \nu) = \sup_u \left| \int u \, d\mu - \int u \, d\nu \right|. \tag{3.1}$$

Here, the supremum is taken over all functions $u : E \to \mathbb{R}$ such that $|u(x) - u(y)| \le \rho(x, y)$, $x, y \in E$, i.e., with Lipschitz semi-norm $\|u\|_{\mathrm{Lip}} \le 1$ (cf. [10]).

Definition 3.1 In the case of the Euclidean space $E = \mathbb{R}^d$, the Zolotarev distance $\zeta_\alpha(\mu, \nu)$ of order $\alpha \in (0, 1]$ is defined to be the right-hand side of (3.1) with supremum taken over all $u : E \to \mathbb{R}$ such that $|u(x) - u(y)| \le |x - y|^\alpha$, $x, y \in \mathbb{R}^d$.

The latter inequality describes the Lipschitz property of the function u with respect to the metric $\rho(x, y) = |x - y|^\alpha$ on $\mathbb{R}^d$. Hence $\mathrm{W}_\omega = \zeta_\alpha$ for the modulus of continuity $\omega(\delta) = \delta^\alpha$. The Zolotarev distance is also defined for $\alpha > 1$, but we do not consider it here (cf. [22,23]).

Some further generalization of the Kantorovich distance (1.1) between the probability measures on $\mathbb{R}^d$ is given by

$$\mathrm{T}_\alpha(\mu, \nu) = \inf_\lambda \iint |x - y|^\alpha \, d\lambda(x, y), \qquad \alpha > 0.$$

As before, the infimum is running over all probability measures λ on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. Thus, $\zeta_\alpha = \mathrm{T}_\alpha$ for $0 < \alpha \le 1$. In the case $\alpha > 1$, $\mathrm{T}_\alpha^{1/\alpha}$ represents a metric in the space of all probability measures μ on $\mathbb{R}^d$ with finite absolute moment $\int |x|^\alpha \, d\mu(x)$. It is called a minimal distance, or the Kantorovich power distance, and also (not quite correctly) the Wasserstein distance of order α (with a standard notation W_α which we avoid here). The relationship between ζ_α and T_α for $\alpha > 1$ for one dimensional measures was investigated by Rio [15] in his study of transport central limit theorems.

When $\omega(\delta) = \delta^\alpha$, $0 < \alpha \le 1$, the transport bounds (2.4)–(2.5) may easily be simplified by optimizing the right-hand sides over $t > 0$ and $T > 0$. Let us start with dimension $d = 1$ and apply (2.4). If $\alpha > \frac{1}{2}$, one may let $t \to 0$, and then (2.4) leads to the standard rate

$$\mathbb{E}\left(\zeta_\alpha(\mu_n, \nu_n)\right) \le \frac{c_1}{\sqrt{n}} \left(\sum_{m=1}^\infty \frac{1}{m^{2\alpha}} \right)^{1/2} \le \frac{c_2}{\sqrt{2\alpha - 1}} \, \frac{1}{\sqrt{n}}$$

with some absolute constants $c_j > 0$.

If $\alpha \le \frac{1}{2}$, one may replace the sum in (2.4) with an integral. Note that $m^{-2\alpha} e^{-tm^2}$ is decreasing in m, so that

$$\sum_{m=1}^\infty \frac{1}{m^{2\alpha}} e^{-tm^2} \le 1 + \int_1^\infty \frac{1}{x^{2\alpha}} e^{-tx^2} \, dx = 1 + \frac{1}{t^{\frac{1}{2} - \alpha}} \int_{\sqrt{t}}^\infty \frac{1}{y^{2\alpha}} e^{-y^2} \, dy. \tag{3.2}$$

If $\alpha < \frac{1}{2}$ and $0 < t \le 1$, the last integral is bounded, up to a constant, by $\frac{1}{1-2\alpha}\, t^{\alpha-\frac{1}{2}}$, and then (3.2) gives, after replacing $s = \sqrt{t}$,

$$c\,\mathbb{E}\left(\zeta_\alpha(\mu_n,\nu_n)\right) \le \frac{1}{\sqrt{n}}\,\frac{1}{\sqrt{1-2\alpha}}\,s^{\alpha-\frac{1}{2}} + s^\alpha.$$

Choosing $s = 1/n$, we get

$$\mathbb{E}\left(\zeta_\alpha(\mu_n,\nu_n)\right) \le \frac{c}{n^\alpha\sqrt{1-2\alpha}}.$$

If $\alpha = \frac{1}{2}$ and $0 < t \le 1$, the last integral in (3.2) is bounded, up to a constant, by $\log(2/\sqrt{t})$, and then (3.2) gives, after replacing $s = \sqrt{t}$,

$$c\,\mathbb{E}\left(\zeta_\alpha(\mu_n,\nu_n)\right) \le \frac{1}{\sqrt{n}}\sqrt{\log(2/s)} + \sqrt{s}.$$

Choosing $s = 1/n$, we can summarize.

Corollary 3.2 *In the setting of Proposition 2.1 for the one dimensional case we have*

$$\begin{aligned}
\sqrt{2\alpha-1}\,\mathbb{E}\left(\zeta_\alpha(\mu_n,\nu_n)\right) &\le \frac{c}{\sqrt{n}}, && \text{if } \frac{1}{2} < \alpha \le 1,\\
\mathbb{E}\left(\zeta_\alpha(\mu_n,\nu_n)\right) &\le \frac{c}{\sqrt{n}}\sqrt{\log(2n)}, && \text{if } \alpha = \frac{1}{2},\\
\sqrt{1-2\alpha}\,\mathbb{E}\left(\zeta_\alpha(\mu_n,\nu_n)\right) &\le \frac{c}{n^\alpha}, && \text{if } 0 < \alpha < \frac{1}{2},
\end{aligned}$$

where c is a positive absolute constant. If all X_k, Y_l are independent, similar bounds also hold for the ψ_2-norm of the random variable $\zeta_\alpha(\mu_n,\nu_n)$.

Thus, the value $\alpha = \frac{1}{2}$ is critical, meaning that the rate in the upper bound is changing for smaller values of the parameter α. This interesting phenomenon is deeply connected with the following well-known theorem due to S. Bernstein (cf. [24]): If a (2π)-periodic function u on the line belongs to the Lipschitz class $\mathrm{Lip}(\alpha)$ with $\alpha > \frac{1}{2}$, then this function can be expanded into the absolutely convergent Fourier series

$$u(x) = \frac{a_0}{2} + \sum_{m=1}^{\infty}\left(a_m\cos(mx) + b_m\sin(mx)\right).$$

In fact, $\sum_{m=1}^{\infty}\left(|a_m| + |b_m|\right) \le C_\alpha\,\|u\|_{\mathrm{Lip}(\alpha)}$ where the constant C_α depends on α only. Moreover, this property is no longer true for $\alpha = \frac{1}{2}$.

Now, let $d \geq 2$. We apply the bound (2.5), which in the polynomial case $\omega(\delta) = \delta^\alpha$ takes the form

$$c_d\, \mathbb{E}\left(\zeta_\alpha(\mu_n, \nu_n)\right) \leq \frac{1}{\sqrt{n}} \left(\sum_{1 \leq \|m\|_\infty \leq T} \frac{1}{|m|^{2\alpha}} \right)^{1/2} + \frac{1}{T^\alpha}, \qquad T > 0, \tag{3.3}$$

with some d-dependent constant $c_d > 0$ (possibly varying from line to line below). Clearly, if $T \geq 1$ and $d \geq 3$,

$$\sum_{1 \leq \|m\|_\infty \leq T} \frac{1}{|m|^{2\alpha}} \leq C_d \sum_{1 \leq k \leq T} k^{d-1-2\alpha} \leq C_d' T^{d-2\alpha},$$

and (3.3) is simplified to

$$c_d\, \mathbb{E}\left(\zeta_\alpha(\mu_n, \nu_n)\right) \leq \frac{1}{\sqrt{n}} T^{\frac{d}{2}-\alpha} + \frac{1}{T^\alpha} = \frac{2}{n^{\alpha/d}},$$

where we choose $T = n$ on the last step. Similarly, if $d = 2$ and $\alpha < 1$,

$$\sum_{1 \leq \|m\|_\infty \leq T} \frac{1}{|m|^{2\alpha}} \leq C \sum_{1 \leq k \leq T} k^{1-2\alpha} \leq \frac{C'}{1-\alpha} T^{2-2\alpha},$$

so that

$$c\, \mathbb{E}\left(\zeta_\alpha(\mu_n, \nu_n)\right) \leq \frac{1}{\sqrt{n}} \frac{C'}{1-\alpha} T^{1-\alpha} + \frac{1}{T^\alpha} \leq \frac{C''}{1-\alpha} \frac{1}{n^{\alpha/2}}.$$

Also, if $d = 2$ and $\alpha = 1$, we recover the upper bound in the AKT theorem (2.3).

Collecting these statements together, we arrive at:

Corollary 3.3 *In the setting of Proposition 2.1 for the dimension $d \geq 2$, we have*

$$\mathbb{E}\left(\zeta_\alpha(\mu_n, \nu_n)\right) \leq C_d \frac{A_n(\alpha)}{n^{\alpha/d}},$$

where the constant C_d depends on d only, and

$$A_n(\alpha) = \begin{cases} \frac{1}{1-\alpha} & \text{if } d = 2,\ \alpha < 1, \\ \sqrt{\log(2n)} & \text{if } d = 2,\ \alpha = 1, \\ 1 & \text{if } d \geq 3,\ \alpha \leq 1. \end{cases}$$

If all X_k, Y_l are independent, a similar bound also holds for the ψ_2-norm of $\zeta_\alpha(\mu_n, \nu_n)$.

For independent random vectors $X_1, \dots, X_n$ having a common distribution μ, similar bounds on $\mathbb{E}\,(\mathrm{T}_\alpha(\mu_n, \mu))$ have been derived by Fournier and Guillin [11].

They considered a general multidimensional case of samples with not necessarily compactly supported distributions, by involving moment conditions. In particular, if the moment $M_{2\alpha} = \int |x|^{2\alpha}\,d\mu(x)$ is finite for a fixed $\alpha > 0$, it was shown that

$$c\,\mathbb{E}\left(\mathrm{T}_\alpha(\mu_n,\mu)\right) \le M_{2\alpha}^{1/2} \times \begin{cases} \frac{1}{\sqrt{n}} & \text{if } \alpha > d/2, \\ \frac{1}{\sqrt{n}}\log(2n) & \text{if } \alpha = d/2, \\ n^{-\alpha/d} & \text{if } \alpha < d/2, \end{cases}$$

where the constant $c > 0$ depends only on α and d. Thus, in the case of a compactly supported measure μ, this statement coincides with Corollary 3.3 for $d \ge 3$ with an arbitrary $\alpha \in (0, 1]$ and for $d = 2$ with $\alpha \in (0, 1)$ (without specifying the type of dependence on α when this parameter approaches 1). In the case $d = 2$ and $\alpha = 1$, the assertion is not optimal, while Corollary 3.3 yields the AKT theorem. Also, when $d = 1$, Corollary 3.2 provides an improvement for $\alpha = \frac{1}{2}$. Another advantage of Corollaries 3.2–3.3 is that the random vectors X_k are not required to be equally distributed (and fully independent).

The authors of [11] have also studied concentration properties. Specializing their Theorem 2 to the case where μ is compactly supported, it was shown that, if $(X_1,\dots,X_n)$ is a sample drawn from μ, then

$$\left\|\mathrm{T}_\alpha(\mu_n,\mu)\right\|_{\psi_2} \le \frac{c}{\sqrt{n}} \quad \text{for } \alpha > d/2.$$

This is consistent with the second part of Corollary 3.2 when $d = 1$ and $\frac{1}{2} < \alpha \le 1$. If $\alpha < d/2$, Theorem 2 of [11] provides the bound

$$\left\|\mathrm{T}_\alpha(\mu_n,\mu)\right\|_{\psi_{d/\alpha}} \le \frac{c}{n^\alpha},$$

which is stronger than the concentration part of Corollary 3.2 when $d = 1$, $\alpha < \frac{1}{2}$, since the Orlicz norm for the Young function $\psi_{d/\alpha}(t) = \exp\{|t|^{d/\alpha}\} - 1$ is stronger than the one for ψ_2.

Let us also emphasize that the $\frac{1}{\sqrt{n}}$-rate for $\mathbb{E}\,(T_\alpha(\mu_n,\mu))$ with $\alpha > 1$ may essentially be improved for a large family of probability distributions μ. Staying on the real line, introduce the functional

$$J_\alpha(\mu) = \int_{-\infty}^{\infty} \left(F(x)(1-F(x))\right)^{\frac{\alpha}{2}} p(x)^{1-\alpha}\,dx,$$

where F is the distribution function associated to the probability measure μ, and where p denotes the density of its absolutely continuous component (with respect to the Lebesgue measure on the line). For $\alpha = 1$, this integral thus becomes the one from the characterization (2.2). As was shown in [6], cf. Corollary 5.9, for any fixed $\alpha > 1$,

$$\left(\mathbb{E}\,(\mathrm{T}_\alpha(\mu_n,\mu))\right)^{1/\alpha} \sim \frac{1}{\sqrt{n}} \quad \text{if and only if} \quad J_\alpha(\mu) < \infty.$$

If $1 < \alpha < 2$, the integral $J_\alpha(\mu)$ is finite, as long as, for example, μ represents a transform of the two-sided exponential distribution with density $\frac{1}{2}\,e^{-|x|}$ under a map with finite Lipschitz semi-norm (equivalently, μ has a positive Cheeger isoperimetric constant).

If $d \geq 2$, an almost sure behaviour of $\mathrm{T}_\alpha(\mu_n, \nu_n)$ was considered by Barthe and Bordenave [2] (under certain constraints on α and d). In particular, if μ represents a uniform distribution over a bounded set $\Omega \subset \mathbb{R}^d$ with volume $|\Omega| > 0$, and μ_n and ν_n are independent samples drawn from μ of size n, it was shown that, with probability one,

$$\lim_{n\to\infty} \left[n^{\alpha/d}\,\mathrm{T}_\alpha(\mu_n, \nu_n)\right] = c_d(\alpha)\,|\Omega|, \qquad 0 < \alpha < d/2$$

(for some unknown constant $c_d(\alpha) > 0$).

4 Minimax Grid Matching

Following [17], for two collections of points $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_n\}$ in the cube $[0, 1]^d$, define the minimax matching length

$$\mathrm{L}(X, Y) = \min_\sigma \max_{1\leq k\leq n} |x_k - y_{\sigma(k)}|,$$

where the minimum is running over all permutations σ of $\{1, \dots, n\}$. This quantity represents a metric in the space of all unordered collections with fixed n. Equivalently,

$$\mathrm{L}(X, Y) = \lim_{\alpha\to\infty} (\mathrm{T}_\alpha(\mu_n, \nu_n))^{1/\alpha}$$

for the associated empirical measures μ_n and ν_n as in (2.1). Note that $\mathrm{L}(X, Y) \geq \mathrm{W}(\mu_n, \nu_n)$.

If X and Y are independent samples drawn from a given distribution μ, the minimax grid matching problem is to find the rate of $\mathbb{E}\,(\mathrm{L}(X, Y))$ at which it tends to zero as $n \to \infty$. When μ is a uniform distribution on the cube, it was shown by Shor and Yukich [17] that, for $d \geq 3$,

$$\mathbb{E}\left(\mathrm{L}(X, Y)\right) \sim \left(\frac{\log n}{n}\right)^{1/d}.$$

If $d = 2$, the rate is somewhat different,

$$\mathbb{E}\left(\mathrm{L}(X, Y)\right) \sim \frac{(\log n)^{3/4}}{\sqrt{n}},$$

which was proved earlier by Leighton and Shor [13]. Like for W, this case turns out to be very similar to the one dimensional case ($d = 1$), when

$$\mathbb{E}\,\big(\mathrm{L}(X, Y)\big) \sim \frac{1}{\sqrt{n}}$$

(cf. e.g. [6]). The improvement is thus only in the logarithmic term.

Nevertheless, using Corollary 3.2, one can improve the rate in dimension one, if counting not all, but most of the points in perfect matching. Namely, for a (non-empty) subset I of $\{1, \dots, n\}$, define the restricted minimax matching length

$$\mathrm{L}_I(X, Y) = \min_{\sigma} \max_{k \in I} |x_k - y_{\sigma(k)}|,$$

still assuming that the minimum is running over all permutations σ of $\{1, \dots, n\}$. In the next observation, there is no need to keep the assumption that the distributions of the components X_k are identical.

Proposition 4.1 *Let $Y = (Y_1, \dots, Y_n)$ be an independent copy of the random vector $X = (X_1, \dots, X_n)$ which has independent coordinates with values in* $[0, 1]$. *With high probability, for each $\varepsilon > 0$, there is a (random) set $I \subset \{1, \dots, n\}$ of cardinality $|I| \geq (1 - \varepsilon)\, n$ such that*

$$\mathrm{L}_I(X, Y) \leq C_\varepsilon\, \frac{\log^2(2n)}{n},$$

where one may take $C_\varepsilon = C/\varepsilon^2$ with an absolute constant C.

Here, "with high probability" means the probability $1 - n^{-p}$, where p can be chosen as large as we wish by a proper choice of the constant C. We thus obtain a much better $\frac{1}{n}$-rate (modulo a logarithmic term) by removing a small proportion of "bad" points in both samples X and Y and constructing a perfect matching between remaining "good" points.

Proof We apply Corollary 3.2 in its stronger ψ_2-version with the modulus of continuity $\omega(\delta) = \sqrt{\delta}$. It is telling us that the random variable

$$\xi_n = \frac{1}{n} \inf_{\sigma} \sum_{k=1}^{n} |X_k - Y_{\sigma(k)}|^{1/2},$$

where the infimum is running over all permutations σ of $\{1, \dots, n\}$, has the norm

$$\|\xi_n\|_{\psi_2} \leq \frac{c}{\sqrt{n}} \log^{1/2}(2n)$$

with some absolute constant $c > 0$. Hence, the event $A_n = \{\xi_n \geq c\sqrt{p}\, \frac{\log(2n)}{\sqrt{n}}\}$ has probability

$$\mathbb{P}(A_n) \leq 2\, \exp\big\{-p \log(2n)\big\} = \frac{2}{(2n)^p}, \tag{4.1}$$

for any $p \geq 1$, the complementary set, we thus have an opposite inequality $\xi_n < c\sqrt{p}\,\frac{\log(2n)}{\sqrt{n}}$, which means that for some σ,

$$\frac{1}{n}\sum_{k=1}^{n} |X_k - Y_{\sigma(k)}|^{1/2} < c\sqrt{p}\,\frac{\log(2n)}{\sqrt{n}}.$$

□

Let π_n denote the uniform discrete probability measure on $\{1, \dots, n\}$, so that to recognize the above left-hand side as the expectation with respect to π_n. By Markov's inequality, it follows that, for any $\varepsilon \in (0, 1)$,

$$\pi_n \left\{ k \leq n : |X_k - Y_{\sigma(k)}|^{1/2} \geq c\sqrt{p}\,\frac{\log(2n)}{\varepsilon\sqrt{n}} \right\} < \varepsilon,$$

that is,

$$\pi_n \left\{ k \leq n : |X_k - Y_{\sigma(k)}| \geq c^2 p\,\frac{\log^2(2n)}{\varepsilon^2 n} \right\} < \varepsilon.$$

In other words, the set

$$I = \left\{ k \leq n : |X_k - Y_{\sigma(k)}| < c^2 p\,\frac{\log^2(2n)}{\varepsilon^2 n} \right\}$$

has cardinality $|I| \geq (1-\varepsilon)n$, and, by the construction, $\mathrm{L}_I(X, Y) < c^2 p\,\frac{\log^2(2n)}{\varepsilon^2 n}$. So, by (4.1),

$$\mathbb{P}\left\{ \inf_{|I| \geq (1-\varepsilon)n} \mathrm{L}_I(X, Y) \geq c^2 p\,\frac{\log^2(2n)}{\varepsilon^2 n} \right\} \leq \frac{1}{n^p}.$$

□

5 Reduction to the Torus

Before turning to the proof of Theorem 1.1, we need some preparation. By the Kantorovich-Rubinstein theorem, applied to the Euclidean space $\mathbb{R}^d$ with metric $\rho(x, y) = \omega(|x - y|)$, where ω is a modulus of continuity, the transport distance (1.4) between given (Borel) probability measures μ and ν on $\mathbb{R}^d$ admits a dual description (3.1), i.e.,

$$\mathrm{W}_\omega(\mu, \nu) = \sup_u \left| \int_{\mathbb{R}^d} u\,d\mu - \int_{\mathbb{R}^d} u\,d\nu \right|. \tag{5.1}$$

Here, the supremum is taken over all functions u such that $|u(x) - u(y)| \leq \rho(x, y)$, $x, y \in \mathbb{R}^d$, which for short we denote as the Lipschitz property $\|u\|_{\mathrm{Lip}(\omega)} \leq 1$. To bound W_ω via (5.1) by means of the Fourier analysis, it is desirable to require some additional properties of u like periodicity. To this aim, let us consider a related quantity

$$\widetilde{\mathrm{W}}_\omega(\mu,\nu) = \sup_u \left| \int_{Q^d} u\,d\mu - \int_{Q^d} u\,d\nu \right|, \tag{5.2}$$

assuming that μ and ν are supported on the cube $Q^d = (-\pi,\pi]^d$, and where the supremum is taken over all (2π)-periodic functions u on $\mathbb{R}^d$ with Lipschitz seminorm $\|u\|_{\mathrm{Lip}(\omega)} \le 1$. By the definition, $\widetilde{\mathrm{W}}_\omega \le \mathrm{W}_\omega$. On the other hand, we have:

Lemma 5.1 *If μ and ν are supported on $[0,\pi]^d$, then $\mathrm{W}_\omega(\mu,\nu) = \widetilde{\mathrm{W}}_\omega(\mu,\nu)$.*

Proof Introduce a canonical functional

$$\|z\| = \mathrm{dist}(z, 2\pi\mathbb{Z}^d) = \min_{m\in\mathbb{Z}^d} |z - 2\pi m|, \qquad z \in \mathbb{R}^n,$$

describing the (shortest) distance from a point to the lattice. It is subadditive, i.e. $\|z_1 + z_2\| \le \|z_1\| + \|z_2\|$ for all $z_1, z_2 \in \mathbb{R}^d$. Hence

$$\tilde{\rho}(x,y) = \omega(\|x-y\|)$$

is non-negative, symmetric, and satisfies the triangle inequality on $\mathbb{R}^d$, by the subadditivity of ω. The only axiom of a metric which is not satisfied by $\tilde{\rho}$ is the separation: we have $\tilde{\rho}(x,y) = 0$ if and only if $x - y \in 2\pi\mathbb{Z}^d$.

If $u : \mathbb{R}^d \to \mathbb{R}$ is (2π)-periodic and $\|u\|_{\mathrm{Lip}(\omega)} \le 1$, then, for all $x, y \in \mathbb{R}^d$ and $m \in \mathbb{Z}^d$,

$$|u(x) - u(y)| = |u(x) - u(y + 2\pi m)| \le \omega\big(|(x-y) - 2\pi m|\big),$$

so,

$$|u(x) - u(y)| \le \tilde{\rho}(x,y). \tag{5.3}$$

Conversely, this inequality readily implies that u is (2π)-periodic with $\|u\|_{\mathrm{Lip}(\omega)} \le 1$. Thus, (5.3) describes the class of all functions u on $\mathbb{R}^d$ participating in the definition (5.2).

Being restricted to the cube $Q^d = (-\pi,\pi]^d$, $\tilde{\rho}$ becomes a metric: For $x, y \in Q^d$, necessarily $x - y \in (-2\pi, 2\pi)^d$, so that $\tilde{\rho}(x,y) = 0 \Rightarrow x = y$. Hence, the property (5.3) being applied to the functions u on Q^d describes the Lipschitz property with respect to the metric $\tilde{\rho}$. Let us also note that, if u is defined on Q^d and satisfies (5.3) for all $x, y \in Q^d$, its (2π)-periodic extension to $\mathbb{R}^d$ satisfies the same inequality for all $x, y \in \mathbb{R}^d$ (since the right-hand side of (5.3) is invariant under adding to x and y elements of $2\pi\mathbb{Z}^d$).

Thus, when the measures μ and ν are supported on Q^d, the quantity $\widetilde{\mathrm{W}}_\omega(\mu,\nu)$ represents the Kantorovich distance between μ and ν with respect to the metric $\tilde{\rho}$. By the Kantorovich-Rubinstein theorem (3.1), one can therefore conclude that

$$\widetilde{\mathrm{W}}_\omega(\mu, \nu) \ = \ \inf_\lambda \iint \omega(\|x-y\|)\, d\lambda(x, y), \tag{5.4}$$

where the infimum is running over all Borel probability measures λ on $Q^d \times Q^d$ with marginals μ and ν. Moreover, if μ and ν are supported on $[0, \pi]^d$, then all such λ's have to be supported on $[0, \pi]^d \times [0, \pi]^d$. But, for $x, y \in [0, \pi]^d$, necessarily $x - y \in [-\pi, \pi]^d$, and for this point the origin is the closest element in the lattice $2\pi\mathbb{Z}^d$. Hence, $\|x-y\| = |x-y|$, and the right-hand side of (5.4) becomes $\mathrm{W}_\omega(\mu, \nu)$, according to (1.4). □

Remark 5.2 In the definition (5.2) one may additionally require that the function u be of class $C^\infty(\mathbb{R}^d)$. Indeed, for $\varepsilon > 0$, consider the convolution of u with a Gaussian density

$$u_\varepsilon(x) = \frac{1}{(2\pi\varepsilon^2)^{d/2}} \int u(x-y)\, e^{-|y|^2/2\varepsilon^2}\, dy, \qquad x \in \mathbb{R}^d.$$

Clearly, u is also (2π)-periodic, of class $C^\infty(\mathbb{R}^d)$ and with $\|u_\varepsilon\|_{\mathrm{Lip}(\omega)} \le 1$. Writing

$$u_\varepsilon(x) - u(x) = \frac{1}{(2\pi)^{d/2}} \int (u(x-\varepsilon y) - u(x))\, e^{-|y|^2/2}\, dy,$$

it follows that

$$\begin{aligned} \sup_x |u_\varepsilon(x) - u(x)| &\le \frac{1}{(2\pi)^{d/2}} \int \omega(\varepsilon|y|)\, e^{-|y|^2/2}\, dy \\ &= \mathbb{E}\left(\omega(\varepsilon|Z|)\right) \ \le \ 3\,\omega(\varepsilon\sqrt{d}) \ \to \ 0 \quad \text{as } \varepsilon \to 0, \end{aligned}$$

where Z is a standard normal random vector in $\mathbb{R}^d$. Thus, in (5.2) the functions u may be replaced with u_ε's.

Here, on the last step we applied the following Jensen-type inequality, which will be needed later on in the smoothing argument.

Lemma 5.3 *For any random variable $\xi \ge 0$,*

$$\mathbb{E}\,(\omega(\xi)) \ \le \ 3\,\omega(\mathbb{E}(\xi)). \tag{5.5}$$

Proof One may assume that ξ is bounded, say $0 \le \xi \le M$, and $\mathbb{E}(\xi) = a$. Given this linear constraint, the inequality (5.5) is affine with respect to the distribution of ξ. The collection of all probability distributions μ on $[0, M]$ with $\int x\, d\mu(x) = a$ is a convex compact space (for the weak topology) and has as extreme "points" the probability measures with at most two atoms. This means that we only need to check (5.5) for Bernoulli distributions μ, that is, for random variables ξ taking two values, say $x > 0$

and $y > 0$ with probabilities p and $q = 1 - p$ respectively ($0 \le p \le 1$), and such that $px + qy = a$.

For definiteness, let $x \ge y$, so that $x \ge a \ge y$. Then, since $px \le a$,

$$\mathbb{E}\,(\omega(\xi)) \,=\, p\omega(x) + q\omega(y) \,\le\, \frac{a}{x}\,\omega(x) + \omega(a). \tag{5.6}$$

On the other hand, by the subadditivity property of moduli of continuity, $\omega(n\delta) \le n\omega(\delta)$ for all $\delta \ge 0$ and all positive integers n. Introducing $n = [\frac{x}{a}]$, we therefore have

$$\omega(x) \,=\, \omega\Big(\frac{x}{a}\cdot a\Big) \,\le\, \omega((n+1)\cdot a) \,\le\, (n+1)\,\omega(a) \,\le\, 2n\,\omega(a) \,\le\, \frac{2x}{a}\,\omega(a).$$

Plugging this bound in (5.6), we arrive at (5.5). □

Note that any non-decreasing concave function ω on the half-axis $[0, \infty)$ such that $\omega(0) = 0$ (and which is not identically zero) represents a modulus of continuity. In that case, the constant 3 in (5.5) may be removed, by the usual Jensen inequality.

6 Fourier Analytic Inequalities

Let us fix a modulus of continuity ω and the associated metric $\rho(x, y) = \omega(|x - y|)$ on $\mathbb{R}^d$.

The integrals in (5.2) may be explored in terms of the Fourier-Stieltjes transforms f_μ and f_ν. As in Remark 5.2, suppose that a function u participating in the supremum (5.2) is of class $C^\infty(\mathbb{R}^d)$, so that it has an absolutely convergent Fourier series

$$u(x) = \sum_{m\in\mathbb{Z}^d} a_m\, e^{im\cdot x} \tag{6.1}$$

with $\sum_{m\in\mathbb{Z}^d} |a_m| < \infty$. Integrating this equality over the measures μ and ν, we get

$$\int_{Q^d} u\,d\mu - \int_{Q^d} u\,d\nu \,=\, \sum_{m\in\mathbb{Z}^d} a_m\,(f_\mu(m) - f_\nu(m)). \tag{6.2}$$

To bound the right-hand side, one may follow a standard argument used in the proof of Bernstein's theorem on the absolute convergence of Fourier series, cf. e.g. [24], Ch. VI.

Given $h \in \mathbb{R}^d$, it follows from (6.1) that

$$u(x+h) - u(x-h) \,=\, 2i \sum_{m\in\mathbb{Z}^d} a_m\, \sin(m\cdot h)\, e^{im\cdot x},$$

so that, by the Parseval identity,

$$\frac{1}{(2\pi)^d}\int_{Q^d}|u(x+h)-u(x-h)|^2\,dx \;=\; 4\sum_{m\in\mathbb{Z}^d}|a_m|^2\,\sin^2(m\cdot h).$$

Using the Lipschitz assumption $|u(x+h)-u(x-h)|\le\omega(2|h|)$, it implies

$$\sum_{m\in\mathbb{Z}^d}|a_m|^2\,\sin^2(m\cdot h)\;\le\;\frac{1}{4}\,\omega^2(2|h|). \tag{6.3}$$

We choose here $h=\pi\,2^{-k-1}\,\theta$ with $\theta\in\mathbb{R}^d$, $|\theta|=1$, and $k=1,2,\dots$ This gives

$$\sum_{m\in\mathbb{Z}^d}|a_m|^2\,\sin^2\left(\pi\,2^{-k-1}\,m\cdot\theta\right)\;\le\;\frac{1}{4}\,\omega^2\left(\pi\,2^{-k}\right).$$

In particular, if $|m|\le 2^k$, the expression under the sine function does not exceed $\pi/2$ in absolute value, so,

$$\sin^2\left(\pi\,2^{-k-1}m\cdot\theta\right)\ge\frac{4}{\pi^2}\left(\pi\,2^{-k-1}m\cdot\theta\right)^2=4^{-k}\,(m\cdot\theta)^2,$$

which thus gives

$$\sum_{|m|\le 2^k}|a_m|^2\,4^{-k}\,(m\cdot\theta)^2\;\le\;\frac{1}{4}\,\omega^2\left(\pi\,2^{-k}\right). \tag{6.4}$$

Averaging the left-hand side over θ with respect to the uniform measure σ_{d-1} on the unit sphere S^{d-1} in $\mathbb{R}^d$, we get

$$\sum_{|m|\le 2^k}|a_m|^2\,4^{-k}\,|m|^2\;\le\;\frac{d}{4}\,\omega^2\left(\pi\,2^{-k}\right). \tag{6.5}$$

It makes sense to restrict the summation on the left to the integral values of m with length $|m|\ge 2^{k-1}$, under which $4^{-k}\,|m|^2\ge\frac{1}{4}$, and then

$$\sum_{2^{k-1}\le|m|<2^k}|a_m|^2\;\le\;d\,\omega^2\left(\pi\,2^{-k}\right).$$

Hence, by Cauchy's inequality,

$$\sum_{2^{k-1}\le|m|<2^k}|a_m|\,|f_\mu(m)-f_\nu(m)|\;\le\;\sqrt{d}\,\omega(\pi\,2^{-k})\Bigg(\sum_{2^{k-1}\le|m|<2^k}|f_\mu(m)-f_\nu(m)|^2\Bigg)^{1/2}.$$

Performing summation over all $k \geq 1$ and applying Cauchy's inequality once more, we get

$$\left(\sum_{m\neq 0} |a_m|\,|f_\mu(m) - f_\nu(m)|\right)^2 \leq \sum_{k=1}^{\infty} \sum_{2^{k-1}\leq |m|<2^k} d\omega^2(\pi\, 2^{-k})|f_\mu(m) - f_\nu(m)|^2$$
$$\leq d \sum_{m\neq 0} \omega^2\left(\frac{\pi}{|m|}\right) |f_\mu(m) - f_\nu(m)|^2.$$

Thus, the last expression may be used as an upper bound for the square of the left-hand side of (6.2). Taking the supremum over all admissible functions u as in (5.2), we arrive at the following Fourier analytic inequality.

Proposition 6.1 *Given two probability measures μ and ν on $Q^d = (-\pi, \pi]^d$,*

$$\widetilde{\mathrm{W}}_\omega(\mu, \nu) \leq \sqrt{d} \left(\sum_{m\neq 0} \omega^2\left(\frac{\pi}{|m|}\right) |f_\mu(m) - f_\nu(m)|^2\right)^{1/2}. \tag{6.6}$$

It would be interesting to explore whether or not the coefficient $\sqrt{d}$ may be removed from the right-hand side of (6.6) (which is indeed possible for the Euclidean metric corresponding to the linear modulus of continuity $\omega(\delta) = \delta$). As was noticed to us by S. Sodin, some improvement might be achieved, if we perform averaging over θ not in the inequality (6.4), but on the earlier step such as (6.3) with $h = r\theta$. To this aim, we need the following:

Lemma 6.2 *If $|v|^2 \leq d/2$, $v \in \mathbb{R}^d$, then*

$$\int_{S^{d-1}} \sin^2(v\cdot\theta)\, d\sigma_{d-1}(\theta) \geq \frac{|v|^2}{2d}. \tag{6.7}$$

Proof Let ξ be a random variable with $\mathbb{E}\xi^2 = 1$ and $\beta_4 = \mathbb{E}\xi^4 < \infty$. The function $u(t) = \mathbb{E}\sin^2(t\xi)$ satisfies $u(0) = u'(0) = u'''(0) = 0$, $u''(0) = 2$. Since also $u^{(\mathrm{iv})}(t) = -8\,\mathbb{E}\,\xi^4\cos(2t\xi)$, necessarily $|u^{(\mathrm{iv})}(t)| \leq 8\beta_4$. Hence, by Taylor's formula, for some point t_1 between zero and t,

$$u(t) = t^2 + \frac{u^{(\mathrm{iv})}(t_1)}{4!}\,t^4 \geq t^2 - \frac{8\beta_4}{4!}\,t^4 \geq \frac{1}{2}\,t^2, \tag{6.8}$$

where in the last inequality we assume that $t^2 \leq 3/(2\beta_4)$. We apply this inequality on the unit sphere to $\xi = \sqrt{d}\,\theta_1$ in which case $\beta_4 = \frac{3d}{d+2} < 3$. Since $v\cdot\theta$ has the same distribution under σ_{d-1} as $|v|\,\xi/\sqrt{d}$, the inequality (6.7) follows from (6.8) by choosing $t = |v|/\sqrt{d}$. □

Putting $h = r\theta$, $r > 0$, and taking the average over $\theta \in S^{d-1}$ in (6.3), we get, by (6.7), that

$$\frac{r^2}{2d} \sum_{r|m| \leq \sqrt{d/2}} |a_m|^2\, |m|^2 \leq \frac{1}{4}\, \omega^2(\pi r).$$

This bound may be used for the values $r = \sqrt{d/2}\, 2^{-k}$, when it takes the form as in (6.5)

$$\sum_{|m| \leq 2^k} |a_m|^2\, 4^{-k}\, |m|^2 \leq \omega^2\left(\pi \sqrt{d/2}\, 2^{-k}\right).$$

Continuing to argue as before, we therefore arrive at the following slight sharpening of (6.6).

Proposition 6.3 *Given two probability measures μ and ν on Q^d,*

$$\widetilde{\mathrm{W}}_\omega(\mu, \nu) \leq 2 \left(\sum_{m \neq 0} \omega^2\left(\frac{\pi\sqrt{d/2}}{|m|}\right) |f_\mu(m) - f_\nu(m)|^2 \right)^{1/2}. \tag{6.9}$$

7 Smoothed Fourier Analytic Inequalities: Proof of Theorem 1.1

The inequality (6.6) may be useless, especially for discrete measures, when the sum on the right-hand side is not convergent. However, in typical approximation problems, $f_\mu(m)$ and $f_\nu(m)$ are close to each other on large integer m-balls. In this case, smoothing arguments are useful, which we first describe in a somewhat general form.

Identifying $Q^d = (-\pi, \pi]^d$ with the torus $(S^1)^d = S^1 \times \cdots \times S^1$ via the map

$$x \to \left(e^{ix_1}, \ldots, e^{ix_d}\right), \qquad x = (x_1, \ldots, x_d),$$

one may introduce the convolution operation. Namely, for any two probability measures μ and κ on Q^d, their convolution is defined as the probability measure $\tilde{\mu} = \mu * \kappa$ on Q^d satisfying

$$\int u\, d\tilde{\mu} = \iint u(x+y)\, d\mu(x)\, d\kappa(y) \tag{7.1}$$

in the class of all (2π)-periodic continuous functions u on $\mathbb{R}^d$. Applying this equality to $u(x) = e^{im\cdot x}$, it is expressed in terms of the Fourier-Stieltjes transforms as

$$f_{\tilde{\mu}}(m) = f_\mu(m)\, f_\kappa(m), \qquad m \in \mathbb{Z}^d, \tag{7.2}$$

which may also be taken as the definition of convolution.

Being supported on Q^d, the measure $\tilde{\mu}$ is close to μ in a weak sense, as long as κ is close to the delta measure at the origin. More precisely, if ρ is the metric generated by the modulus of continuity ω, and $\|u\|_{\mathrm{Lip}(\omega)} \leq 1$ (in addition to the (2π)-periodicity), it follows from (7.1) and the Jensen-type inequality (5.5) that

$$\begin{aligned}\left|\int u\,d\tilde{\mu} - \int u\,d\mu\right| &= \left|\iint (u(x+y) - u(x))\,d\mu(x)\,d\kappa(y)\right| \\ &\leq \iint \rho(x+y, x)\,d\mu(x)\,d\kappa(y) \\ &= \int \omega(|y|)\,d\kappa(y) \ \leq\ 3\,\omega\big(\mathbb{E}\,(|K|)\big),\end{aligned}$$

where the random vector K is distributed according to κ. Here, the factor 3 may be removed once ω is a concave function.

Now, let us start with a random vector $H = (H_1, \dots, H_d)$ with characteristic function $h(s) = \mathbb{E}\,(e^{is\cdot H})$, $s \in \mathbb{R}^d$, and apply the above inequality to the random vector $K = U_d(H) = (U(H_1), \dots, U(H_d))$, where the map $U : \mathbb{R} \to (-\pi, \pi]$ is defined by $U(x) = x - 2\pi k$ for $\pi(2k-1) < x \leq \pi(2k+1)$ ($k \in \mathbb{Z}$). In particular, $|U(x)| \leq |x|$ for all $x \in \mathbb{R}$ and thus $|U_d(x)| \leq |x|$ for all $x \in \mathbb{R}^d$. Hence $\mathbb{E}\,(|K|) \leq \mathbb{E}\,(|H|)$, so that

$$\left|\int u\,d\tilde{\mu} - \int u\,d\mu\right| \ \leq\ 3\,\omega\big(\mathbb{E}\,(|H|)\big). \tag{7.3}$$

On the other hand, since $U_d(x) - x = -2\pi k(x)$ for each $x \in \mathbb{R}^d$ with some $k(x) \in \mathbb{Z}^d$, the characteristic functions of H and K coincide on the lattice $\mathbb{Z}^d$:

$$\mathbb{E}\left(e^{im\cdot K}\right) = \mathbb{E}\left(e^{im\cdot U_d(H)}\right) = \mathbb{E}\left(e^{im\cdot(H - 2\pi k(H))}\right) = h(m).$$

Hence, according to (7.2), $f_{\tilde{\mu}}(m) = f_\mu(m)\,h(m)$ for all $m \in \mathbb{Z}^d$.

Now, given two probability measures μ and ν on Q^d, one can apply the triangle inequality for the metric $\widetilde{\mathrm{W}}_\omega$, to derive from (7.3) a smoothing inequality

$$\widetilde{\mathrm{W}}_\omega(\mu, \nu) \leq \widetilde{\mathrm{W}}_\omega\,(\tilde{\mu}, \tilde{\nu}) + 6\,\omega(\mathbb{E}\,|H|).$$

Here, the left distance coincides with $\mathrm{W}_\omega(\mu, \nu)$ as long as μ and ν are supported on $[0, \pi]^d$ (Lemma 5.1). Applying Proposition 6.1 to the measures $\tilde{\mu}$ and $\tilde{\nu}$, we thus obtain the following smoothed Fourier analytic inequality for the metric W_ω.

Proposition 7.1 *Let ω be a modulus of continuity. Given probability measures μ and ν on $[0, \pi]^d$, for any random vector H in $\mathbb{R}^d$ with characteristic function h,*

$$\mathrm{W}_\omega(\mu, \nu) \ \leq\ \sqrt{d}\,\bigg(\sum_{m \neq 0} \omega^2\left(\frac{\pi}{|m|}\right)\,|f_\mu(m) - f_\nu(m)|^2\,|h(m)|^2\bigg)^{1/2} + 6\,\omega\,(\mathbb{E}\,(|H|))\,. \tag{7.4}$$

Here, the constant 6 may be replaced with 2, if ω is a concave function. The factor $\sqrt{d}$ may also be removed for the canonical case $\omega(\delta) = \delta$, as was shown in [7] with a different argument.

Choosing a characteristic function h which decays sufficiently fast, the series in (7.4) will be convergent. In practice, one applies this inequality to the random vectors $H = \beta Z$ with a parameter $\beta > 0$, and then tries to optimize in $\beta > 0$ the bound

$$\mathrm{W}_\omega(\mu,\nu) \le \sqrt{d}\,\bigg(\sum_{m\neq 0} \omega^2\Big(\frac{\pi}{|m|}\Big)\,|f_\mu(m) - f_\nu(m)|^2\,|h(\beta m)|^2\bigg)^{1/2} + 6\,\omega\big(\beta\,\mathbb{E}\,(|Z|)\big), \tag{7.5}$$

where now $h(s)$ denotes the characteristic function of Z.

Proof of Theorem 1.1. As a classical smoothing probability distribution, one may take for Z in (7.5) a standard normal random vector in $\mathbb{R}^d$. In this case $h(s) = e^{-|s|^2/2}$ and $\mathbb{E}\,|Z| \le \sqrt{d}$. Replacing $\beta = \sqrt{t}$, we then get the desired inequality (1.5). □

A different choice of the smoothing distribution leads to the inequality (1.6). Assuming that the characteristic function $h(s)$ of the random vector H in $\mathbb{R}^d$ is supported on the cube $\|s\|_\infty \le T$, from (7.4) we immediately obtain that

$$\mathrm{W}_\omega(\mu,\nu) \le \sqrt{d}\,\bigg(\sum_{1\le\|m\|_\infty\le T} \omega^2\Big(\frac{\pi}{|m|}\Big)\,|f_\mu(m) - f_\nu(m)|^2\bigg)^{1/2} + 6\,\omega\big(\mathbb{E}\,(|H|)\big). \tag{7.6}$$

Our aim is now to indicate a specific h with this property and to estimate the expectation on the right-hand side of (7.6).

Let us start with $v(s) = (1 - 2\,|s|)^+$, $s \in \mathbb{R}$, which represents the characteristic function of the probability measure on the real line with density

$$q(x) = \frac{2(1-\cos(x/2))}{\pi x^2}, \quad x \in \mathbb{R}.$$

Although v is compactly supported, the second moment of the measure with density q is infinite (and the first moment is also infinite). So, we consider the convolution

$$w(s) = \frac{1}{c}\,(v * v)(s) = \frac{1}{c}\int_{-\infty}^{\infty} v(s-r)v(r)\,dr, \tag{7.7}$$

where c is a normalizing constant corresponding to the condition $w(0) = 1$, i.e.,

$$c = \int_{-\infty}^{\infty} v(s)^2\,ds = 2\int_0^{1/2} (1-2s)^2\,ds = \frac{1}{3}.$$

Clearly, w is also a characteristic function, namely – for the density $\frac{2\pi}{c}\,q(x)^2$ (by Plancherel's theorem), which has a finite second moment.

Now, since v is supported on $(-\frac{1}{2}, \frac{1}{2})$, the function w is supported on the interval $(-1, 1)$. If ξ is a random variable with characteristic function w, then $\mathbb{E}(\xi^2) =$

$-w''(0)$. Using the property that the function $v'(s) = -2\,\mathrm{sign}(s)\,1_{\{|s|<1/2\}}$ serves as a Radon-Nikodym derivative of v, from (7.7) we obtain that

$$w'(s) = 3\int_{-\infty}^{\infty} v'(s-r)v(r)\,dr = 3\int_{-\infty}^{\infty} v(s-r)v'(r)\,dr$$

and

$$w''(s) = 3\int_{-\infty}^{\infty} v'(s-r)v'(r)\,dr.$$

Hence

$$\mathbb{E}(\xi^2) = -3\int_{-\infty}^{\infty} v'(-r)v'(r)\,dr = 3\int_{-\infty}^{\infty} v'(r)^2\,dr = 3\int_{-1/2}^{1/2} 4\,dr = 12.$$

Finally, consider the random vector $H = \frac{1}{T}\,(\xi_1, \dots, \xi_d)$, where $\xi_1, \dots, \xi_d$ are independent copies of ξ. It has characteristic function

$$h(s) = w(s_1/T) \cdots w(s_d/T), \qquad s = (s_1, \dots, s_d) \in \mathbb{R}^d,$$

which is vanishing in the cube $\|s\|_\infty \le T$. Hence, (7.6) is applicable, and in this case, we have $\mathbb{E}\,(|H|^2) = \frac{d}{T^2}\,\mathbb{E}\,(\xi^2) = \frac{12\,d}{T^2}$, thus proving (1.6).

Remark 7.2 A slight sharpening of (1.5) may be achieved by applying Proposition 6.3, which leads to the bound

$$\mathrm{W}_\omega(\mu, \nu) \;\le\; 2\left(\sum_{m \neq 0} \omega^2\left(\frac{\pi\sqrt{d/2}}{|m|}\right) e^{-t|m|^2}\,|f_\mu(m) - f_\nu(m)|^2\right)^{1/2} + 6\,\omega(\sqrt{dt}).$$

8 Matching and Order Statistics

Returning to the one dimensional Corollary 3.1, let us recall that, for empirical measures μ_n and ν_n as in (2.1), the Zolotarev distance is reduced to the representation

$$\xi_n \;\equiv\; \zeta_\alpha(\mu_n, \nu_n) \;=\; \frac{1}{n}\,\inf_\sigma \sum_{k=1}^{n} |X_k - Y_{\sigma(k)}|^\alpha, \tag{8.1}$$

where the infimum is running over all permutations σ of $\{1, \dots, n\}$. Moreover, if $\alpha = 1$, one may use the general formula (1.2) which further simplifies (8.1) to the explicit expression

$$\zeta_1(\mu_n, \nu_n) = \mathrm{W}(\mu_n, \nu_n) = \frac{1}{n}\sum_{k=1}^{n} |X_k^* - Y_k^*|$$

in terms of the order statistics, that is, the non-decreasing rearrangements $X_1^* \le \cdots \le X_n^*$ and $Y_1^* \le \cdots \le Y_n^*$ inside each collection of points (the randomness of μ_n and ν_n is irrelevant). One may therefore wonder, whether or not a similar description continues to hold for the expression (8.1) in the case $\alpha < 1$. However, this case turns out to be quite different. To see this, define

$$\xi_n^* = \frac{1}{n} \sum_{k=1}^n |X_k^* - Y_k^*|^\alpha.$$

By the construction, $\xi_n^* \ge \xi_n$, and we have $\xi_n^* = \xi_n$ for $\alpha = 1$. The next statement shows that ξ_n^* is asymptotically much larger than ξ_n for independent samples.

Proposition 8.1 *Let* $0 < \alpha < 1$*, and let the random variables* $X_1, \dots, X_n, Y_1, \dots, Y_n$ *be independent and have a common log-concave density on* $(0, 1)$*. If n is large enough, then* $\xi_n^* > \xi_n$ *with positive probability.*

Proof It is sufficient to see that $\mathbb{E}(\xi_n^*) > \mathbb{E}(\xi_n)$ for all n large enough. We use the following two-sided bound

$$\frac{1}{12} \le \sup_{a<x<b} p^2(x) \operatorname{Var}(X) \le 1,$$

which holds true for any random variable X, whose distribution μ is supported on an interval (a, b) and has there a log-concave density p (cf. e.g. [5], Proposition 2.1). One may rewrite these inequalities as

$$\frac{1}{\sqrt{12}\, \sup_{0<t<1} I(t)} \le \sqrt{\operatorname{Var}(X)} \le \frac{1}{\sup_{0<t<1} I(t)} \tag{8.2}$$

in terms of the associated function $I(t) = p(F^{-1}(t))$ defined via the inverse F^{-1} : $(0, 1) \to (a, b)$ of the distribution function $F(x) = \mathbb{P}\{X \le x\}$ restricted to (a, b).

If $X_1, \dots, X_n$ are drawn independently from a common law μ with density $p(x)$, the k-th order statistic X_k^* has density

$$p_k(x) = \frac{n!}{(k-1)!\,(n-k)!}\, F(x)^{k-1} \big(1 - F(x)\big)^{n-k}\, p(x), \qquad x \in \mathbb{R},$$

where $F(x) = \mu((-\infty, x])$, $x \in \mathbb{R}$, is the associated distribution function. This formula implies that, all the functions p_k are log-concave as well. Hence, by (8.2), cf. also [6], Lemma 6.1, we have

$$\frac{1}{\sqrt{12}\, \sup_{0<t<1} J_{k,n}(t) I(t)} \le \sqrt{\operatorname{Var}(X_k^*)} \le \frac{1}{\sup_{0<t<1} J_{k,n}(t) I(t)}, \tag{8.3}$$

where

$$J_{k,n}(t) \;=\; \frac{n!}{(k-1)!\,(n-k)!}\, t^{k-1}(1-t)^{n-k}.$$

For example, if $U_1, \dots, U_n$ is another sample drawn from the uniform distribution on (0, 1), then the associated I-function is equal to 1, so that for the k-th order statistic U_k^*, (8.3) gives

$$\frac{1}{\sqrt{12}\ \sup_{0<t<1} J_{k,n}(t)} \;\le\; \sqrt{\mathrm{Var}(U_k^*)} \;\le\; \frac{1}{\sup_{0<t<1} J_{k,n}(t)}. \tag{8.4}$$

To further estimate from below the left-hand side in (8.3), one may apply the upper bound in (8.2) with $X = X_1$, which leads to

$$\sqrt{\mathrm{Var}(X_k^*)} \;\ge\; \frac{1}{\sqrt{12}\ \sup_{0<t<1} J_{k,n}(t)}\, \sqrt{\mathrm{Var}(X_1)}.$$

In view of (8.4), this implies that

$$\mathrm{Var}(X_k^*) \;\ge\; \frac{1}{12}\, \mathrm{Var}(X_1)\, \mathrm{Var}(U_k^*).$$

Thus, using the log-concavity hypothesis, we have reduced the problem of the estimation of the variance of the k-th order statistic to the simple case where the sample is drawn from the uniform distribution on (0, 1). In that case, U_k^* has a beta distribution with parameters $(k, n-k+1)$, so,

$$\mathrm{Var}(U_k^*) \;=\; \frac{k(n-k+1)}{(n+1)^2\,(n+2)}.$$

Hence

$$\mathrm{Var}(X_k^*) \;\ge\; \frac{k(n-k+1)}{12\,(n+1)^2\,(n+2)}\, \mathrm{Var}(X_1) \;\ge\; \frac{c}{n}\, \mathrm{Var}(X_1),$$

where the last bound holds with an absolute constant $c > 0$ in the region $\frac{n}{4} \le k \le \frac{3n}{4}$.

Now, since the distributions of X_k^* are log-concave, so are the distributions of $X_k^* - Y_k^*$. Hence, the quantities $\|X_k^* - Y_k^*\|_q = (\mathbb{E}\,(|X_k^* - Y_k^*|^q))^{1/q}$ are equivalent to each other within an absolute factor for the region $0 \le q \le 2$. That is, with some absolute constants

$$\|X_k^* - Y_k^*\|_\alpha \;\ge\; \|X_k^* - Y_k^*\|_0 \;\ge\; c\,\|X_k^* - Y_k^*\|_2 \;=\; c\sqrt{2\,\mathrm{Var}(X_k^*)} \;\ge\; \frac{c'}{\sqrt{n}}\, \sqrt{\mathrm{Var}(X_1)}$$

for $\frac{n}{4} \leq k \leq \frac{3n}{4}$ in the last inequality. It follows that, for $n \geq 2$,

$$\begin{aligned}\mathbb{E}\,(\xi_n^*) &= \frac{1}{n}\sum_{k=1}^{n} \mathbb{E}\left(|X_k^* - Y_k^*|^\alpha\right) \;=\; \frac{1}{n}\sum_{k=1}^{n} \|X_k^* - Y_k^*\|_\alpha^\alpha \\ &\geq \frac{c}{n}\left(\mathrm{Var}(X_1)\right)^{\alpha/2} \sum_{\frac{n}{4}\leq k\leq \frac{3n}{4}} \frac{1}{n^{\alpha/2}} \;\geq\; \frac{c'}{n^{\alpha/2}}\left(\mathrm{Var}(X_1)\right)^{\alpha/2}.\end{aligned}$$

As a result, $\mathbb{E}(\xi_n^*) > \mathbb{E}(\xi_n)$ for all n large enough in view of the upper bounds on $\mathbb{E}(\xi_n)$ in Corollary 3.1. □

Remark 8.2 From another point of view, matching for one-dimensional concave costs are investigated in [8].

Acknowledgements We would like to thank S. Sodin for his note on possible improvement of the bound (1.5) with respect to the growing parameter d (as indicated in Remark 7.2; cf also Proposition 6.3). We are also grateful to referees for careful reading of the manuscript and useful comments.

References

1. Ajtai, M., Komlós, J., Tusnády, G.: On optimal matchings. Combinatorica **4**, 259–264 (1984)
2. Barthe, F., Bordenave, C.: Combinatorial optimization over two random point sets. Sémin. Probab. **45**, 483–535 (2013)
3. Bogachev, V. I., Kolesnikov, A. V.: The Monge-Kantorovich problem: achievements, connections, and prospects. (Russian) Uspekhi Mat. Nauk **67**, no. 5(407), 3–110 (2012); translation in Rus. Math. Surveys **67**(5), 785–890 (2012)
4. Bobkov, S. G.: Closeness of probability distributions in terms of Fourier-Stieltjes transforms. (Russian). Uspekhi Mat. Nauk **71**, no. 6 (432), 37—98(2016); translation in Rus. Math. Surveys 71 (2016), no. 6, 1021–1079
5. Bobkov, S.G., Chistyakov, G.: On concentration functions of random variables. J. Theor. Probab. **28**(3), 976–988 (2015)
6. Bobkov, S.G., Ledoux, M.: One-dimensional empirical measures, order statistics and Kantorovich transport distances. Memoirs AMS **261**(1259), 126 (2019)
7. Bobkov, S. G., Ledoux, M.: A simple Fourier analytic proof of the AKT optimal matching theorem. Preprint (2019)
8. Caracciolo, S., D'Achille, M. P., Erba, V., Sportiello, A.: The Dyck bound in the concave 1-dimensional random assignment model. arXiv:1904.10867 [cond-mat.dis-nn] (2019)
9. Dereich, S., Scheutzow, M., Schottstedt, R.: Constructive quantization: approximation by empirical measures. Ann. Inst. Henri Poincaré Probab. Stat. **49**, 1183–1203 (2013)
10. Dudley, R.: Real analysis and probability. Revised reprint of the 1989 original. In: Cambridge Studies in Advanced Mathematics 74. Cambridge University Press (2002)
11. Fournier, N., Guillin, A.: On the rate of convergence in Wasserstein distance of the empirical measure. Probab. Theory Relat. Fields **162**(3–4), 707–738 (2015)
12. Hahn, M., Shao, Y.: An exposition of Talagrand's mini-course on matching theorems. Probability in Banach spaces 8. Progr. Probab. **30**, 3–38 (1992)
13. Leighton, T., Shor, P.: Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. Combinatorica **9**(2), 161–187 (1989)
14. Rachev, S.T., Rüschendorf, L.: Mass Transportation Problems. Springer, Berlin (1998)
15. Rio, E.: Upper bounds for minimal distances in the central limit theorem. Ann. Inst. Henri Poincaré Probab. Stat. **45**(3), 802–817 (2009)
16. Shor, P.: How to pack better than Best Fit: Tight bounds for average-case on-line bin packing. Proc. 32nd Annual Symposium on Foundations of Computer Sciences, pp. 752–759 (1991)

17. Shor, P.W., Yukich, J.E.: Minimax grid matching and empirical measures. Ann. Probab. **19**(3), 1338–1348 (1991)
18. Talagrand, M.: The Ajtai-Komlós-Tusnády matching theorem for general measures. Probability in Banach spaces 8. Progr. Probab. **30**, 39–54 (1992)
19. Talagrand, M.: Matching theorems and empirical discrepancy computations using majorizing measures. J. Am. Math. Soc. **7**, 455–537 (1994)
20. Talagrand, M., Yukich, J.: The integrability of the square exponential transportation cost. Ann. Appl. Probab. **3**, 1100–1111 (1993)
21. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003)
22. Zolotarev, V. M.: Metric distances in spaces of random variables and of their distributions. (Russian) Mat. Sb. (N.S.) 101(143), no. 3, 416–454, 456 (1976)
23. Zolotarev, V. M.: Modern theory of summation of random variables. In: Modern Probability and Statistics. VSP, Utrecht, x+412 pp (1997)
24. Zygmund, A.: Trigonometric series: Vols. I, II. Second edition, reprinted with corrections and some additions Cambridge University Press, London-New York Vol. I. xiv+383 pp (1968)