WASSERSTEIN METRIC AND QUANTITATIVE EQUIDISTRIBUTION, I THE CASE OF TORI

E. KOWALSKI AND T. UNTRAU

Abstract. The Wasserstein distance between probability measures on tori provides a natural "invariant" quantitative measure of equidistribution, similar to the classical Erdős–Turán–Koksma inequalities, but is a more intrinsic quantity. We recall the basic properties of the Wasserstein distance and present some applications to quantitative forms of our previous work on the equidistribution of ultra-short exponential sums, and on the equidistribution of sums of additive characters over very small multiplicative subgroups.

Contents

1. Quantitative equidistribution and Wasserstein distances

Starting from the work of Weyl about a century ago, equidistribution has been a major theme of modern number theory. Besides the qualitative aspect, there is considerable interest in having quantitative versions of equidistribution theorems. In the most classical case which concerns the equidistribution modulo 1 of a sequence $(x_n)_{n\geqslant 1}$ of real numbers in $[0,1]$, equidistribution is the property that

$$\lim_{\mathrm{N}\to+\infty} \frac{1}{\mathrm{N}} |\{n \leqslant \mathrm{N} \mid a < x_n < b\}| = b - a \tag{1}$$

for any real numbers a and b, with $0 \leqslant a < b \leqslant 1$. The Weyl Criterion states that this is equivalent to the fact that the Weyl sums

$$\mathrm{W}_h(\mathrm{N}) = \frac{1}{\mathrm{N}} \sum_{n\leqslant \mathrm{N}} e(hx_n), \qquad h \in \mathbf{Z}$$

converge to 0 as $\mathrm{N} \to +\infty$ for any non-zero integer h.

The simplest quantitative forms of this equidistribution property are simply quantitative estimates for the decay to 0 of the Weyl sums, which should ideally be uniform in terms of h and N. Such estimates also provide a quantitative version of (1), by means of the classical *Erdős–Turán inequality*: if we denote

$$\mathrm{E_N}(a,b) = \frac{1}{\mathrm{N}}|\{n \leqslant \mathrm{N} \mid a < x_n < b\}|,$$

then we have

$$\sup_{0\leqslant a<b\leqslant 1} |\mathrm{E_N}(a,b) - (b-a)| \ll \frac{1}{\mathrm{T}} + \sum_{0<|h|\leqslant \mathrm{T}} \frac{|\mathrm{W}_h(\mathrm{N})|}{|h|}$$

for any parameter $\mathrm{T} \geqslant 1$.

The left-hand side of this inequality is called the *discrepancy* of the sequence $(x_n)_{1\leqslant n\leqslant \mathrm{N}}$, and is a natural measure of the distance between the probability measure

$$\frac{1}{\mathrm{N}}\sum_{n\leqslant \mathrm{N}} \delta_{x_n}$$

and the Lebesgue measure.

On the other hand, many problems give rise to multi-dimensional equidistribution results, where the sequence (x_n) takes values in $[0,1]^d$ for some integer $d \geqslant 1$. Equidistribution (with respect to the Lebesgue measure) means that

$$\lim_{\mathrm{N}\to+\infty} \frac{1}{\mathrm{N}}|\{n \leqslant \mathrm{N} \mid a_i < x_{n,i} < b_i \text{ for all } i\}| = \prod_i (b_i - a_i)$$

whenever $0 \leqslant a_i < b_i \leqslant 1$ for all i, and is again equivalent to the convergence towards 0 of the Weyl sums

$$\mathrm{W}_{\boldsymbol{h}}(\mathrm{N}) = \frac{1}{\mathrm{N}}\sum_{n\leqslant \mathrm{N}} e(\boldsymbol{h}\cdot x_n)$$

for $\boldsymbol{h} \in \mathbf{Z}^d - \{0\}$, where $x \cdot y$ denotes the usual scalar product on $\mathbf{R}^d$. The analogue of the Erdős–Turán inequality is due to Koksma and states that the "box discrepancy"

$$\Delta_d = \sup_{\substack{0\leqslant a_i\leqslant b_i\leqslant 1\\ 1\leqslant i\leqslant d}} \Bigl|\frac{1}{\mathrm{N}}|\{n \leqslant \mathrm{N} \mid x_n \in \prod_{i=1}^d [a_i,b_i]\}| - \prod_{i=1}^d (b_i - a_i)\Bigr|$$

satisfies

$$\Delta_d \ll \frac{1}{\mathrm{T}} + \sum_{1\leqslant \|\boldsymbol{h}\|_\infty \leqslant \mathrm{T}} \frac{1}{\mathrm{M}(\boldsymbol{h})}|\mathrm{W}_{\boldsymbol{h}}(\mathrm{N})| \tag{2}$$

for any parameter $\mathrm{T} \geqslant 1$, where $\mathrm{M}(\boldsymbol{h}) = \prod \max(1, |h_i|)$ and the implied constant depends only on d (see e.g. [11, Th. 1.21]).

However, this inequality is somewhat unsatisfactory when $d \geqslant 2$, due to its lack of "invariance". By this, we mean that if we apply to the sequence (x_n) a continuous map $f\colon [0,1]^d \to [0,1]^d$ (or $f\colon [0,1]^d \to \mathbf{C}$), then the equidistribution of (x_n) with respect to some measure μ implies that $(f(x_n))$ is equidistributed with respect to the image measure $f_*(\mu)$, and one naturally wants to have a quantitative version of this other convergence statement. Such a statement cannot be obtained from (2) without some analysis of the way the map f

transforms boxes with sides parallel to the coordinate axes. Note that this is a problem even if f is (say) a linear transformation in $\mathrm{SL}_d(\mathbf{Z})$, and μ is the Lebesgue measure. In particular, if the space $[0,1]^d$ arises abstractly without specific choices of coordinates (which may well happen), then the generalized Erdős–Turán inequality imposes the choice of a coordinate system, which might be artificial and awkward for other purposes. Moreover, if we consider the functional definition of equidistribution, namely the fact that

$$\lim_{\mathrm{N}\to+\infty} \frac{1}{\mathrm{N}} \sum_{n\leqslant \mathrm{N}} f(x_n) = \int_{[0,1]^d} f(x)dx$$

for any continuous function $f\colon [0,1]^d \to \mathbf{C}$, it is difficult to pass from (2) to quantitative bounds for the difference

$$\Big|\frac{1}{\mathrm{N}} \sum_{n\leqslant \mathrm{N}} f(x_n) - \int_{[0,1]^d} f(x)dx\Big|,$$

which may be equally relevant to applications.

From the probabilistic point of view, the goal of measuring the distance between probability measures is very classical, and has appeared in many forms since the beginning of modern probability theory. In recent years, there has been increasing interest in using *Wasserstein metrics* for this purpose. Indeed, these distances have considerable impact in probability theory, statistics, the theory of PDEs and numerical analysis. However, they only appeared very recently in works related to equidistribution questions, including in works of Bobkov and Ledoux [3], Steinerberger [19], Brown and Steinerberger [9], Graham [14] and Borda [6, 7]. Although not yet well-established in the analytic number theory community, we believe that the Wasserstein metrics provide a particularly well-suited approach to quantitative equidistribution.

To explain this, we recall the definition of the Wasserstein distances on a compact metric space (M, d). Let $p \geqslant 1$. For (Radon) probability measures μ and ν on M, let $\Pi(\mu,\nu)$ be the set of probability measures on $\mathrm{M}\times\mathrm{M}$ with marginals μ and ν (which is not empty since $\mu\otimes\nu$ belongs to it). The p-Wasserstein distance is then defined by

$$\mathscr{W}_p(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \Big(\int_{\mathrm{M}\times\mathrm{M}} d(x,y)^p d\pi(x,y)\Big)^{1/p}.$$

Remark 1.1. (1) This quantity depends on the choice of the metric d. When needed, we will write $\mathscr{W}_p^{(d)}(\mu,\nu)$ to indicate which metric is used.

(2) In probabilistic terms, we have

$$\mathscr{W}_p(\mu,\nu)^p = \inf_{\substack{\mathrm{X}\sim\mu \\ \mathrm{Y}\sim\nu}} \mathbf{E}(d(\mathrm{X},\mathrm{Y})^p),$$

where the infimum is taken over families of random variables (X,Y) with values in M such that the law of X is μ and that of Y is ν.

The Wasserstein distances are particularly important in the theory of optimal transport: indeed, they measure the cost of "moving" μ to ν (see, e.g. the book [22] of Villani for an introduction to optimal transport). Crucially from our point of view, the definition of $\mathscr{W}_p$ is intrinsic and does not suffer from the same invariance issues as the box discrepancy. The key points are the following:

(1) the Wasserstein distances are metrics on the set of probability measures on M, and the topology that they define on this set is the topology of convergence in law;
(2) the Wasserstein metrics satisfy simple inequalities under various operations, such as pushforward by a Lipschitz map;
(3) the Wasserstein metrics satisfy inequalities in terms of Weyl sums (in contexts much more general than that of $\mathrm{M} = (\mathbf{R}/\mathbf{Z})^d$ above) which are comparable to (2);
(4) for $p = 1$, the Wasserstein metric $\mathscr{W}_1$ admits a very clean functional interpretation, known as the Kantorovich–Rubinstein Theorem.

We summarize these basic properties.

Theorem 1.2 (Wasserstein distance properties)**.** *Let* (M, d) *be a compact metric space.*

(1) *For* $p \geqslant 1$*, the Wasserstein metric is a metric on the space of probability measures on* M*, and the topology it defines is the topology of convergence in law:* $\mathscr{W}_p(\mu_n, \mu) \to 0$ *if and only if, for all (bounded) continuous functions* $f \colon \mathrm{M} \to \mathbf{C}$*, we have*

$$\lim_{n \to +\infty} \int_{\mathrm{M}} f d\mu_n = \int_{\mathrm{M}} f d\mu.$$

(2) *For probability measures* μ *and* ν *on* M*, we have*

$$\mathscr{W}_p(\mu, \nu) \leqslant \mathscr{W}_q(\mu, \nu)$$

if $p \leqslant q$*.*
(3) *Let* (N, δ) *be a compact metric space. Let* $c \geqslant 0$ *be a real number and let* $f \colon \mathrm{M} \to \mathrm{N}$ *be a c-Lipschitz map. For any probability measures* μ *and* ν *on* M*, we have*

$$\mathscr{W}_p(f_*\mu, f_*\nu) \leqslant c\mathscr{W}_p(\mu, \nu).$$

(4) *If* $\mathrm{N} \subset \mathrm{M}$ *is a compact subset with inclusion* $i \colon \mathrm{N} \to \mathrm{M}$*, then for probability measures* μ *and* ν *on* N*, we have*

$$\mathscr{W}_p(i_*\mu, i_*\nu) = \mathscr{W}_p(\mu, \nu),$$

where the right-hand side is a Wasserstein distance on N*.*
(5) *Let* (N, δ) *be a compact metric space. Let* Δ *be a metric on* $\mathrm{M} \times \mathrm{N}$ *such that there exists* $c > 0$ *such that*

$$\Delta((x, y), (x', y')) \leqslant c(d(x, x') + \delta(y, y'))$$

for (x, y) *and* (x', y') *in* $\mathrm{M} \times \mathrm{N}$*. For any probability measures* μ *and* ν *on* M*,* μ' *and* ν' *on* N*, we have*

$$\mathscr{W}_p^{(\Delta)}(\mu \otimes \mu', \nu \otimes \nu') \leqslant c2^{1/q}(\mathscr{W}_p(\mu, \nu) + \mathscr{W}_p(\mu', \nu'))$$

for $p \geqslant 1$*, where* q *is the dual exponent with* $1/p + 1/q = 1$*, and we use the convention* $2^{1/\infty} = 1$*.*
(6) *For probability measures* μ *and* ν *on* M*, we have*

$$\mathscr{W}_1(\mu, \nu) = \sup_u \Big| \int_{\mathrm{M}} u d\mu - \int_{\mathrm{M}} u d\nu \Big|$$

where the supremum is over functions $u \colon \mathrm{M} \to \mathbf{R}$ *which are* 1*-Lipschitz.*

(7) *Suppose that* $\mathrm{M} = (\mathbf{R}/\mathbf{Z})^k$ *with its standard metric for some integer* $k \geqslant 1$*. For any probability measures* μ *and* ν *on* M *and for all* $\mathrm{T} > 0$*, we have*

$$\mathscr{W}_1(\mu,\nu) \leqslant \frac{4\sqrt{3}\sqrt{k}}{\mathrm{T}} + \Big(\sum_{1\leqslant|\boldsymbol{h}|_\infty\leqslant \mathrm{T}} \frac{1}{\|\boldsymbol{h}\|^2}|\widehat{\mu}(\boldsymbol{h}) - \widehat{\nu}(\boldsymbol{h})|^2\Big)^{1/2}, \tag{3}$$

where for a probability measure μ *on* M *and* $\boldsymbol{h} \in \mathbf{Z}^k$*, we denote by* $\widehat{\mu}(\boldsymbol{h})$ *the Fourier coefficient of* μ*, i.e.*

$$\widehat{\mu}(\boldsymbol{h}) = \int_{\mathrm{M}} e(\boldsymbol{h}\cdot x)d\mu(x).$$

Proof. (1) This is proved, e.g., in Villani's book [22, Th. 7.3, Th. 7.12].

(2) This follows easily from Hölder's inequality and the definition of the Wasserstein metric.

(3) This is also a formal consequence of the definition and the fact that, given $\pi \in \Pi(\mu,\nu)$, we have $(f\times f)_*\pi \in \Pi(f_*\mu, f_*\nu)$. Then, for $\pi \in \Pi(\mu,\nu)$, we get

$$\int_{\mathrm{N}\times\mathrm{N}} \delta(x,y)^p d(f\times f)_*\pi = \int_{\mathrm{M}\times\mathrm{M}} \delta(f(x),f(y))^p d\pi \leqslant c^p \int_{\mathrm{M}\times\mathrm{M}} d(x,y)^p d\pi,$$

and taking the infimum over $\pi \in \Pi(\mu,\nu)$ gives the inequality.

(4) This follows from (3) and the fact that any measure in $\Pi(i_*\mu, i_*\nu)$ has support in $\mathrm{N}\times\mathrm{N}$, hence is of the form $(i\times i)_*\pi$ for some $\pi \in \Pi(\mu,\nu)$.

(5) This is also elementary: if $\pi \in \Pi(\mu,\nu)$ and $\pi' \in \Pi(\mu',\nu')$, then $\pi\otimes\pi' \in \Pi(\mu\otimes\mu', \nu\otimes\nu')$, with

$$\int_{\mathrm{M}^2\times\mathrm{N}^2} \Delta((x,y),(x',y'))^p d(\pi\otimes\pi') \leqslant c^p \int_{\mathrm{M}^2\times\mathrm{N}^2} (d(x,x') + \delta(y,y'))^p d(\pi\otimes\pi')$$

by assumption. Using the Hölder inequality in the form $(a+b)^p \leqslant 2^{p/q}(a^p + b^p)$ for a, $b \geqslant 0$, we obtain

$$\int_{\mathrm{M}^2\times\mathrm{N}^2} \Delta((x,y),(x',y'))^p d(\pi\otimes\pi') \leqslant c^p 2^{p/q}\Big(\int_{\mathrm{M}^2} d(x,x')^p d\pi + \int_{\mathrm{N}^2} \delta(y,y')^p d\pi'\Big),$$

hence the result.

(6) This is much deeper and is a special case of the Kantorovich–Rubinstein duality (see, e.g., [22, Th. 1.14]).

(7) This is (essentially) an application of a statement due to Bobkov and Ledoux [3]. For completeness, we will give a proof in the Appendix starting on page 18. ∎

Thus, as an alternative to the classical inequality (2), we have the following corollary.

Corollary 1.3. *Let* $k \geqslant 1$ *be an integer. Let* $(x_n)_{n\geqslant 1}$ *be a sequence in* $(\mathbf{R}/\mathbf{Z})^k$*. Define the probability measures*

$$\mu_{\mathrm{N}} = \frac{1}{\mathrm{N}}\sum_{n=1}^{\mathrm{N}} \delta_{x_n},$$

for $\mathrm{N} \geqslant 1$*. For any* $\mathrm{N} \geqslant 1$, $\mathrm{T} > 0$*, we have*

$$\mathscr{W}_1(\mu_{\mathrm{N}}, \lambda_k) \leqslant \frac{4\sqrt{3}\sqrt{k}}{\mathrm{T}} + \Big(\sum_{1 \leqslant |\boldsymbol{h}|_\infty \leqslant \mathrm{T}} \frac{1}{\|\boldsymbol{h}\|^2} |\mathrm{W}_{\boldsymbol{h}}(\mathrm{N})|^2 \Big)^{1/2}.$$

In the sequel to this paper, we will further use the Wasserstein metric in statements of quantitative equidistribution in compact Lie groups, with applications to exponential sums over finite fields.

Example 1.4. For illustration, it is worth looking concretely at the meaning of the Wasserstein distance between a measure of the type

$$\mu_{\mathrm{N}} = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \delta_{x_i}$$

on $\mathbf{R}/\mathbf{Z}$ and the uniform Lebesgue measure λ. We assume for simplicity that the x_i's are different, and denote $\mathscr{X} = \{x_i,\ 1 \leqslant i \leqslant \mathrm{N}\}$. Any measure $\pi \in \Pi(\mu_{\mathrm{N}}, \lambda)$ must be supported in $\mathscr{X} \times \mathbf{R}/\mathbf{Z} \subset (\mathbf{R}/\mathbf{Z})^2$, since $\mathscr{X}$ is the support of μ_{N}. Hence we may write

$$\pi = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \delta_{x_i} \otimes \pi_i,$$

for some probability measure π_i on $\mathbf{R}/\mathbf{Z}$. Any such measure has projection on the first coordinate equal to μ_{N}. On the other hand, the projection on the second coordinate is the combination

$$\frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \pi_i,$$

and thus $\pi \in \Pi(\mu_{\mathrm{N}}, \lambda)$ is equivalent with

$$\frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \pi_i = \lambda.$$

In particular, suppose that $\pi_i = f_i(y)dy$ for some measurable function $f_i \geqslant 0$ with integral equal to 1. Then the condition above becomes

$$\frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} f_i = 1.$$

The corresponding integral for the p-Wasserstein distance is

$$\int_{(\mathbf{R}/\mathbf{Z})^2} d(x,y)^p d\pi = \frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \int_{\mathbf{R}/\mathbf{Z}} d(x_i, y)^p f_i(y) dy.$$

As a simple illustration, let $\mathrm{N} \geqslant 2$ and consider $x_i = (i-1)/\mathrm{N}$ for $1 \leqslant i \leqslant \mathrm{N}$. Let φ_i be the indicator function of the (image modulo $\mathbf{Z}$ of the) interval

$$\Big[\frac{i-1}{\mathrm{N}} - \frac{1}{2\mathrm{N}}, \frac{i-1}{\mathrm{N}} + \frac{1}{2\mathrm{N}}\Big[,$$

and let $f_i = \mathrm{N}\varphi_i$. Then f_i has integral 1, and

$$\sum_i f_i = \mathrm{N}$$

since any $x \in \mathbf{R}/\mathbf{Z}$ belongs to a single interval, where the corresponding f_i takes the value N. For the measure π described above, using the usual circular distance on $\mathbf{R}/\mathbf{Z}$, we get

$$\int_{(\mathbf{R}/\mathbf{Z})^2} d(x,y)^p d\pi = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\int_{\mathbf{R}/\mathbf{Z}} d(x_i,y)^p f_i(y)dy = \sum_{i=1}^{\mathrm{N}}\int_{(i-1)/\mathrm{N}-1/(2\mathrm{N})}^{(i-1)/\mathrm{N}+1/(2\mathrm{N})}\Bigl|\frac{i-1}{\mathrm{N}}-y\Bigr|^p dy.$$

By a simple integration, this is

$$\frac{2\mathrm{N}}{p+1}\Bigl(\frac{1}{2\mathrm{N}}\Bigr)^{p+1} = \frac{1}{(p+1)2^p}\frac{1}{\mathrm{N}^p},$$

proving that

$$\mathscr{W}_p\Bigl(\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}\delta_{(i-1)/\mathrm{N}},\lambda\Bigr) \leqslant \frac{1}{2(p+1)^{1/p}}\frac{1}{\mathrm{N}} \leqslant \frac{1}{2\mathrm{N}}$$

for all $p \geqslant 1$. For $p = 1$, this result is of the same quality as the outcome of Corollary 1.3, since in this case the Weyl sum $\mathrm{W}_h(\mathrm{N})$ is zero unless $\mathrm{N} \mid h$, in which case it is equal to 1, so that the estimate in loc. cit. is

$$\mathscr{W}_1(\mu_{\mathrm{N}},\lambda) \leqslant \frac{4\sqrt{3}}{\mathrm{T}} + \Bigl(\sum_{\substack{1\leqslant |h|\leqslant \mathrm{T}\\ \mathrm{N}\mid h}}\frac{1}{|h|^2}\Bigr)^{1/2} \ll \frac{1}{\mathrm{T}}+\frac{1}{\mathrm{N}},$$

and taking $\mathrm{T} = \mathrm{N}$ gives the result. This is of course comparable to the well-known fact that the discrepancy of equally-spaced points is of size $1/\mathrm{N}$.

Note that as a consequence of Theorem 1.2, (5), we deduce also that for the uniform measure $\mu_{\mathrm{N}}^{(k)}$ on the finite grid

$$\{0, 1/\mathrm{N}, \ldots, (\mathrm{N}-1)/\mathrm{N}\}^k \subset (\mathbf{R}/\mathbf{Z})^k,$$

we have

$$\mathscr{W}_p(\mu_{\mathrm{N}}^{(k)},\lambda_k) \ll \frac{1}{\mathrm{N}}$$

where the implied constant depends on k only (and the Wasserstein distance is computed for the standard metric). Here again, taking $\mathrm{T} = \mathrm{N}$ in Theorem 1.2, (7) gives a bound of the same shape for $p = 1$.

2. Ultra-short sums of trace functions

Our previous paper [16] considers an equidistribution problem on a torus for which the quantitative features of the Wasserstein metric are particularly useful. This concerns the distribution properties of sums of the form

$$\mathrm{S}_g(q,a) = \sum_{\substack{x\in\mathbf{F}_q\\ g(x)\equiv 0 \,(\mathrm{mod}\, q)}} e\Bigl(\frac{ax}{q}\Bigr), \tag{4}$$

where $g \in \mathbf{Z}[\mathrm{X}]$ is a fixed monic polynomial and q is a prime number (subject to suitable conditions).

We associated to these sums the probability measures

$$\nu_q = \frac{1}{q} \sum_{a \in \mathbf{F}_q} \delta_{\mathrm{S}_g(q,a)}$$

on $\mathbf{C}$. In [16, Th. 1.1], we proved that these measures converge weakly to an explicit probability measure μ_g which is related to the additive relations among the complex roots of g, as recalled in Section 3. In this work, we refine these results and prove the following rates of convergence in 1-Wasserstein metric:

Theorem 2.1 (Cor. 3.5)**.** *Let $g \in \mathbf{Z}[\mathrm{X}]$ be a monic and separable polynomial of degree $d \geqslant 1$, with splitting field K_g over $\mathbf{Q}$. Then for all prime numbers q totally split in K_g that do not divide the discriminant of g, the upper bound*

$$\mathscr{W}_1(\nu_q, \mu_g) \ll_g q^{-\frac{1}{[\mathrm{K}_g:\mathbf{Q}]}}$$

holds.

The quantitative results in Wasserstein metrics also allow us to consider some problems involving *varying* polynomials. As an example, in Section 4 we exploit the explicit dependency on k in the inequality of Theorem 1.2, (7) to study the distribution of exponential sums over very small multiplicative subgroups of prime order, i.e. sums of the form

$$\sum_{x \in \mathrm{H}} e\left(\frac{ax}{q}\right)$$

where H is a subgroup of $\mathbf{F}_q^\times$ whose cardinality is a *prime* divisor of $q-1$ which is very small compared to q. Precisely, we prove in Theorem 4.1 that the sums

$$\frac{1}{\sqrt{|\mathrm{H}|}} \sum_{x \in \mathrm{H}} e\left(\frac{ax}{q}\right)$$

become equidistributed with respect to a standard complex normal distribution as q tends to infinity and $|\mathrm{H}|$ tends to infinity while satisfying

$$|\mathrm{H}| = o\left(\frac{\log q}{\log\log q}\right)$$

as $q \to +\infty$.

Notation. For $x = (x_1, \ldots, x_d) \in \mathbf{R}^d$, we denote by $|x|$ its euclidean norm.

For a random vector $\mathrm{X} = (\mathrm{X}_1, \ldots, \mathrm{X}_d)$ in $\mathbf{R}^d$, we denote its characteristic function by φ_{X}, and we recall that it is defined for all $s = (s_1, \ldots, s_d) \in \mathbf{R}^d$ by $\varphi_{\mathrm{X}}(s) = \mathbf{E}\left(e^{i\mathrm{X}\cdot s}\right)$ where $\mathrm{X} \cdot s$ denotes the usual dot product on $\mathbf{R}^d$.

We say that a map $u \colon \mathbf{R}^d \to \mathbf{R}$ is 2π-periodic if for all $m \in \mathbf{Z}^d$, for all $x \in \mathbf{R}^d$, $u(x+2\pi m) = u(x)$.

For any set X, we denote by $\mathrm{C}(\mathrm{Z}_g;\mathrm{X})$ the set of maps from Z_g to X; it is a group if X itself is a group. For all $\alpha \in \mathrm{C}(\mathrm{Z}_g;\mathbf{Z})$, we set

$$\|\alpha\|_1 = \sum_{x\in\mathrm{Z}_g} |\alpha(x)|, \quad \|\alpha\|_\infty = \max_{x\in\mathrm{Z}_g} |\alpha(x)|, \quad \text{and } |\alpha| = \left(\sum_{x\in\mathrm{Z}_g} \alpha(x)^2\right)^{1/2}.$$

Acknowledgements. We thank G. Tenenbaum for suggesting that the properties of the Weyl sums in our equidistribution problem would lead to strong quantitative versions.

We are thankful to Bence Borda for pointing out the two versions of the inequality of Bobkov and Ledoux and for very helpful comments on an earlier version of this article. The second author wishes to thank Igor Shparlinski for sending an unpublished note that contained interesting ideas and led him to study the question of Section 4. He would also like to thank Florent Jouve, Guillaume Ricotta, Robert Tichy for encouragements, as well as François Bolley, Thomas Cavallazzi, Baptiste Huguet and Thibault Modeste for taking the time to answer his questions on probability theory.

3. Proof of Theorem 2.1

We recall more precisely the setting and results of the article [16], starting with the sums $\mathrm{S}_g(q,a)$ of (4).

Let g be a fixed monic and separable polynomial $g \in \mathbf{Z}[\mathrm{X}]$ of degree $d \geqslant 1$. We denote by Z_g the set of complex roots of g and by $\mathrm{K}_g = \mathbf{Q}(\mathrm{Z}_g)$ the splitting field of g over $\mathbf{Q}$. Its ring of integers is denoted by $\mathbf{O}_g$, and since g is monic, the set Z_g is contained in $\mathbf{O}_g$.

For primes q that are totally split in K_g, we proved that the measures

$$\nu_q = \frac{1}{q} \sum_{a\in\mathbf{F}_q} \delta_{\mathrm{S}_g(q,a)}$$

associated to the sums (4) converge, as q goes to infinity and a varies in $\mathbf{F}_q$, to a limiting measure μ_g defined as follows.

Let $\mathrm{R}_g \subset \mathrm{C}(\mathrm{Z}_g;\mathbf{Z})$ be the subgroup of maps $\alpha\colon \mathrm{Z}_g \to \mathbf{Z}$ such that

$$\sum_{x\in\mathrm{Z}_g} \alpha(x)x = 0,$$

and let

$$\mathrm{H}_g = \left\{ f \in \mathrm{C}(\mathrm{Z}_g;\mathbf{S}^1) \mid \text{for all } \alpha \in \mathrm{R}_g, \text{ we have } \prod_{x\in\mathrm{Z}_g} f(x)^{\alpha(x)} = 1 \right\} \tag{5}$$

be the orthogonal of R_g in the sense of Pontryagin duality. Let λ_g denote the probability Haar measure on the compact group H_g. Then we showed that $\mu_g = \sigma_*\lambda_g$, where $\sigma\colon \mathrm{C}(\mathrm{Z}_g;\mathbf{S}^1) \to \mathbf{C}$ is defined by

$$\sigma(f) = \sum_{x\in\mathrm{Z}_g} f(x).$$

We will apply Theorem 1.2 to estimate the Wasserstein distances $\mathscr{W}_1(\mu_p, \lambda_g)$ (between measures on $\mathrm{C}(\mathrm{Z}_g; \mathbf{S}^1)$, or H_g) and $\mathscr{W}_1(\sigma_*\mu_p, \sigma_*\lambda_g)$ (between measures on $\mathbf{C}$), where the measures μ_p are introduced in §3.1. These distances are defined with respect to the following metrics on the underlying spaces:

(1) We consider on $\mathbf{C}$ the usual metric given by the complex modulus.
(2) We consider on the compact group $\mathrm{C}(\mathrm{Z}_g; \mathbf{S}^1)$ the metric $\varrho = \varrho_g$ obtained by transport of structure from the usual flat Riemannian metric on the torus $(\mathbf{R}/\mathbf{Z})^{\mathrm{Z}_g}$ using the isomorphism

$$\iota \colon (\mathbf{R}/\mathbf{Z})^{\mathrm{Z}_g} \to \mathrm{C}(\mathrm{Z}_g; \mathbf{S}^1)$$

such that $\iota(\boldsymbol{\xi})$ is the map $\mathrm{Z}_g \to \mathbf{S}^1$ which maps $x \in \mathrm{Z}_g$ to $e(\boldsymbol{\xi}_x)$. This metric on the torus is also the quotient metric of the euclidean distance using the projection $\mathbf{R}^{\mathrm{Z}_g} \to (\mathbf{R}/\mathbf{Z})^{\mathrm{Z}_g}$.

(Concretely, it can also be described as follows when identifying further $\mathbf{R}/\mathbf{Z}$ with $[0, 1[$: if ℓ denotes the "circular" metric on $[0, 1[$, identified with $\mathbf{R}/\mathbf{Z}$, such that $\ell(x, y) = \min(|x - y|, 1 - |x - y|)$, then the metric ϱ can be identified with the metric on $[0, 1[^{\mathrm{Z}_g}$ given by

$$(x, y) \mapsto \Big(\sum_{j=1}^{d} \ell(x_j, y_j)\Big)^{1/2}. \tag{6}$$

3.1. **The Weyl sums.** The key property of our equidistribution problem is that the relevant Weyl sums, for the equidistribution at the level of the space $\mathrm{C}(\mathrm{Z}_g; \mathbf{S}^1)$, *vanish* when q is suitably large, as we observed in [16, Rem. 2.3].

We recall some further notation from [16]. If g is a polynomial as above, we denote by $\mathscr{S}_g$ the set of prime ideals $p \subset \mathbf{O}_g$ which do not divide the discriminant of g and have residual degree one. For $p \in \mathscr{S}_g$, the norm $q = |p| = |\mathbf{O}_g/p|$ is a prime number, and the restriction $\mathbf{Z} \to \mathbf{O}_g/p$ of the restriction map $\varpi_p \colon \mathbf{O}_g \to \mathbf{O}_g/p$ induces a field isomorphism $\mathbf{F}_q \to \mathbf{O}_g/p$, which we use to identify these two fields.

We view $\mathbf{O}_g/p$ as a finite probability space with the uniform probability measure and we consider the random variables U_p on $\mathbf{O}_g/p$, taking values in $\mathrm{C}(\mathrm{Z}_g; \mathbf{S}^1)$, which are defined by

$$\mathrm{U}_p(a)(x) = e\Big(\frac{a\varpi_p(x)}{|p|}\Big)$$

for all $a \in \mathbf{O}_g/p$ and all $x \in \mathrm{Z}_g$ (the element $a\varpi_p(x)$ of $\mathbf{O}_g/p$ being identified with an element of $\mathbf{F}_{|p|}$ as explained before).

Let μ_p be the law of U_p; concretely, this is the measure

$$\mu_p = \frac{1}{|p|} \sum_{a \in \mathbf{O}_g/p} \delta_{\mathrm{U}_p(a)}$$

on $\mathrm{C}(\mathrm{Z}_g; \mathbf{S}^1)$. We also view λ_g as a measure on $\mathrm{C}(\mathrm{Z}_g; \mathbf{S}^1)$ by identifying it with $j_*\lambda_g$, where $j \colon \mathrm{H}_g \to \mathrm{C}(\mathrm{Z}_g; \mathbf{S}^1)$ is the inclusion.

The next two lemmas compare the Fourier coefficients of the measures μ_p and λ_g.
For $\alpha \in \mathrm{C}(\mathrm{Z}_g; \mathbf{Z})$, we will denote by η_α the associated character of $\mathrm{C}(\mathrm{Z}_g; \mathbf{S}^1)$. For a measure ν

on $\mathrm{C}(\mathrm{Z}_g;\mathbf{S}^1)$, the corresponding Fourier coefficient is

$$\widehat{\nu}(\alpha)=\int\eta_\alpha(x)d\nu(x).$$

Lemma 3.1. *Let $\alpha\in\mathrm{C}(\mathrm{Z}_g;\mathbf{Z})$. If η_α is trivial on H_g, then $\widehat{\lambda}_g(\alpha)=\widehat{\mu}_p(\alpha)=1$.*

Proof. The result about λ_g simply follows from the fact that it has total mass 1. The case of μ_p follows because it is straightforward from the definition that the random variables U_p take values in H_g. ☐

Lemma 3.2. *There exists a positive real number C_g, depending only on the polynomial g, with the following property: for all $\alpha\in\mathrm{C}(\mathrm{Z}_g;\mathbf{Z})$ such that η_α is non-trivial on H_g and for all $p\in\mathscr{S}_g$ such that*

$$|p|>\mathrm{C}_g\|\alpha\|_1^{[\mathrm{K}_g:\mathbf{Q}]},$$

we have $\widehat{\lambda}_g(\alpha)=\widehat{\mu}_p(\alpha)=0$.

More precisely, the value

$$\mathrm{C}_g=\prod_{\sigma\in\mathrm{Gal}(\mathrm{K}_g/\mathbf{Q})}\max_{x\in\mathrm{Z}_g}|\sigma(x)| \tag{7}$$

has this property.

Proof. Let $\alpha\in\mathrm{C}(\mathrm{Z}_g;\mathbf{Z})$ be such that η_α is non-trivial on H_g. It is then a classical property of the Haar measure that $\widehat{\lambda}_g(\alpha)=0$.

Let $p\in\mathscr{S}_g$. As in [16, Prop. 2.2], we have

$$\widehat{\mu}_p(\alpha)=\frac{1}{|p|}\sum_{a\in\mathbf{O}_g/p}e\Big(\frac{a}{|p|}\varpi_p\Big(\sum_{x\in\mathrm{Z}_g}\alpha(x)x\Big)\Big),$$

and therefore

$$\widehat{\mu}_p(\alpha)=\begin{cases}1 & \text{if } \sum_{x\in\mathrm{Z}_g}\alpha(x)x\in p,\\ 0 & \text{otherwise.}\end{cases}$$

by orthogonality of characters. We now analyze this condition further.

Let

$$\gamma(\alpha)=\sum_{x\in\mathrm{Z}_g}\alpha(x)x.$$

Note that since η_α is non-trivial on H_g, we have $\alpha\notin\mathrm{R}_g$ (by definition), and therefore $\gamma(\alpha)\neq 0$.

If $\widehat{\mu}_p(\alpha)\neq 0$, then $\gamma(\alpha)\in p$, so $\gamma(\alpha)\mathbf{O}_g\subset p$, and in particular $|p|\mid\mathrm{N}(\gamma(\alpha)\mathbf{O}_g)$. In particular, since $\gamma(\alpha)$ is non-zero, we obtain

$$|\mathrm{N}_{\mathrm{K}_g/\mathbf{Q}}(\gamma(\alpha))|\geqslant|p|. \tag{8}$$

On the other hand, we have

$$\mathrm{N}_{\mathrm{K}_g/\mathbf{Q}}(\gamma(\alpha)) = \prod_{\sigma\in \mathrm{Gal}(\mathrm{K}_g/\mathbf{Q})} \sigma(\gamma(\alpha)) = \prod_{\sigma\in \mathrm{Gal}(\mathrm{K}_g/\mathbf{Q})} \Big(\sum_{x\in \mathrm{Z}_g} \alpha(x)\sigma(x)\Big),$$

and with the value C_g given by (7), this implies that

$$|\mathrm{N}_{\mathrm{K}_g/\mathbf{Q}}(\gamma(\alpha))| \leqslant \mathrm{C}_g\Big(\sum_{x\in \mathrm{Z}_g} |\alpha(x)|\Big)^{[\mathrm{K}_g:\mathbf{Q}]} = \mathrm{C}_g\|\alpha\|_1^{[\mathrm{K}_g:\mathbf{Q}]}. \tag{9}$$

Combining (8) and (9) we deduce that

$$|p| \leqslant \mathrm{C}_g\|\alpha\|_1^{[\mathrm{K}_g:\mathbf{Q}]}$$

if $\widehat{\mu}_p(\alpha)\neq 0$, which is the desired conclusion. □

3.2. **Quantitative equidistribution.**

Proposition 3.3. *There exists an explicit constant* $\mathrm{C}_g^{'} > 0$ *(depending only on* g*) such that for all* $p \in \mathscr{S}_g$ *we have*

$$\mathscr{W}_1(\mu_p, \lambda_g) \leqslant \mathrm{C}_g'|p|^{-\frac{1}{[\mathrm{K}_g:\mathbf{Q}]}}.$$

Proof. let $\mathrm{T} > 0$ be an auxiliary parameter to be fixed below. By Theorem 1.2, (7), the inequality

$$\mathscr{W}_1(\mu_p, \lambda_g) \leqslant \frac{4\sqrt{3}\sqrt{d}}{\mathrm{T}} + \Big(\sum_{\substack{\alpha\in \mathrm{C}(\mathrm{Z}_g;\mathbf{Z})\\ 0<\|\alpha\|_\infty\leqslant \mathrm{T}}} \frac{1}{|\alpha|^2}|\widehat{\mu_p}(\alpha) - \widehat{\lambda}_g(\alpha)|^2\Big)^{1/2} \tag{10}$$

holds. We take

$$\mathrm{T} = \frac{1}{d+1}\left(\frac{|p|}{\mathrm{C}_g}\right)^{\frac{1}{[\mathrm{K}_g:\mathbf{Q}]}}.$$

Lemma 3.1 and Lemma 3.2 together imply that the sum on the right-hand side of (10) is zero (using the inequality $\|\alpha\|_1 \leqslant d\|\alpha\|_\infty$), and therefore we obtain

$$\mathscr{W}_1(\mu_p, \lambda_g) \leqslant \frac{4\sqrt{3}\sqrt{d}}{\mathrm{T}} = 4\sqrt{3}\sqrt{d}(d+1)\mathrm{C}_g^{\frac{1}{[\mathrm{K}_g:\mathbf{Q}]}}|p|^{-\frac{1}{[\mathrm{K}_g:\mathbf{Q}]}},$$

which immediately implies that result. □

Remark. This result matches the rate of convergence obtained in [21, Th. 5.30], where the 1-Wasserstein metric was replaced by a notion of φ-discrepancy, which had the disadvantage of being non-intrinsic.

We can easily deduce Theorem 2.1 using Theorem 1.2, (3). First we compute a Lipschitz constant for the summation map σ.

Lemma 3.4. *The map* $\sigma\colon \mathrm{C}(\mathrm{Z}_g;\mathbf{S}^1) \to \mathbf{C}$ *is* $\sqrt{d}$*-Lipschitz.*

Proof. Let f and g be elements of $\mathrm{C}(\mathrm{Z}_g;\mathbf{S}^1)$. We have

$$|\sigma(f)-\sigma(g)| = \Big|\sum_{x\in\mathrm{Z}_g}(f(x)-g(x))\Big| \leqslant \sum_{x\in\mathrm{Z}_g}\Big|f(x)-g(x)\Big|.$$

The euclidean distance on $\mathbf{S}^1\subset\mathbf{C}$ is bounded above by the arc length (Riemannian) distance ℓ, hence $|f(x)-g(x)|\leqslant \ell(f(x),g(x))$. Applying the Cauchy–Schwarz inequality, we obtain

$$\sum_{x\in\mathrm{Z}_g}|f(x)-g(x)| \leqslant \sum_{x\in\mathrm{Z}_g}\ell(f(x),g(x)) \leqslant \sqrt{d}\Big(\sum_{x\in\mathrm{Z}_g}\ell(f(x),g(x))^2\Big)^{1/2} = \sqrt{d}\times\varrho(f,g),$$

as desired. ☐

Corollary 3.5. *Let $d\geqslant 1$ and let $g\in\mathbf{Z}[\mathrm{X}]$ be a monic and separable polynomial of degree d. For all prime numbers q totally split in K_g that do not divide the discriminant of g, define the measures*

$$\nu_q = \frac{1}{q}\sum_{a\in\mathbf{F}_q}\delta_{\mathrm{S}_g(q,a)}$$

and $\mu_g=\sigma_*\lambda_g$ (the pushforward measure via σ of the probability Haar measure on H_g). Then for all such prime numbers q, we have

$$\mathscr{W}_1(\nu_q,\mu_g)\ll_g q^{-\frac{1}{[\mathrm{K}_g:\mathbf{Q}]}}.$$

Proof. Let $p\in\mathscr{S}_g$. By Proposition 3.3, we have $\mathscr{W}_1(\mu_p,\lambda_g)\ll_g |p|^{-\frac{1}{[\mathrm{K}_g:\mathbf{Q}]}}$. It follows from Theorem 1.2, (3) and Lemma 3.4 that

$$\mathscr{W}_1(\sigma_*\mu_p,\sigma_*\lambda_g)\ll_g |p|^{-\frac{1}{[\mathrm{K}_g:\mathbf{Q}]}}$$

(where the implied constant on the right-hand side incorporates the factor $\sqrt{d}$). Then the conclusion follows from the fact that for all q totally split in K_g and such that it does not divide the discriminant of g, any $p\in\mathbf{O}_g$ lying above q is a prime ideal that belongs to $\mathscr{S}_g$ and $\sigma_*\mu_p=\nu_q$. ☐

This corollary refines [16, Cor. 2.4], since the latter only stated the weak convergence of ν_q to μ_g, whereas here we obtain a quantitative rate of convergence.

4. Sums of additive characters over growing multiplicative subgroups

So far, we have only been dealing with weak convergence of measures in compact groups since we were essentially working in $(\mathbf{S}^1)^d$ for a *fixed* d. However, the Wasserstein metric metrizes weak convergence in a much more general context, and in this section we give an application in a non-compact setting.

We consider the sums

$$\mathrm{S}_d(q,a) = \sum_{x\in\mu_d(\mathbf{F}_q)} e\left(\frac{ax}{q}\right)$$

which are a special case of the sums $\mathrm{S}_g(q,a)$ with $g=\mathrm{X}^d-1$. For a *fixed* integer d and q going to infinity among the primes congruent to 1 modulo d, their asymptotic behaviour attracted interest partly because of the beautiful visual aspect of the plots: see e.g. [10, 12, 13, 18, 20] for examples and generalizations. In particular, if d is prime [12, Th. 6.3] states that they become equidistributed (as q goes to infinity and a varies in $\mathbf{F}_q$) with respect to the pushforward measure of the Haar probability measure on $(\mathbf{S}^1)^{d-1}$ via the Laurent polynomial

$$\begin{array}{cccc} g_d & : & (\mathbf{S}^1)^{d-1} & \to & \mathbf{C} \\ & & (z_1, \dots, z_{d-1}) & \mapsto & z_1 + \cdots + z_{d-1} + \frac{1}{z_1 \dots z_{d-1}} \end{array}$$

This result explains why the sums appeared to fill out the region of the complex plane delimited by a d-cusps hypocycloid.

We are now interested in studying the case where d is allowed to vary with q. In more probabilistic terms, the previous result says that when q is large the subsets

$$\{\mathrm{S}_d(q, a),\ a \in \mathbf{F}_q\}$$

of the complex plane look like q independent values taken by a sum of the form

$$\mathrm{Z}_1 + \cdots + \mathrm{Z}_{d-1} + \frac{1}{\mathrm{Z}_1 \dots \mathrm{Z}_{d-1}}$$

of independent and identically distributed Steinhaus random variables Z_i (i.e. uniform on $\mathbf{S}^1$). Thanks to the multidimensional Central Limit Theorem, the random variables

$$\frac{1}{\sqrt{d}} \left(\mathrm{Z}_1 + \cdots + \mathrm{Z}_{d-1} + \frac{1}{\mathrm{Z}_1 \dots \mathrm{Z}_{d-1}} \right)$$

converge in law to a two dimensional gaussian $\mathscr{N}(0, \frac{1}{2}\mathrm{Id})$ (the coefficient $\frac{1}{2}$ just comes from the value of the variance of the real and imaginary parts of a uniform random variable on the circle). Therefore, if we denote by $\mu_{q,d}$ the measure

$$\frac{1}{q} \sum_{a \in \mathbf{F}_q} \delta_{\frac{1}{\sqrt{d}} \mathrm{S}_d(q,a)}$$

then $\lim_{d\to\infty} (\lim_{q\to\infty} \mu_{q,d}) = \mathscr{N}(0, \frac{1}{2}\mathrm{Id})$. However $\lim_{q\to\infty} (\lim_{d\to\infty} \mu_{q,d}) = \delta_0$, so we are interested in intermediate regimes, in which both q and d tend to infinity, and the limit of the sequence of measures $\mu_{q,d}$ can be determined. Equivalently, this means that we are interested in the distribution of the sums

$$\frac{1}{\sqrt{d}} \sum_{x \in \mu_d(\mathbf{F}_q)} e\left(\frac{ax}{q}\right) \tag{11}$$

as a varies in $\mathbf{F}_q$ and both q and d tend to infinity (with q and d prime and $q \equiv 1 \,(\mathrm{mod}\, d)$). These sums of additive characters over multiplicative subgroups whose cardinality grows with q have been studied before, mostly with the aim of proving non-trivial upper bounds. In particular, when d grows at least like a small power of q, the groundbreaking work of Bourgain, Glibichuk and Konyagin [8] shows a power saving bound for $\mathrm{S}_d(q, a)$. On the other hand, [15, Th. 1.8] shows that if $d \ll \log(q)$, it is impossible to obtain a non-trivial bound.

Combining the results of Section 3.2 and the Central Limit Theorem, we will prove the following result in a similar setting where d is very small with respect to q.

Theorem 4.1. *For every odd prime q, we let $d = d(q)$ be a prime divisor of $q-1$. If $d(q) \underset{q\to+\infty}{\longrightarrow} +\infty$ and*

$$d(q) \underset{q\to\infty}{=} o\left(\frac{\log q}{\log\log q}\right)$$

then as q tends to infinity and a varies in $\mathbf{F}_q$, the sums (11) become equidistributed in the complex plane with respect to a normal distribution $\mathscr{N}(0, \frac{1}{2}\mathrm{Id})$.

The first step of the proof consists in using the following quantitative form of the convergence already obtained in [12, Th. 6.3].

Lemma 4.2. *Let d and q be two prime numbers such that $q \equiv 1 \,(\mathrm{mod}\, d)$. Denote by γ_d the pushforward measure of the Haar probability measure on $(\mathbf{S}^1)^{d-1}$ via the Laurent polynomial $\frac{1}{\sqrt{d}} g_d$ (in other words it is the law of the random variable $\frac{1}{\sqrt{d}}\left(\mathrm{Z}_1 + \cdots + \mathrm{Z}_{d-1} + \frac{1}{\mathrm{Z}_1 \ldots \mathrm{Z}_{d-1}}\right)$ where the Z_i are independent and identically distributed uniform random variables on $\mathbf{S}^1$). Then $\mathscr{W}_1(\mu_{q,d}, \gamma_d) \leqslant 2\sqrt{12}\sqrt{d}(d+1)q^{-\frac{1}{d-1}}$.*

Proof. We first use Proposition 3.3 in the particular case where $g = \mathrm{X}^d - 1$. Then $\mathrm{Z}_g = \boldsymbol{\mu}_d$ (the set of d–th roots of unity in $\mathbf{C}$) and K_g is the cyclotomic field $\mathbf{Q}(\boldsymbol{\mu}_d)$. In particular, $[\mathrm{K}_g : \mathbf{Q}] = d-1$ and we can take $\mathrm{C}_g = 1$ in Lemma 3.2 because the roots of unity have modulus 1. We recall that for all $p \in \mathscr{S}_{\mathrm{X}^d-1}$, the measure μ_p denotes the law of the random variable U_p (as introduced in Section 3.1), and λ_{X^d-1} is the Haar probability measure on $\mathrm{H}_{\mathrm{X}^d-1}$. Then using the explicit constant C'_g obtained at the end of the proof of Proposition 3.3, we deduce that

$$\mathscr{W}_1(\mu_p, \lambda_{\mathrm{X}^d-1}) \leqslant 4\sqrt{3}\sqrt{d}(d+1)|p|^{-\frac{1}{d-1}}.$$

Then we pushforward via the map $\sigma_d = d^{-1/2}\sigma$, defined by

$$\sigma_d(f) = \frac{1}{\sqrt{d}} \sum_{x \in \boldsymbol{\mu}_d} f(x).$$

This map is 1-Lipschitz thanks to Lemma 3.4. This gives

$$\mathscr{W}_1\left((\sigma_d)_* \mu_p, (\sigma_d)_* \lambda_{\mathrm{X}^d-1}\right) \leqslant 4\sqrt{3}\sqrt{d}(d+1)|p|^{-\frac{1}{d-1}}. \tag{12}$$

Finally, in [16, §3] we showed that when d is prime the $\mathbf{Z}$-module $\mathrm{R}_{\mathrm{X}^d-1}$ of additive relations among the roots of $\mathrm{X}^d - 1$ is generated by the constant map equal to 1, so $\mathrm{H}_{\mathrm{X}^d-1}$ can be identitified with $(\mathbf{S}^1)^{d-1}$. Then it follows formally from the definitions that for $q \equiv 1 \,(\mathrm{mod}\, d)$ the measure $\mu_{q,d}$ coincides with the measure $(\sigma_d)_*\mu_p$ for all p lying above q (and such p belong to $\mathscr{S}_{\mathrm{X}^d-1}$ thanks to [17, Cor. 10.4]), and the measure γ_d coincides with $(\sigma_d)_*\lambda_{\mathrm{X}^d-1}$. Therefore, (12) actually says that $\mathscr{W}_1(\mu_{q,d}, \gamma_d) \leqslant 4\sqrt{3}\sqrt{d}(d+1)q^{-\frac{1}{d-1}}$. ☐

We will also need a form of the Central Limit Theorem in Wasserstein metric, which we state in the next Lemma.

Lemma 4.3. *Let $k \geqslant 1$ and let $(\mathrm{X}_i)_{i\geqslant 1}$ be a sequence of independent and identically distributed random variables taking values in $\mathbf{R}^k$. Assuming further that they admit a moment* *of order 2, we denote by*

$$m = \mathbf{E}(\mathrm{X}_1) = \begin{pmatrix} \mathbf{E}(\mathrm{X}_{1,1}) \\ \dots \\ \mathbf{E}(\mathrm{X}_{1,k}) \end{pmatrix}$$

the mean value of X_1 *and by* $\Sigma = (\sigma_{i,j})_{1\leqslant i,j\leqslant k}$ *the covariance matrix of* X_1*, meaning that for all* $i,j \in \{1,\dots,k\}$*,* $\sigma_{i,j} = \mathbf{E}((\mathrm{X}_{1,i} - \mathbf{E}(\mathrm{X}_{1,i}))(\mathrm{X}_{1,j} - \mathbf{E}(\mathrm{X}_{1,j})))$*. Then if* μ_n *denotes the law of* $\frac{(\mathrm{X}_1+\cdots\mathrm{X}_n)-nm}{\sqrt{n}}$ *we have*

$$\mathscr{W}_1(\mu_n, \mathscr{N}(0,\Sigma)) \underset{n\to\infty}{\longrightarrow} 0.$$

Proof. Thanks to [22, Th. 7.12], convergence with respect to the p-Wasserstein metric is equivalent to the weak convergence of measures and the convergence of absolute moments of order p. However, in the setting of the Central Limit Theorem we have (denoting by N a random variable with distribution $\mathscr{N}(0,\Sigma)$):

$$\mathbf{E}\Big(\Big|\frac{(\mathrm{X}_1+\cdots\mathrm{X}_n)-nm}{\sqrt{n}}\Big|^2\Big) = \mathbf{E}|\mathrm{N}|^2 = \mathrm{Tr}(\Sigma),$$

hence the convergence of absolute moments of order 2 is automatically satisfied. Therefore, the usual Central Limit Theorem (see, e.g., [1, Th. 29.5]) which states the weak convergence of μ_n to $\mathscr{N}(0,\Sigma)$ immediately gives the apparently stronger statement

$$\mathscr{W}_2(\mu_n, \mathscr{N}(0,\Sigma)) \underset{n\to\infty}{\longrightarrow} 0.$$

Since $\mathscr{W}_1 \leqslant \mathscr{W}_2$, the conclusion follows. □

Remark. We stated a qualitative result that suffices for our application, but there are several articles investigating the rate of convergence in the Central Limit Theorem with respect to Wasserstein metrics. Under the assumption that $\mathbf{E}(|\mathrm{X}_1|^4)$ is finite, [4, Th. 1] states that $\mathrm{W}_2(\mu_n, \mathscr{N}(0,\Sigma)) \ll_k \frac{1}{\sqrt{n}}$. The dependence of the implicit constant with respect to the dimension k is a more subtle question, we refer to [5] and the references therein for a recent account.

Proof of Theorem 4.1. Let q be an odd prime, and $d = d(q)$ a prime divisor of $q-1$. By the triangle inequality for the metric $\mathscr{W}_1$ we have

$$\mathscr{W}_1(\mu_{q,d}, \mathscr{N}(0,\Sigma)) \leqslant \mathscr{W}_1(\mu_{q,d}, \gamma_d) + \mathscr{W}_1(\gamma_d, \mathscr{N}(0,\Sigma)).$$

The second term converges to zero when d tends to infinity thanks to the central limit theorem (the term $(\sqrt{d}\mathrm{Z}_1\dots\mathrm{Z}_{d-1})^{-1}$ does not cause any issue because it has modulus $1/\sqrt{d}$ so it converges almost surely to zero). Moreover, thanks to Lemma 4.2 the first term is upper bounded by $4\sqrt{3}\sqrt{d}(d+1)q^{-\frac{1}{d-1}}$, so it suffices to show that the condition

$$d \underset{q\to+\infty}{=} o\left(\frac{\log q}{\log\log q}\right)$$

implies that this upper bound converges to zero as q goes to infinity. This comes from the fact that Lambert W_0 function (which is the inverse bijection to $x \mapsto xe^x$ on $\left[-\frac{1}{e}, +\infty\right[$)

satisfies $\mathrm{W}_0(x) \sim_{x\to+\infty} \log(x)$, so the condition above may be rewritten as

$$d \underset{q\to\infty}{=} o\left(\frac{\log q}{\mathrm{W}_0(\log q)}\right).$$

Therefore, $d = \frac{\log q}{\mathrm{W}_0(\log q)}\varepsilon(q)$ for some function ε that tends to zero as q tends to infinity, so

$$d\log d = \frac{\log q}{\mathrm{W}_0(\log q)}\varepsilon(q)\log\left(\frac{\log q}{\mathrm{W}_0(\log q)}\varepsilon(q)\right)$$

and for q large enough this is upper bounded by

$$\frac{\log q}{\mathrm{W}_0(\log q)}\varepsilon(q)\log\left(\frac{\log q}{\mathrm{W}_0(\log q)}\right),$$

but by definition of W_0 this is equal to $\log(q)\varepsilon(q)$. This shows that $d\log(d) \underset{q\to\infty}{=} o(\log q)$ and elementary manipulations show that this implies $d^{3/2} \underset{q\to\infty}{=} o\left(q^{\frac{1}{d-1}}\right)$, concluding the proof. □

Another type of factorization of d for which the limiting distribution can be determined is when d is a power of fixed prime. Indeed, when d is of the form r^b where r is a prime number and $b \geqslant 1$, the Laurent polynomial g_d of [12, 13] can be made more explicit (this comes from the fact that the coefficients of the cyclotomic polynomial Φ_{r^b} are known). Precisely, [13, Cor. 1] states that the sums $\mathrm{S}_d(a,q)$ become equidistributed with respect to the pushforward measure of the Haar measure on $(\mathbf{S}^1)^{\varphi(r^b)}$ with respect to the Laurent polynomial g_{r^b} defined by

$$g_{r^b}\left(z_1, z_2, \ldots, z_{\varphi(r^b)}\right) = \sum_{j=1}^{\varphi(r^b)} z_j + \sum_{m=1}^{r^{b-1}} \prod_{\ell=0}^{r-2} z_{m+\ell r^{b-1}}^{-1}.$$

Rearranging the terms according to their residue classes modulo r^{b-1} we can rephrase that statement as follows: the sums $\frac{1}{\sqrt{d}}\mathrm{S}_d(q,a)$ become equidistributed with respect to a measure γ_d which the law of a random variable

$$\frac{1}{r^{b/2}}\sum_{i=1}^{r^{b-1}} \mathrm{Z}_{i,1} + \cdots + \mathrm{Z}_{i,r-1} + \frac{1}{\mathrm{Z}_{i,1}\ldots\mathrm{Z}_{i,r-1}}. \tag{13}$$

where $(\mathrm{Z}_{i,j})_{1\leqslant i\leqslant r^{b-1},\ 1\leqslant j\leqslant r-1}$ is a family of independent and identically distributed Steinhaus random variables. Thanks to the proof of Lemma 4.2 (we only used the fact that d is prime to make γ_d more explicit, but the lemma holds for arbitrary d, except for the description of the Laurent polynomial g_d), we have

$$\mathscr{W}_1(\mu_{q,d}, \gamma_d) \leqslant 4\sqrt{3}\sqrt{d}(d+1)q^{-\frac{1}{d-1}}.$$

This upper bound converges to zero as d and q tend to infinity provided $d = o\left(\frac{\log q}{\log\log q}\right)$. Now, if r is fixed and only b varies, the sum (13) may be rewritten as

$$\frac{1}{\sqrt{r}}\left(\frac{1}{\sqrt{r^{b-1}}}\sum_{i=1}^{r^{b-1}} \mathrm{X}_i\right)$$

where the $\mathrm{X}_i = \mathrm{Z}_{i,1} + \cdots + \mathrm{Z}_{i,r-1} + \frac{1}{\mathrm{Z}_{i,1}\ldots\mathrm{Z}_{i,r-1}}$ are independent and identically distributed random variables which have mean 0. Thanks to the Central Limit Theorem, we have

$$\frac{1}{\sqrt{r^{b-1}}}\sum_{i=1}^{r^{b-1}} \mathrm{X}_i \xrightarrow{\text{law}} \mathscr{N}(0,\Sigma)$$

where $\Sigma = \frac{r}{2}\mathrm{Id}$ is the covariance matrix of X_1 (viewed as a random variable with values in $\mathbf{R}^2$). Taking into account the factor $1/\sqrt{r}$ in front of the sum, we obtain the following result:

Theorem 4.4. *Let r be a fixed prime. For all integers b and all prime numbers q such that $d = r^b$ divides $q-1$, we define the sums $\frac{1}{\sqrt{d}}\mathrm{S}_d(q,a)$ as above. Then as d and q both tend to infinity with $d = o\left(\frac{\log q}{\log\log q}\right)$, they become equidistributed in the complex plane with respect to the normal distribution $\mathscr{N}(0,\frac{1}{2}\mathrm{Id})$.*

Appendix: proof of the Bobkov–Ledoux inequality

In this section, we reproduce and combine the arguments of Bobkov and Ledoux in [2, 3] to obtain the variant of [2, Eq. (1.6)] stated in Theorem 1.2, (7). Following the original source, for an integer $d \geqslant 1$, we identify here the torus $(\mathbf{R}/\mathbf{Z})^d$ with $\mathrm{Q}_d = [0, 2\pi[^d \subset \mathbf{R}^d$ with the distance ϱ_d defined in (6).

Lemma 4.5. *Denote by $\mathrm{Lip}_1^{2\pi}(\mathbf{R}^d,\mathbf{R})$ the set of maps $v\colon \mathbf{R}^d \to \mathbf{R}$ that are 2π-periodic and 1-Lipschitz (with respect to the euclidean norm on $\mathbf{R}^d$). Then we have that for all Borel probability measures μ and ν on $(\mathrm{Q}^d, \varrho_d)$,*

$$\sup_{v\in \mathrm{Lip}_1^{2\pi}(\mathbf{R}^d,\mathbf{R})}\Big|\int_{\mathrm{Q}^d} v d\mu - \int_{\mathrm{Q}^d} v d\nu\Big| = \sup_{w\in \mathrm{Lip}_1^{2\pi}(\mathbf{R}^d,\mathbf{R})\cap \mathscr{C}^\infty}\Big|\int_{\mathrm{Q}^d} w d\mu - \int_{\mathrm{Q}^d} w d\nu\Big|.$$

Proof. This is a standard smoothing argument by convolution. If $v \in \mathrm{Lip}_1^{2\pi}(\mathbf{R}^d,\mathbf{R})$ then one easily checks that for all $\varepsilon > 0$,

$$v_\varepsilon(x) = \frac{1}{(2\pi\varepsilon^2)^{d/2}}\int_{\mathbf{R}^d} v(x-y)e^{-\frac{|y|^2}{2\varepsilon^2}}dy$$

defines a function in $\mathrm{Lip}_1^{2\pi}(\mathbf{R}^d,\mathbf{R}) \cap \mathscr{C}^\infty$ that satisfies $\|v_\varepsilon - v\|_{\infty,\mathbf{R}^d} \underset{\varepsilon\to 0}{\longrightarrow} 0$. Then the result follows from this approximation by a smooth function. ☐

Lemma 4.6. *Let μ and ν be two Borel probability measures on $(\mathrm{Q}^d, \varrho_d)$. The following inequality holds:*

$$\sup_{w\in \mathrm{Lip}_1^{2\pi}(\mathbf{R}^d,\mathbf{R})\cap \mathscr{C}^\infty}\Big|\int_{\mathrm{Q}^d} w d\mu - \int_{\mathrm{Q}^d} w d\nu\Big| \leqslant \Big(\sum_{m\in\mathbf{Z}^d-\{0\}} \frac{|\widehat{\mu}(m)-\widehat{\nu}(m)|^2}{|m|^2}\Big)^{1/2}.$$

Proof. Let $w \in \mathrm{Lip}_1^{2\pi}(\mathbf{R}^d, \mathbf{R}) \cap \mathscr{C}^\infty$. Since w is smooth, it admits a Fourier series expansion that converges absolutely:

$$w(x) = \sum_{m \in \mathbf{Z}^d} a_m e^{im\cdot x}.$$

Moreover, one can differentiate term by term, so that for all $k \in \{1, \dots, d\}$,

$$\frac{\partial w}{\partial x_k}(x) = \sum_{m \in \mathbf{Z}^d} i m_k a_m e^{im\cdot x}.$$

Then thanks to Parseval's equality (the L^2-norm of $\frac{\partial w}{\partial x_\ell}$ equals the ℓ^2-norm of its sequence of Fourier coefficients):

$$\frac{1}{(2\pi)^d}\int_{\mathrm{Q}^d}\Big|\frac{\partial w}{\partial x_k}(x)\Big|^2 dx = \sum_{m \in \mathbf{Z}^d} |m_k|^2 |a_m|^2.$$

Summing over $k \in \{1, \dots, d\}$ yields

$$\frac{1}{(2\pi)^d}\int_{\mathrm{Q}^d}\Big|\nabla w(x)\Big|^2 dx = \sum_{m \in \mathbf{Z}^d} |m|^2 |a_m|^2.$$

Finally, we use the fact that w is 1-Lipschitz to deduce that the norm of its gradient is always bounded above by 1. Therefore,

$$\sum_{m \in \mathbf{Z}^d} |m|^2 |a_m|^2 \leqslant 1. \tag{14}$$

To conclude, we first write

$$\begin{aligned}\Big|\int_{\mathrm{Q}^d} w d\mu - \int_{\mathrm{Q}^d} w d\nu\Big|^2 &= \Big|\sum_{m \in \mathbf{Z}^d} a_m \Big(\int_{\mathrm{Q}^d} e^{im\cdot x} d\mu(x) - \int_{\mathrm{Q}^d} e^{im\cdot x} d\nu(x)\Big)\Big|^2 \\ &= \Big|\sum_{m \in \mathbf{Z}^d} a_m \Big(\widehat{\mu}(m) - \widehat{\nu}(m)\Big)\Big|^2\end{aligned}$$

then we observe that the two Fourier coefficients coincide at $m = 0$, so the right-hand side may be rewritten as

$$\Big|\sum_{m \in \mathbf{Z}^d - \{0\}} |m| a_m \Big(\frac{\widehat{\mu}(m) - \widehat{\nu}(m)}{|m|}\Big)\Big|^2$$

and the conclusion follows from the Cauchy–Schwarz inequality and (14). ☐

One deduces quickly the following corollary:

Corollary 4.7. *Let μ, ν be two Borel probability measures on $(\mathrm{Q}^d, \varrho_d)$. The following inequality holds:*

$$\mathscr{W}_1(\mu, \nu) \leqslant \Big(\sum_{m \in \mathbf{Z}^d - \{0\}} \frac{|\widehat{\mu}(m) - \widehat{\nu}(m)|^2}{|m|^2}\Big)^{1/2}.$$

Proof. Thanks to the dual formulation of Th. 1.2 (6), we have

$$\mathscr{W}_1(\mu,\nu) = \sup_{u\in \mathrm{Lip}_1(\mathrm{Q}^d,\mathbf{R})} \Big|\int_{\mathrm{Q}^d} ud\mu - \int_{\mathrm{Q}^d} ud\nu\Big|.$$

Now, if $u\colon (\mathrm{Q}^d,\varrho_d)\to\mathbf{R}$ is a 1-Lipschitz map (with respect to ϱ_d) then its 2π-periodic extension to $\mathbf{R}^d$ is a 1-Lipschitz map with respect to the euclidean norm on $\mathbf{R}^d$. Conversely, any $v\in \mathrm{Lip}_1^{2\pi}(\mathbf{R}^d,\mathbf{R})$ satisfies that $v_{|\mathrm{Q}^d}$ is 1-Lipschitz with respect to ϱ_d. Therefore,

$$\sup_{u\in \mathrm{Lip}_1(\mathrm{Q}^d,\mathbf{R})} \Big|\int_{\mathrm{Q}^d} ud\mu - \int_{\mathrm{Q}^d} ud\nu\Big| = \sup_{v\in \mathrm{Lip}_1^{2\pi}(\mathbf{R}^d,\mathbf{R})} \Big|\int_{\mathrm{Q}^d} vd\mu - \int_{\mathrm{Q}^d} vd\nu\Big|$$

and thanks to Lemma 4.5 and 4.6, the right-hand side is bounded by

$$\Big(\sum_{m\in\mathbf{Z}^d-\{0\}} \frac{|\widehat{\mu}(m)-\widehat{\nu}(m)|^2}{|m|^2}\Big)^{1/2}.$$

□

This corollary is not suitable for all applications because the series on the right-hand side may well diverge. In order to obtain a more useful inequality, the method is the usual application of convolution by a measure whose sequence of Fourier coefficients is compactly supported. Recall that if μ and ν are Borel probability measures on $(\mathbf{R}/\mathbf{Z})^d$, the convolution $\mu*\nu$ is the unique Borel probability measure on $(\mathbf{R}/\mathbf{Z})^d$ such that

$$\int_{(\mathbf{R}/\mathbf{Z})^d} ud(\mu*\nu) = \int_{(\mathbf{R}/\mathbf{Z})^{2d}} u(x+y)d\mu(x)d\nu(y)$$

for all continuous functions $u\colon (\mathbf{R}/\mathbf{Z})^d\to\mathbf{C}$. As usual, one can identify μ, ν and $\mu*\nu$ with measures on Q^d, and then

$$\int_{\mathbf{R}^d} ud(\mu*\nu) = \int_{\mathbf{R}^d\times\mathbf{R}^d} u(x+y)d\mu(x)d\nu(y)$$

for all functions $u\colon\mathbf{R}^d\to\mathbf{C}$ which are continuous and 2π-periodic.

Lemma 4.8. *Let μ and ν be two Borel probability measures on (Q^d,ϱ_d) and let $\mathrm{H} = (\mathrm{H}_1,\dots,\mathrm{H}_d)$ be a random vector in $\mathbf{R}^d$. For $x\in\mathbf{R}$, denote by $\mathrm{M}(x)$ the unique element of $(x+2\pi\mathbf{Z})\cap(-\pi,\pi]$. Let N be the random vector $(\mathrm{M}(\mathrm{H}_1),\dots,\mathrm{M}(\mathrm{H}_d))$ and let η denote the law of N. Then*

$$\mathscr{W}_1(\mu,\nu)\leqslant \mathscr{W}_1(\mu*\eta,\nu*\eta)+2\mathbf{E}(|\mathrm{H}|).$$

Proof. Let $u\in \mathrm{Lip}_1^{2\pi}(\mathbf{R}^d,\mathbf{R})$. By the triangle inequality, we have

$$\Big|\int_{\mathrm{Q}^d} ud\mu - \int_{\mathrm{Q}^d} ud\nu\Big| \leqslant \Big|\int_{\mathrm{Q}^d} ud\mu - \int_{\mathrm{Q}^d} ud(\mu*\eta)\Big| + \Big|\int_{\mathrm{Q}^d} ud(\mu*\eta) - \int_{\mathrm{Q}^d} ud(\nu*\eta)\Big| + \Big|\int_{\mathrm{Q}^d} ud(\nu*\eta) - \int_{\mathrm{Q}^d} ud\nu\Big|$$

and thanks to Theorem 1.2 (6),

$$\Big|\int_{\mathrm{Q}^d} ud(\mu*\eta) - \int_{\mathrm{Q}^d} ud(\nu*\eta)\Big| \leqslant \mathscr{W}_1(\mu*\eta,\nu*\eta).$$

Thus, it suffices to show that

$$\Big|\int_{\mathrm{Q}^d} u d\mu - \int_{\mathrm{Q}^d} u d(\mu * \eta)\Big| \leqslant \mathbf{E}(|\mathrm{H}|),$$

since the same proof will also show the inequality for the term with ν and $\nu * \eta$. Now, by definition of the convolution and the fact that η is a probability measure, we have

$$\begin{aligned}\Big|\int_{\mathrm{Q}^d} u d(\mu * \eta) - \int_{\mathrm{Q}^d} u d\mu\Big| &= \Big|\int_{\mathrm{Q}^d}\Big(\int_{\mathrm{Q}^d} u(x+y) - u(x) d\mu(x)\Big) d\eta(y)\Big| \\ &\leqslant \int_{\mathrm{Q}^d}\Big(\int_{\mathrm{Q}^d} |u(x+y) - u(x)| d\mu(x)\Big) d\eta(y).\end{aligned}$$

Using the fact that $u \in \mathrm{Lip}_1^{2\pi}(\mathbf{R}^d, \mathbf{R})$, we deduce that

$$\Big|\int_{\mathrm{Q}^d} u d(\mu * \eta) - \int_{\mathrm{Q}^d} u d\mu\Big| \leqslant \int_{\mathrm{Q}^d} |y| d\eta(y) = \mathbf{E}(|\mathrm{N}|).$$

Finally, since for all $x \in \mathbf{R}$, $|\mathrm{M}(x)| \leqslant |x|$, we have $\mathbf{E}(|\mathrm{N}|) \leqslant \mathbf{E}(|\mathrm{H}|)$, hence the conclusion. □

Lemma 4.9. *For all integers* $\mathrm{T} \geqslant 1$*, there exists a random variable* H *with values in* $\mathbf{R}^d$ *whose characteristic function is supported on the cube* $[-\mathrm{T}, \mathrm{T}]^d$ *and such that* $\mathbf{E}(|\mathrm{H}|) \leqslant \frac{2\sqrt{3}\sqrt{d}}{\mathrm{T}}$.

Proof. As a first attempt (that will require some adjustments) let us use what is known on the Fourier transform of the triangle function: it is a classic fact that if X is a real random variable whose density function is

$$f(x) = \frac{2(1 - \cos(x/2))}{\pi x^2}$$

then its characteristic function $\varphi_{\mathrm{X}}(\bullet) = \mathbf{E}(e^{i\mathrm{X}\cdot\bullet})$ is given by

$$\varphi_{\mathrm{X}}(s) = (1 - 2|s|)^+$$

where $(-)^+$ denotes the positive part. In particular, φ_{X} is supported on $\left[-\frac{1}{2}, \frac{1}{2}\right]$, so that suitable renormalizations of X will allow us to construct random variables with support in $[-\mathrm{T}, \mathrm{T}]$ for all T. However, the issue is that $\mathbf{E}(|\mathrm{X}|) = +\infty$, which would make the inequality of Lemma 4.8 useless. This is why the following adjustment is needed: rather than working with X, we will work with the random variable ξ whose characteristic function is given by

$$w(s) = 3(\varphi_{\mathrm{X}} * \varphi_{\mathrm{X}})(s) = 3\int_{\mathbf{R}} \varphi_{\mathrm{X}}(s-t)\varphi_{\mathrm{X}}(t) dt. \tag{15}$$

Here, the factor 3 is just a normalization factor to ensure that $w(0) = 1$, which is a necessary condition in order to be a characteristic function. Then thanks to the convolution theorem for the Fourier transform, w is the characteristic function φ_ξ of a random variable ξ with density

$$g(x) = 6\pi f(x)^2 = \frac{24(1 - \cos(x/2))^2}{\pi x^4}.$$

Since $w = 3(\varphi_{\mathrm{X}} * \varphi_{\mathrm{X}})$, it is supported on $[-1, 1]$, and this time $\mathbf{E}(|\xi|) < +\infty$. Even better, $\mathbf{E}(\xi^2)$ is also finite, and can be explicitly computed! Indeed, since w is the characteristic function of ξ, the second moment of ξ is equal to $-w''(0)$, and this can be calculated from (15), yielding $\mathbf{E}(\xi^2) = 12$. To conclude, it suffices to define H as $\frac{1}{\mathrm{T}}(\xi_1, \ldots, \xi_d)$ for independent random variables ξ_i having the same distribution as ξ. Indeed, $\mathbf{E}(|\mathrm{H}|^2) = \frac{12d}{\mathrm{T}^2}$ and $(\mathbf{E}|\mathrm{H}|)^2 \leqslant \mathbf{E}(|\mathrm{H}|^2)$ so $\mathbf{E}(|\mathrm{H}|) \leqslant \frac{2\sqrt{3}\sqrt{d}}{\mathrm{T}}$. Moreover, for all $s = (s_1, \dots, s_d) \in \mathbf{R}^d$,

$$\varphi_{\mathrm{H}}(s) = \prod_{j=1}^{d} w(s_j/\mathrm{T})$$

so the choice of a w supported on $[-1, 1]$ implies that φ_{H} is supported on $[-\mathrm{T}, \mathrm{T}]^d$. □

Remark. Among all random vectors H such that φ_{H} is supported on $[-\mathrm{T}, \mathrm{T}]^d$, what is the lowest $\mathbf{E}(|\mathrm{H}|)$ one can hope for? In this remark, we show that the result of Lemma 4.9 is close to optimal.

Let us denote by $(e_1, \dots, e_d)$ the canonical basis of $\mathbf{R}^d$ and by

$$\begin{array}{cccc} \varphi_{\mathrm{H}} & : & \mathbf{R}^d & \to & \mathbf{C} \\ & & t = (t_1, \dots, t_d) & \mapsto & \mathbf{E}(e^{it\cdot \mathrm{H}}) \end{array}$$

the characteristic function of H. Then for all $j \in \{1, \dots, d\}$,

$$\frac{\partial \varphi_{\mathrm{H}}}{\partial t_j}(t) = i\mathbf{E}(\mathrm{H}_j e^{it\cdot \mathrm{H}}).$$

Therefore, if φ_{H} is supported on $[-\mathrm{T}, \mathrm{T}]^d$, we have $\varphi_{\mathrm{H}}(\mathrm{T}e_j) = 0$, hence

$$1 = \Big|\varphi_{\mathrm{H}}(\mathrm{T}e_j) - \varphi_{\mathrm{H}}(0)\Big| = \Big|\int_0^{\mathrm{T}} \frac{\partial \varphi_{\mathrm{H}}}{\partial t_j}(se_j)ds\Big| \leqslant \mathrm{T}\mathbf{E}(|\mathrm{H}_j|)$$

thanks to the triangle inequality. Summing over j and using Cauchy–Schwarz inequality yields

$$d \leqslant \mathrm{T}\mathbf{E}(|\mathrm{H}_1| + \cdots + |\mathrm{H}_d|) \leqslant \mathrm{T}\sqrt{d}\mathbf{E}(|\mathrm{H}|)$$

hence

$$\frac{\sqrt{d}}{\mathrm{T}} \leqslant \mathbf{E}(|\mathrm{H}|).$$

Therefore the construction of H in Lemma 4.9 gives the best possible dependence with respect to T and d. The only question that remains is whether one can obtain $\mathbf{E}(|\mathrm{H}|) = c\frac{\sqrt{d}}{\mathrm{T}}$ for some $1 \leqslant c < 2\sqrt{3}$.

Proof of Theorem 1.2, (7). Let $\mathrm{T} \geqslant 1$ and let H be a random vector as given by the previous lemma. If μ and ν are two Borel measures on $(\mathrm{Q}^d, \varrho_d)$, then thanks to Lemma 4.8 we have

$$\mathscr{W}_1(\mu, \nu) \leqslant \mathscr{W}_1(\mu * \eta, \nu * \eta) + 2\mathbf{E}(|\mathrm{H}|)$$

The choice of H ensures that $\mathbf{E}(|\mathrm{H}|) \leqslant \frac{2\sqrt{3}\sqrt{d}}{\mathrm{T}}$, so it just remains to prove that

$$\mathscr{W}_1(\mu * \eta, \nu * \eta) \leqslant \Big(\sum_{\substack{m \in \mathbf{Z}^d \\ 0 < \|m\|_\infty \leqslant \mathrm{T}}} \frac{|\widehat{\mu}(m) - \widehat{\nu}(m)|^2}{|m|^2} \Big)^{1/2}.$$

This follows from Corollary 4.7 and the fact that $\widehat{\eta}(m) = 0$ for all $m \notin [-\mathrm{T}, \mathrm{T}]^d$. □

References

[1] Patrick Billingsley, *Probability and measure*, Third, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1995. A Wiley-Interscience Publication.

[2] Sergey G. Bobkov and Michel Ledoux, *Transport inequalities on Euclidean spaces for non-Euclidean metrics*, J. Fourier Anal. Appl. **26** (2020), no. 4, 27.

[3] ———, *A simple Fourier analytic proof of the AKT optimal matching theorem*, Ann. Appl. Probab. **31** (2021), no. 6, 2567–2584.

[4] Thomas Bonis, *Stein's method for normal approximation in Wasserstein distances with application to the multivariate central limit theorem*, Probab. Theory Relat. Fields **178** (2020), no. 3-4, 827–860.

[5] ———, *Improved rates of convergence for the multivariate central limit theorem in Wasserstein distance*, Electron. J. Probab. **29** (2024).

[6] Bence Borda, *Berry-Esseen smoothing inequality for the Wasserstein metric on compact Lie groups*, J. Fourier Anal. Appl. **27** (2021), no. 2, 24.

[7] ———, *Equidistribution of random walks on compact groups. II: The Wasserstein metric*, Bernoulli **27** (2021), no. 4, 2598–2623.

[8] J. Bourgain, A. A. Glibichuk, and S. V. Konyagin, *Estimates for the number of sums and products and for exponential sums in fields of prime order*, J. London Math. Soc. (2) **73** (2006), no. 2, 380–398.

[9] L. Brown and S. Steinerberger, *On the Wasserstein distance between classical sequences and the Lebesgue measure*, Trans. Am. Math. Soc. **373** (2020), 8943–8962.

[10] Paula Burkhardt, Alice Zhuo-Yu Chan, Gabriel Currier, Stephan Ramon Garcia, Florian Luca, and Hong Suh, *Visual properties of generalized Kloosterman sums*, J. Number Theory **160** (2016), 237–253.

[11] Michael Drmota and Robert F. Tichy, *Sequences, discrepancies and applications*, Lect. Notes Math., vol. 1651, Berlin: Springer, 1997.

[12] William Duke, Stephan Ramon Garcia, and Bob Lutz, *The graphic nature of Gaussian periods*, Proc. Am. Math. Soc. **143** (2015), no. 5, 1849–1863.

[13] Stephan Ramon Garcia, Trevor Hyde, and Bob Lutz, *Gauss's hidden menagerie: from cyclotomy to supercharacters*, Notices Am. Math. Soc. **62** (2015), no. 8, 878–888.

[14] C. Graham, *Irregularity of distribution in Wasserstein distance*, J. Fourier Anal. Appl. **26–75** (2020).

[15] S. V. Konyagin, *Exponential sums over multiplicative groups in fields of prime order and related combinatorial problems*, Lecture notes available at https://www.mathtube.org/sites/default/files/lecture-notes/Konyagin_Lectures.pdf.

[16] Emmanuel Kowalski and Théo Untrau, *Ultra-short sums of trace functions*, Acta Arith. **210** (2023), 367–390.

[17] Jürgen Neukirch, *Algebraic number theory. Transl. from the German by Norbert Schappacher*, Grundlehren Math. Wiss., vol. 322, Berlin: Springer, 1999.

[18] Samantha Platt, *Visual aspects of Gaussian periods and analogues*, Int. J. Number Theory **20** (2024), no. 9, 2227–2266.

[19] Stefan Steinerberger, *Wasserstein distance, Fourier series and applications*, Monatsh. Math. **194** (2021), no. 2, 305–338.

[20] Théo Untrau, *Equidistribution of exponential sums indexed by a subgroup of fixed cardinality*, Math. Proc. Camb. Philos. Soc. **176** (2024), no. 1, 65–94.

[21] Théo Untrau, *Study of the distribution of some short exponential sums*, Ph.D. thesis, Université de Bordeaux, https://theses.hal.science/tel-04230646 (2023).

[22] Cédric Villani, *Topics in optimal transportation*, Grad. Stud. Math., vol. 58, Providence, RI: American Mathematical Society (AMS), 2003.

ETH Zürich – D-MATH, Rämistrasse 101, 8092 Zürich, Switzerland

Email address: kowalski@math.ethz.ch

Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France

Email address: theo.untrau@ens-rennes.fr