Journal of Fourier Analysis and Applications (2021) 27:13
https://doi.org/10.1007/s00041-020-09803-0

Berry–Esseen Smoothing Inequality for the Wasserstein Metric on Compact Lie Groups

Bence Borda[1,2]

Received: 18 March 2020 / Accepted: 23 May 2020 / Published online: 3 March 2021

Abstract
We prove a sharp general inequality estimating the distance of two probability measures on a compact Lie group in the Wasserstein metric in terms of their Fourier transforms. We use a generalized form of the Wasserstein metric, related by Kantorovich duality to the family of functions with an arbitrarily prescribed modulus of continuity. The proof is based on smoothing with a suitable kernel, and a Fourier decay estimate for continuous functions. As a corollary, we show that the rate of convergence of random walks on semisimple groups in the Wasserstein metric is necessarily almost exponential, even without assuming a spectral gap. Applications to equidistribution and empirical measures are also given.

Keywords Transport metric · Fourier transform · Compact group · Random walk · Spectral gap · Erdős–Turán inequality · Equidistribution

Mathematics Subject Classification 43A77 · 60B15

1 Introduction

If the Fourier transform of two Borel probability measures on $\mathbb{R}$ are equal, then the measures themselves are also equal. The celebrated Berry–Esseen smoothing inequality is a quantitative form of this fundamental fact of classical Fourier analysis. Given two Borel probability measures ν_1 and ν_2 on $\mathbb{R}$, let $F_j(x) = \nu_j((-\infty, x])$, $j = 1, 2$, and define

$$\delta_{\mathrm{unif}}(\nu_1, \nu_2) = \sup_{x \in \mathbb{R}} |F_1(x) - F_2(x)| .$$

Communicated by Fulvio Ricci.

✉ Bence Borda
borda@math.tugraz.at

1 Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria

2 Alfréd Rényi Institute of Mathematics, Reáltanoda utca 13–15, 1053 Budapest, Hungary

Theorem A (Berry–Esseen smoothing inequality) *Assume that* $|F_2(x) - F_2(y)| \leq K|x-y|$ *for all* $x, y \in \mathbb{R}$ *with some constant* $K > 0$. *Then for any real number* $T > 0$,

$$\delta_{\mathrm{unif}}(\nu_1, \nu_2) \ll \frac{K}{T} + \int_{-T}^{T} \frac{|\widehat{\nu_1}(t) - \widehat{\nu_2}(t)|}{|t|} \, \mathrm{d}t$$

with a universal implied constant.

In the terminology of probability theory, $F_j(x)$ is the distribution function of ν_j; the Fourier transform $\widehat{\nu_j}(t) = \int_{\mathbb{R}} e^{itx} \, \mathrm{d}\nu_j(x)$ is the characteristic function of ν_j; finally, δ_{unif} is the uniform metric (or Kolmogorov metric) on the set of probability distributions. For somewhat sharper forms of Theorem A see Petrov [28, Chapter 5.1]. Throughout the paper $A \ll B$ and $A = O(B)$ mean that $A \leq CB$ with some implied constant $C > 0$.

Similar smoothing inequalities are known for several other probability metrics on $\mathbb{R}$, see Bobkov [4] for a survey. Some, but not all require a smoothness assumption on one of the distributions; for instance, Theorem A is usually formulated under the assumption that F_2 is differentiable and $|F_2'(x)| \leq K$. A common feature of such results is that the distance of ν_1 and ν_2 in some probability metric is bounded above by the sum of two terms depending on a free parameter $T > 0$: one term decays as T increases, and the other term depends on the Fourier transforms of ν_1 and ν_2 only on the interval $[-T, T]$.

Berry–Esseen type smoothing inequalities are known in other spaces as well. The first multidimensional version, an upper bound for the uniform metric on $\mathbb{R}^d$ is due to von Bahr [36]. Niederreiter and Philipp proved an analogous result for two Borel probability measures ν_1 and ν_2 on the torus $\mathbb{R}^d/\mathbb{Z}^d$. By identifying $\mathbb{R}^d/\mathbb{Z}^d$ with the unit cube $[0,1)^d$, we can define the uniform metric as $\delta_{\mathrm{unif}}(\nu_1, \nu_2) = \sup_{x \in [0,1]^d} |\nu_1(B(x)) - \nu_2(B(x))|$, where $B(x) = [0, x_1) \times \cdots \times [0, x_d)$. The Fourier transform is now $\widehat{\nu_j}(m) = \int_{\mathbb{R}^d/\mathbb{Z}^d} e^{-2\pi i \langle m, x \rangle} \, \mathrm{d}\nu_j(x)$, $m \in \mathbb{Z}^d$. Let $\mu_{\mathbb{R}^d/\mathbb{Z}^d}$ be the normalized Haar measure, and let $\|m\|_\infty = \max_{1 \leq k \leq d} |m_k|$.

Theorem B (Niederreiter–Philipp [25]) *Assume that* $\nu_2(B) \leq K \mu_{\mathbb{R}^d/\mathbb{Z}^d}(B)$ *for all axis parallel boxes* $B \subseteq [0,1)^d$ *with some constant* $K > 0$. *Then for any real number* $M > 0$,

$$\delta_{\mathrm{unif}}(\nu_1, \nu_2) \ll \frac{K}{M} + \sum_{\substack{m \in \mathbb{Z}^d \\ 0 < \|m\|_\infty < M}} \frac{|\widehat{\nu_1}(m) - \widehat{\nu_2}(m)|}{\prod_{k=1}^{d} \max\{|m_k|, 1\}}$$

with an implied constant depending only on d.

The goal of this paper is to prove a Berry–Esseen type smoothing inequality in more general compact groups. In this more general setting only those probability metrics remain meaningful whose definition does not rely on concepts such as axis parallel boxes and distribution functions. One of the most important such metrics is the p-Wasserstein metric W_p. Given a compact metric space (X, ρ) and two Borel probability measures ν_1 and ν_2 on X, we define

$$W_p(\nu_1, \nu_2) = \inf_{\vartheta \in \mathrm{Coup}(\nu_1, \nu_2)} \int_{X \times X} \rho(x, y)^p \, \mathrm{d}\vartheta(x, y) \qquad (0 < p \leq 1),$$

and

$$W_p(\nu_1, \nu_2) = \inf_{\vartheta \in \mathrm{Coup}(\nu_1, \nu_2)} \left(\int_{X \times X} \rho(x, y)^p \, \mathrm{d}\vartheta(x, y) \right)^{1/p} \qquad (1 < p < \infty).$$

Here $\mathrm{Coup}(\nu_1, \nu_2)$ is the set of couplings of ν_1 and ν_2; that is, the set of Borel probability measures ϑ on $X \times X$ with marginals $\vartheta(B \times X) = \nu_1(B)$ and $\vartheta(X \times B) = \nu_2(B)$, $B \subseteq X$ Borel. Recall that for any $p > 0$, W_p is a metric on the set of Borel probability measures on X, and it generates the topology of weak convergence. Observe also the general inequalities $W_p(\nu_1, \nu_2) \leq W_1(\nu_1, \nu_2)^p$, $0 < p \leq 1$ and $W_1(\nu_1, \nu_2) \leq W_p(\nu_1, \nu_2)$, $1 \leq p < \infty$. The Wasserstein metric originates in the theory of optimal transportation, see Villani [35].

Respecting the philosophy of the Berry–Esseen inequality, we wish to find an upper bound to $W_p(\nu_1, \nu_2)$ depending on the Fourier transform of ν_1 and ν_2 only up to a certain "level". For this reason we chose to work with compact Lie groups, where the theory of highest weights provides a suitable framework to formalize the meaning of "level". More precisely, our main result applies to any compact, connected Lie group G; classical examples include $\mathbb{R}^d/\mathbb{Z}^d$, $\mathrm{U}(d)$, $\mathrm{SU}(d)$, $\mathrm{SO}(d)$, $\mathrm{Sp}(d)$ and $\mathrm{Spin}(d)$. Let $\widehat{G}$ denote the unitary dual, and let d_π, λ_π and κ_π denote the dimension, the highest weight and the negative Laplace eigenvalue of the representation $\pi \in \widehat{G}$, respectively. Further, let $\|A\|_{\mathrm{HS}} = \sqrt{\mathrm{tr}(A^* A)}$ be the Hilbert–Schmidt norm of a matrix A. For a more formal setup we refer to Sect. 2.1. In this paper we prove the following Berry–Esseen type smoothing inequality for W_p on compact Lie groups.

Theorem 1 *Let ν_1 and ν_2 be Borel probability measures on a compact, connected Lie group G. For any $0 < p \leq 1$ and any real number $M > 0$,*

$$W_p(\nu_1, \nu_2) \ll \frac{1}{M^p} + M^{1-p} \left(\sum_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| < M}} \frac{d_\pi}{\kappa_\pi} \|\widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi)\|_{\mathrm{HS}}^2 \right)^{1/2} \tag{1}$$

with an implied constant depending only on G.

The result holds without any smoothness assumption on ν_1 and ν_2. As the applications in Sect. 2.3 will show, the inequality is sharp up to a constant factor depending on G.

In the simplest case of the torus $G = \mathbb{R}^d/\mathbb{Z}^d$ equation (1) reads

$$W_p(\nu_1, \nu_2) \ll \frac{1}{M^p} + M^{1-p} \left(\sum_{\substack{m\in\mathbb{Z}^d \\ 0<|m|<M}} \frac{|\widehat{\nu_1}(m) - \widehat{\nu_2}(m)|^2}{|m|^2} \right)^{1/2} \tag{2}$$

with an implied constant depending on d, where $|m|$ is the Euclidean norm of m; this should be compared to Theorem B. For a detailed proof with explicit constants we refer to our earlier paper [8, Proposition 3] and to Bobkov and Ledoux [5,6].

Our methods do not work when $p > 1$; the reason is that the proof is based on Kantorovich duality for W_p. Recall that the Kantorovich duality theorem states that for any $0 < p \leq 1$,

$$W_p(\nu_1, \nu_2) = \sup_{f\in\mathcal{F}_p} \left| \int_G f \,\mathrm{d}\nu_1 - \int_G f \,\mathrm{d}\nu_2 \right|,$$

where, with ρ denoting the geodesic distance on G,

$$\mathcal{F}_p = \left\{ f : G \to \mathbb{R} \, : \, |f(x) - f(y)| \leq \rho(x, y)^p \text{ for all } x, y \in G \right\}$$

is the set of p-Hölder functions on G, with Hölder constant 1. Theorem 1 thus estimates the difference of the integrals of f with respect to ν_1 and ν_2 uniformly in $f \in \mathcal{F}_p$. From our methods it also follows that for any $f \in \mathcal{F}_p$,

$$\left| \int_G f \,\mathrm{d}\nu_1 - \int_G f \,\mathrm{d}\nu_2 \right| \ll \frac{1}{M^p} + \sum_{\substack{\pi\in\widehat{G} \\ 0<|\lambda_\pi|<M}} d_\pi \|\widehat{f}(\pi)\|_{\mathrm{HS}} \cdot \|\widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi)\|_{\mathrm{HS}}$$

with an implied constant depending only on G; see Proposition 5. Hence for a given $f \in \mathcal{F}_p$ whose Fourier transform decays fast enough, the results of Theorem 1 can be improved. Fast Fourier decay follows e.g. from suitable smoothness assumptions on f, see Sugiura [32]. We mention that by prescribing a higher order modulus of continuity for f, the term $1/M^p$ can also be improved. Note that Fourier decay rates play a role in classical Berry–Esseen type inequalities as well: the coefficient $|t|^{-1}$ (resp. $\prod_{k=1}^d \max\{|m_k|, 1\}^{-1}$) in Theorem A (resp. Theorem B) is explained by the fact that the Fourier transform of the indicator function of an interval (resp. axis parallel box) decays at this rate.

The most straightforward application of Theorem 1 is estimating the rate of convergence of random walks in the W_p metric. Let ν^{*k} denote the k-fold convolution power of ν, and let μ_G be the Haar measure on G. Recall that $\nu^{*k} \to \mu_G$ weakly as $k \to \infty$ if and only if the support of ν is contained neither in a proper closed subgroup, nor in a coset of a proper closed normal subgroup of G; see Stromberg [31]. Using a nonuniform spectral gap result of Varjú, we prove the following application of Theorem 1.

Corollary 2 *Let ν be a Borel probability measure on a compact, connected, semisimple Lie group G. If $\nu^{*k} \to \mu_G$ weakly as $k \to \infty$, then*

$$W_1(\nu^{*k}, \mu_G) \ll e^{-ck^{1/3}},$$

where the constant $c > 0$ and the implied constant depend only on G and ν.

The condition of semisimplicity cannot be removed. The rate of convergence $W_p \ll \exp(-pck^{1/3})$, $0 < p \le 1$ immediately follows. The main motivation came from our recent paper [8] on quantitative ergodic theorems for random walks. Given independent, identically distributed G-valued random variables $\zeta_1, \zeta_2, \dots$ with distribution ν, we showed that for any $f \in \mathcal{F}_p$ the sum $\sum_{k=1}^N f(\zeta_1\zeta_2\cdots\zeta_k)$ satisfies the central limit theorem and the law of the iterated logarithm, provided that $\sum_{k=1}^\infty W_p(\nu^{*k}, \mu_G) < \infty$. Corollary 2 thus provides a large class of examples of random walks with fast enough convergence in W_p, and consequently to which our quantitative ergodic theorems apply. We do not know whether $W_p(\nu^{*k}, \mu_G) \ll e^{-ck^{1/3}}$ remains true for $p > 1$.

Another possible application is in uniform distribution theory, where the goal is finding finite sets $\{a_1, a_2, \dots, a_N\} \subset G$ which make the integration error $|N^{-1}\sum_{k=1}^N f(a_k) - \int_G f \,\mathrm{d}\mu_G|$ small for a suitable class of test functions. Applying Theorem 1 to $\nu_1 = N^{-1}\sum_{k=1}^N \delta_{a_k}$ (where δ_a is the Dirac measure concentrated at $a \in G$) and $\nu_2 = \mu_G$, we can quantitatively measure how well distributed a finite set is with respect to test functions $f \in \mathcal{F}_p$. Note that in this case we have

$$\|\widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi)\|_{\mathrm{HS}}^2 = \frac{1}{N^2}\sum_{k,\ell=1}^N \chi_\pi(a_k^{-1}a_\ell), \tag{3}$$

where $\chi_\pi(x) = \operatorname{tr}\pi(x)$ is the character of the representation $\pi \in \widehat{G}$. Theorem 1 thus becomes an abstract version of the Erdős–Turán inequality, estimating the distance of a finite set from uniformity in terms of character sums. As an illustration, consider the classical construction of Lubotzky, Phillips and Sarnak [23,24] of a finite point set in SO(3) with optimal spectral gap. We will prove that this point set is also optimally close to the Haar measure in the W_p metric up to a constant factor; see Sect. 2.3.2 for a more precise formulation.

Corollary 3 *The finite point set $\{a_1, a_2, \dots, a_N\} \subset \mathrm{SO}(3)$ with optimal spectral gap constructed in [23,24] satisfies, for all $0 < p \le 1$,*

$$W_p\left(\frac{1}{N}\sum_{k=1}^N \delta_{a_k}, \mu_{\mathrm{SO}(3)}\right) \ll N^{-p/3}$$

with a universal implied constant.

Estimating W_p is more difficult in the $p > 1$ case, when Kantorovich duality is not available. Graham [21] gave an analogue of (2) for all $p \ge 1$ on $\mathbb{R}/\mathbb{Z}$ and on the interval $[0, 1]$. His arguments do not generalize to higher dimensions; see, however, [30] for connections between W_p, $p \ge 1$ and Fourier analysis on $\mathbb{R}^d/\mathbb{Z}^d$. Fourier methods were used to estimate W_p, $p \ge 1$ in the optimal matching problem in [1,5,6,14].

Brown and Steinerberger [13,29] considered the more abstract setting of compact Riemannian manifolds, and estimated the distance in W_2 from $N^{-1}\sum_{k=1}^{N}\delta_{a_k}$ to the Riemannian volume in terms of character sums (3), and also in terms of the Green function of the Laplace–Beltrami operator. A similar Erdős–Turán inequality on compact Riemannian manifolds with respect to sufficiently nice test sets was proved by Colzani, Gigante and Travaglini [17]. Numerical results for certain finite point sets on the orthogonal group $\mathrm{O}(d)$ and on Grassmannian manifolds were obtained by Pausinger [27].

For probability distributions on $\mathbb{R}$, Esseen [19], [4, Corollary 8.3] used a smoothing inequality for $W_1(\nu_1,\nu_2)$ to estimate the rate of convergence in W_1 in the central limit theorem; our Theorem 1 and Corollary 2 are far reaching analogues of these classical results in the compact setting. In the $p>1$ case the only known smoothing inequality on $\mathbb{R}$ applies to $W_2(\nu_1,\nu_2)$ with ν_2 a Gaussian distribution [4, Theorem 11.1].

The discussion above can be generalized from $\mathcal{F}_p$ to the class of functions with an arbitrarily prescribed modulus of continuity, and we will actually work out the details in this generality. In particular, our results apply to any given $f\in C(G)$. The formal setup and notation are given in Sect. 2.1; we state the general form of Theorem 1 with explicit constants in Sect. 2.2; applications to random walks and to uniform distribution theory are discussed in more detail, and the proofs of Corollaries 2 and 3 are given in Sect. 2.3. The proof of the main result, Theorem 4 is given in Sect. 3.

2 Results

2.1 Notation

For the general theory of compact Lie groups we refer to Bourbaki [9, Chapter 9]. Throughout the paper G denotes a compact, connected Lie group with identity element $e\in G$ and Lie algebra $\mathfrak{g}$. Let $\exp:\mathfrak{g}\to G$ and $n=\dim G$ denote the exponential map and the dimension of G as a real smooth manifold. Fix an Ad-invariant inner product $(\cdot,\cdot)$ on $\mathfrak{g}$, and let $|X|=\sqrt{(X,X)}$, $X\in\mathfrak{g}$. This inner product defines a Riemannian metric on G; let ρ denote the corresponding geodesic metric on G. The Laplace–Beltrami operator on G is $\Delta=\sum_{k=1}^{n}X_kX_k$ (as an element of the universal enveloping algebra of $\mathfrak{g}$), where $X_1,\dots,X_n$ is an orthonormal base in $\mathfrak{g}$ with respect to $(\cdot,\cdot)$; this does not depend on the choice of the orthonormal base.

Let r denote the rank of G, and fix a maximal torus T in G with Lie algebra $\mathfrak{t}$. Let $\mathfrak{t}^*=\mathrm{Hom}(\mathfrak{t},\mathbb{R})$ denote the dual vector space. The sets

$$\begin{aligned}\Gamma&=\{X\in\mathfrak{t}\,:\,\exp(2\pi X)=e\},\\ \Gamma^*&=\left\{\lambda\in\mathfrak{t}^*\,:\,\lambda(X)\in\mathbb{Z}\text{ for all }X\in\Gamma\right\}\end{aligned}$$

are dual lattices of full rank in $\mathfrak{t}$ and $\mathfrak{t}^*$, respectively. The inner product on $\mathfrak{g}$ naturally defines an inner product on $\mathfrak{t}^*$, which we also denote by $(\cdot,\cdot)$; we also write $|\lambda|=\sqrt{(\lambda,\lambda)}$, $\lambda\in\mathfrak{t}^*$. The inner product defines a normalized Lebesgue measure m on $\mathfrak{t}$.

The weights will be considered elements of Γ^*; the character of T corresponding to $\lambda \in \Gamma^*$ is $\exp(2\pi X) \mapsto e^{2\pi i \lambda(X)}$, $X \in \mathfrak{t}$. Let R be the set of roots, and choose a set of positive roots R^+; we have $|R| = n - r$ and $|R^+| = (n-r)/2$. Let $\Gamma^*_{\mathrm{root}} \subseteq \Gamma^*$ be the lattice spanned by the roots. Further, let

$$C^+ = \left\{ \lambda \in \mathfrak{t}^* \, : \, (\lambda, \alpha) \geq 0 \text{ for all } \alpha \in R^+ \right\}$$

be the dominant Weyl chamber; the set of dominant weights is thus $\Gamma^* \cap C^+$. The Weyl group of G with respect to T will be denoted by $W(G,T) = N_G(T)/T$.

Let $\widehat{G}$ be the unitary dual of G. For any $\pi \in \widehat{G}$, let d_π and λ_π denote the dimension and the highest weight of π. The map $\pi \mapsto \lambda_\pi$ is a bijection from $\widehat{G}$ to the set of dominant weights $\Gamma^* \cap C^+$. Let $\kappa_\pi \geq 0$ denote the negative Laplace eigenvalue of π; that is, $\Delta\pi = -\kappa_\pi \pi$ where Δ acts entrywise. Recall that

$$\kappa_\pi = |\lambda_\pi|^2 + 2(\lambda_\pi, \rho^+) \qquad \text{and} \qquad d_\pi = \frac{\prod_{\alpha \in R^+} (\lambda_\pi + \rho^+, \alpha)}{\prod_{\alpha \in R^+} (\rho^+, \alpha)},$$

where $\rho^+ = \sum_{\alpha \in R^+} \alpha/2$ is the half-sum of positive roots; in particular,

$$|\lambda_\pi|^2 \leq \kappa_\pi \leq |\lambda_\pi|^2 + O(|\lambda_\pi|) \qquad \text{and} \qquad d_\pi \ll |\lambda_\pi|^{(n-r)/2}.$$

Let μ_G (resp. μ_T) denote the normalized Haar measure on G (resp. T). The Fourier transform of a function $f : G \to \mathbb{C}$ is $\widehat{f}(\pi) = \int_G f(x)\pi(x)^* \, \mathrm{d}\mu_G(x)$, $\pi \in \widehat{G}$; that of a Borel probability measure ν on G is $\widehat{\nu}(\pi) = \int_G \pi(x)^* \, \mathrm{d}\nu(x)$, $\pi \in \widehat{G}$. Here $\pi(x)^*$ denotes the adjoint of $\pi(x)$, and the integrals are taken entrywise.

Let $g : [0, \infty) \to [0, \infty)$ be a nondecreasing and subadditive[1] function such that $\lim_{t \to 0^+} g(t) = 0$, and define

$$W_g(\nu_1, \nu_2) = \inf_{\vartheta \in \mathrm{Coup}(\nu_1, \nu_2)} \int_{G \times G} g(\rho(x,y)) \, \mathrm{d}\vartheta(x,y),$$

where $\mathrm{Coup}(\nu_1, \nu_2)$ is the set of couplings, as before. Letting

$$\mathcal{F}_g = \{ f : G \to \mathbb{R} \, : \, |f(x) - f(y)| \leq g(\rho(x,y)) \text{ for all } x, y \in G \},$$

the Kantorovich duality theorem states

$$W_g(\nu_1, \nu_2) = \sup_{f \in \mathcal{F}_g} \left| \int_G f \, \mathrm{d}\nu_1 - \int_G f \, \mathrm{d}\nu_2 \right|.$$

Note that W_g is a metric on the set of Borel probability measures on G and it generates the topology of weak convergence, unless g is constant zero. In the special case

[1] That is, $g(t+u) \leq g(t) + g(u)$ for all $t, u \geq 0$.

$g(t)=t^p$, $0<p\le 1$ we write W_p (resp. $\mathcal{F}_p$) instead of W_g (resp. $\mathcal{F}_g$). We mention that given $f\in C(G)$, the function

$$g_f(t)=\sup\{|f(x)-f(y)|\,:\,x,y\in G,\ \rho(x,y)\le t\}$$

is nondecreasing and subadditive, and $\lim_{t\to 0^+} g_f(t)=0$; in fact, g_f is the smallest g for which $f\in\mathcal{F}_g$.

Remark Kantorovich duality is usually stated for $g(t)=t$, i.e. for Lipschitz functions. To see the general case, note that $g(\rho(x,y))$ is another metric on G generating the topology of G, unless g is constant zero; the subadditivity of g is needed for the triangle inequality. Kantorovich duality for Lipschitz functions in the $g(\rho(x,y))$ metric thus implies Kantorovich duality for W_g as claimed. Further, since the usual 1-Wasserstein metric with respect to $g(\rho(x,y))$ generates the topology of weak convergence, so does W_g.

2.2 Berry–Esseen Inequality on Compact Lie Groups

Let $\eta:\mathfrak{t}\to\mathbb{R}$ be a $W(G,T)$-invariant smooth function such that $\eta(X)=0$ for all $|X|\ge 1$, and $0\le\eta(X)\le\eta(0)=1$ for all $X\in\mathfrak{t}$. Since $W(G,T)$ acts by orthogonal transformations on $\mathfrak{t}$, $W(G,T)$-invariance can be ensured e.g. if $\eta(X)$ depends only on $|X|$. For instance, we can use the "bump function"

$$\eta(X)=\begin{cases}\exp\left(-\frac{|X|^2}{1-|X|^2}\right) & \text{if } |X|<1,\\ 0 & \text{if } |X|\ge 1.\end{cases}$$

Let $F:\mathfrak{t}\to\mathbb{C}$, $F(X)=\int_{\mathfrak{t}}\eta(Y)e^{2\pi i(X,Y)}\,\mathrm{d}m(Y)$; note that $F(X)$ is a Schwarz function, thus $|F(X)|$ decays at an arbitrary polynomial rate as $|X|\to\infty$. The main result of the paper is the following Berry–Esseen type inequality.

Theorem 4 *Let ν_1 and ν_2 be Borel probability measures on a compact, connected Lie group G, and let $g:[0,\infty)\to[0,\infty)$ be nondecreasing and subadditive such that $\lim_{t\to 0^+}g(t)=0$. Let*

$$\psi(t)=\frac{2}{|W(G,T)|}\int_{\mathfrak{t}} g\left(\frac{2\pi|X|}{\lfloor t/(|2\rho^+|+a)\rfloor}\right)a^r|F(aX)|\left(\prod_{\alpha\in R^+}|e^{2\pi i\alpha(X)}-1|^2\right)\mathrm{d}m(X)$$

where $a=\min_{\lambda\in\Gamma^*_{\mathrm{root}}}|\lambda|/2=\min_{\alpha\in R}|\alpha|/2$ and $\rho^+=\sum_{\alpha\in R^+}\alpha/2$, and let

$$\phi(t)=\inf_{0<c<2(\sqrt{n^2+n}-n)}\sqrt{\frac{n}{1-c-c^2/(4n)}}\cdot\frac{g\left(\frac{c}{nt}\right)}{\frac{c}{nt}}.$$

Then for any real number $M \geq |2\rho^+| + a$,

$$W_g(\nu_1, \nu_2) \leq \psi(M) + \phi(M) \left(\sum_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| < M}} \frac{d_\pi}{\kappa_\pi} \|\widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi)\|_{\mathrm{HS}}^2 \right)^{1/2}.$$

Observe that $\psi(M) \ll g(1/M)$ and $\phi(M) \ll Mg(1/M)$ with implied constants depending only on G. Theorem 1 thus follows from Theorem 4 with $g(t) = t^p$, $0 < p \leq 1$.

2.3 Applications

2.3.1 Spectral Gaps and Random Walks

Given Borel probability measures ν_1 and ν_2 on G, let $\nu_1 * \nu_2$ denote their convolution, and let $\nu_1^*(B) = \nu_1(B^{-1})$, $B \subseteq G$ Borel. If ζ_1 and ζ_2 are independent G-valued random variables with distribution ν_1 and ν_2, then $\nu_1 * \nu_2$ (resp. ν_1^*) is the distribution of $\zeta_1\zeta_2$ (resp. ζ_1^{-1}).

Let $L_0^2(G, \mu_G)$ be the orthogonal complement of the space of constant functions in $L^2(G, \mu_G)$; that is, the set of all $f \in L^2(G, \mu_G)$ with $\int_G f \, \mathrm{d}\mu_G = 0$. Given a Borel probability measure ν on G, let $T_\nu : L_0^2(G, \mu_G) \to L_0^2(G, \mu_G)$,

$$(T_\nu f)(x) = \int_G f(xy) \, \mathrm{d}\nu(y)$$

be its associated Markov operator. Observe that $T_{\nu_1 * \nu_2} = T_{\nu_1} T_{\nu_2}$ and $T_{\nu^*} = T_\nu^*$; in particular, T_ν is self-adjoint (resp. normal) if and only if $\nu = \nu^*$ (resp. $\nu * \nu^* = \nu^* * \nu$).

We start with a trivial estimate for $W_p(\nu, \mu_G)$ in terms of T_ν. It is not difficult to see that

$$q_\nu := \|T_\nu\|_{\mathrm{op}} = \sup_{\substack{\pi \in \widehat{G} \\ \pi \neq \pi_0}} \|\widehat{\nu}(\pi)\|_{\mathrm{op}},$$

where $\pi_0 \in \widehat{G}$ denotes the trivial representation, and $\| \cdot \|_{\mathrm{op}}$ is the operator norm. Let $f \in \mathcal{F}_p$ with $\int_G f \, \mathrm{d}\mu_G = 0$ be arbitrary, and note that $T_\nu f \in \mathcal{F}_p$. Since $|T_\nu f| \geq |(T_\nu f)(e)|/2$ on the ball centered at e with radius $r = (|(T_\nu f)(e)|/2)^{1/p}$, we have

$$\|T_\nu f\|_2^2 \geq \left(\frac{(T_\nu f)(e)}{2} \right)^2 \mu_G(B(e, r)) \gg |(T_\nu f)(e)|^{2+n/p},$$
$$\|T_\nu f\|_2^2 \leq \|T_\nu\|_{\mathrm{op}}^2 \cdot \|f\|_2^2 \ll \|T_\nu\|_{\mathrm{op}}^2.$$

Therefore $|\int_G f \,\mathrm{d}\nu| = |(T_\nu f)(e)| \ll q_\nu^{2p/(n+2p)}$, and consequently

$$W_p(\nu, \mu_G) \ll q_\nu^{2p/(n+2p)}. \tag{4}$$

We now deduce a sharp improvement on the trivial estimate (4). Recall that $\|A\|_{\mathrm{HS}} \le \sqrt{d_\pi}\|A\|_{\mathrm{op}}$ for any $d_\pi \times d_\pi$ matrix A. With $\nu_1 = \nu$ and $\nu_2 = \mu_G$,

$$\begin{aligned} \sum_{\substack{\pi \in \widehat{G} \\ 0<|\lambda_\pi|<M}} \frac{d_\pi}{\kappa_\pi} \|\widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi)\|_{\mathrm{HS}}^2 &\le \sum_{\substack{\pi \in \widehat{G} \\ 0<|\lambda_\pi|<M}} \frac{d_\pi^2}{\kappa_\pi} \|\widehat{\nu}(\pi)\|_{\mathrm{op}}^2 \\ &\ll \sum_{\substack{\pi \in \widehat{G} \\ 0<|\lambda_\pi|<M}} |\lambda_\pi|^{n-r-2} q_\nu^2 \\ &\ll \begin{cases} q_\nu^2 & \text{if } n=1, \\ (\log(M+2)) q_\nu^2 & \text{if } n=2, \\ M^{n-2} q_\nu^2 & \text{if } n \ge 3. \end{cases} \end{aligned} \tag{5}$$

Optimizing the value of the free parameter $M > 0$, in dimension $n \ge 3$ Theorem 1 thus gives that for any $0 < p \le 1$,

$$W_p(\nu, \mu_G) \ll q_\nu^{2p/n} \tag{6}$$

with an implied constant depending only on G. Using Theorem 4 instead, we get $W_g(\nu, \mu_G) \ll g(q_\nu^{2/n})$. Similar estimates can be deduced in dimensions $n = 1$ and 2. Clearly $q_\nu \le 1$, and $q_{\nu_1 * \nu_2} \le q_{\nu_1} q_{\nu_2}$; in particular, (6) gives the upper bound for the rate of convergence of random walks $W_p(\nu_1 * \cdots * \nu_N, \mu_G) \ll \prod_{k=1}^N q_{\nu_k}^{2p/n}$.

We say that ν has a *spectral gap*, if the spectral radius of T_ν is strictly less than 1; note that this is a direct generalization of Cramér's condition in classical probability theory. Assuming T_ν is normal, having a spectral gap is equivalent to $q_\nu < 1$; for general T_ν, it is equivalent to $q_{\nu^{*m}} < 1$ for some integer $m \ge 1$. Deciding whether a given ν has a spectral gap is a highly nontrivial problem. Generalizing results of Bourgain and Gamburd [10,11] on SU(2) and SU(d), Benoist and Saxcé considered a Borel probability measure ν on a compact, connected, simple Lie group G. They proved [3, Theorem 3.1] that if the support of ν is not contained in any proper closed subgroup, and each element of the support (as a matrix) has algebraic entries, then ν has a spectral gap. The same authors also conjectured that the condition that the matrix entries are algebraic can be dropped.

Using (6) (or even just (4)), $W_p(\nu^{*k}, \mu_G) \to 0$ exponentially fast as $k \to \infty$ whenever ν has a spectral gap. Corollary 2 is thus basically an unconditional (i.e. not assuming the conjecture of Benoist and Saxcé), weaker form of this fact. In contrast to the (semi)simple case, $W_p(\nu^{*k}, \mu_G) \to 0$ polynomially fast for certain finitely supported measures ν on the torus $\mathbb{R}^d/\mathbb{Z}^d$ [8].

So far we have only discussed the relationship between $W_p(\nu, \mu_G)$ and the spectral gap of ν. Theorem 1, however, provides a quantitative relationship between $W_p(\nu, \mu_G)$ and the spectrum of the self-adjoint operator $T_\nu^* T_\nu$ itself. Indeed, by the Peter–Weyl theorem $L_0^2(G, \mu_G) = \oplus_{\pi \in \widehat{G}, \pi \neq \pi_0} V_\pi$, where V_π is the vector space spanned by the entries of $\pi(x)$. Since $(T_\nu \pi)(x) = \pi(x)\widehat{\nu}(\pi)^*$, the action of T_ν on V_π is determined by $\widehat{\nu}(\pi)$; in particular, $d_\pi \|\widehat{\nu}(\pi)\|_{\mathrm{HS}}^2$ is simply the sum of all spectrum points of $T_\nu^* T_\nu$ on V_π. The proof of Corollary 2 is based on this quantitative relationship.

Proof of Corollary 2 Varjú [34, Theorem 6] proved that for any Borel probability measure ϑ on G and any $M > 0$,

$$1 - \max_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| \leq M}} \|\widehat{\vartheta}(\pi)\|_{\mathrm{op}} \geq c_0 \left(1 - \max_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| \leq M_0}} \|\widehat{\vartheta}(\pi)\|_{\mathrm{op}} \right) \frac{1}{\log^A(M+2)}, \tag{7}$$

where the constants $c_0, M_0 > 0$ and $1 \leq A \leq 2$ depend only on the group G; in fact, the exact value of A was also given. Since $\nu^{*k} \to \mu_G$ weakly, we have $\widehat{\nu}(\pi)^k = \widehat{\nu^{*k}}(\pi) \to 0$ for all $\pi \neq \pi_0$, and hence the spectral radius of $\widehat{\nu}(\pi)$ is less than 1. It follows that for any $\pi \in \widehat{G}$ with $0 < |\lambda_\pi| \leq M_0$, we have $\|\widehat{\nu}(\pi)^m\|_{\mathrm{op}} < 1$ with some positive integer $m = m(G, \nu)$; in particular,

$$b = b(G, \nu) := c_0 \left(1 - \max_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| \leq M_0}} \|\widehat{\nu}(\pi)^m\|_{\mathrm{op}} \right) > 0.$$

Applying (7) to $\vartheta = \nu^{*m}$, we get that for any positive integer k and any $M > 0$,

$$\begin{aligned} \max_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| \leq M}} \|\widehat{\nu^{*k}}(\pi)\|_{\mathrm{op}} \leq \left(\max_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| \leq M}} \|\widehat{\nu}(\pi)^m\|_{\mathrm{op}} \right)^{\lfloor k/m \rfloor} &\leq \left(1 - \frac{b}{\log^A(M+2)} \right)^{(k-m)/m} \\ &\leq e^{-b(k-m)/(m \log^A(M+2))}. \end{aligned}$$

Hence

$$\begin{aligned} \sum_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| < M}} \frac{d_\pi}{\kappa_\pi} \|\widehat{\nu^{*k}}(\pi)\|_{\mathrm{HS}}^2 &\leq \sum_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| < M}} \frac{d_\pi^2}{\kappa_\pi} \|\widehat{\nu^{*k}}(\pi)\|_{\mathrm{op}}^2 \\ &\ll \sum_{\substack{\pi \in \widehat{G} \\ 0 < |\lambda_\pi| < M}} |\lambda_\pi|^{n-r-2} e^{-b(k-m)/(m \log^A(M+2))} \\ &\ll M^n e^{-b(k-m)/(m \log^A(M+2))}. \end{aligned}$$

The first factor is actually 1, $\log(M+2)$, M^{n-2} in the cases $n = 1, n = 2, n \geq 3$, but this will not play an important role. Theorem 1 thus gives that for any $M > 0$,

$$W_1(\nu^{*k}, \mu_G) \ll \frac{1}{M} + M^{n/2} e^{-b(k-m)/(2m \log^A(M+2))}.$$

Choosing $\log^{A+1} M = b(k-m)/(2mn)$, we deduce

$$W_1(\nu^{*k}, \mu_G) \ll k^{\frac{1}{A+1}} \exp\left(-\frac{n}{2}\left(\frac{b(k-m)}{2mn}\right)^{\frac{1}{A+1}}\right).$$

In particular, $W_1(\nu^{*k}, \mu_G) \ll \exp\left(-ck^{\frac{1}{A+1}}\right)$ with any $0 < c < \frac{n}{2} \cdot (\frac{b}{2mn})^{\frac{1}{A+1}}$. □

Remark Using Theorem 4 instead of Theorem 1, we deduce the general form of the conclusion of Corollary 2 as $W_g(\nu^{*k}, \mu_G) \ll g(e^{-ck^{1/3}})$.

2.3.2 Uniform Distribution Theory

Next, we consider applications in uniform distribution theory. It is not difficult to see e.g. directly from the definition of W_g, that for any given nonempty finite set $A \subset G$ and any g as in Sect. 2.1,

$$\inf_{\operatorname{supp} \nu \subseteq A} W_g(\nu, \mu_G) = \int_G g(\operatorname{dist}(A, x))\, \mathrm{d}\mu_G(x), \tag{8}$$

where the infimum is over all probability measures ν whose support is contained in A, and $\operatorname{dist}(A, \cdot)$ denotes distance from the set A. Indeed, the infimum is attained when for any $a \in A$, $\nu(\{a\})$ is the Haar measure of the Voronoi cell

$$\{x \in G \,:\, \operatorname{dist}(A, x) = \rho(a, x)\}.$$

In this case the optimal transport plan from ν to μ_G is to simply spread $\nu(\{a\})$ evenly over the given Voronoi cell. Recall that open balls $B(x, r)$ in G of radius $0 < r < \operatorname{diam} G$ satisfy $r^n \ll \mu_G(B(x, r)) \ll r^n$. A standard ball packing argument (using e.g. the "$3r$ covering lemma" of Vitali) shows that the optimal distance from a probability measure supported on at most N points to the Haar measure is

$$g(N^{-1/n}) \ll \inf_{|\operatorname{supp} \nu| \le N} W_g(\nu, \mu_G) \ll g(N^{-1/n}) \tag{9}$$

with implied constants depending only on G. In particular, (8) and (9) hold for W_p, $0 < p \le 1$. We mention that W_p, $1 \le p < \infty$ also satisfies the same estimates as W_1. For a detailed proof in the case $1 \le p < \infty$ see Kloeckner [22]; the proof for $0 < p < 1$ and for more general g is identical. We refer to the same paper for far reaching generalizations (e.g. to more general measures on Riemannian manifolds).

Lubotzky, Phillips and Sarnak [23,24] considered the problem of finding well distributed finite sets in SO(3), and consequently, on the sphere S^2. For any N such that $2N-1$ is a prime congruent to 1 modulo 4, they constructed a symmetric set $\{a_1, a_2, \ldots, a_{2N}\} \subset \mathrm{SO}(3)$ for which the probability measure $\nu_N = (2N)^{-1} \sum_{k=1}^{2N} \delta_{a_k}$ satisfies $q_{\nu_N} = \sqrt{2N-1}/N$; this spectral gap is in fact optimal among all symmetric sets of size $2N$. Since SO(3) has dimension $n = 3$, (6) yields that for any $0 < p \leq 1$,

$$W_p(\nu_N, \mu_{SO(3)}) \ll N^{-p/3}$$

with a universal implied constant; by (9), this is optimal. This proves Corollary 3. Note that the trivial estimate (4) only yields $W_p(\nu_N, \mu_{SO(3)}) \ll N^{-p/(3+2p)}$. More generally, we have $W_g(\nu_N, \mu_{SO(3)}) \ll g(N^{-1/3})$.

Clozel [16] proved a similar optimal (up to a constant factor) spectral gap estimate in terms of the size of a finite set in $\mathrm{U}(d)$. Less precise estimates on more general compact homogeneous spaces were obtained by Oh [26].

2.3.3 Empirical Measures

Finally, we address the sharpness of Theorems 1 and 4; we do so by deducing a simple estimate on the mean rate of convergence of empirical measures. Let ν be an arbitrary Borel probability measure on G, and let $\zeta_1, \zeta_2, \dots, \zeta_N$ be independent, identically distributed G-valued random variables with distribution ν. The probability measure $\overline{\nu}_N := N^{-1} \sum_{k=1}^N \delta_{\zeta_k}$ is called the corresponding empirical measure. Theorem 1 gives an estimate for $W_p(\overline{\nu}_N, \nu)$ — a random variable! — as follows. Let $E_\pi = \mathbb{E}\pi(\zeta_1) = \widehat{\nu}(\pi)^*$. With $\nu_1 = \overline{\nu}_N$ and $\nu_2 = \nu$ we then have

$$\widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi) = \frac{1}{N} \sum_{k=1}^N (\pi(\zeta_k) - E_\pi)^* ,$$

and by independence, the “variance” satisfies

$$\mathbb{E}\, \| \widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi) \|_{\mathrm{HS}}^2 = \frac{1}{N^2} \sum_{k=1}^N \mathbb{E}\, \mathrm{tr} \left(\pi(\zeta_k)^* \pi(\zeta_k) - E_\pi^* E_\pi \right) \leq \frac{d_\pi}{N}.$$

In the last step we used that $\pi(x)$ is unitary. Following the steps in (5), in dimension $n \geq 3$ Theorem 1 gives that for any $0 < p \leq 1$ and any $M > 0$,

$$\mathbb{E} W_p(\overline{\nu}_N, \nu) \leq \sqrt{\mathbb{E} W_p(\overline{\nu}_N, \nu)^2} \ll \frac{1}{M^p} + M^{1-p} \sqrt{\frac{M^{n-2}}{N}}.$$

Optimizing the value of $M > 0$, we finally obtain that in dimension $n \geq 3$, for any $0 < p \leq 1$,

$$\mathbb{E} W_p(\overline{\nu}_N, \nu) \ll N^{-p/n} \tag{10}$$

with an implied constant depending only on G; more generally, $\mathbb{E} W_g(\overline{\nu}_N, \nu) \ll g(N^{-1/n})$. These are sharp by (9); in particular, Theorems 1 and 4 are also sharp up to a constant factor depending on G. Note that the only compact, connected Lie groups in dimension $n = 1$ and $n = 2$ are $\mathbb{R}/\mathbb{Z}$ and $\mathbb{R}^2/\mathbb{Z}^2$, and the sharpness of Theorem 1 on these groups follows from results in [8].

The rate of convergence of empirical measures in W_p, $p \geq 1$ on more general metric spaces was studied by Bach and Weed [2], and by Boissard and Le Gouic [7]. Instead of Fourier methods, they used a sequence of partitions of the metric space, each refining its predecessor to construct transport plans. It follows e.g. from [7, Corollary 1.2] that our estimate (10) can be improved to $\mathbb{E} W_p(\overline{\nu}_N, \nu) \ll N^{-1/n}$ for all $1 \leq p < n/2$. We refer to [2] for improvements for measures ν supported on sets of lower dimension than the ambient space.

3 Proof of Theorem 4

The proof of Berry–Esseen type inequalities are usually based on smoothing with an approximate identity whose Fourier transform has bounded support. For instance, in the proof of Theorem A this Fourier transform is the "rooftop function" $\max\{1 - |t|/T, 0\}$, supported on $[-T, T]$. The proof of Theorem B uses the discrete version $\prod_{k=1}^d \max\{1 - |m_k|/(M+1), 0\}$, supported on $[-M, M]^d$; in the setting of the torus this is known as the Fejér kernel.

Our proof of Theorem 4 follows the same idea. We will choose a kernel $K_M : G \to \mathbb{C}$ whose Fourier transform satisfies $\widehat{K_M}(\pi) = 0$ whenever $|\lambda_\pi| \geq M$. Clearly,

$$\left| \int_G f \, \mathrm{d}\nu_1 - \int_G f \, \mathrm{d}\nu_2 \right| \leq 2\| f - f * K_M \|_\infty + \left| \int_G f * K_M \, \mathrm{d}\nu_1 - \int_G f * K_M \, \mathrm{d}\nu_2 \right|, \tag{11}$$

where $f * K_M$ denotes convolution. Our goal is to find an upper estimate of the right hand side which is uniform in $f \in \mathcal{F}_g$; by Kantorovich duality, the same upper estimate will hold for $W_g(\nu_1, \nu_2)$. A possible choice for K_M could be a Fejér-like kernel

$$\frac{1}{|B_M|} \Bigg| \sum_{\substack{\pi \in \widehat{G} \\ \lambda_\pi \in B_M}} \chi_\pi \Bigg|^2$$

with some set $B_M \subseteq \{\lambda \in \Gamma^* \cap C^+ \,:\, |\lambda| < M/2\}$. For convergence properties of such Fejér kernels on compact Lie groups we refer to [12] and [33]. We mention that using these kernels it is possible to deduce upper bounds to W_p, $0 < p < 1$ sharp up to a constant factor depending on p, but in the case $p = 1$ we necessarily lose a logarithmic factor in M; the reason is that the Fejér kernel does not approximate Lipschitz functions optimally in the supremum norm. Fixing this shortcoming is easy on the torus; we simply need to use the normalized square of the Fejér kernel instead. By Jackson's theorem we then have the optimal rate of approximation $\| f - f * K_M \|_\infty \ll g(1/M)$, and a sharp Berry–Esseen smoothing inequality on the torus follows [8]. Similar modifications of the Fejér kernel are known to yield Jackson type theorems on certain classical groups, see Gong [20]; however, this approach seems not to have been worked out in full generality. An elegant proof of Jackson's theorem on an arbitrary compact, connected Lie group was nevertheless found by Cartwright and Kucharski [15], and in this paper we will use their kernel.

For the sake of completeness, we include the construction of the kernel in Sect. 3.1; we carry out the smoothing procedure in Sect. 3.2; finally, prove a decay estimate for the Fourier transform of f and finish the proof of Theorem 4 in Sect. 3.3.

3.1 Construction of the Kernel

Recall that the Weyl integral formula [9, p. 338] states that for any central function $\varphi \in L^1(G, \mu_G)$ we have

$$\int_G \varphi \,\mathrm{d}\mu_G = \frac{1}{|W(G,T)|} \int_T \varphi \cdot \delta_G \,\mathrm{d}\mu_T,$$

where the function $\delta_G : T \to \mathbb{R}$ is defined as

$$\delta_G(\exp(2\pi X)) = \prod_{\alpha \in R} (e^{2\pi i \alpha(X)} - 1) = \prod_{\alpha \in R^+} \left| e^{2\pi i \alpha(X)} - 1 \right|^2, \qquad X \in \mathfrak{t}.$$

In particular, $\delta_G \geq 0$ and $\int_T \delta_G \,\mathrm{d}\mu_T = |W(G,T)|$. Expanding the product in its definition, δ_G is thus a $W(G,T)$-invariant trigonometric polynomial on T of the form

$$\delta_G(\exp(2\pi X)) = \sum_{\lambda \in \Gamma^*_{\mathrm{root}}} c_\lambda e^{2\pi i \lambda(X)}, \qquad X \in \mathfrak{t}. \tag{12}$$

The constant term is $c_0 = |W(G,T)|$, and all coefficients satisfy $|c_\lambda| \leq |W(G,T)|$. Observe also that for any $\varphi \in L^1(T, \mu_T)$,

$$\int_T \varphi(t) \,\mathrm{d}\mu_T(t) = \int_{\mathfrak{t}/\Gamma} \varphi(\exp(2\pi X)) \,\mathrm{d}\mu_{\mathfrak{t}/\Gamma}(X), \tag{13}$$

where $\mu_{\mathfrak{t}/\Gamma}$ is the normalized Haar measure on $\mathfrak{t}/\Gamma$. Note that $\mu_{\mathfrak{t}/\Gamma} = m/\mathrm{Vol}(\mathfrak{t}/\Gamma)$, where $\mathrm{Vol}(\mathfrak{t}/\Gamma)$ is the Lebesgue measure of the fundamental domain of Γ.

Following [15] with minor modifications, we now construct a kernel K_M. Let $\eta(X)$ and $F(X)$ be as in Sect. 2.2, and recall the notation $a = \min_{\lambda \in \Gamma^*_{\mathrm{root}}} |\lambda|/2$. Let $M \geq |2\rho^+| + a$ be arbitrary, and set $M_0 = \lfloor M/(|2\rho^+| + a) \rfloor$. Define $P : T \to \mathbb{C}$ as

$$P(\exp(2\pi X)) = \mathrm{Vol}(\mathfrak{t}/\Gamma)(aM_0)^r \sum_{Y \in \Gamma} F(aM_0(X+Y)), \qquad X \in \mathfrak{t}.$$

Note that P is well-defined, smooth and $W(G,T)$-invariant. By (13), its Fourier coefficient with respect to $\lambda \in \Gamma^*$ (i.e. the character $\exp(2\pi X) \mapsto e^{2\pi i \lambda(X)}$ on T) is

$$\begin{aligned}\widehat{P}(\lambda) &= \int_{\mathfrak{t}/\Gamma} P(\exp(2\pi X)) e^{-2\pi i \lambda(X)} \,\mathrm{d}\mu_{\mathfrak{t}/\Gamma}(X)\\ &= \int_{\mathfrak{t}} (aM_0)^r F(aM_0 X) e^{-2\pi i \lambda(X)} \,\mathrm{d}m(X)\end{aligned}$$

$$= \eta(\lambda^*/(aM_0)),$$

where λ^* is the unique element in $\mathfrak{t}$ with $\lambda(X) = (\lambda^*, X)$. By the construction of η, $\widehat{P}(\lambda) = 0$ whenever $|\lambda| \geq aM_0$; consequently, P is a $W(G, T)$-invariant trigonometric polynomial on T with degree $< aM_0$. Observe also that

$$\frac{\delta_G(t^{M_0})}{\delta_G(t)} = \prod_{\alpha \in R} \frac{e^{2\pi i M_0 \alpha(X)} - 1}{e^{2\pi i \alpha(X)} - 1} \qquad (t = \exp(2\pi X))$$

is a $W(G, T)$-invariant trigonometric polynomial on T of degree $\leq |2\rho^+|(M_0 - 1)$. Hence $P(t)\delta_G(t^{M_0})/\delta_G(t)$ is a $W(G, T)$-invariant trigonometric polynomial on T of degree $< aM_0 + |2\rho^+|(M_0 - 1) < M$. It follows (see e.g. [15, Lemma 1]), that there exists a central trigonometric polynomial K_M on G of degree $< M$ — that is, a function $K_M : G \to \mathbb{C}$ of the form $K_M = \sum_{\pi \in \widehat{G}, |\lambda_\pi| < M} a_\pi \chi_\pi$ — such that

$$K_M(t) = P(t)\frac{\delta_G(t^{M_0})}{\delta_G(t)} \qquad \text{for all } t \in T.$$

3.2 The Smoothing Procedure

First, we estimate the coefficients in $K_M = \sum_{\pi \in \widehat{G}, |\lambda_\pi| < M} a_\pi \chi_\pi$. Let $W(\pi)$ denote the set of weights of a representation $\pi \in \widehat{G}$. Since there exists a unitary matrix U such that $U\pi(\exp(2\pi X))U^* = \operatorname{diag}(e^{2\pi i \mu(X)} : \mu \in W(\pi))$, $X \in \mathfrak{t}$, we have

$$\chi_\pi(\exp(2\pi X)) = \sum_{\mu \in W(\pi)} e^{2\pi i \mu(X)}, \qquad X \in \mathfrak{t}.$$

Therefore

$$\begin{aligned}
a_\pi &= \int_G K_M \overline{\chi_\pi}\, \mathrm{d}\mu_G \\
&= \frac{1}{|W(G,T)|} \int_{\mathfrak{t}/\Gamma} P(\exp(2\pi X))\delta_G(\exp(2\pi M_0 X))\overline{\chi_\pi(\exp(2\pi X))}\, \mathrm{d}\mu_{\mathfrak{t}/\Gamma}(X) \\
&= \int_{\mathfrak{t}} (aM_0)^r F(aM_0 X) \Bigg(\sum_{\lambda \in \Gamma^*_{\text{root}}} \frac{c_\lambda}{|W(G,T)|} e^{2\pi i M_0 \lambda(X)} \Bigg) \Bigg(\sum_{\mu \in W(\pi)} e^{-2\pi i \mu(X)} \Bigg) \mathrm{d}m(X) \\
&= \sum_{\mu \in W(\pi)} \sum_{\lambda \in \Gamma^*_{\text{root}}} \frac{c_\lambda}{|W(G,T)|} \eta(-\lambda^*/a + \mu^*/(aM_0)),
\end{aligned}$$

where $\lambda^* \in \mathfrak{t}$ is such that $\lambda(X) = (\lambda^*, X)$. By the construction of η, for any given $\mu \in W(\pi)$ there is at most one $\lambda \in \Gamma^*_{\text{root}}$ for which $\eta(-\lambda^*/a + \mu^*/(aM_0)) \neq 0$. Recalling that $|c_\lambda| \leq |W(G,T)|$ and $0 \leq \eta \leq 1$, it follows that the coefficients of K_M satisfy $|a_\pi| \leq d_\pi$. For the trivial character the only nonvanishing term is $\lambda = 0$; in particular, $\int_G K_M \,\mathrm{d}\mu_G = (c_0/|W(G,T)|)\eta(0) = 1$.

Remark In fact, the only nonvanishing term is $\lambda = 0$ whenever $|\lambda_\pi| \le aM_0$. Assuming in addition, that $\eta(X) = 1$ for all $|X| \le 1/2$, we thus have $a_\pi = d_\pi$ whenever $|\lambda_\pi| \le aM_0/2$. In other words, $f * K_M = f$ for any central trigonometric polynomial of the form $f = \sum_{\pi \in \widehat{G}, |\lambda_\pi| \le aM_0/2} b_\pi \chi_\pi$. Thus K_M is an analogue of the de la Vallée Poussin kernel, although its construction is not based on the Fejér kernel.

Proposition 5 *For any $f \in \mathcal{F}_g$ and any real $M \ge |2\rho^+| + a$,*

$$\left| \int_G f \, \mathrm{d}\nu_1 - \int_G f \, \mathrm{d}\nu_2 \right| \le \psi(M) + \sum_{\substack{\pi \in \widehat{G} \\ |\lambda_\pi| < M}} d_\pi \| \widehat{f}(\pi) \|_{\mathrm{HS}} \cdot \| \widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi) \|_{\mathrm{HS}},$$

where ψ is as in Theorem 4.

Proof Recall (11). We first show that $2\|f - f * K_M\|_\infty \le \psi(M)$; with somewhat weaker constant factors this was proved in [15]. Since the geodesic metric ρ is translation invariant both from the left and from the right, from the Weyl integral formula, (13) and (12) we deduce

$$\begin{aligned}
&\|f - f * K_M\|_\infty \\
&\quad = \sup_{x \in G} \left| \int_G \left(f(x) - f(xy^{-1}) \right) K_M(y) \, \mathrm{d}\mu_G(y) \right| \\
&\quad \le \int_G g(\rho(e, y)) |K_M(y)| \, \mathrm{d}\mu_G(y) \\
&\quad = \frac{1}{|W(G,T)|} \int_{\mathfrak{t}/\Gamma} g(\rho(e, \exp(2\pi X))) |P(\exp(2\pi X))| \delta_G(\exp(2\pi M_0 X)) \, \mathrm{d}\mu_{\mathfrak{t}/\Gamma}(X) \\
&\quad \le \frac{1}{|W(G,T)|} \int_{\mathfrak{t}} g(2\pi|X|)(aM_0)^r |F(aM_0 X)| \left(\prod_{\alpha \in R^+} |e^{2\pi i M_0 \alpha(X)} - 1|^2 \right) \mathrm{d}m(X) \\
&\quad = \psi(M)/2.
\end{aligned}$$

In the penultimate step we used that $g(\rho(e, \exp(2\pi X))) \le g(2\pi|X|)$, as the exponential map is a geodesic of unit speed. Finally, using $|a_\pi| \le d_\pi$ we get

$$\begin{aligned}
\left| \int_G f * K_M \, \mathrm{d}\nu_1 - \int_G f * K_M \, \mathrm{d}\nu_2 \right| &\le \sum_{\substack{\pi \in \widehat{G} \\ |\lambda_\pi| < M}} d_\pi \left| \int_G f * \chi_\pi \, \mathrm{d}\nu_1 - \int_G f * \chi_\pi \, \mathrm{d}\nu_2 \right| \\
&= \sum_{\substack{\pi \in \widehat{G} \\ |\lambda_\pi| < M}} d_\pi \left| \mathrm{tr} \left(\widehat{f}(\pi)^* \left(\widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi) \right) \right) \right| \\
&\le \sum_{\substack{\pi \in \widehat{G} \\ |\lambda_\pi| < M}} d_\pi \| \widehat{f}(\pi) \|_{\mathrm{HS}} \cdot \| \widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi) \|_{\mathrm{HS}},
\end{aligned}$$

which proves the claim. □

3.3 Decay of the Fourier Transform

We prove a decay estimate for the Fourier transform in somewhat greater generality than what we need, and then finish the proof of Theorem 4.

Proposition 6 *Assume that $f \in L^1(G, \mu_G)$ satisfies*

$$\left(\int_G |f(xh) - f(x)|^2 \, \mathrm{d}\mu_G(x) \right)^{1/2} \leq g(\rho(h, e))$$

for all $h \in G$ with some nondecreasing function $g : [0, \infty) \to [0, \infty)$. Then for any real number $M > 0$,

$$\sum_{\substack{\pi \in \widehat{G} \\ |\lambda_\pi| \leq M}} d_\pi \kappa_\pi \| \widehat{f}(\pi) \|_{\mathrm{HS}}^2 \leq \inf_{0 < c < 2(\sqrt{n^2+n} - n)} \frac{n}{1 - c - c^2/(4n)} \cdot \frac{g\left(\frac{c}{nM}\right)^2}{\left(\frac{c}{nM}\right)^2}.$$

If $g(t) = t^p$ with some $0 < p \leq 1$, we can choose e.g. $c = (\sqrt{17} - 3)/2$ (this is optimal in the worst case $p \to 0$, $n = 1$) yielding

$$\sum_{\pi \in \widehat{G}, |\lambda_\pi| \leq M} d_\pi \kappa_\pi \| \widehat{f}(\pi) \|_{\mathrm{HS}}^2 \leq 9 n^{3-2p} M^{2-2p}. \tag{14}$$

In the special case $p = 1$ the factor 9 can be removed, since the optimal choice is then to let $c \to 0$ (and $M \to \infty$). An estimate similar to (14) has recently been proved by Daher, Delgado and Ruzhansky [18], with an unspecified implied constant in the place of $9n^{3-2p}$. Our main improvement is that this implied constant does not depend on f; a crucial feature in the study of the p-Wasserstein metric.

Proof of Proposition 6 We follow ideas in [18]. For the sake of simplicity, we shall think about $\pi \in \widehat{G}$ as a $d_\pi \times d_\pi$ unitary matrix-valued function on G. For any matrix $A \in \mathbb{C}^{d_\pi \times d_\pi}$ let $\|A\|_{\mathrm{op}} = \sup\{|Av| : v \in \mathbb{C}^{d_\pi}, |v| = 1\}$ and $\|A\|_{\mathrm{HS}} = \sqrt{\mathrm{tr}\,(A^* A)}$ denote the operator norm and the Hilbert–Schmidt norm, respectively. The operator norm is submultiplicative; further, for all $A, B \in \mathbb{C}^{d_\pi \times d_\pi}$ we have $\|AB\|_{\mathrm{HS}} \leq \|A\|_{\mathrm{op}} \cdot \|B\|_{\mathrm{HS}}$, and the Cauchy–Schwarz inequality $|\mathrm{tr}\,(A^* B)| \leq \|A\|_{\mathrm{HS}} \cdot \|B\|_{\mathrm{HS}}$.

One readily verifies the identity

$$\left(\pi(h) - I_{d_\pi}\right) \widehat{f}(\pi) = \int_G (f(xh) - f(x)) \, \pi(x)^* \, \mathrm{d}\mu_G(x),$$

where I_{d_π} denotes the $d_\pi \times d_\pi$ identity matrix. By the Parseval formula and the assumption on f, for any $h \in G$ we have

$$\begin{aligned} \sum_{\pi \in \widehat{G}} d_\pi \mathrm{tr}\left(\left(\pi(h) - I_{d_\pi}\right)^* \left(\pi(h) - I_{d_\pi}\right) \widehat{f}(\pi) \widehat{f}(\pi)^* \right) &= \int_G |f(xh) - f(x)|^2 \, \mathrm{d}\mu_G(x) \\ &\leq g(\rho(h, e))^2. \end{aligned}$$

Since the exponential map is a geodesic, we have $\rho(\exp(uX), e) \leq |uX|$ for all $X \in \mathfrak{g}$ and $u \in \mathbb{R}$. For any $h = \exp(uX)$ the previous estimate thus yields

$$\sum_{\pi \in \widehat{G}} d_\pi \operatorname{tr}\left(\left(\pi(h) - I_{d_\pi}\right)^* \left(\pi(h) - I_{d_\pi}\right) \widehat{f}(\pi)\widehat{f}(\pi)^*\right) \leq g(|uX|)^2. \tag{15}$$

Next, we wish to find a lower estimate. For any $X \in \mathfrak{g}$ let

$$\mathrm{d}\pi(X) = \frac{\mathrm{d}}{\mathrm{d}u}\pi(\exp(uX)) \mid_{u=0} \in \mathbb{C}^{d_\pi \times d_\pi}$$

denote the derived representation of π.

Lemma 1 (Taylor expansion of degree 1) *For any $X \in \mathfrak{g}$ and any $u \in \mathbb{R}$,*

$$\left\|\pi(\exp(uX)) - I_{d_\pi} - u \cdot \mathrm{d}\pi(X)\right\|_{\mathrm{op}} \leq \frac{u^2}{2}\|\mathrm{d}\pi(X)\|_{\mathrm{op}}^2.$$

Proof of Lemma 1 We simply apply the usual Taylor formula to the matrix-valued function $F(u) = \pi(\exp(uX))$. Since π is a homomorphism, we have $F'(u) = \pi(\exp(uX))\mathrm{d}\pi(X)$. First, note that for any $u \in \mathbb{R}$,

$$\begin{aligned}
\left\|\pi(\exp(uX)) - I_{d_\pi}\right\|_{\mathrm{op}} &= \left\|\int_0^u \pi(\exp(yX))\mathrm{d}\pi(X)\,\mathrm{d}y\right\|_{\mathrm{op}} \\
&\leq \int_0^{|u|} \|\pi(\exp(yX))\|_{\mathrm{op}} \cdot \|\mathrm{d}\pi(X)\|_{\mathrm{op}}\,\mathrm{d}y \\
&= |u| \cdot \|\mathrm{d}\pi(X)\|_{\mathrm{op}}.
\end{aligned}$$

We used the fact that $\pi(\exp(yX))$ is a unitary matrix and thus has operator norm 1. Therefore

$$\begin{aligned}
\left\|\pi(\exp(uX)) - I_{d_\pi} - u \cdot \mathrm{d}\pi(X)\right\|_{\mathrm{op}} &= \left\|\int_0^u \left(\pi(\exp(yX)) - I_{d_\pi}\right)\mathrm{d}\pi(X)\,\mathrm{d}y\right\|_{\mathrm{op}} \\
&\leq \int_0^{|u|} \left\|\pi(\exp(yX)) - I_{d_\pi}\right\|_{\mathrm{op}} \cdot \|\mathrm{d}\pi(X)\|_{\mathrm{op}}\,\mathrm{d}y \\
&\leq \int_0^{|u|} |y| \cdot \|\mathrm{d}\pi(X)\|_{\mathrm{op}}^2\,\mathrm{d}y \\
&= \frac{u^2}{2}\|\mathrm{d}\pi(X)\|_{\mathrm{op}}^2.
\end{aligned}$$

□

Lemma 2 (Sugiura) *For any $X \in \mathfrak{g}$, we have $\|\mathrm{d}\pi(X)\|_{\mathrm{op}} \leq |\lambda_\pi| \cdot |X|$.*

Proof of Lemma 2 In [32, Theorem 2] Sugiura stated and proved the estimate $\|\mathrm{d}\pi(X)\|_{\mathrm{HS}} \leq \sqrt{d_\pi}|\lambda_\pi| \cdot |X|$. His proof is based on the fact that with some $d_\pi \times d_\pi$

unitary matrix U, we have $U\mathrm{d}\pi(X)U^* = \operatorname{diag}(i\lambda(X) : \lambda \in W(\pi))$, where $W(\pi)$ is the set of weights of π. Further, we have $|\lambda| \le |\lambda_\pi|$ for all $\lambda \in W(\pi)$. Hence Sugiura's proof in fact yields the slightly stronger claim of Lemma 2. □

Lemma 3 *Let $X_1, \ldots, X_n$ be an orthonormal base in $\mathfrak{g}$. For any $u \in \mathbb{R}$, the points $h_k = \exp(uX_k)$ satisfy*

$$\sum_{k=1}^{n} \left(\pi(h_k) - I_{d_\pi}\right)^* \left(\pi(h_k) - I_{d_\pi}\right) = u^2 \kappa_\pi I_{d_\pi} + E$$

with some $E \in \mathbb{C}^{d_\pi \times d_\pi}$, $\|E\|_{\mathrm{op}} \le n|u|^3|\lambda_\pi|^3 + n(u^4/4)|\lambda_\pi|^4$.

Proof of Lemma 3 By Lemma 1 we can write

$$\pi(h_k) - I_{d_\pi} = u \cdot \mathrm{d}\pi(X_k) + E_k$$

with some error matrix E_k satisfying $\|E_k\|_{\mathrm{op}} \le (u^2/2)\|\mathrm{d}\pi(X_k)\|_{\mathrm{op}}^2$. Therefore

$$\sum_{k=1}^{n} \left(\pi(h_k) - I_{d_\pi}\right)^* \left(\pi(h_k) - I_{d_\pi}\right) = u^2 \sum_{k=1}^{n} \mathrm{d}\pi(X_k)^* \mathrm{d}\pi(X_k) + E$$

where

$$\begin{aligned} \|E\|_{\mathrm{op}} &= \left\| \sum_{k=1}^{n} \left(u \cdot \mathrm{d}\pi(X_k)^* E_k + E_k^* u \cdot \mathrm{d}\pi(X_k) + E_k^* E_k \right) \right\|_{\mathrm{op}} \\ &\le \sum_{k=1}^{n} \left(2|u| \cdot \|\mathrm{d}\pi(X_k)\|_{\mathrm{op}} \cdot \frac{u^2}{2} \|\mathrm{d}\pi(X_k)\|_{\mathrm{op}}^2 + \frac{u^4}{4} \|\mathrm{d}\pi(X_k)\|_{\mathrm{op}}^4 \right). \end{aligned}$$

By Lemma 2, the previous estimate yields $\|E\|_{\mathrm{op}} \le n|u|^3|\lambda_\pi|^3 + n(u^4/4)|\lambda_\pi|^4$. On the other hand, we have $\mathrm{d}\pi(X)^* = -\mathrm{d}\pi(X)$, and by the definition of the Laplace–Beltrami operator,

$$\sum_{k=1}^{n} \mathrm{d}\pi(X_k)^* \mathrm{d}\pi(X_k) = -\sum_{k=1}^{n} \mathrm{d}\pi(X_k)\mathrm{d}\pi(X_k) = -(\Delta\pi)(e) = \kappa_\pi I_{d_\pi}.$$ □

We now finish the proof of Proposition 6. Recall that $|\lambda_\pi|^2 \le \kappa_\pi$. From Lemma 3 we deduce that for any $u \in \mathbb{R}$,

$$\begin{aligned} &\operatorname{tr}\left(\sum_{k=1}^{n} \left(\pi(h_k) - I_{d_\pi}\right)^* \left(\pi(h_k) - I_{d_\pi}\right) \widehat{f}(\pi)\widehat{f}(\pi)^* \right) \\ &\qquad = \operatorname{tr}\left(u^2 \kappa_\pi \widehat{f}(\pi)\widehat{f}(\pi)^* \right) + \operatorname{tr}\left(E \widehat{f}(\pi)\widehat{f}(\pi)^* \right) \\ &\qquad \ge u^2 \kappa_\pi \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2 - \|E\widehat{f}(\pi)\|_{\mathrm{HS}} \cdot \|\widehat{f}(\pi)\|_{\mathrm{HS}} \end{aligned} \tag{16}$$

$$
\begin{aligned}
&\geq u^2\kappa_\pi \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2 - \|E\|_{\mathrm{op}} \cdot \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2 \\
&\geq \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2 \left(u^2\kappa_\pi - n|u|^3|\lambda_\pi|^3 - n\frac{u^4}{4}|\lambda_\pi|^4 \right) \\
&\geq \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2 \cdot u^2\kappa_\pi \left(1 - n|u| \cdot |\lambda_\pi| - n\frac{u^2}{4}|\lambda_\pi|^2 \right).
\end{aligned}
$$

Let $M > 0$ and $0 < c < 2(\sqrt{n^2+n} - n)$ be arbitrary, and choose $u = c/(nM)$. For any $|\lambda_\pi| \leq M$ we then have

$$
1 - n|u| \cdot |\lambda_\pi| - n\frac{u^2}{4}|\lambda_\pi|^2 \geq 1 - c - \frac{c^2}{4n} > 0,
$$

and thus (15) and (16) imply

$$
\begin{aligned}
ng\left(\frac{c}{nM}\right)^2 &\geq \sum_{\substack{\pi\in\widehat{G} \\ |\lambda_\pi|\leq M}} d_\pi \operatorname{tr}\left(\sum_{k=1}^n \left(\pi(h_k) - I_{d_\pi}\right)^* \left(\pi(h_k) - I_{d_\pi}\right) \widehat{f}(\pi)\widehat{f}(\pi)^* \right) \\
&\geq \sum_{\substack{\pi\in\widehat{G} \\ |\lambda_\pi|\leq M}} d_\pi \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2 \left(\frac{c}{nM}\right)^2 \kappa_\pi \left(1 - c - \frac{c^2}{4n}\right).
\end{aligned}
$$

Since $0 < c < 2(\sqrt{n^2+n} - n)$ was arbitrary, the claim follows. □

Proof of Theorem 4 From Propositions 5 and 6 and the Cauchy–Schwarz inequality we get that for any $f \in \mathcal{F}_g$ and any real number $M \geq |2\rho^+| + a$,

$$
\begin{aligned}
&\left| \int_G f \,\mathrm{d}\nu_1 - \int_G f \,\mathrm{d}\nu_2 \right| \\
&\quad\leq \psi(M) + \left(\sum_{\substack{\pi\in\widehat{G} \\ 0<|\lambda_\pi|<M}} d_\pi\kappa_\pi \|\widehat{f}(\pi)\|_{\mathrm{HS}}^2 \right)^{1/2} \left(\sum_{\substack{\pi\in\widehat{G} \\ 0<|\lambda_\pi|<M}} \frac{d_\pi}{\kappa_\pi} \|\widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi)\|_{\mathrm{HS}}^2 \right)^{1/2} \\
&\quad\leq \psi(M) + \phi(M) \left(\sum_{\substack{\pi\in\widehat{G} \\ 0<|\lambda_\pi|<M}} \frac{d_\pi}{\kappa_\pi} \|\widehat{\nu_1}(\pi) - \widehat{\nu_2}(\pi)\|_{\mathrm{HS}}^2 \right)^{1/2}.
\end{aligned}
$$

By Kantorovich duality, the same upper bound holds for $W_g(\nu_1, \nu_2)$. □

Acknowledgements The author is supported by the Austrian Science Fund (FWF), project Y-901. I would like to thank Daniel El-Baz, Florian Pausinger and Szilárd Révész for helpful discussions. I am also grateful to the referees for valuable comments, and for bringing the papers [1,5,6,14,21] to my attention.

References

1. Ambrosio, L., Stra, F., Trevisan, D.: A PDE approach to a 2-dimensional matching problem. Probab. Theory Related Fields **173**, 433–477 (2019)
2. Bach, F., Weed, J.: Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance. Bernoulli **25**, 2620–2648 (2019)
3. Benoist, Y., de Saxcé, N.: A spectral gap theorem in simple Lie groups. Invent. Math. **205**, 337–361 (2016)
4. Bobkov, S.: Closeness of probability distributions in terms of Fourier–Stieltjes transforms. Russ. Math. Surveys **71**(6), 1021–1079 (2016)
5. Bobkov, S., Ledoux, M.: A simple Fourier analytic proof of the AKT optimal matching theorem. To appear in Ann. Appl. Probab
6. Bobkov, S., Ledoux, M.: Transport inequalities on Euclidean spaces for non-Euclidean metrics. J. Fourier Anal. Appl. **26**, 60 (2020)
7. Boissard, E., Le Gouic, T.: On the mean speed of convergence of empirical and occupation measures in Wasserstein distance. Ann. Inst. Henri Poincaré Probab. Stat. **50**, 539–563 (2014)
8. Borda, B.: Equidistribution of random walks on compact groups II. The Wasserstein metric. To appear in Bernoulli
9. Bourbaki, N.: Lie Groups and Lie Algebras. Chapters 7–9. Translated from the 1975 and 1982 French originals by Andrew Pressley. Elements of Mathematics (Berlin). Springer, Berlin (2005)
10. Bourgain, J., Gamburd, A.: On the spectral gap for finitely-generated subgroups of $\mathrm{SU}(2)$. Invent. Math. **171**, 83–121 (2008)
11. Bourgain, J., Gamburd, A.: A spectral gap theorem in $\mathrm{SU}(d)$. J. Eur. Math. Soc. **14**, 1455–1511 (2012)
12. Brandolini, L., Travaglini, G.: Pointwise convergence of Fejer type means. Tohoku Math. J. **49**, 323–336 (1997)
13. Brown, L., Steinerberger, S.: On the Wasserstein distance between classical sequences and the Lebesgue measure. Trans. Am. Math. Soc. **373**, 8943–8962 (2020)
14. Caracciolo, S., Lucibello, C., Parisi, G., Sicuro, G.: Scaling hypothesis for the Euclidean bipartite matching problem. Phys. Rev. E **90**, 012118 (2014)
15. Cartwright, D., Kucharski, K.: Jackson's theorem for compact connected Lie groups. J. Approx. Theory **55**, 352–359 (1988)
16. Clozel, L.: Automorphic forms and the distribution of points on odd-dimensional spheres. Isr. J. Math. **132**, 175–187 (2002)
17. Colzani, L., Gigante, G., Travaglini, G.: Trigonometric approximation and a general form of the Erdős Turán inequality. Trans. Am. Math. Soc. **363**, 1101–1123 (2011)
18. Daher, R., Delgado, J., Ruzhansky, M.: Titchmarsh theorems for Fourier transforms of Hölder–Lipschitz functions on compact homogeneous manifolds. Monatsh. Math. **189**(1), 23–49 (2019)
19. Esseen, C.: On mean central limit theorems. Kungl. Tekn. Högsk. Handl. Stockholm **121**, 31 (1958)
20. Gong, S.: Harmonic Analysis on Classical Groups. Revised edition of the original Chinese edition published by Science Press Beijing 1983 as the 12th volume in the Series in Pure and Applied Mathematics. Springer, Beijing (1991)
21. Graham, C.: Irregularity of distribution in Wasserstein distance. J. Fourier Anal. Appl. **26**, 75 (2020)
22. Kloeckner, B.: Approximation by finitely supported measures. ESAIM Control Optim. Calc. Var. **18**, 343–359 (2012)
23. Lubotzky, A., Phillips, R., Sarnak, P.: Hecke operators and distributing points on the sphere I. Commun. Pure Appl. Math. **39**, 149–186 (1986)
24. Lubotzky, A., Phillips, R., Sarnak, P.: Hecke operators and distributing points on S^2 II. Commun. Pure Appl. Math. **40**, 401–420 (1987)
25. Niederreiter, H., Philipp, W.: Berry–Esseen bounds and a theorem of Erdős and Turán on uniform distribution mod 1. Duke Math. J. **40**, 633–649 (1973)
26. Oh, H.: The Ruziewicz problem and distributing points on homogeneous spaces of a compact Lie group. Isr. J. Math. **149**, 301–316 (2005)
27. Pausinger, F.: Uniformly distributed sequences in the orthogonal group and on the Grassmannian manifold. Math. Comput. Simul. **160**, 13–22 (2019)
28. Petrov, V.: Limit Theorems of Probability Theory. Sequences of Independent Random Variables. Oxford Studies in Probability. Oxford University Press, New York (1995)

29. Steinerberger, S.: A Wasserstein inequality and minimal Green energy on compact manifolds. arXiv:1907.09023
30. Steinerberger, S.: Wasserstein distance, Fourier series and applications. Monatsh. Math. **194**, 305–338 (2021)
31. Stromberg, K.: Probabilities on a compact group. Trans. Am. Math. Soc. **94**, 295–309 (1960)
32. Sugiura, M.: Fourier series of smooth functions on compact Lie groups. Osaka Math. J. **8**, 33–47 (1971)
33. Travaglini, G.: Fejér kernels for Fourier series on T^n and on compact Lie groups. Math. Z. **216**, 265–281 (1994)
34. Varjú, P.: Random walks in compact groups. Doc. Math. **18**, 1137–1175 (2013)
35. Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003)
36. von Bahr, B.: Multi-dimensional integral limit theorems. Ark. Mat. **7**, 71–88 (1967)