The Annals of Applied Probability
2021, Vol. 31, No. 6, 2567–2584
https://doi.org/10.1214/20-AAP1656

A SIMPLE FOURIER ANALYTIC PROOF OF THE AKT OPTIMAL MATCHING THEOREM

BY SERGEY G. BOBKOV[1] AND MICHEL LEDOUX[2]

[1]*School of Mathematics, University of Minnesota, bobkov@math.umn.edu*
[2]*Institut de Mathématiques de Toulouse, Université de Toulouse—Paul-Sabatier, ledoux@math.univ-toulouse.fr*

We present a short and elementary proof of the Ajtai–Komlós–Tusnády (AKT) optimal matching theorem in dimension 2 via Fourier analysis and a smoothing argument. The upper bound applies to more general families of samples, including dependent variables, of interest in the study of rates of convergence for empirical measures. Following the recent pde approach by L. Ambrosio, F. Stra and D. Trevisan, we also adapt a simple proof of the lower bound.

Given two samples $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$ of independent random variables uniformly distributed on the unit square $[0, 1]^2$, the famous Ajtai–Komlós–Tusnády (AKT) optimal matching theorem [2] establishes that, with high probability,

$$\inf_\sigma \frac{1}{n} \sum_{k=1}^n |X_k - Y_{\sigma(k)}| \sim \sqrt{\frac{\log n}{n}}.$$

Here the infimum is taken over all permutations σ of $\{1, \dots, n\}$, $|\cdot|$ is the Euclidean norm in $\mathbb{R}^2$, and $A \sim B$ means that $A \le CB$ and $B \le CA$ for some constant $C > 0$ independent of n (≥ 2). The $\frac{1}{n}$ normalization is for convenience with the further statements and formulations, and for the purpose of this note, with high probability is simply translated by an equivalence on the average

$$\mathbb{E}\biggl(\inf_\sigma \frac{1}{n} \sum_{k=1}^n |X_k - Y_{\sigma(k)}|\biggr) \sim \sqrt{\frac{\log n}{n}} \tag{1}$$

(concentration arguments allowing for quantitative probabilistic estimates, cf. [3, 5]).

The AKT theorem is proved in [2] with combinatorial dyadic decompositions, where it is also mentioned that the analogous statement with the Euclidean norm at the power p, $1 \le p < \infty$, holds similarly. Further proofs, still based on the same principle, and with improved conclusions, have been provided in [20, 21] or [29]. M. Talagrand [13, 27] undertook a deep investigation of optimal matching with the tool of the ellipsoid theorem from the generic chaining (majorizing measure) theory, with significant strengthenings and further, still open, conjectures (cf. the monograph [28]). In particular, with this approach, he extended in [25] the upper bound in (1) to samples with arbitrary distribution on $[0, 1]^2$ (which may be then further extended to distributions on $\mathbb{R}^2$ under moment conditions [31]). Grid matching corresponding to $p = \infty$ has been investigated simultaneously [16, 22, 28].

For the specific uniform distribution, alternate approaches have been developed recently, such as gravitational allocation in [14]. A major breakthrough is the investigation [3] by L. Ambrosio, F. Stra and D. Trevisan who used pde methods towards exact asymptotics of the optimal matching for $p = 2$.

Received September 2019; revised December 2020.
MSC2020 subject classifications. Primary 60D05, 49Q22; secondary 62G30, 60H15, 58J35, 49J55.
Key words and phrases. Optimal matching, Ajtai–Komlós–Tusnády theorem, Fourier analysis, heat kernel smoothing, empirical measure.

The optimal matching problem may be formulated similarly for samples on the cube $[0,1]^d$ for any dimension d. The value $d = 2$ is actually known to be the critical and most delicate one (see the discussion in [28]), since when $d = 1$ monotone rearrangement arguments show that the order is $\frac{1}{\sqrt{n}}$ (cf. [5]), while when $d \geq 3$, easy tools produce the rate $\frac{1}{n^{1/d}}$; see, for example, [10] for some early achievements and [9, 12, 26] for recent more general developments concerning $d \geq 3$. We refer in particular to the latter [12] and to [28, 32] for further bibliographical references on the topic of optimal matching.

The purpose of this note is to present an elementary Fourier analytic proof of the AKT theorem (1), with in particular a very simple argument towards the upper bound, valid for any underlying distribution on $[0,1]^2$. While the use of Fourier transform is also the first step in the Talagrand investigation [27, 28] (inspired from [8]), we replace the delicate generic chaining analysis by a standard smoothing procedure.[1] This smoothing procedure is also part of the pde analysis developed in [3] towards exact asymptotics. We borrow from the latter work [3] the Lusin approximation theorem of Sobolev functions towards a simplified proof of the lower bound. The simplicity of the approach developed in this note allows for several extensions, and should potentially be useful in the study of related issues. Some applications of Fourier analysis and heat kernel smoothing in the study of Kantorovich metrics have been proposed recently in [23].

The note is structured as follows. In Section 1, we reformulate the optimal matching theorem in suitable Kantorovich metrics adapted to Fourier analysis. Next, the main Fourier analytic argument is developed, while in Section 3 the smoothing procedure is presented by means of standard Gaussian kernel regularization. The proof of the upper bound in the AKT theorem is then immediately deduced in Section 4, and shown to apply to more general samples, including dependent structures. In this formulation, the optimal matching problem is part of the study of rates of convergence of empirical measures in Kantorovich metrics. Empirical measures with nonrandom atoms are considered in Section 5, producing in particular new instances of the AKT theorem. The lower bound is established in the next paragraph. In the final Section 7, we derive more precise quantitative upper bounds taking care of the dependence of the constants as the dimension d grows, essentially recovering some claims from [24].

1. Kantorovich metric. To present the approach, it is convenient to recast the optimal matching problem in terms of the Kantorovich metric W_1. We mention, for example, [11, 18, 30] as standard references on the Kantorovich transport distances.

Given two probability measures μ and ν on the Borel sets of $\mathbb{R}^d$ with a first absolute moment, the Kantorovich transport distance $\mathrm{W}_1(\mu,\nu)$ between μ and ν is defined as

$$\mathrm{W}_1(\mu,\nu) = \inf_{\lambda} \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} |x-y|\, d\lambda(x,y), \tag{2}$$

where the infimum is running over all probability measures λ on $\mathbb{R}^d \times \mathbb{R}^d$ with respective marginals μ and ν, and $|x-y|$ represents the Euclidean distance between $x, y \in \mathbb{R}^d$. It is a standard consequence of the Birkhoff theorem on the extreme points of the set of bi-stochastic matrices that whenever $x_1,\dots,x_n, y_1,\dots,y_n$ are points in $\mathbb{R}^d$, and $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{x_k}$, $\nu_n = \frac{1}{n}\sum_{k=1}^n \delta_{y_k}$, then

$$\mathrm{W}_1(\mu,\nu) = \inf_{\sigma} \frac{1}{n}\sum_{k=1}^n |x_k - y_{\sigma(k)}|,$$

[1] Since the paper was made available, M. Talagrand gave its own presentation of this proof in the forthcoming new edition of [28].

connecting therefore with the optimal matching formulation. In particular, we will study and state below the AKT results using this Kantorovich metric W_1.

By the Kantorovich–Rubinstein theorem, the distance W_1 has another description as

$$\mathrm{W}_1(\mu,\nu)=\sup_u\left|\int_{\mathbb{R}^d}u\,d\mu-\int_{\mathbb{R}^d}u\,d\nu\right|, \tag{3}$$

where the supremum is taken over all (real-valued) Lipschitz functions u on $\mathbb{R}^d$ with Lipschitz semi-norm $\|u\|_{\mathrm{Lip}}\le 1$ with respect to the Euclidean distance on $\mathbb{R}^d$.

The aim is to bound the distance $\mathrm{W}_1(\mu,\nu)$ by means of Fourier analysis for probability measures supported on a bounded set, say $Q^d=(-\pi,\pi]^d$, which requires some additional properties of u like periodicity. This is possible, at least when μ and ν are supported on a smaller part of Q^d. In that case, any Lipschitz map u on $\mathbb{R}^d$ can indeed be modified outside the sub-cube to become periodic and to still be Lipschitz (thus not changing the difference of the integrals in (3)).

As an alternate approach, one may consider a similar problem on the torus $\mathbb{T}^d=(\mathbb{S}^1)^d$ where $\mathbb{S}^1=\{z\in\mathbb{C};|z|=1\}$ denotes the unit circle on the complex plane, endowed with the geodesic distance. The circle may be identified with the semi-open interval $(-\pi,\pi]$ with metric

$$\rho(x,y)=\min\{|x-y|,2\pi-|x-y|\},\quad x,y\in(-\pi,\pi],$$

via the isometric mapping $U(x)=e^{ix}$. In that case, $\mathbb{T}^d$ should be identified with Q^d with metric

$$\rho_d(x,y)=\left(\sum_{\ell=1}^d\rho(x_\ell,y_\ell)^2\right)^{1/2},\quad x=(x_1,\dots,x_d),y=(y_1,\dots,y_d)\in Q^d.$$

For probability measures μ and ν on Q^d, the Kantorovich transport distance with respect to ρ_d is defined similar to (2) as

$$\widetilde{\mathrm{W}}_1(\mu,\nu)=\inf_\lambda\int_{Q^d}\int_{Q^d}\rho_d(x,y)\,d\lambda(x,y). \tag{4}$$

The general Kantorovich–Rubinstein theorem holds true for the metric space (Q^d,ρ_d) as well (cf. [11, 18, 30]) and may be restated similar to (3): For any two Borel probability measures μ and ν on Q^d,

$$\widetilde{\mathrm{W}}_1(\mu,\nu)=\sup_u\left|\int_{Q^d}u\,d\mu-\int_{Q^d}u\,d\nu\right|, \tag{5}$$

where the supremum is taken over all (real-valued) maps u on Q^d with Lipschitz semi-norm $\|u\|_{\mathrm{Lip}}\le 1$ with respect to ρ_d. The 2π-periodic extension of any such function u satisfies $|u(x)-u(y)|\le\mathrm{dist}(x-y,2\pi\mathbb{Z}^d)$ for all $x,y\in\mathbb{R}^d$. In particular, u is continuous on $\mathbb{R}^d$ and has Lipschitz semi-norm at most 1 in the sense of the Euclidean distance. Conversely, any 2π-periodic function u on $\mathbb{R}^d$ with Euclidean Lipschitz semi-norm $\|u\|_{\mathrm{Lip}}\le 1$ has ρ_d-Lipschitz semi-norm at most 1 on Q^d. Indeed, using the isometric map U, it is sufficient to note that the Lipschitz property of a function on the torus is a local property, while locally the difference between the geodesic and the Euclidean metrics is negligible. Thus, the supremum in (5) may be taken over all 2π-periodic u on $\mathbb{R}^d$ with $\|u\|_{\mathrm{Lip}}\le 1$ with respect to the Euclidean distance.

It should also be clear that the supremum in the Kantorovich–Rubinstein representations may be restricted to C^∞-functions. Once u is 2π-periodic on $\mathbb{R}^d$ and 1-Lipschitz, that is, $\|u\|_{\mathrm{Lip}}\le 1$, the convolutions

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

of u with Gaussian densities represent 2π-periodic, C^∞-smooth, and 1-Lipschitz functions for any $\varepsilon > 0$. Since $\max_x |u_\varepsilon(x) - u(x)| \le d\varepsilon \to 0$ as $\varepsilon \to 0$, the function u in (5) may be replaced with u_ε's. A similar remark applies to the supremum in (3) as well.

Since ρ_d is dominated by the usual Euclidean distance, it follows from (2) and (4) that $\widetilde{\mathrm{W}}_1 \le \mathrm{W}_1$. On the other hand, $\widetilde{\mathrm{W}}_1(\mu,\nu) = \mathrm{W}_1(\mu,\nu)$ as long as both μ and ν are supported on a smaller part of the cube Q^d such as $[0,\pi]^d$ (suitable for the applications). In this case all measures λ with marginals μ and ν have to be supported on $[0,\pi]^d \times [0,\pi]^d$, and since $\rho_d(x,y) = |x-y|$ in this sub-cube, the right-hand sides of (2) and (4), and therefore the right-hand sides of (3) and (5), do coincide.

It is a consequence of this analysis, together with elementary scaling, that we may investigate the AKT theorem via the metric $\widetilde{\mathrm{W}}_1$ described by (4) and (5). This observation will be used implicitly throughout the exposition.

2. Fourier transform. For a probability measure μ on the cube Q^d, its Fourier–Stieltjes transform is defined as the multi-indexed sequence

$$f_\mu(m) = \int_{Q^d} e^{i\langle m,x\rangle}\, d\mu(x), \quad m \in \mathbb{Z}^d,$$

where $\langle m,x\rangle = \sum_{\ell=1}^d m_\ell x_\ell$, $m = (m_1,\dots,m_d) \in \mathbb{Z}^d$, $x \in (x_1,\dots,x_d) \in \mathbb{R}^d$, which determines μ in a unique way. Equivalently, f_μ represents the characteristic function of a random vector distributed according to μ, which is restricted to the lattice $\mathbb{Z}^d$. Therefore, when bounding various distances between two probability measures μ and ν on Q^d, it is sufficient to examine closeness of their Fourier transforms f_μ and f_ν.

If a 2π-periodic function u on $\mathbb{R}^d$ is sufficiently smooth, one may expand it as an absolutely convergent Fourier series

$$u(x) = \sum_{m\in\mathbb{Z}^d} a_m e^{i\langle m,x\rangle}, \quad x \in \mathbb{R}^d,$$

which can be differentiated term by term. Differentiating this equality with respect to the ℓth coordinate, we have $\partial_\ell u(x) = i\sum_{m\in\mathbb{Z}^d} m_\ell a_m e^{i\langle m,x\rangle}$, which, according to the Parseval identity, yields

$$\frac{1}{(2\pi)^d}\int_{Q^d} |\partial_\ell u(x)|^2\, dx = \sum_{m\in\mathbb{Z}^d} m_\ell^2 |a_m|^2.$$

Summing over $\ell = 1,\dots,d$,

$$\frac{1}{(2\pi)^d}\int_{Q^d} |\nabla u(x)|^2\, dx = \sum_{m\in\mathbb{Z}^d} |m|^2 |a_m|^2,$$

where $|m|^2 = \langle m,m\rangle$. Moreover, if (additionally) u is 1-Lipschitz, the modulus $|\nabla u|$ of its gradient is everywhere less than or equal to 1, hence the left-hand side of the preceding is bounded by 1 so that

$$\sum_{m\in\mathbb{Z}^d} |m|^2 |a_m|^2 \le 1. \tag{6}$$

Now, by integration,

$$\int_{Q^d} u\, d\mu - \int_{Q^d} u\, d\nu = \sum_{m\ne 0} a_m\big[f_\mu(m) - f_\nu(m)\big].$$

At this point, the analysis of [28] makes use of tools from the study of stochastic processes. We follow a simpler direct route. Applying Cauchy's inequality on the basis of (6), we arrive at

$$\left|\int_{Q^d} u\,d\mu - \int_{Q^d} u\,d\nu\right|^2 \leq \sum_{m\neq 0} \frac{1}{|m|^2}\big|f_\mu(m) - f_\nu(m)\big|^2.$$

Take then the supremum over all (sufficiently) smooth 2π-periodic Lipschitz functions u on the left-hand side with $\|u\|_{\mathrm{Lip}} \leq 1$ to reach the following statement.

LEMMA 1. *Given two probability measures μ and ν on Q^d with Fourier–Stieltjes transforms f_μ and f_ν,*

$$\widetilde{\mathrm{W}}_1(\mu,\nu)^2 \leq \sum_{m\neq 0} \frac{1}{|m|^2}\big|f_\mu(m) - f_\nu(m)\big|^2.$$

A similar inequality holds for $\mathrm{W}_1(\mu,\nu)$ if μ and ν are supported on $[0,\pi]^d$.

It could be mentioned that if μ and ν have respective smooth densities φ and ψ with respect to dx, then

$$\sum_{m\neq 0} \frac{1}{|m|^2}\big|f_\mu(m) - f_\nu(m)\big|^2 = \frac{1}{(2\pi)^d}\int_{Q^d}\big|\nabla(-\Delta)^{-1}(\varphi-\psi)\big|^2\,dx, \tag{7}$$

where, for a convergent Fourier series $g = \sum_{m\in\mathbb{Z}^d} a_m e^{i\langle m,x\rangle}$ such that $a_0 = 0$,

$$(-\Delta)^{-1}g = \sum_{m\neq 0}\frac{1}{|m|^2}a_m e^{i\langle m,x\rangle}.$$

The quantity on the right-hand side of (7) may be identified as an inverse Sobolev-type norm (cf. [30]). In case one of the measures is the Lebesgue measure, for example, $\nu = dx$, it may be shown (see [15, 17, 19]) that

$$\widetilde{\mathrm{W}}_2(\mu,\nu)^2 = \inf_\lambda \int_{Q^d}\int_{Q^d} \rho_d(x,y)^2\,d\lambda(x,y) \leq \frac{4}{(2\pi)^d}\int_{Q^d}\big|\nabla(-\Delta)^{-1}(\varphi-1)\big|^2\,dx.$$

In this instance ($\nu = dx$), the argument and upper bound developed next for the $\widetilde{\mathrm{W}}_1$ distance will therefore apply simultaneously to the quadratic Kantorovich distance $\widetilde{\mathrm{W}}_2$, and thus to the AKT theorem (1) for $p = 2$.

One negative issue about the inequality of Lemma 1 is that the sum therein may be divergent. To settle the problem, one may use a smoothing operation by suitable convolutions of μ and ν.

3. Smoothing. We make use of the simple Gaussian heat kernel smoothing, along the line of what is developed in [3] (towards more ambitious aims), although other convolution kernels might be used to this task.

On Q^d, consider the heat kernel

$$p_t(x) = \frac{1}{(2\pi)^d}\sum_{m\in\mathbb{Z}^d} e^{i\langle m,x\rangle - |m|^2 t}, \quad t > 0, x \in Q^d.$$

In other words, p_t is the density (with respect to the Lebesgue measure) of the probability measure γ_t supported on Q^d whose Fourier–Stieltjes transform is given by

$$f_{\gamma_t}(m) = e^{-|m|^2 t}, \quad m \in \mathbb{Z}^d.$$

In particular, $\int_{Q^d} p_t(x)\,dx = 1$.

If μ is a probability measure supported on Q^d, the heat kernel smoothed (probability) measure μ_t, $t > 0$, is defined as the convolution $\mu * \gamma_t$ via the equality

$$\int_{Q^d} g\,d\mu_t = \int_{Q^d}\int_{Q^d} g(x+y)\,p_t(y)\,dy\,d\mu(x)$$

holding for all 2π-periodic continuous functions g on $\mathbb{R}^d$. Therefore, if f_μ is the characteristic function of μ, for every $m \in \mathbb{Z}^d$,

$$f_{\mu_t}(m) = \int_{Q^d} e^{i\langle m,x\rangle}\,d\mu_t(x) = e^{-|m|^2 t} f_\mu(m). \tag{8}$$

The task is now to control the cost in regularization for the Kantorovich metric. If $u : \mathbb{R}^d \to \mathbb{R}$ is 1-Lipschitz and 2π-periodic, consider

$$\int_{Q^d} u\,d\mu_t - \int_{Q^d} u\,d\mu = \int_{Q^d}\int_{Q^d} \big[u(x+y) - u(x)\big]\,p_t(y)\,dy\,d\mu(x).$$

Hence

$$\left|\int_{Q^d} u\,d\mu_t - \int_{Q^d} u\,d\mu\right| \le \int_{Q^d} |y|\,p_t(y)\,dy$$

and, taking the supremum over all such Lipschitz functions u,

$$\widetilde{\mathrm{W}}_1(\mu,\mu_t) \le \int_{Q^d} |y|\,p_t(y)\,dy \le \left(\int_{Q^d} |y|^2\,p_t(y)\,dy\right)^{1/2}.$$

The decay as $t \to 0$ of the expression on the right-hand side actually turns out to be of the order of $\sqrt{t}$. To verify this claim, note that γ_t can be recognized as the product measure whose marginals are the image of the Gaussian measure on the real line with mean zero and variance $2t$ under the map

$$M(y) = y - 2\pi k, \qquad \pi(2k-1) < y \le \pi(2k+1), \quad k \in \mathbb{Z}.$$

Indeed, this map pushes forward any probability measure η on $\mathbb{R}$ to a probability measure $\tilde{\eta}$ on $(-\pi,\pi]$. By the construction, $M(y) - y$ is a multiple of 2π, so $f_{\tilde{\eta}}(m) = f_\eta(m)$ for all $m \in \mathbb{Z}$. In addition, $|M(y)| \le |y|$ for all $y \in \mathbb{R}$, so that

$$\int_{-\infty}^{\infty} |y|^2\,d\tilde{\eta}(y) \le \int_{-\infty}^{\infty} |y|^2\,d\eta(y).$$

Choosing for η the centered Gaussian measure on the real line with variance $2t$, we obtain in this way the one-dimensional marginal measure with density p_t on $(-\pi,\pi]$. Moreover, as a consequence of the preceding comparison along each coordinate,

$$\int_{Q^d} |y|^2\,p_t(y)\,dy \le \int_{\mathbb{R}^d} |y|^2\,d\eta^{\otimes d}(y) = 2dt.$$

As a conclusion of this analysis, for any μ supported on Q^d and any $t > 0$,

$$\widetilde{\mathrm{W}}_1(\mu,\mu_t) \le \sqrt{2dt}. \tag{9}$$

We next combine the various steps. By the triangle inequality for $\widetilde{\mathrm{W}}_1$ and (9), for any $t > 0$,

$$\widetilde{\mathrm{W}}_1(\mu,\nu) \le \widetilde{\mathrm{W}}_1(\mu_t,\nu_t) + 2\sqrt{2dt}.$$

It remains to apply the Fourier bound from Lemma 1 to μ_t and ν_t which satisfy (8) to reach the following conclusion.

PROPOSITION 2. *Given two probability measures μ and ν on Q^d with Fourier–Stieltjes transforms f_μ and f_ν, for any $t > 0$,*

$$\widetilde{\mathrm{W}}_1(\mu, \nu) \le \Big(\sum_{m \neq 0} \frac{1}{|m|^2} e^{-2|m|^2 t} \big| f_\mu(m) - f_\nu(m) \big|^2 \Big)^{1/2} + 2\sqrt{2dt}.$$

A similar inequality holds for $\mathrm{W}_1(\mu, \nu)$ *if μ and ν are supported on* $[0, \pi]^d$.

4. Application to the AKT theorem. This section describes the application of the preceding Fourier analytic approach to the upper bound in the AKT theorem. It actually applies to a somewhat extended probabilistic setting, a form of which having already been emphasized in [25].

Namely, consider random variables $X_1, \dots, X_n, Y_1, \dots, Y_n$ on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with values in $[0, 1]^d$ such that the couples $(X_1, Y_1), \dots, (X_n, Y_n)$ are pairwise independent and, for every $k = 1, \dots, n$, X_k and Y_k have the same distribution. Apply Proposition 2 to the empirical measures $\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$ and $\nu_n = \frac{1}{n} \sum_{k=1}^n \delta_{Y_k}$ (supported on $[0, 1]^d \subset [0, \pi]^d$) to get that, after averaging and use of Jensen's inequality,

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n, \nu_n)\big) \le \Big(\sum_{m \neq 0} \frac{1}{|m|^2} e^{-2|m|^2 t} \mathbb{E}\big(\big| f_{\mu_n}(m) - f_{\nu_n}(m) \big|^2 \big) \Big)^{1/2} + 2\sqrt{2dt}$$

for any $t > 0$. Now, by the independence and equidistribution assumptions on the variables $X_1, \dots, X_n, Y_1, \dots, Y_n$,

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

for every $m \in \mathbb{Z}^d$, so that

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n, \nu_n)\big) \le \frac{2}{\sqrt{n}} \Big(\sum_{m \neq 0} \frac{1}{|m|^2} e^{-2|m|^2 t} \Big)^{1/2} + 2\sqrt{2dt}. \tag{10}$$

From a (crude) comparison between series and integral, it should be clear without computations that, up to d-dependent factors,

$$S_d(t) = \sum_{m \neq 0} \frac{1}{|m|^2} e^{-2|m|^2 t} \sim \int_{|x| \ge 1} \frac{1}{|x|^2} e^{-2t|x|^2} \, dx \sim \int_1^\infty r^{d-3} e^{-2tr^2} \, dr. \tag{11}$$

For the small values of $t > 0$, the latter integral is of order 1 if $d = 1$, $\log(\frac{1}{t})$ if $d = 2$ and $t^{-(d/2)+1}$ if $d \ge 3$. After optimization in $t > 0$ in (10), we thus conclude to the following statement which covers the upper bound in the AKT theorem when $d = 2$, providing at the same time the optimal rates for $d = 1$ and $d \ge 3$.

THEOREM 3. *Let $X_1, \dots, X_n, Y_1, \dots, Y_n$ be random variables with values in $[0, 1]^d$ such that the couples $(X_1, Y_1), \dots, (X_n, Y_n)$ are pairwise independent and, for every $k = 1, \dots, n$, X_k and Y_k have the same distribution. For the empirical measures $\mu_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$ and $\nu_n = \frac{1}{n} \sum_{k=1}^n \delta_{Y_k}$ associated to the samples $(X_1, \dots, X_n)$ and $(Y_1, \dots, Y_n)$, it holds true that*

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n, \nu_n)\big) = \begin{cases} O\Big(\dfrac{1}{\sqrt{n}}\Big) & \text{if } d = 1, \\ O\Big(\sqrt{\dfrac{\log n}{n}}\Big) & \text{if } d = 2, \\ O\Big(\dfrac{1}{n^{1/d}}\Big) & \text{if } d \ge 3. \end{cases}$$

In the last Section 7, we develop a more careful analysis of the function $S_d(t)$ of (11) to reach more explicit quantitative bounds, in particular with respect to dependence as the dimension d increases. Namely, Proposition 6 below with $\delta = \frac{2}{\sqrt{n}}$ yields the following quantitative statement of Theorem 3:

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n, \nu_n)\big) \le \begin{cases} \dfrac{2}{\sqrt{n}} & \text{if } d = 1, \\ 10\sqrt{\dfrac{1+\log n}{n}} & \text{if } d = 2, \\ \dfrac{16\sqrt{d}}{n^{1/d}} & \text{if } d \ge 3. \end{cases} \tag{12}$$

The numerical constants are not sharp, but the order of growth as $d \to \infty$ matches the first order asymptotics of [24].

If the random variables $X_1, \dots, X_n, Y_1, \dots, Y_n$ are independent and have the same law μ, then by Jensen's inequality

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n, \nu_n)\big) \ge \mathbb{E}\big(\mathrm{W}_1(\mu_n, \mu)\big)$$

since $\mathbb{E}(\nu_n) = \mu$. The upper bounds of Theorem 3 and (12) thus apply to $\mathbb{E}(\mathrm{W}_1(\mu_n, \mu))$. As such, the conclusions enter the framework of rates of convergence for empirical measures.

As an example illustrating Theorem 3, one may consider two sequences

$$X_k(\omega) = U(k\omega_1 + \omega_2), \qquad Y_k(\omega) = V(k\omega_1 + \omega_2), \quad \omega = (\omega_1, \omega_2) \in \Omega, k \ge 1,$$

defined for given 1-periodic Borel measurable functions $U, V : [0,1] \to [0,1]^d$ on the square $\Omega = [0,1] \times [0,1]$, which we equip with the normalized Lebesgue measure $\mathbb{P}$. As easy to check, $(X_k)_{k \ge 1}$ forms a strictly stationary sequence of pairwise independent random variables (which, however, are not independent), and the same is true for $(Y_k)_{k \ge 1}$. If U and V have equal distributions under the Lebesgue measure on $[0,1]$, then Theorem 3 is applicable, so that one can make the conclusion about the closeness of the associated empirical measures.

One may even further generalize Theorem 3 to the setting of weakly dependent random variables. Recall that, given a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and two σ-algebras $\mathcal{A}_1, \mathcal{A}_2 \subset \mathcal{A}$, the Rosenblatt coefficient, which quantifies the strength of dependence between $\mathcal{A}_1$ and $\mathcal{A}_2$, is defined to be

$$\alpha(\mathcal{A}_1, \mathcal{A}_2) = \sup\big\{\big|\mathbb{P}(A_1 \cap A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)\big|; A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\big\}.$$

It is one of eight well-known measures of dependence (and the weakest one) which is used in the theory of strong mixing conditions (cf. [6, 7]). Clearly,

$$\alpha(\mathcal{A}_1, \mathcal{A}_2) = \sup\big|\mathrm{Cov}(\varphi, \psi)\big|,$$

where the supremum is running over all $\mathcal{A}_1$- and respectively $\mathcal{A}_2$-measurable functions φ and ψ on Ω with values in $[0,1]$. If φ and ψ are complex-valued with $|\varphi| \le 1$ and $|\psi| \le 1$, then, by the bilinearity of the covariance functional, $\mathrm{Cov}(\varphi, \psi) = \mathbb{E}((\varphi - \mathbb{E}(\varphi))(\overline{\psi} - \mathbb{E}(\overline{\psi})))$ is bounded in absolute value by $16\alpha(\mathcal{A}_1, \mathcal{A}_2)$.

In practice, one is given a sequence of σ-algebras $\mathcal{A}_k$ generated by random elements Z_k, $k \ge 1$, defined on the same probability space Ω, with which one associates the characteristics

$$\alpha(\ell) = \sup_{|j-k| \ge \ell} \alpha(\mathcal{A}_j, \mathcal{A}_k), \quad \ell \ge 1.$$

Repeating the arguments in the proof of Theorem 3, we have:

COROLLARY 4. *Let* $Z_k = (X_k, Y_k)$, $k \geq 1$, *be random variables with values in* $[0,1]^d \times [0,1]^d$ *such that* X_k *and* Y_k *have the same distribution for every* $k \geq 1$. *If the associated mixing sequence* $\alpha(\ell)$, $\ell \geq 1$, *is summable, then the asymptotic bounds of Theorem 3 remain to hold for the empirical measures* $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ *and* $\nu_n = \frac{1}{n}\sum_{k=1}^n \delta_{Y_k}$.

Indeed, for every $m \in \mathbb{Z}^d$,

$$\begin{aligned}\mathbb{E}\big(\big|f_{\mu_n}(m) - f_{\nu_n}(m)\big|^2\big) &= \frac{1}{n^2}\sum_{j,k=1}^n \mathrm{Cov}\big(e^{i\langle m,X_j\rangle} - e^{i\langle m,Y_j\rangle}, e^{i\langle m,X_k\rangle} - e^{i\langle m,Y_k\rangle}\big)\\ &\leq \frac{4}{n} + \frac{64}{n^2}\sum_{1\leq j\neq k\leq n} \alpha\big(|j-k|\big)\\ &= \frac{4}{n} + \frac{128}{n^2}\sum_{\ell=1}^{n-1}(n-\ell)\alpha(\ell)\\ &\leq \frac{4}{n} + \frac{128}{n}\sum_{\ell=1}^{n-1}\alpha(\ell).\end{aligned}$$

It therefore remains to apply Proposition 2 as for Theorem 3.

5. Empirical measures with nonrandom atoms. As another application of the preceding approach, fix a collection of points in the unit cube $[0,1]^d$, say $x_1, \dots, x_N$, $N \geq 2$. One may use various selections of indices to construct (deterministic) empirical measures with atoms at x_j (repetition of the points in the sequence is allowed). Namely, for $1 \leq n \leq N$, let $\mathcal{G}_n$ denote the collection of all subsets τ of $\{1, \dots, N\}$ of cardinality $|\tau| = n$ equipped with the uniform probability measure π_n. With every $\tau \in \mathcal{G}_n$, we associate an "empirical" measure

$$\mu_\tau = \frac{1}{n}\sum_{j\in\tau}\delta_{x_j},$$

which may be treated as a random measure on the probability space $(\mathcal{G}_n, \pi_n)$. The goal is to show that most of μ_τ's are concentrated around the average measure

$$\mu = \mathbb{E}_{\pi_n}(\mu_\tau) = \int_{\mathcal{G}_n} \mu_\tau \, d\pi_n(\tau) = \frac{1}{N}\sum_{j=1}^N \delta_{x_j} \tag{13}$$

as long as n is large (in the sense of the distance W_1). For simplicity, we skip the parameter N since interest in the final estimates is concerned with the dependence with respect to the growing n, while N may be arbitrarily large. To this aim, consider the functional

$$L_u(\tau) = \int_{\mathcal{G}_n} u(x)\, d\mu_\tau(x) = \frac{1}{n}\sum_{j\in\tau} u(x_j), \qquad \tau \in \mathcal{G}_n,$$

associated to a given complex-valued function u on the cube $[0,1]^d$. As is easy to check,

$$\mathrm{Var}_{\pi_n}(L) = \mathbb{E}_{\pi_n}\big(\big|L - \mathbb{E}_{\pi_n}(L)\big|^2\big) = \frac{N-n}{2nN^2(N-1)}\sum_{i,j=1}^N \big|u(x_i) - u(x_j)\big|^2.$$

If $|u| \leq 1$, it follows that $\mathrm{Var}_{\pi_n}(L_u) \leq \frac{2}{n}$. Since the Fourier–Stieltjes transform $f_{\mu_\tau}(\pi m)$ corresponds to $L_u(\tau)$ with $u(x) = e^{i\pi\langle m,x\rangle}$, the analysis of the preceding section may be developed in the same way. Together with the more quantitative estimates from Proposition 6 below, the following corollary holds true.

COROLLARY 5. *Given a collection of points $x_1, \dots, x_N$ in $[0,1]^d$, for any integer $1 \le n \le N$, the empirical measures μ_τ satisfy*

$$\mathbb{E}_{\pi_n}\big(\mathrm{W}_1(\mu_\tau, \mu)\big) \le \begin{cases} \sqrt{\dfrac{2}{n}} & \text{if } d = 1, \\ 8\sqrt{\dfrac{1+\log(2n)}{n}} & \text{if } d = 2, \\ \dfrac{13\sqrt{d}}{n^{1/d}} & \text{if } d \ge 3. \end{cases}$$

Note in particular that if $N = 2n$,

$$\mathbb{E}_{\pi_n}\big(\mathrm{W}_1(\mu_\tau, \mu)\big) = \frac{1}{2n}\mathbb{E}\left(\inf \sum_{k=1}^n |x_{i_k} - x_{j_k}|\right),$$

where the averaging on the right is performed over all choices of indices $1 \le i_1 < \cdots < i_n \le 2n$, while the infimum is taken over all permutations $j_1, \dots, j_n$ of the remaining integers in the set $\{1, \dots, 2n\} \setminus \{i_1, \dots, i_n\}$.

In fact, the preceding corollary easily implies Theorem 3 specialized to the i.i.d. case. This is achieved by averaging (13) over $x_1, \dots, x_{2n}$ according to the product measure $\mu^{\otimes 2n}$ for a fixed probability distribution μ on $[0,1]^d$. Actually, the argument extends to more general classes. Namely, if the joint distribution of the random vectors $X_1, \dots, X_{2n}$ with values in $[0,1]^d$ is invariant under permutations of the indices, then for the empirical measures $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ and μ the distribution of X_1, $\mathbb{E}(\mathrm{W}_1(\mu_n, \mu))$ is controlled as in Corollary 5.

6. Lower bound. While the AKT upper bound may be extended to families of samples with arbitrary (compactly supported) distributions, it is well known (cf. e.g., [4, 28]) that the lower bound requires distributions with enough regularity, for example, absolutely continuous with respect to Lebesgue measure. A pde proof of the lower bound in the AKT theorem has been provided recently in the paper [3], relying on a somewhat heavy analysis involving in particular Riesz transform bounds. We extract here the necessary argument in our framework via a simple fourth moment computation, thereby producing a rather mild proof.

Let $X_1, \dots, X_n, Y_1, \dots, Y_n$, on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, be independent with uniform distribution $d\mu = \frac{dx}{(2\pi)^d}$ on Q^d. For any $t > 0$, contractivity of the Kantorovich metric shows that

$$\widetilde{\mathrm{W}}_1(\mu_n, \nu_n) \ge \widetilde{\mathrm{W}}_1(\mu_{n,t}, \nu_{n,t}).$$

This is actually immediate from the definition of $\widetilde{\mathrm{W}}_1$ and the heat kernel regularization since

$$\int_{Q^d} u\, d\mu_{n,t} - \int_{Q^d} u\, d\mu = \int_{Q^d}\left[\int_{Q^d} u(x+y)\, d\mu_n(x) - \int_{Q^d} u(x+y)\, d\mu(x)\right] p_t(y)\, dy$$

and $\int_{Q^d} p_t(y)\, dy = 1$.

From here, the principle of the proof will be to lower bound, by the Kantorovich–Rubinstein theorem, $\widetilde{\mathrm{W}}_1(\mu_{n,t}, \nu_{n,t})$ by

$$\frac{1}{\alpha}\left|\int_{Q^d} u\, d\mu_{n,t} - \int_{Q^d} u\, d\nu_{n,t}\right|$$

for a well chosen α-Lipschitz function u (i.e., $\|u\|_{\mathrm{Lip}} \le \alpha$). This function u will be constructed from the absolutely convergent random Fourier series

$$h(x) = \sum_{m \ne 0} \frac{1}{|m|^2} e^{-|m|^2 t}\left(\frac{1}{n}\sum_{k=1}^n \big[e^{i\langle m, X_k\rangle} - e^{i\langle m, Y_k\rangle}\big]\right) e^{-i\langle m, x\rangle}, \quad x \in \mathbb{R}^d.$$

This equality defines a 2π-periodic, real-valued, C^∞-smooth function, whose Laplacian

$$\Delta h(x) = -\sum_{m\neq 0} e^{-|m|^2 t}\left(\frac{1}{n}\sum_{k=1}^{n}\left[e^{i\langle m,X_k\rangle} - e^{i\langle m,Y_k\rangle}\right]\right)e^{-i\langle m,x\rangle}$$

represents the multiple Fourier series for the density of $\nu_{n,t} - \mu_{n,t}$ (with respect to the Lebesgue measure on Q^d). Hence, the integration by parts formula for a smooth 2π-periodic function $v : \mathbb{R}^d \to \mathbb{R}$ yields

$$\int_{Q^d} v\,d\mu_{n,t} - \int_{Q^d} v\,d\nu_{n,t} = -\int_{Q^d} v\Delta h\,d\mu = \int_{Q^d}\langle\nabla h, \nabla v\rangle\,d\mu. \tag{14}$$

For $\alpha > 0$, denote then by $u : Q^d \to \mathbb{R}$ the α-Lipschitz Lusin extension of h on the torus (Q^d, ρ_d) such that

$$\mu(\{h \neq u\}) \leq \frac{K}{\alpha^2}\int_{Q^d}|\nabla h|^2\,d\mu, \tag{15}$$

where $K > 0$ only depends on d ([1], cf. Lemma 5.1 in [3]). Hence, as announced, by the Kantorovich–Rubinstein theorem,

$$\widetilde{\mathrm{W}}_1(\mu_n, \nu_n) \geq \widetilde{\mathrm{W}}_1(\mu_{n,t}, \nu_{n,t}) \geq \frac{1}{\alpha}\left|\int_{Q^d} u\,d\mu_{n,t} - \int_{Q^d} u\,d\nu_{n,t}\right|.$$

We next carefully investigate the difference $\int_{Q^d} u\,d\mu_{n,t} - \int_{Q^d} u\,d\nu_{n,t}$, decomposing the integrals over the set $E = \{h \neq u\}$ and its complement. Namely, by (14),

$$\begin{aligned}\int_{Q^d} u\,d\mu_{n,t} - \int_{Q^d} u d\nu_{n,t} &= \int_{Q^d}\langle\nabla h, \nabla u\rangle\,d\mu \\ &= \int_{Q^d}|\nabla h|^2\,d\mu - \int_E\langle\nabla h, \nabla h - \nabla u\rangle\,d\mu.\end{aligned}$$

In this step, it is used, as a version of Sard's lemma, that $\int_{E^c}\langle\nabla h, \nabla h - \nabla u\rangle\,d\mu = 0$ since $\nabla(h-u) = 0$ on the level set $E^c = \{h - u = 0\}$. Hence, after integration,

$$\alpha\mathbb{E}\big(\widetilde{\mathrm{W}}_1(\mu_n, \nu_n)\big) \geq \mathbb{E}\left(\int_{Q^d}|\nabla h|^2\,d\mu\right) - \mathbb{E}\left(\left|\int_E\langle\nabla h, \nabla h - \nabla u\rangle\,d\mu\right|\right). \tag{16}$$

The two expectations on the right-hand side of this inequality are examined separately. First, since the X_k, Y_k's are independent and uniformly distributed on Q^d,

$$\begin{aligned}\mathbb{E}\left(\int_{Q^d}|\nabla h|^2\,d\mu\right) &= \sum_{m\neq 0}\frac{1}{|m|^2}e^{-2|m|^2 t}\frac{1}{n^2}\mathbb{E}\left(\left|\sum_{k=1}^{n}\left[e^{i\langle m,X_k\rangle} - e^{i\langle m,Y_k\rangle}\right]\right|^2\right) \\ &= \frac{1}{n}\sum_{m\neq 0}\frac{2}{|m|^2}e^{-2|m|^2 t}.\end{aligned}$$

Denote by $c(n,t)$ this quantity (that is $\frac{2}{n}S_d(t)$ in the notation of (11)), where $t = t(n) \in (0,1)$ will be specified.

Turning to the second term on the right-hand side of (16), note first that since u is α-Lipschitz,

$$\left|\int_E\langle\nabla h, \nabla h - \nabla u\rangle\,d\mu\right| \leq \int_E|\nabla h|^2\,d\mu + \alpha\int_E|\nabla h|\,d\mu.$$

Taking expectation, by the use of the Cauchy–Schwarz and Hölder inequalities for the product measure $\mathbb{P} \otimes \mu$,

$$\mathbb{E}\left(\left|\int_E \langle \nabla h, \nabla h - \nabla u\rangle \, d\mu\right|\right)$$
$$\leq [\mathbb{E}(\mu(E))]^{1/2}\left[\mathbb{E}\left(\int_{Q^d} |\nabla h|^4 \, d\mu\right)\right]^{1/2} + \alpha[\mathbb{E}(\mu(E))]^{3/4}\left[\left(\mathbb{E}\int_{Q^d} |\nabla h|^4 \, d\mu\right)\right]^{1/4}.$$

Now, by the Lusin approximation (15),

$$\mathbb{E}(\mu(E)) \leq \frac{K}{\alpha^2} c(n,t).$$

Setting $d(n,t) = \mathbb{E}(\int_{Q^d} |\nabla h|^4 \, d\mu)$, we have therefore obtained at this stage that

$$\alpha \mathbb{E}(\widetilde{\mathrm{W}}_1(\mu_n, \mu)) \geq c(n,t) - \frac{1}{\alpha}(Kc(n,t))^{1/2} d(n,t)^{1/2} - \frac{1}{\sqrt{\alpha}}(Kc(n,t))^{3/4} d(n,t)^{1/4}. \tag{17}$$

The final task is now to suitably evaluate $d(n,t)$, before optimization of the choice of $\alpha > 0$. By the triangle inequality,

$$d(n,t) \leq 8\mathbb{E}\left(\int_{Q^d} |\nabla \widetilde{h}|^4 \, d\mu\right),$$

where $\widetilde{h}(x) = \sum_m b_m e^{-i\langle m, x\rangle}$ with

$$b_m = \frac{1}{|m|^2} e^{-|m|^2 t}\left(\frac{1}{n}\sum_{k=1}^n e^{i\langle m, X_k\rangle}\right)$$

for $m \neq 0$ and $b_0 = 0$. It holds that

$$|\nabla \widetilde{h}(x)|^2 = -\sum_{m_1, m_2 \in \mathbb{Z}^d} \langle m_1, m_2\rangle b_{m_1} b_{m_2} e^{-i\langle m_1 + m_2, x\rangle}$$

and

$$\int_{Q^d} |\nabla \widetilde{h}|^4 \, d\mu = \sum \langle m_1, m_2\rangle b_{m_1} b_{m_2} \langle m_3, m_4\rangle b_{m_3} b_{m_4},$$

where the sum is taken over $m_1, m_2, m_3, m_4 \in \mathbb{Z}^d$ such that $m_1 + m_2 + m_3 + m_4 = 0$. Now

$$\mathbb{E}(b_{m_1} b_{m_2} b_{m_3} b_{m_4}) = \frac{1}{|m_1|^2 |m_2|^2 |m_3|^2 |m_4|^2} e^{-(|m_1|^2 + |m_2|^2 + |m_3|^2 + |m_4|^2)t}$$
$$\times \frac{1}{n^4} \sum_{k_1, k_2, k_3, k_4 = 1}^n \mathbb{E}\left(e^{i\langle m_1, X_{k_1}\rangle} e^{i\langle m_2, X_{k_2}\rangle} e^{i\langle m_3, X_{k_3}\rangle} e^{i\langle m_4, X_{k_4}\rangle}\right).$$

Since the relevant indices satisfy $m_\ell \neq 0$, $\ell = 1, 2, 3, 4$, and $m_1 + m_2 + m_3 + m_4 = 0$, the last expectation is nonzero, equal to 1, only if $k_1 = k_2 = k_3 = k_4$ or if

$$\begin{cases} k_1 = k_2, & k_3 = k_4 \quad \text{and} \quad m_1 + m_2 = 0, & m_3 + m_4 = 0, \\ k_1 = k_3, & k_2 = k_4 \quad \text{and} \quad m_1 + m_3 = 0, & m_2 + m_4 = 0, \\ k_1 = k_4, & k_2 = k_3 \quad \text{and} \quad m_1 + m_4 = 0, & m_2 + m_3 = 0. \end{cases}$$

These respective contributions yield the upper bound

$$\mathbb{E}\left(\int_{Q^d} |\nabla\widetilde{h}|^4\,d\mu\right) \le \frac{1}{n^3}\sum \frac{1}{|m_1||m_2||m_3||m_4|}e^{-(|m_1|^2+|m_2|^2+|m_3|^2+|m_4|^2)t}$$
$$+\frac{3}{n^2}\sum_{m_1,m_3\neq 0}\frac{1}{|m_1|^2|m_3|^2}e^{-2(|m_1|^2+|m_3|^2)t}$$
$$= e(n,t)+\frac{3}{4}c(n,t)^2,$$

where the first sum on the right-hand side is over all $m_1, m_2, m_3, m_4 \in \mathbb{Z}^d \setminus \{0\}$.

It is easily seen that, for $0 < t \le \frac{1}{2}$,

$$e(n,t)=\frac{1}{n^3}\left(\sum_{m\neq 0}\frac{1}{|m|}e^{-|m|^2t}\right)^4 \sim \frac{1}{n^3}\left(\frac{1}{t^{(d-1)/2}}\int_{\sqrt{t}}^{\infty} r^{d-2}e^{-r^2}\,dr\right)^4$$

while (recall (11))

$$c(n,t)\sim\frac{1}{n}\frac{1}{t^{(d-2)/2}}\int_{\sqrt{2t}}^{\infty} r^{d-3}e^{-r^2}\,dr.$$

In the following, take $d=2$. Hence $e(n,t)$ is of the order of $\frac{1}{n^3t^2}$ and $c(n,t)$ of the order of $\frac{1}{n}\log(\frac{1}{t})$. Choosing $t=t(n)=\frac{1}{2\sqrt{n}}$ (e.g.,), $e(n,t)$ is negligible with respect to the square of

$$c_n = c\big(n,t(n)\big)\sim\frac{\log n}{n}. \tag{18}$$

So for this choice of $t=t(n)$, for some constant $K'>0$,

$$d\big(n,t(n)\big)=\mathbb{E}\left(\int_{Q^d}|\nabla h|^4\,d\mu\right)\le K'c_n^2. \tag{19}$$

The argument may now be concluded via optimization in $\alpha>0$. Implementing the preceding estimate (19) on $d(n,t(n))$ in (17), it follows that

$$\alpha\mathbb{E}\big(\widetilde{\mathrm{W}}_1(\mu_n,\nu_n)\big)\ge c_n-\frac{1}{\alpha}(Kc_n)^{1/2}\big(K'c_n^2\big)^{1/2}-\frac{1}{\sqrt{\alpha}}(Kc_n)^{3/4}\big(K'c_n^2\big)^{1/4}.$$

For the choice of $\alpha=\beta\sqrt{c_n}$ with $\beta>0$ large enough, it follows that

$$\mathbb{E}\big(\widetilde{\mathrm{W}}_1(\mu_n,\nu_n)\big)\ge c\sqrt{c_n}\sim\sqrt{\frac{\log n}{n}}$$

by (18). This is therefore the expected lower bound in the AKT theorem (1).

To conclude, it should be mentioned that in dimension one, the lower bound of the order of $\frac{1}{\sqrt{n}}$ is easily achieved via the monotone representation

$$\mathrm{W}_1(\mu_n,\nu_n)=\int_0^1\frac{1}{n}\left|\sum_{k=1}^n\big(\mathbb{1}_{\{X_k\le x\}}-\mathbb{1}_{\{Y_k\le x\}}\big)\right|dx$$

for independent uniform random variables $X_1,\dots,X_n,Y_1,\dots,Y_n$ on [0, 1] (cf. [5]). Hence

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n,\nu_n)\big)\ge\frac{1}{n}\mathbb{E}\left(\left|\sum_{k=1}^n(X_k-Y_k)\right|\right)$$

from which the claim follows by convergence of moments in the central limit theorem.

When $d \ge 3$, a standard argument (cf. e.g., [10]) goes as follows: for independent random variables $X_1, \dots, X_n$ with common uniform distribution μ on $[0,1]^d$. By the Kantorovich–Rubinstein representation (3) of W_1,

$$\mathrm{W}_1(\mu_n, \mu) \ge \int_{[0,1]^d} \mathrm{dist}\big(x, \{X_1, \dots, X_n\}\big)\, d\mu(x).$$

Let C_ℓ, $\ell = 1, \dots, n$, be a partition of $[0,1]^d$ into n cubes with length of order $\frac{1}{n^{1/d}}$, so that

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n, \mu)\big) \ge \sum_{\ell=1}^n \mathbb{E}\bigg(\int_{C_\ell} \mathrm{dist}\big(x, \{X_1, \dots, X_n\}\big)\, d\mu(x)\bigg).$$

If D_ℓ is the collection of cubes surrounding C_ℓ, then $\mathbb{P}(\forall k = 1, \dots, n; X_k \notin D_\ell) \ge c$ for some $c > 0$ only depending on d. As a result,

$$\mathbb{E}\big(\mathrm{W}_1(\mu_n, \mu)\big) \ge \sum_{\ell=1}^n \frac{c}{n^{1/d}} \mu(C_\ell) = \frac{c}{n^{1/d}}.$$

7. Quantitative bounds. In this last section, we briefly investigate quantitative bounds, in particular with respect to dependence on the dimensional constant d in the main statement (Theorem 3) in the form of (12). We somewhat expand the framework to cover at the same time Corollary 5.

PROPOSITION 6. *Let μ and ν be two random probability measures on the cube* $[0,1]^d$ *such that their characteristic functions satisfy* $\mathbb{E}(|f_\mu(\pi m) - f_\nu(\pi m)|^2) \le \delta^2$ *for all* $m \in \mathbb{Z}^d$ *for some* $0 \le \delta \le 2$. *Then*

$$\mathbb{E}\big(\mathrm{W}_1(\mu, \nu)\big) \le \begin{cases} \delta & \text{if } d = 1, \\ 5\delta\sqrt{1 + \log\Big(\dfrac{4}{\delta^2}\Big)} & \text{if } d = 2, \\ 10\sqrt{d}\,\delta^{2/d} & \text{if } d \ge 3. \end{cases}$$

This proposition applied with $\mu = \mu_n$, $\nu = \nu_n$ and $\delta = \frac{2}{\sqrt{n}}$ in the setting of Section 4 yields (12). Applied to $\mu = \mu_\tau$, $\nu = \mathbb{E}_{\pi_n}(\mu_\tau)$ and $\delta = \sqrt{\frac{2}{n}}$ in the setting of Section 5, it yields Corollary 5.

PROOF. Using the homogeneity of the distance W_1, one may equivalently formulate Proposition 6 for random measures μ and ν supported on the cube $[0,\pi]^d$ as

$$\mathbb{E}\big(\mathrm{W}_1(\mu, \nu)\big) \le \begin{cases} \pi\delta & \text{if } d = 1, \\ 5\pi\delta\sqrt{1 + \log\Big(\dfrac{4}{\delta^2}\Big)} & \text{if } d = 2, \\ 10\pi\sqrt{d}\,\delta^{2/d} & \text{if } d \ge 3, \end{cases}$$

under the assumption that

$$\mathbb{E}\big(\big|f_\mu(m) - f_\nu(m)\big|^2\big) \le \delta^2 \quad \text{for all } m \in \mathbb{Z}^d.$$

That is, the resulting inequalities for measures supported on the standard cube $[0,1]^d$ rather than on $[0,\pi]^d$ are obtained with numerical factors divided by π.

Given thus two random measures μ and ν on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ supported on the cube $[0,\pi]^d$ satisfying the latter, averaging the inequality of Proposition 2 yields for every $t > 0$,

$$\mathbb{E}\big(\mathrm{W}_1(\mu,\nu)\big) \le \delta\bigg(\sum_{|m|>0} \frac{1}{|m|^2} e^{-2|m|^2 t}\bigg)^{1/2} + 2\sqrt{2dt}. \tag{20}$$

The task is therefore to suitably optimize in $t > 0$.

First note that when $d = 1$, the smoothing operation is actually not needed and we may simply take $t \to 0$ to get that

$$\mathbb{E}\big(\mathrm{W}_1(\mu,\nu)\big) \le \frac{\pi}{\sqrt{3}}\delta \le \pi\delta.$$

Let us then examine more specifically the cases $d = 2$ and $d \ge 3$ analyzing the sum

$$\widetilde{S}_d(t) = S_d\bigg(\frac{t}{2}\bigg) = \sum_{|m|>0} \frac{1}{|m|^2} e^{-|m|^2 t}, \quad t > 0.$$

The function $\widetilde{S}_d(t)$ is decreasing in $t > 0$, vanishing at infinity, and

$$T_d(t) = -\widetilde{S}'_d(t) = \sum_{|m|>0} e^{-|m|^2 t} = \big(1 + T_1(t)\big)^d - 1. \tag{21}$$

In view of the monotonicity of the function $x \to e^{-tx^2}$ for $x > 0$, we have

$$\sum_{m=2}^{\infty} e^{-m^2 t} \le \int_1^{\infty} e^{-tx^2}\,dx = \frac{1}{\sqrt{2t}} \int_{\sqrt{2t}}^{\infty} e^{-y^2/2}\,dy \le \frac{\sqrt{\pi}}{2\sqrt{t}} e^{-t}.$$

Hence, for any $t > 0$,

$$T_1(t) = \sum_{m \in \mathbb{Z}\setminus\{0\}} e^{-m^2 t} = 2e^{-t} + 2\sum_{m=2}^{\infty} e^{-m^2 t} \le \bigg(2 + \sqrt{\frac{\pi}{t}}\bigg) e^{-t}.$$

Putting $a = \sqrt{t}$ and $b = (2\sqrt{t} + \sqrt{\pi})e^{-t}$, it holds that

$$\begin{aligned} t^{d/2}\big[\big(1 + T_1(t)\big)^d - 1\big] &\le (a+b)^d - a^d \\ &= \sum_{\ell=0}^{d-1} \binom{d}{\ell} a^\ell b^{d-\ell} \\ &\le (2\sqrt{t} + \sqrt{\pi})^d \sum_{\ell=0}^{d-1} \binom{d}{\ell} e^{-(d-\ell)t} \\ &\le 2^d (2\sqrt{t} + \sqrt{\pi})^d e^{-t}. \end{aligned}$$

Hence, from (21),

$$T_d(t) \le 2^d \bigg(2 + \frac{\sqrt{\pi}}{\sqrt{t}}\bigg)^d e^{-t}.$$

It follows that, in the range $t \ge \pi$, $T_d(t) \le 6^d e^{-t}$ and thus

$$\widetilde{S}_d(t) = \int_t^{\infty} T_d(s)\,ds \le 6^d e^{-t} \le 6^d e^{-\pi}.$$

On the other hand, if $t \le \pi$, then

$$T_d(t) \le 2^d \left(\frac{3\sqrt{\pi}}{\sqrt{t}}\right)^d e^{-t} \le (6\sqrt{\pi})^d t^{-d/2}$$

so that, in the case $d \ge 3$,

$$\begin{aligned}\widetilde{S}_d(t) &= \widetilde{S}_d(\pi) + \int_t^{\pi} T_d(s)\,ds \\ &\le 6^d e^{-\pi} + 2(6\sqrt{\pi})^d t^{1-(d/2)} \\ &\le 36(e^{-\pi} + 2\pi)\left(\frac{36\pi}{t}\right)^{(d/2)-1}\end{aligned}$$

while for $d = 2$,

$$\widetilde{S}_2(t) = \widetilde{S}_2(\pi) + \int_t^{\pi} T_2(s)\,ds \le 36e^{-\pi} + 36\pi \log\left(\frac{\pi}{t}\right).$$

Let us now return to (20) which states that for any $t > 0$,

$$\mathbb{E}(\mathrm{W}_1(\mu,\nu)) \le \delta\sqrt{\widetilde{S}_d(2t)} + 2\sqrt{2dt}.$$

When $d \ge 3$, choose $t = 18\pi\delta^{4/d}$ which is less than or equal to $\frac{\pi}{2}$ whenever $\delta^{2/d} \le \frac{1}{6}$ in which case

$$\begin{aligned}\mathbb{E}(\mathrm{W}_1(\mu,\nu)) &\le 6\big(\sqrt{e^{-\pi} + 2\pi} + 2\sqrt{\pi d}\big)\delta^{2/d} \\ &\le 6\big(\sqrt{e^{-\pi} + 2\pi}\sqrt{d/3} + 2\sqrt{\pi d}\big)\delta^{2/d} \\ &\le 30\sqrt{d}\,\delta^{2/d}.\end{aligned}$$

On the other hand $\mathrm{W}_1(\mu,\nu) \le \pi\sqrt{d}$ for all probability measures μ and ν supported on $[0,\pi]^d$ so that if $\delta^{2/d} \ge \frac{1}{6}$, the latter inequality is still true. A similar analysis in the case $d = 2$ yields that

$$\mathbb{E}(\mathrm{W}_1(\mu,\nu)) \le 14\delta\sqrt{1 + \log\left(\frac{4}{\delta^2}\right)}.$$

The proof of the proposition is therefore complete. □

Acknowledgments. The authors thank the reviewer for relevant suggestions of improvements in the exposition, and for pointing out a technical issue in one proof.

Funding. Research of S.B. was partially supported by NSF Grant DMS-1855575.

REFERENCES

[1] ACERBI, E. and FUSCO, N. (1984). Semicontinuity problems in the calculus of variations. *Arch. Ration. Mech. Anal.* **86** 125–145. MR0751305 https://doi.org/10.1007/BF00275731

[2] AJTAI, M., KOMLÓS, J. and TUSNÁDY, G. (1984). On optimal matchings. *Combinatorica* **4** 259–264. MR0779885 https://doi.org/10.1007/BF02579135

[3] AMBROSIO, L., STRA, F. and TREVISAN, D. (2019). A PDE approach to a 2-dimensional matching problem. *Probab. Theory Related Fields* **173** 433–477. MR3916111 https://doi.org/10.1007/s00440-018-0837-x

[4] BARTHE, F. and BORDENAVE, C. (2013). Combinatorial optimization over two random point sets. In *Séminaire de Probabilités XLV*. *Lecture Notes in Math.* **2078** 483–535. Springer, Cham. MR3185927 https://doi.org/10.1007/978-3-319-00321-4_19

[5] BOBKOV, S. and LEDOUX, M. (2019). One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.* **261** v+126. MR4028181 https://doi.org/10.1090/memo/1259

[6] BRADLEY, R. C. (2005). Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** 107–144. MR2178042 https://doi.org/10.1214/154957805100000104

[7] BRADLEY, R. C. (2007). *Introduction to Strong Mixing Conditions. Vols.* 1–3. Kendrick Press, Heber City, UT. MR2325295

[8] COFFMAN, E. G. JR. and SHOR, P. W. (1991). A simple proof of the $O(\sqrt{n}\log^{3/4} n)$ upright matching bound. *SIAM J. Discrete Math.* **4** 48–57. MR1090288 https://doi.org/10.1137/0404005

[9] DEREICH, S., SCHEUTZOW, M. and SCHOTTSTEDT, R. (2013). Constructive quantization: Approximation by empirical measures. *Ann. Inst. Henri Poincaré Probab. Stat.* **49** 1183–1203. MR3127919 https://doi.org/10.1214/12-AIHP489

[10] DUDLEY, R. M. (1968). The speed of mean Glivenko–Cantelli convergence. *Ann. Math. Stat.* **40** 40–50. MR0236977 https://doi.org/10.1214/aoms/1177697802

[11] DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge Univ. Press, Cambridge. MR1932358 https://doi.org/10.1017/CBO9780511755347

[12] FOURNIER, N. and GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Probab. Theory Related Fields* **162** 707–738. MR3383341 https://doi.org/10.1007/s00440-014-0583-7

[13] HAHN, M. G. and SHAO, Y. (1992). An exposition of Talagrand's mini-course on matching theorems. In *Probability in Banach Spaces*, 8 (*Brunswick*, *ME*, 1991). *Progress in Probability* **30** 3–38. Birkhäuser, Boston, MA. MR1227607

[14] HOLDEN, N., PERES, Y. and ZHAI, A. (2018). Gravitational allocation on the sphere. *Proc. Natl. Acad. Sci. USA* **115** 9666–9671. MR3860784 https://doi.org/10.1073/pnas.1720804115

[15] LEDOUX, M. (2017). On optimal matching of Gaussian samples. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov.* (*POMI*) **457** 226–264. MR3723584

[16] LEIGHTON, T. and SHOR, P. (1989). Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica* **9** 161–187. MR1030371 https://doi.org/10.1007/BF02124678

[17] PEYRE, R. (2018). Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and localization of Wasserstein distance. *ESAIM Control Optim. Calc. Var.* **24** 1489–1501. MR3922440 https://doi.org/10.1051/cocv/2017050

[18] RACHEV, S. T. and RÜSCHENDORF, L. (1998). *Mass Transportation Problems. Vols. I & II.* Springer, New York.

[19] SANTAMBROGIO, F. (2015). *Optimal Transport for Applied Mathematicians*: *Calculus of Variations*, *PDEs*, *and Modeling*. *Progress in Nonlinear Differential Equations and Their Applications* **87**. Birkhäuser/Springer, Cham. MR3409718 https://doi.org/10.1007/978-3-319-20828-2

[20] SHOR, P. (1991). How to pack better than Best Fit: Tight bounds for average-case on-line bin packing. In *Proc. 32nd Annual Symposium on Foundations of Computer Sciences* 752–759.

[21] SHOR, P. W. (1985). Random planar matching and bin packing. Ph.D. Thesis, M.I.T.

[22] SHOR, P. W. and YUKICH, J. E. (1991). Minimax grid matching and empirical measures. *Ann. Probab.* **19** 1338–1348. MR1112419

[23] STEINERBERGER, S. (2018). Wasserstein distance, Fourier series and applications. Available at arXiv:1803.08011v3.

[24] TALAGRAND, M. (1992). Matching random samples in many dimensions. *Ann. Appl. Probab.* **2** 846–856. MR1189420

[25] TALAGRAND, M. (1992). The Ajtai–Komlós–Tusnády matching theorem for general measures. In *Probability in Banach Spaces*, 8 (*Brunswick*, *ME*, 1991). *Progress in Probability* **30** 39–54. Birkhäuser, Boston, MA. MR1227608

[26] TALAGRAND, M. (1994). The transportation cost from the uniform measure to the empirical measure in dimension ≥ 3. *Ann. Probab.* **22** 919–959. MR1288137

[27] TALAGRAND, M. (1994). Matching theorems and empirical discrepancy computations using majorizing measures. *J. Amer. Math. Soc.* **7** 455–537. MR1227476 https://doi.org/10.2307/2152764

[28] TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes*: *Modern Methods and Classical Problems. Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics* [*Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics*] **60**. Springer, Heidelberg. MR3184689 https://doi.org/10.1007/978-3-642-54075-2

[29] TALAGRAND, M. and YUKICH, J. E. (1993). The integrability of the square exponential transportation cost. *Ann. Appl. Probab.* **3** 1100–1111. MR1241036

[30] VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

[31] YUKICH, J. (1992). Some generalizations of the Euclidean two-sample matching problem. In *Probability in Banach Spaces*, 8 (*Brunswick*, *ME*, 1991). *Progress in Probability* **30** 55–66. Birkhäuser, Boston, MA. MR1227609

[32] YUKICH, J. E. (1998). *Probability Theory of Classical Euclidean Optimization Problems*. *Lecture Notes in Math*. **1675**. Springer, Berlin. MR1632875 https://doi.org/10.1007/BFb0093472