Empirical measures and random walks on compact spaces in the quadratic Wasserstein metric

Bence Borda,

Graz University of Technology, Steyrergasse 30, 8010 Graz, Austria. E-mail: borda@math.tugraz.at

Abstract. Estimating the rate of convergence of the empirical measure of an i.i.d. sample to the reference measure is a classical problem in probability theory. Extending recent results of Ambrosio, Stra and Trevisan on 2-dimensional manifolds, in this paper we prove sharp asymptotic and nonasymptotic upper bounds for the mean rate in the quadratic Wasserstein metric $W_2$ on a $d$-dimensional compact Riemannian manifold. Under a smoothness assumption on the reference measure, our bounds match the classical rate in the optimal matching problem on the unit cube due to Ajtai, Komlós, Tusnády and Talagrand. The i.i.d. condition is relaxed to stationary samples with a mixing condition. As an example of a nonstationary sample, we also consider the empirical measure of a random walk on a compact Lie group. Surprisingly, on semisimple groups random walks attain almost optimal rates even without a spectral gap assumption. The proofs are based on Fourier analysis, and in particular on a Berry–Esseen smoothing inequality for $W_2$ on compact manifolds, a result of independent interest with a wide range of applications.

MSC2020 subject classifications: 60B05, 60B15, 60G10, 49Q22

Keywords: Riemannian manifold, Lie group, optimal transportation, Berry–Esseen inequality, heat kernel, occupation measure

1. Introduction

Given a sequence of random variables $X_1, X_2, \ldots$, each with distribution $\mu$, the empirical measure $\mu_N = N^{-1} \sum_{n=1}^{N} \delta_{X_n}$ converges weakly to $\mu$ under very general circumstances. Estimating the rate of convergence in the Wasserstein metric $W_p$ has received considerable attention. Classical results of Ajtai, Komlós, Tusnády [1] and Talagrand [36] on optimal matchings concern i.i.d. random variables uniformly distributed on the unit cube $[0,1]^d$, in which case

$$\mathbb{E} W_1(\mu_N, \mu) \ll \begin{cases} N^{-1/2} & \text{if } d = 1, \\ (\log N)^{1/2} N^{-1/2} & \text{if } d = 2, \\ \sqrt{d} N^{-1/d} & \text{if } d \geq 3 \end{cases}$$

(1)

with a universal implied constant, and this is sharp for all $d \geq 1$. Several deep results on $W_p(\mu_N, \mu)$ have since appeared for more general probability measures $\mu$ on Euclidean spaces, including moment estimates, concentration inequalities and central limit theorems, see [7] for a comprehensive account of the 1-dimensional case. The i.i.d. condition can also be relaxed: the empirical measure of stationary $m$-dependent or $\rho$-mixing sequences, and that of Markov chains have been shown to behave similarly to the i.i.d. case under suitable conditions [21, 22].

Weak convergence of empirical measures remains perfectly meaningful on more general metric spaces, with the Wasserstein metric providing a natural way to quantify the rate of convergence. Boissard and Le Gouic [10] showed that an i.i.d. sample with an arbitrary distribution $\mu$ taking values in a compact metric space satisfies the sharp estimate

$$\mathbb{E} W_p(\mu_N, \mu) \ll N^{-1/d} \quad \text{provided that } 1 \leq p < d/2,$$

(2)

and that the i.i.d. condition can be relaxed to suitable stationary $\rho$-mixing sequences and to Markov chains. Here $d > 0$ is the “dimension”, in the sense that for all small $R > 0$, the metric space can be covered by $\ll R^{-d}$ balls of radius $R$. See also [4, 9].

1Throughout, $a_N \ll b_N$ and $a_N = O(b_N)$ mean that there exists an implied constant $C > 0$ and $N_0 \in \mathbb{N}$ such that $|a_N| \leq C b_N$ for all $N \geq N_0$. Similar notation is used for functions.
In this paper, we study the convergence rate of empirical measures on a compact, connected, smooth Riemannian manifold (without boundary) \( M \) of dimension \( d \), with Riemannian volume \( \text{Vol} \). We shall only use the quadratic Wasserstein metric \( W_2 \), defined in terms of the geodesic distance \( \rho \) on \( M \). The case of a uniformly distributed (i.e., \( \mu = \text{Vol}/\text{Vol}(M) \)) i.i.d. sample in dimension 2 has recently been settled by Ambrosio, Stra and Trevisan [2], who showed the remarkable asymptotic relation

\[
\sqrt{\mathbb{E} W_2^2(\mu_N, \mu)} \sim \sqrt{\frac{\text{Vol}(M)}{4\pi}} \cdot \sqrt{\frac{\log N}{N}} \quad \text{as } N \to \infty.
\]

One of the goals of the present paper is to extend this result to higher dimensions, and to relax the i.i.d. condition. For the latter purpose, we shall use the pairwise mixing coefficients \( \alpha(X_m, X_n) \) and \( \beta(X_m, X_n) \) (see Section 2 for definitions).

To state our results, let \( X_1, X_2, \ldots \) be a sequence of \( M \)-valued random variables, each with distribution \( \mu \) (a Borel probability measure on \( M \)), and let \( \mu_N = \sum_{n=1}^N \delta_{X_n} \).

**Theorem 1.** If \( N^{-2} \sum_{1 \leq m < n \leq N} \alpha(X_m, X_n) \to 0 \), then \( \mathbb{E} W_2^2(\mu_N, \mu) \to 0 \) as \( N \to \infty \).

The following nonasymptotic result concerns a finite sample \( X_1, X_2, \ldots, X_N \) under a smoothness assumption on \( \mu \).

**Theorem 2.** Assume that \( \mu \geq c \text{Vol}/\text{Vol}(M) \) and that \( \sum_{1 \leq m < n \leq N} \beta(X_m, X_n) \leq BN \) with some constants \( c > 0 \) and \( B \geq 0 \).

(i) If \( d = 2 \), then

\[
\sqrt{\mathbb{E} W_2^2(\mu_N, \mu)} \leq \sqrt{\frac{\text{Vol}(M) + BC(M)}{\pi c}} \cdot \sqrt{\frac{\log N}{N}} + C(M) \sqrt{\frac{\text{Vol}(M)}{cN}}
\]

with some constant \( C(M) > 0 \) depending only on the manifold.

(ii) If \( d \geq 3 \), then

\[
\sqrt{\mathbb{E} W_2^2(\mu_N, \mu)} \leq \left( \frac{\text{Vol}(M) + BC(M)}{c} \right)^{1/d} \frac{\kappa \sqrt{d}}{N^{1/d}} + C(M) \left( 1 + \frac{B}{cN} \right)^{3/(2d)}
\]

with

\[
\kappa = \frac{1}{\sqrt{\pi}} \left( 1 + (1 - \phi)^{1/2} \right)^{1 - 2/d} \left( \frac{8}{d(d-2)} \right)^{1/d} \leq \frac{2}{\sqrt{\pi}}
\]

and some constant \( C(M) > 0 \) depending only on the manifold.

If there exists an orthonormal basis \( \{ \phi_k : k \geq 0 \} \) of \( L^2(M, \text{Vol}/\text{Vol}(M)) \) of eigenfunctions of the Laplace–Beltrami operator such that \( \sup_{k \geq 0} \sup_{x \in M} |\phi_k(x)| < \infty \), then the same holds with \( \beta(X_m, X_n) \) replaced by \( \alpha(X_m, X_n) \).

The constants \( C(M) \) depend only on simple geometric and spectral properties of the manifold \( M \). In fact, we shall derive Theorems 1 and 2 in a completely explicit form, see Theorem 6 in Section 4. The additional condition on the existence of an orthonormal basis of Laplace eigenfunctions with bounded sup-norms is satisfied by the flat torus \( \mathbb{R}^d/\mathbb{Z}^d \) and certain other flat manifolds [37], and is an open problem e.g. on the 2-dimensional unit sphere [38].

In addition to i.i.d. samples, Theorem 2 applies just as well to pairwise independent samples with \( B = 0 \). For a general survey and the history of various mixing coefficients we refer to Bradley [16]. The notion of the \( \alpha \)-mixing (also known as strong mixing) coefficient goes back to Rosenblatt, while the \( \beta \)-mixing coefficient (also known as coefficient of absolute regularity) was first introduced in [42, 43], where it was attributed to Kolmogorov. Although we do not use it in this paper, we also mention the perhaps better known \( \phi \)-mixing coefficient. These are related by the general inequalities

\[ 2\alpha(X,Y) \leq \beta(X,Y) \leq \phi(X,Y), \]

hence our results apply in particular to \( \phi \)-mixing samples. We refer to [16, 30] and references therein for quantitative bounds for the \( \beta \)-mixing coefficients of Markov chains on a general state space. Every strictly stationary, aperiodic, Harris recurrent Markov chain \( X_1, X_2, \ldots \) satisfies \( \beta(X_m, X_n) \to 0 \) as \( |m-n| \to \infty \). If the strictly stationary Markov chain is geometrically ergodic, then the convergence \( \beta(X_m, X_n) \to 0 \) is at least exponentially fast, whereas subgeometric ergodicity leads to slower convergence rates. (Sub)geometric ergodicity in turn is ensured by suitable drift conditions. Quantitative bounds for the \( \beta \)-mixing coefficients have also been established in the setting of SDEs [29, 40].

In the case of a uniformly distributed, pairwise independent sample in dimension 2, Theorem 2 (i) with \( c = 1 \) and \( B = 0 \) recovers the upper bound part of (3) up to a factor of 2. Theorem 2 (ii) seems to be new even in the i.i.d. case in dimensions 3 and 4, and is comparable to (2) in dimension \( d \geq 5 \). The only 1-dimensional compact, connected Riemannian manifold is the unit circle; we include it for the sake of completeness.
Theorem 3. Let $M = \mathbb{R}/\mathbb{Z}$, normalized so that $\text{Vol}(\mathbb{R}/\mathbb{Z}) = 1$. Assume that $\mu \geq c\text{Vol}$ and $\sum_{1 \leq m < n \leq N} \alpha(X_m, X_n) \leq BN$ with some constants $c > 0$ and $B \geq 0$. Then
\[
\sqrt{\mathbb{E}W_2^2(\mu_N, \mu)} \leq \sqrt{\frac{2 + 16B}{3cN}}.
\]

The smoothness assumption on $\mu$ cannot be removed from Theorems 2 and 3 in dimensions $d = 1, 2, 3$ even in the i.i.d. case, but (2) suggests that it might be superfluous in dimension $d \geq 5$ under a suitable mixing condition. Dimension $d = 4$ (when $p = d/2$) seems the most delicate case in this regard. The necessity of the smoothness assumption in low dimensions follows from an observation originally made for Euclidean spaces in [22], but which applies equally well on compact metric spaces, and which we now recall. Let $A, B \subseteq M$ be two Borel sets with $\text{dist}(A, B) > 0$. If $\mu$ be a Borel probability measure on $M$ such that $\mu(A \cup B) = 1$ and $\mu(A), \mu(B) > 0$. If the sample is i.i.d. with distribution $\mu$, then the indicators $\xi_n = I_{\{X_n \in A\}}$, $1 \leq n \leq N$ are i.i.d. Bernoulli variables with $\mathbb{E}\xi_n = \mu(A)$. It is not difficult to see that
\[
W_2^2(\mu_N, \mu) \geq \text{dist}(A, B)^2 \left| \frac{1}{N} \sum_{n=1}^{N} (\xi_n - \mathbb{E}\xi_n) \right|,
\]
since at least $|\mu_N(A) - \mu(A)| = |N^{-1} \sum_{n=1}^{N} (\xi_n - \mathbb{E}\xi_n)|$ amount of mass has to be transported either from $A$ to $B$, or from $B$ to $A$. Consequently $\sqrt{\mathbb{E}W_2^2(\mu_N, \mu)} \gg N^{-1/4}$, thus Theorems 2 and 3 do not apply to $\mu$ in dimensions $d = 1, 2, 3$. Note that the measure $\mu$ constructed above can be e.g. absolutely continuous with bounded density; the problem comes instead from the fact that its support is not connected.

It would be interesting to extend Theorem 2 to higher moments, concentration inequalities and almost sure asymptotics of $W_2(\mu_N, \mu)$ with weakly dependent samples. It seems likely that such results will hold only under stronger mixing assumptions.

Our approach is a simplified version of that in [2]. In a nutshell, the Benamou–Brenier formula relates the linearization of the quadratic Wasserstein metric to a suitable negative Sobolev norm $H_{-1}$, and the latter has a natural Fourier analytic interpretation. This leads to a Berry–Esseen type smoothing inequality for $W_2$ on compact Riemannian manifolds, a result we believe to be of independent interest, see Theorem 5 in Section 3. Estimating the convergence rate of empirical measures is, as we will see, one of many applications of such a smoothing inequality. For further interplay between Fourier analysis and the Wasserstein metric, see [11, 17, 18, 33, 34].

The same approach of using a smoothing inequality to estimate the rate of convergence of empirical measures has recently been used by Bobkov and Ledoux [6, 8] on the flat torus $\mathbb{R}^d/\mathbb{Z}^d$, who showed that the classical estimate (1) for $\mathbb{E}W_1(\mu_N, \mu)$ remains true for an identically distributed sample with arbitrary distribution $\mu$ under the mixing condition $\sum_{1 \leq m < n \leq N} \alpha(X_m, X_n) \ll N$. Our results show that $\sqrt{\mathbb{E}W_2^2(\mu_N, \mu)}$ satisfies the same classical estimate, under an additional smoothness assumption on $\mu$.

Our results can be extended to nonstationary processes, such as Markov chains on the state space $M$. In this paper we work out the details in the special case of random walks on compact groups. Let $G$ be a compact, connected Lie group of dimension $d$ equipped with an invariant Riemannian metric, normalized so that $\text{Vol}(G) = 1$. Note that the Riemannian volume $\text{Vol}$ is now the Haar measure. Let $Y_1, Y_2, \ldots$ be a sequence of i.i.d. $G$-valued random variables, each with distribution $\nu$, and let $S_n = Y_1 Y_2 \cdots Y_n$ be the corresponding random walk, whose distribution is thus the $n$-fold convolution power $\nu^n$. A classical theorem of Kawada and Itô [24, 35] states that $\nu^n \to \text{Vol}$ weakly as $n \to \infty$ if and only if the support of $\nu$ is contained neither in a proper closed subgroup, nor in a coset of a proper closed normal subgroup of $G$.

It is not surprising that under a spectral gap condition, $W_2(\nu^n, \text{Vol}) \to 0$ exponentially fast, and $\sqrt{\mathbb{E}W_2^2(\mu_N, \text{Vol})}$ with the empirical measure $\mu_N = N^{-1} \sum_{n=1}^{N} \delta_{S_n}$ satisfies the classical bound (1). See Theorems 7 and 8 in Section 5.1 for an explicit and more general form of this fact. We mention that explicit constructions of discrete $\nu$ with a spectral gap are known on $\text{SU}(n)$ and other simple groups [5, 14, 15]. In fact, the so-called spectral gap conjecture, a deep unsolved problem in the theory of compact groups, predicts that on a semisimple, compact, connected Lie group $G$ (such as $\text{SU}(n)$, $n \geq 2$ or $\text{SO}(n)$, $n \geq 3$), a probability measure $\nu$ has a spectral gap whenever $\nu^n \to \text{Vol}$ weakly. Recall that every such group is of dimension $d \geq 3$. The spectral gap conjecture would thus imply the remarkable fact that the purely qualitative (and trivially necessary) Kawada–Itô condition automatically improves to optimal rates for the random walk $S_n$. Using the best known partial result on the spectral gap conjecture due to Varjú [39], we show that this remarkable self-improving property holds unconditionally in a somewhat weaker form.
Theorem 4. Let \( G \) be a semisimple, compact, connected Lie group of dimension \( d \), and let \( \nu \) be a Borel probability measure on \( G \). If \( \nu^n \to \text{Vol} \) weakly as \( n \to \infty \), then

\[
W_2(\nu^n, \text{Vol}) \leq e^{-an^{1/3}} \quad \text{and} \quad \sqrt{\mathbb{E}W_2^2(\mu_n, \text{Vol})} \leq \frac{(\log N)^{2/d}}{N^{1/d}}
\]

with some constant \( a > 0 \) and implied constants depending only on \( \nu \) and \( G \).

The bound on the empirical rate is new, whereas the almost exponential bound on the rate of weak convergence \( \nu^n \to \text{Vol} \) improves our recent result in [11] from \( W_1 \) to \( W_2 \). The condition of semisimplicity cannot be removed. Indeed, in a recent paper [12] we constructed certain discrete random walks on the flat torus \( \mathbb{R}^d/\mathbb{Z}^d \), and found the precise rate of convergence of \( W_p(\nu^n, \text{Vol}) \), \( 0 < p \leq 1 \) to be polynomial instead of (almost) exponential. We mention that using the methods of this paper, these results can also be improved to \( 0 < p \leq 2 \), and in particular [12, Theorem 6] remains true verbatim with \( W_1 \) replaced by \( W_2 \). We refer to [12] also for the closely related problem of additive functionals of the process \( S_n \), such as the functional central limit theorem and the functional law of the iterated logarithm.

Finally, we comment on the problem of quantization of measure, where the goal is to approximate a given probability measure by a finitely supported one. For the sake of simplicity, we only consider the Haar measure on a compact, connected Lie Group \( G \). It follows from a standard ball packing argument that for any \( 1 \leq p < \infty \),

\[
\frac{1}{N^{1/d}} \leq \inf_{|\text{supp}(\nu)| \leq N} W_p(\nu, \text{Vol}) \leq \frac{1}{N^{1/d}},
\]

where the infimum is over all Borel probability measures \( \nu \) on \( G \) supported on at most \( N \) points (with arbitrary weights). See [23, 26] for more precise and general results. The main message of the theory of empirical measures is that random point sets, such as weakly dependent random variables or random walks, attain this optimum in dimension \( d \geq 3 \), but not in dimensions \( d = 1, 2 \).

It is also an important problem to explicitly construct (deterministic) point sets attaining the optimum in (4). As an application of our smoothing inequality, in Section 5.1 we deduce a sharp upper bound for \( W_2(\nu, \text{Vol}) \) in terms of the spectral radius of the Markov operator associated to \( \nu \). In particular, we will show that if \( \nu \) is supported on at most \( N \) points and its Markov operator has spectral radius \( \ll N^{-1/2} \), then \( \nu \) attains the optimal distance \( W_2(\nu, \text{Vol}) \ll N^{-1/d} \) from the uniform distribution. Finitely supported measures with equal weights having optimally small spectral radius were first constructed by Lubotzky, Phillips and Sarnak [27, 28] on \( SU(2) \) and \( SO(3) \); by our results, these consequently also attain the optimum in (4) for \( p = 2 \). See [20, 31] for similar constructions of finite point sets on more general spaces.

2. Notation

Throughout the paper, \( M \) is a compact, connected, smooth Riemannian manifold without boundary. Let \( d \) denote its dimension, \( \text{Vol} \) the Riemannian volume, \( \rho \) the geodesic distance, and \( \Delta \) the Laplace–Beltrami operator. Let \( 0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \) be the spectrum of \( M \), each value repeated according to its multiplicity, and let \( \phi_k, k \geq 0 \) be the corresponding (real-valued) smooth eigenfunctions; that is, \( \Delta \phi_k = -\lambda_k \phi_k \). We normalize the eigenfunctions so that \( \{ \phi_k : k \geq 0 \} \) is an orthonormal basis in \( L^2(M, \text{Vol}/\text{Vol}(M)) \). Let

\[
P(t, x, y) = \sum_{k=0}^{\infty} e^{-\lambda_k t} \phi_k(x)\phi_k(y), \quad t > 0, \ x, y \in M
\]

denote the heat kernel. The series is uniformly convergent on \( [t_0, \infty) \times M \times M \) for any \( t_0 > 0 \), and is normalized as

\[
\frac{1}{\text{Vol}(M)} \int_M P(t, x, y) \text{dVol}(x) = \frac{1}{\text{Vol}(M)} \int_M P(t, x, y) \text{dVol}(y) = 1.
\]

Further, \( P(t, x, y) > 0 \), it is \( C^1 \) in the variable \( t > 0 \), smooth in \( (x, y) \in M \times M \), and satisfies the heat equation \( \frac{\partial}{\partial t} P(t, x, y) = \Delta_x P(t, x, y) = \Delta_y P(t, x, y) \); here \( \Delta_x \) resp. \( \Delta_y \) denote the Laplacian with respect to the variable \( x \) resp. \( y \).

Let \( \mathcal{P}(M) \) denote the set of Borel probability measures on \( M \). For any \( f \in L^1(M, \text{Vol}) \) and \( \mu \in \mathcal{P}(M) \), let \( \hat{f}(k) = \text{Vol}(M)^{-1} \int_M f \phi_k \text{dVol} \) and \( \hat{\mu}(k) = \int_M \phi_k \text{d}\mu \) denote the Fourier coefficients.
The Wasserstein metric \( W_p, 0 < p < \infty \) is defined as

\[
W_p(\mu, \nu) = \inf_{\pi \in \text{Coup}(\mu, \nu)} \left( \int_{M \times M} \rho(x, y)^p \, d\pi(x, y) \right)^{\min\{1/p, 1\}}, \quad \mu, \nu \in \mathcal{P}(M),
\]

where \( \text{Coup}(\mu, \nu) \) is the set of couplings, i.e. the set of all \( \pi \in \mathcal{P}(M \times M) \) whose marginals are \( \pi(B \times M) = \mu(B) \) and \( \pi(M \times B) = \nu(B) \), \( B \subseteq M \) Borel. For any \( 0 < p < \infty \), \( W_p \) is a metric on \( \mathcal{P}(M) \) and it metrizes weak convergence.

The \( \alpha \)-mixing coefficient of two \( M \)-valued random variables \( X \) and \( Y \) is defined as

\[
\alpha(X, Y) = \sup_{A, B} |\Pr(X \in A, Y \in B) - \Pr(X \in A) \Pr(Y \in B)|
\]

where the supremum is over all Borel sets \( A, B \subseteq M \) resp. all Borel measurable functions \( f, g : M \to [0, 1] \). The \( \beta \)-mixing coefficient is defined as

\[
\beta(X, Y) = \frac{1}{2} \sup_{I, J \in \mathbb{N}} \sup_{B_1, \ldots, B_J \subseteq M} \sum_{i=1}^I \sum_{j=1}^J |\Pr(X \in A_i, Y \in B_j) - \Pr(X \in A_i) \Pr(Y \in B_j)|,
\]

where the supremum is over all partitions \( A_1, \ldots, A_I \) and \( B_1, \ldots, B_J \) of \( M \) into Borel sets. Letting \( \vartheta \in \mathcal{P}(M \times M) \), \( \mu, \nu \in \mathcal{P}(M) \) denote the distributions of \( (X, Y), X, Y \), respectively, we can also express it as the total variation distance

\[
\beta(X, Y) = \frac{1}{2} \|\vartheta - \mu \otimes \nu\|_{TV} = \sup_{C} |\vartheta(C) - (\mu \otimes \nu)(C)| = \sup_{0 \leq F \leq 1} \left| \int_{M \times M} F \, d(\vartheta - \mu \otimes \nu) \right|,
\]

the supremum being over all Borel sets \( C \subseteq M \times M \) resp. all Borel measurable functions \( F : M \times M \to [0, 1] \). In comparison, we mention that the \( \phi \)-mixing coefficient is defined as

\[
\phi(X, Y) = \sup_{A, B, \Pr(X \in A) > 0} |\Pr(Y \in B | X \in A) - \Pr(Y \in B)|.
\]


3. A smoothing inequality for \( W_2 \)

The main tool in the present paper is the following Berry–Esseen type inequality, whose proof is based on a smoothing procedure using the heat kernel. It generalizes and makes explicit a result in [17], where the case \( \mu = \text{Vol}/\text{Vol}(M) \) and \( \nu \) a finitely supported probability measure with equal weights was considered. We have recently proved a similar Berry–Esseen inequality for \( W_1 \) on compact Lie groups without any smoothness assumption on the measures [11].

**Theorem 5.** Let \( \mu, \nu \in \mathcal{P}(M) \), and assume that \( \mu \geq c \text{Vol}/\text{Vol}(M) \) with some constant \( c \geq 0 \). For any real \( t > 0 \),

\[
W_2(\mu, \nu) \leq c_1(\mu) \left( d + K(M) t^{3/2} \right)^{1/2} + \frac{2}{c_2(\mu, t)} \left( \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} [\mu(k) - \bar{\nu}(k)]^2 \right)^{1/2}
\]

with

\[
c_1(\mu) = 1 + (1 - c)^{1/2}, \quad c_2(\mu, t) = \left( \inf_{P \in M} \int_M P(t/2, x, y) \, d\mu(x) \right)^{1/2} \geq c^{1/2}
\]

and some constant \( K(M) \) depending only on the manifold. If the Ricci curvature of \( M \) is \( \geq -A \) with some constant \( A \geq 0 \), then

\[
K(M) = \frac{2(d-1)\sqrt{A}}{3} \left( d + (d-1)\sqrt{A} \text{diam } M \right)^{1/2}
\]

is a suitable choice.
In particular, $K(M) = 0$ whenever the Ricci curvature is positive semidefinite, which is the case e.g. for the $d$-dimensional unit sphere and for any compact, connected Lie group equipped with an invariant Riemannian metric. Note that if $c = 0$ (i.e. without any smoothness assumption on $\mu$), we still have $c_2(\mu, t) \geq (\inf_{x,y \in M} P(t/2, x, y))^{1/2} > 0$, but we might not have a positive lower bound independent of $t$.

3.1. Dispersion rate

We now prove an estimate on the dispersion rate, and a simple fact about Sobolev norms.

**Lemma 1.** For any real $t > 0$ and any $x \in M$,

$$\frac{1}{\text{Vol}(M)} \int_M P(t, x, y) \rho(x, y)^2 \, d\text{Vol}(y) \leq 2dt + K(M)(2t)^{3/2}$$

with some constant $K(M)$ depending only on the manifold. If the Ricci curvature of $M$ is $\geq -A$ with some constant $A \geq 0$, then (6) is a suitable choice.

**Proof.** Fix $x \in M$, and consider the function

$$F(t) := \frac{1}{\text{Vol}(M)} \int_M P(t, x, y) \rho(x, y)^2 \, d\text{Vol}(y), \quad t > 0.$$ 

Clearly, $\lim_{t \to 0^+} F(t) = \rho(x, x)^2 = 0$, and Green’s identity shows that

$$F'(t) = \frac{1}{\text{Vol}(M)} \int_M \frac{\partial}{\partial t} P(t, x, y) \rho(x, y)^2 \, d\text{Vol}(y) = \frac{1}{\text{Vol}(M)} \int_M (\Delta_{y} P(t, x, y)) \rho(x, y)^2 \, d\text{Vol}(y)$$

$$= \frac{1}{\text{Vol}(M)} \int_M P(t, x, y) (\Delta_{y} \rho(x, y)^2) \, d\text{Vol}(y).$$

By the Laplacian comparison theorem, here

$$\Delta_{y} \rho(x, y)^2 \leq 2d + 2(d - 1) \sqrt{A} \rho(x, y), \quad (7)$$

and we deduce

$$F'(t) \leq 2d + 2(d - 1) \sqrt{A} \int_M P(t, x, y) \rho(x, y) \, d\text{Vol}(y). \quad (8)$$

To be more precise, since $\rho(x, y)^2$ is not smooth at the cut locus of the point $x$, we used global Laplacian comparison, i.e. the fact that (7) remains true on all of $M$ in the sense of distributions [44].

The trivial bound $\rho(x, y) \leq \text{diam } M$ in (8) gives $F'(t) \leq 2d + 2(d - 1) \sqrt{A} \text{diam } M$, and by integrating we immediately deduce the preliminary estimate

$$F(t) \leq (2d + 2(d - 1) \sqrt{A} \text{diam } M)t.$$ 

On the other hand, applying Cauchy–Schwarz in (8),

$$F'(t) \leq 2d + 2(d - 1) \sqrt{A} \sqrt{F(t)} \leq 2d + 2(d - 1) \sqrt{A} \left(2d + 2(d - 1) \sqrt{A} \text{diam } M \right)^{1/2} t^{1/2}.$$ 

By integrating, we deduce

$$F(t) \leq 2dt + \frac{4(d - 1) \sqrt{A}}{3} \left(2d + 2(d - 1) \sqrt{A} \text{diam } M \right)^{1/2} t^{3/2},$$

as claimed. \hfill \Box

**Lemma 2.** If $f \in L^2(M, \text{Vol})$ is differentiable in the sense of distributions, then

$$\sum_{k=1}^{\infty} \lambda_k |\hat{f}(k)|^2 \leq \frac{1}{\text{Vol}(M)} \int_M |\nabla f|^2 \, d\text{Vol}.$$
Proof. In the special case \( f \in C^2(M) \), the claim easily follows from Green’s identity and the Parseval formula:

\[
\frac{1}{\text{Vol}(M)} \int_M |\nabla f|^2 \text{dVol} = \frac{1}{\text{Vol}(M)} \int_M (-\Delta f) \text{dVol} = \sum_{k=1}^{\infty} \lambda_k |\widehat{f}(k)|^2.
\]

Recall that \( C^\infty(M) \) is dense in the Sobolev space of all \( f \in L^2(M, \text{Vol}) \) differentiable in the sense of distributions such that \( \nabla f \in L^2(M, \text{Vol}) \), with the Sobolev norm

\[
\left( \frac{1}{\text{Vol}(M)} \int_M |f|^2 \text{dVol} \right)^{1/2} + \left( \frac{1}{\text{Vol}(M)} \int_M |\nabla f|^2 \text{dVol} \right)^{1/2}.
\]

Now let \( f \in L^2(M, \text{Vol}) \) be differentiable in the sense of distributions; we may assume that \( \nabla f \in L^2(M, \text{Vol}) \). We thus have a sequence \( f_n \in C^\infty(M) \) converging to \( f \) in the Sobolev norm above. In particular,

\[
\sum_{k=1}^{\infty} \lambda_k |\widehat{f}_n(k)|^2 = \frac{1}{\text{Vol}(M)} \int_M |\nabla f_n|^2 \text{dVol} \to \frac{1}{\text{Vol}(M)} \int_M |\nabla f|^2 \text{dVol} \quad \text{as } n \to \infty.
\]

Since \( f_n \to f \) in \( L^2(M, \text{Vol}) \), we also have \( \widehat{f}_n(k) \to \widehat{f}(k) \) as \( n \to \infty \) for any fixed \( k \geq 1 \), and the claim follows. \( \square \)

3.2. Proof of Theorem 5

Given \( \mu \in \mathcal{P}(M) \), let \( \mathcal{H}_1(\mu) \) be the set of all functions \( f \in L^2(M, \text{Vol}) \) differentiable in the sense of distributions such that

\[
\|f\|_{\mathcal{H}_1(\mu)} := \left( \int_M |\nabla f|^2 \text{d}\mu \right)^{1/2} < \infty.
\]

For any signed Borel measure \( \vartheta \) on \( M \), let

\[
\|\vartheta\|_{\mathcal{H}_{-1}(\mu)} := \sup \left\{ \int_M f \text{d}\vartheta : f \in \mathcal{H}_1(\mu), \|f\|_{\mathcal{H}_1(\mu)} \leq 1 \right\}.
\]

The proof of Theorem 5 relies on a result of Peyre [32], who showed that for any \( \mu, \nu \in \mathcal{P}(M) \),

\[
W_2(\mu, \nu) \leq 2\|\mu - \nu\|_{\mathcal{H}_{-1}(\mu)}.
\]  

(9)

His argument is based on the Benamou–Brenier formula

\[
W_2(\mu, \nu) = \inf_{\gamma} \int_0^1 \|d\gamma(t)\|_{\mathcal{H}_{-1}(\gamma(t))},
\]

where the infimum is over suitable curves \( \gamma : [0, 1] \to \mathcal{P}(M) \) with \( \gamma(0) = \mu \) and \( \gamma(1) = \nu \); choosing \( \gamma(t) = (1-t)\mu + t\nu \) gives (9).

Proof of Theorem 5. Convoluting the measures \( \mu \) and \( \nu \) with the heat kernel leads to the smoothed measures

\[
\mu_t(B) := \frac{1}{\text{Vol}(M)} \int_M \int_M P(t, x, y) I_B(y) \text{d}\mu(x) \text{dVol}(y) \quad (B \subseteq M \text{ Borel}),
\]

\[
\nu_t(B) := \frac{1}{\text{Vol}(M)} \int_M \int_M P(t, x, y) I_B(y) \text{d}\nu(x) \text{dVol}(y) \quad (B \subseteq M \text{ Borel}).
\]

The triangle inequality for \( W_2 \) gives

\[
W_2(\mu, \nu) \leq W_2(\mu, \mu_t) + W_2(\mu_t, \nu_t) + W_2(\nu_t, \nu).
\]  

(10)

To estimate the last term in (10), consider \( \pi \in \mathcal{P}(M \times M) \),

\[
\pi(C) := \frac{1}{\text{Vol}(M)} \int_M \int_M P(t, x, y) I_C(x, y) \text{d}\nu(x) \text{dVol}(y) \quad (C \subseteq M \times M \text{ Borel}).
\]
Observe that the marginals of \( \pi \) are \( \pi(B \times M) = \nu(B) \) and \( \pi(M \times B) = \nu_t(B) \). We can bound the cost of the transport plan \( \pi \) using Lemma 1:

\[
W_2(\nu_t) \leq \left( \frac{1}{\text{Vol}(M)} \int_M \int_M P(t, x, y) \rho(x, y) \, d\nu(x) \, d\text{Vol}(y) \right)^{1/2}.
\]

Finally, let us apply Peyrè’s estimate (9) to the second term in (10):

\[
W_2(\mu_t, \nu_t) \leq 2 \sup \left\{ \left| \int_M f \, d(\mu_t - \nu_t) \right| : f \in \dot{H}_1(\mu_t), \|f\|_{\dot{H}_1(\mu_t)} \leq 1 \right\}.
\]

Fix a function \( f \in \dot{H}_1(\mu_t), \|f\|_{\dot{H}_1(\mu_t)} \leq 1 \), and observe that

\[
\left| \int_M f \, d(\mu_t - \nu_t) \right| = \frac{1}{\text{Vol}(M)} \int_M \int_M P(t, x, y) f(y) \, d(\mu - \nu)(x) \, d\text{Vol}(y)
\]

\[
= \left| \sum_{k=0}^{\infty} e^{-\lambda_k t} \frac{1}{\text{Vol}(M)} \int_M \int_M \phi_k(x) \phi_k(y) f(y) \, d(\mu - \nu)(x) \, d\text{Vol}(y) \right|
\]

\[
= \left| \sum_{k=1}^{\infty} e^{-\lambda_k t} \tilde{f}(k) (\tilde{\mu}(k) - \tilde{\nu}(k)) \right|
\]

\[
\leq \left( \sum_{k=1}^{\infty} \lambda_k |\tilde{f}(k)|^2 \right)^{1/2} \left( \sum_{k=1}^{\infty} \frac{e^{-2\lambda_k t}}{\lambda_k} |\tilde{\mu}(k) - \tilde{\nu}(k)|^2 \right)^{1/2}.
\]

It is easy to see that \( \lambda_k \geq c_2(\mu, 2t)^2 \text{Vol}(M) \) with \( c_2(\mu, t) > 0 \) as in (5). Lemma 2 thus shows that here

\[
\sum_{k=1}^{\infty} \lambda_k |\tilde{f}(k)|^2 \leq \frac{1}{\text{Vol}(M)} \int_M |\nabla f|^2 \, d\text{Vol} \leq \frac{1}{c_2(\mu, 2t)^2} \int_M |\nabla f|^2 \, d\mu_t \leq \frac{1}{c_2(\mu, 2t)^2},
\]

and we obtain

\[
W_2(\mu_t, \nu_t) \leq \frac{2}{c_2(\mu, 2t)} \left( \sum_{k=1}^{\infty} \frac{e^{-2\lambda_k t}}{\lambda_k} |\tilde{\mu}(k) - \tilde{\nu}(k)|^2 \right)^{1/2}.
\]

The estimates above for the terms in (10) give that for any real \( t > 0 \),

\[
W_2(\mu, \nu) \leq c_1(\mu) \left( 2dt + K(M)(2t)^{3/2} \right)^{1/2} + \frac{2}{c_2(\mu, 2t)} \left( \sum_{k=1}^{\infty} \frac{e^{-2\lambda_k t}}{\lambda_k} |\tilde{\mu}(k) - \tilde{\nu}(k)|^2 \right)^{1/2},
\]

as claimed.
4. Weakly dependent random variables

We now prove our main result on the empirical measure of weakly dependent random variables, and then derive Theorems 1, 2 and 3. Let $X_1, X_2, \ldots, X_N$ be identically distributed $M$-valued random variables, each with distribution $\mu \in \mathcal{P}(M)$, and let $\mu_N = N^{-1} \sum_{n=1}^N \delta_{X_n}$.

**Theorem 6.** Assume that $\mu \geq c \operatorname{Vol}/\operatorname{Vol}(M)$ with some constant $c \geq 0$. For any real $t > 0$,

$$E^2 W_2^2(\mu_N, \mu) \leq c_1(\mu) \left( dt + K(M)t^{3/2} \right)^{1/2} + \frac{2}{c_2(\mu, t)} \left( \frac{1}{N} E + \frac{2}{N^2} \sum_{1 \leq m < n \leq N} E_m,n \right)^{1/2},$$

where $c_1(\mu)$, $c_2(\mu, t)$ and $K(M)$ are as in (5) and (6), and

$$E = \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} \left( \int_M \phi_k^2 \, d\mu - \left( \int_M \phi_k \, d\mu \right)^2 \right),$$

$$E_m,n = \min \left\{ 2\beta(X_m, X_n) \sup_{x,y \in M} \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t \phi_k(x) \phi_k(y)}}{\lambda_k} \bigg| 4\alpha(X_m, X_n) \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t \phi_k(x) \phi_k(y)}}{\lambda_k} \bigg| \right\}.$$ 

**Proof.** Applying Theorem 5 to $W_2(\mu_N, \mu)$ and the triangle inequality for the $L^2$-norm leads to

$$E^2 W_2^2(\mu_N, \mu) \leq c_1(\mu) \left( dt + K(M)t^{3/2} \right)^{1/2} + \frac{2}{c_2(\mu, t)} \left( \frac{1}{N} E + \frac{2}{N^2} \sum_{1 \leq m < n \leq N} E_m,n \right)^{1/2}.$$ 

Here

$$\mathbb{E} \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} |\hat{\mu}_N(k) - \hat{\mu}(k)|^2 = \mathbb{E} \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} \left( \frac{1}{N} \sum_{n=1}^{N} (\phi_k(X_n) - \hat{\mu}(k)) \right)^2$$

$$= \frac{1}{N} \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} \left( \int_M \phi_k^2 \, d\mu - \left( \int_M \phi_k \, d\mu \right)^2 \right)$$

$$+ \frac{2}{N^2} \sum_{1 \leq m < n \leq N} \mathbb{E} \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} (\phi_k(X_m) - \hat{\mu}(k))(\phi_k(X_n) - \hat{\mu}(k)).$$

It remains to estimate the last line of the previous formula in two different ways: in terms of the $\beta$-mixing and the $\alpha$-mixing coefficients. By the interpretation of the $\beta$-mixing coefficient as a total variation distance on $M \times M$, we have

$$\left| \mathbb{E} \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} (\phi_k(X_m) - \hat{\mu}(k))(\phi_k(X_n) - \hat{\mu}(k)) \right| \leq 2\beta(X_m, X_n) \sup_{x,y \in M} \left| \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} \phi_k(x) \phi_k(y) \right|.$$ 

On the other hand, using the $\alpha$-mixing coefficients to estimate the covariance of $\phi_k(X_m)$ and $\phi_k(X_n)$, we obtain

$$\left| \mathbb{E} \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} (\phi_k(X_m) - \hat{\mu}(k))(\phi_k(X_n) - \hat{\mu}(k)) \right| \leq \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} 4\alpha(X_m, X_n) \sup_{x \in M} |\phi_k(x)|^2.$$ 

The previous two estimates establish the formula for $E_m,n$. 

**Proof of Theorem 1.** Theorem 6 shows that for any $N \geq 1$ and any $t > 0$,

$$\sqrt{\mathbb{E} W_2^2(\mu_N, \mu)} \leq C(M)t^{1/2} + \psi(t) \left( \frac{1}{N} + \frac{1}{N^2} \sum_{1 \leq m < n \leq N} \alpha(X_m, X_n) \right)^{1/2}.$$
with some constant \( C(M) > 0 \) and some function \( \psi(t) > 0 \) depending only on the manifold (see the remark made on \( c_2(\mu, t) \) after Theorem 5). Letting \( t \to 0 \) slowly enough in terms of \( N \), we deduce \( \sqrt{E_2}^2(\mu_N, \mu) \to 0 \) as \( N \to \infty \), as claimed.

**Proof of Theorem 2.** In this proof \( C(M) > 0 \) denotes a constant depending only on the manifold, whose value changes from line to line. The Minakshisundaram–Pleijel asymptotic expansion for the heat kernel [19, p. 154] implies the diagonal estimate

\[
P(t, x, x) = \frac{\text{Vol}(M)}{(4\pi t)^{d/2}} (1 + O(t)) \quad \text{as } t \to 0^+
\]

with an implied constant depending only on \( M \), as well as the off-diagonal estimate

\[
P(t, x, y) \leq \frac{C(M)}{t^{d/2}} \quad t > 0, \ x, y \in M.
\]

To apply Theorem 6, observe that for all \( 0 < t \leq 1 \),

\[
\left| \sum_{k=1}^{\infty} e^{-\lambda_k t} \phi_k(x) \phi_k(y) \right| = \left| \int_1^{\infty} \sum_{k=1}^{\infty} e^{-\lambda_k u} \phi_k(x) \phi_k(y) \, du + \sum_{k=1}^{\infty} e^{-\lambda_k t} \phi_k(x) \phi_k(y) \right| \\
\leq \int_1^{\infty} \left| P(u, x, y) - 1 \right| \, du + C(M).
\]

By the diagonal resp. off-diagonal heat kernel estimate above, \( E \) resp. \( E_{m,n} \) in Theorem 6 thus satisfy

\[
E \leq \int_M \sum_{k=1}^{\infty} e^{-\lambda_k t} |\phi_k(x)|^2 \, d\mu(x) \leq \int_1^{1} \left( \frac{\text{Vol}(M)}{4\pi u^{d/2}} + \frac{C(M)}{u^{d/2}} \right) \, du + C(M)
\]

and

\[
E_{m,n} \leq \beta(X_m, X_n) \left( \int_1^{1} \frac{C(M)}{u^{d/2}} \, du + C(M) \right).
\]

First, assume that \( d = 2 \). Then (11) and (12) give

\[
E \leq \frac{\text{Vol}(M)}{4\pi} \log \frac{1}{t} + C(M) \quad \text{and} \quad E_{m,n} \leq \beta(X_m, X_n) C(M) \left( 1 + \log \frac{1}{t} \right).
\]

Hence Theorem 6 leads to

\[
\sqrt{E_2}^2(\mu_N, \mu) \leq C(M)t^{1/2} + \frac{2}{C^{1/2} N^{1/2}} \left( \frac{\text{Vol}(M)}{4\pi} \log \frac{1}{t} + C(M) + B C(M) \left( 1 + \log \frac{1}{t} \right) \right)^{1/2}.
\]

Choosing \( t = 1/N \) proves claim (i).

Next, assume that \( d \geq 3 \). Then (11) and (12) give

\[
E \leq \frac{\text{Vol}(M)}{(4\pi)^{d/2}} \cdot \frac{1}{(d/2 - 1)t^{d/2 - 1}} + C(M) R_d(t) \quad \text{and} \quad E_{m,n} \leq \beta(X_m, X_n) C(M) \left( \frac{1}{t^{d/2 - 1}} \right),
\]

where

\[
R_d(t) = \begin{cases} 
1 & \text{if } d = 3, \\
1 + \log \frac{1}{t} & \text{if } d = 4, \\
\frac{1}{t^{d-2}} & \text{if } d \geq 5.
\end{cases}
\]

Hence Theorem 6 leads to

\[
\sqrt{E_2}^2(\mu_N, \mu) \leq c_1(\mu)(dt)^{1/2} + C(M)t^{3/4} \\
+ \frac{2}{C^{1/2} N^{1/2}} \left( \frac{\text{Vol}(M)}{4\pi} \cdot \frac{1}{(d/2 - 1)t^{d/2 - 1}} + C(M) R_d(t) + \frac{BC(M)}{t^{d/2 - 1}} \right)^{1/2}.
\]
The optimal choice is
\[
e^{t^{d/4}} = \min \left\{ 1, \frac{2}{c_1(\mu)d^{1/2}c_1^{1/2}N^{1/2}} \left( \frac{\text{Vol}(M)}{(4\pi)^{d/2}(d/2 - 1)} + BC(M) \right)^{1/2} \right\}.
\]
If \( t = 1 \), then claim (ii) follows from the trivial estimate \( W_2(\mu_N, \mu) \leq C(M) \). If \( t < 1 \), then
\[
e^{t^{3/4}} \leq C(M) \left( \frac{1 + B}{cN} \right)^{3/(2d)}^3 \quad \text{and} \quad \frac{R_d(t)^{1/2}}{c^{1/2}N^{1/2}} \leq C(M) \left( \frac{1 + B}{cN} \right)^{3/(2d)},
\]
and claim (ii) follows once again.

If there exists an orthonormal basis \( \phi_k, k \geq 0 \) of Laplace eigenfunctions such that \( \sup_{k \geq 0} \sup_{x \in M} |\phi_k(x)| < \infty \), then we can use the second estimate for \( E_{m,n} \) in Theorem 6 to show that (12), and consequently claims (i) and (ii) hold with \( \beta(X_m, X_n) \) replaced by \( \alpha(X_m, X_n) \).

\( \square \)

**Proof of Theorem 3.** The spectrum of \( M = \mathbb{R}/\mathbb{Z} \) is the set of values \( 4\pi^2 k^2, k \geq 0 \). For each \( k \neq 0 \), the multiplicity is 2, and the corresponding orthonormal eigenfunctions \( \sqrt{2} \sin(2\pi k x) \) and \( \sqrt{2} \cos(2\pi k x) \) have sup-norm \( \sqrt{2} \). Hence \( E \) and \( E_{m,n} \) in Theorem 6 satisfy
\[
E \leq 2 \sum_{k=1}^{\infty} \frac{e^{-4\pi^2 k^2 t}}{4\pi^2 k^2} \leq \sum_{k=1}^{\infty} \frac{1}{\pi^2 k^2} = \frac{1}{6},
\]
and similarly
\[
E_{m,n} \leq 4\alpha(X_m, X_n) \sum_{k=1}^{\infty} \frac{e^{-4\pi^2 k^2 t}}{4\pi^2 k^2} \leq \frac{2}{3}\alpha(X_m, X_n).
\]
Theorem 6 thus gives that for any real \( t > 0 \),
\[
\sqrt{E} W_2^2(\mu_N, \mu) \leq c_1(\mu)t^{1/2} + \frac{2}{c^{1/2}N^{1/2}} \left( \frac{1}{6} + \frac{4}{3}B \right)^{1/2},
\]
and the claim follows from letting \( t \to 0 \).

\( \square \)

5. Random walks on compact Lie groups

Let \( G \) be a compact, connected Lie group of (real) dimension \( d \). Fixing an Ad-invariant inner product on the Lie algebra gives rise to an invariant Riemannian metric on \( G \), which we normalize so that \( \text{Vol}(G) = 1 \). The Wasserstein metric is defined in terms of the corresponding geodesic distance \( \rho \) on \( G \).

Let \( \bar{G} \) denote the unitary dual, and \( d_x \) resp. \( \lambda_x \) the dimension resp. Laplace eigenvalue of an irreducible unitary representation \( \pi \in \bar{G} \); that is, \( \Delta \pi = -\lambda \pi \), where \( \Delta \) acts entrywise. Let \( \pi_0 \in \bar{G} \) be the trivial representation. The Fourier coefficients of \( f \in L^1(G, \text{Vol}) \) and \( \nu \in \mathcal{P}(G) \) are \( \hat{f}(\pi) = \int_G f \pi^* d\text{Vol} \) and \( \hat{\nu}(\pi) = \int_G \pi^* d\nu \), \( \pi \in \bar{G} \). Throughout, \( T^* \) denotes the adjoint of an operator \( T \), and \( \text{SRad}(T), \|T\|_{\text{op}}, \|T\|_{\text{HS}} \) = \( \sqrt{\text{Tr}(T^*T)} \) its spectral radius, operator norm, Hilbert–Schmidt norm, respectively.

Our main tool is still the Berry–Esseen inequality in Theorem 5, now applied with \( \mu = \text{Vol} \), and hence \( c_1(\mu) = c_2(\mu, t) = 1 \). Recall also that \( \chi(G) = 0 \), as the Ricci curvature is positive semidefinite. The inequality now reads, for any \( \nu \in \mathcal{P}(G) \) and any real \( t > 0 \),
\[
W_2(\nu, \text{Vol}) \leq (dt)^{1/2} + 2 \left( \sum_{\pi \in \bar{G}} e^{-\lambda \pi t} \lambda \pi \|\hat{\nu}(\pi)\|_{\text{HS}}^2 \right)^{1/2}.
\]

Indeed, we know that the entries \( \{\sqrt{d_x} \pi_m : 1 \leq m, n \leq d_x, \pi \in \bar{G}\} \) form an orthonormal basis of \( L^2(G, \text{Vol}) \) of complex-valued eigenfunctions of the Laplacian, and to apply Theorem 5 it remains to construct an orthonormal basis \( \phi_k \).
\[ k \geq 0 \] of real-valued eigenfunctions. If \( \pi \in \widehat{G} \) and its contragredient (complex conjugate) \( \overline{\pi} \) are unitarily inequivalent, i.e. \( \pi, \overline{\pi} \in \widehat{G} \) and \( \pi \neq \overline{\pi} \), then the \( 2d_\pi^2 \) functions
\[
\frac{\sqrt{d_\pi}}{\sqrt{2}} (\pi_{mn} + \overline{\pi}_{mn}) \quad \text{and} \quad \frac{\sqrt{d_\pi}}{\sqrt{2}} \cdot \frac{i (\pi_{mn} - \overline{\pi}_{mn})}{i}, \quad 1 \leq m, n \leq d_\pi
\]
is an orthonormal system of real-valued eigenfunctions of the Laplacian with the same eigenvalue \( \lambda_\pi \), spanning the same subspace as the entries of \( \pi \) and \( \overline{\pi} \). With the notation of Section 2, the corresponding \( 2d_\pi^2 \) Fourier coefficients in the sense of manifolds satisfy
\[
\sum_k |\hat{\nu}(k)|^2 = 2d_\pi \sum_{m,n=1}^{d_\pi} \left| \int_G \frac{\sqrt{d_\pi}}{\sqrt{2}} (\pi_{mn} + \overline{\pi}_{mn}) \, d\nu \right|^2 + 2d_\pi \sum_{m,n=1}^{d_\pi} \left| \int_G \frac{\sqrt{d_\pi}}{\sqrt{2}} \cdot \frac{i (\pi_{mn} - \overline{\pi}_{mn})}{i} \, d\nu \right|^2
\]
\[
= 2d_\pi \sum_{m,n=1}^{d_\pi} \left| \int_G \pi_{mn} \, d\nu \right|^2
\]
\[
= 2d_\pi \|\hat{\nu}(\pi)\|^2_{HS}.
\]
If \( \pi \in \widehat{G} \) and its contragredient \( \overline{\pi} \) are unitarily equivalent, then the entries of \( \pi \) and those of \( \overline{\pi} \) span the same \( d_\pi^2 \)-dimensional subspace, which is consequently closed under complex conjugation. We can thus choose an orthonormal basis of real-valued eigenfunctions of the Laplacian spanning the same subspace. With the notation of Section 2, the corresponding \( d_\pi^2 \) Fourier coefficients in the sense of manifolds satisfy
\[
\sum_k |\hat{\nu}(k)|^2 = d_\pi \|\hat{\nu}(\pi)\|^2_{HS}.
\]
In particular, the spectrum \( \lambda_k, k \geq 0 \) of \( G \) consists of the values \( \lambda_\pi \) repeated \( d_\pi^2 \) times, \( \pi \in G \). This reduces (13) to Theorem 5.

5.1. Spectral gaps

Let \( L^2_0(G, \text{Vol}) = \{ f \in L^2(G, \text{Vol}) : \int_G f \, \text{Vol} = 0 \} \) denote the orthogonal complement of the 1-dimensional subspace of constant functions. Given \( \nu \in \mathcal{P}(G) \), let \( T_\nu : L^2_0(G, \text{Vol}) \to L^2_0(G, \text{Vol}) \),
\[
(T_\nu f)(x) = \int_G f(xy) \, d\nu(y)
\]
denote the corresponding Markov operator. Note the identity \( T_\mu * \nu = T_\mu T_\nu \), and in particular, \( T_\nu^n = T_\nu^m \), where \( \mu * \nu \) denotes convolution. Let
\[
q(\nu) := \sqrt{\text{SRad}(T_\nu^* T_\nu)} \leq \|T_\nu\|_{op}.
\]
We say that \( \nu \in \mathcal{P}(G) \) has a spectral gap if \( \text{SRad}(T_\nu) < 1 \). Since \( \text{SRad}(T_\nu) = \lim_{m \to \infty} \|T_\nu^m\|_{op}^{1/m} \), the existence of a spectral gap implies \( q(\nu)^m \leq 1 \) for some \( m \geq 1 \). In case \( T_\nu \) is a normal operator, we have \( q(\nu) = \text{SRad}(T_\nu) = \|T_\nu\|_{op} \).

We first prove an estimate for \( W_2(\nu, \text{Vol}) \) in terms of \( q(\nu) \). For the sake of completeness, we include \( \mathbb{R}^d/\mathbb{Z}^d, d = 1, 2, \) the only compact, connected Lie groups of dimension \( d \leq 2 \).

**Theorem 7.** Let \( \nu \in \mathcal{P}(G) \).

(i) If \( G = \mathbb{R}/\mathbb{Z} \), then
\[
W_2(\nu, \text{Vol}) \leq q(\nu) \sqrt{\frac{2}{\pi}} + 3q(\nu).
\]

(ii) If \( G = \mathbb{R}^2/\mathbb{Z}^2 \), then
\[
W_2(\nu, \text{Vol}) \leq q(\nu) \sqrt{\frac{2}{\pi}} \log \frac{1}{q(\nu)} + 3q(\nu).
\]

(iii) If \( d \geq 3 \), then
\[
W_2(\nu, \text{Vol}) \leq \left( \frac{8}{d(d-2)} \right)^{1/d} \frac{\sqrt{d}}{\sqrt{\pi}} q(\nu)^{2/d} + C(G)q(\nu)^{3/d}
\]
with some constant $C(G) > 0$ depending only on $G$.

Theorem 7 immediately yields an upper bound to $W_2(\nu_1 \ast \nu_2 \ast \cdots \ast \nu_n, \Vol)$, i.e. the rate of weak convergence of a random walk with independent steps, in terms of $q(\nu_1 \ast \nu_2 \ast \cdots \ast \nu_n)$, \( \forall n \geq 1 \). In particular, if \( \|T_{\nu_1}\|_{\text{op}} < 1 \), then \( W_2(\nu_1 \ast \nu_2 \ast \cdots \ast \nu_n, \Vol) \to 0 \) exponentially fast. As for random walks with i.i.d. steps, we have \( W_2(\nu^{*n}, \Vol) \to 0 \) exponentially fast whenever $\nu$ has a spectral gap (even without assuming normality of the Markov operator $T_\nu$).

We also mention an application to deterministic point sets with a spectral gap. Assume that $\nu \in \mathcal{P}(G)$ is supported on at most $N$ points, and satisfies $q(\nu) \leq BN^{-1/2}$ with some constant $B > 0$; note that such point sets exist only in noncommutative groups, that is, in dimension $n \geq 3$. Theorem 7 then yields

$$W_2(\nu, \Vol) \leq \left( \frac{8}{d(d-2)} \right)^{1/d} \frac{\sqrt{dB^{2/d}}}{\sqrt{\pi}N^{1/d}} + \frac{C(G)B^{3/d}}{N^{3/(2d)}},$$

which shows that $\nu$ achieves the optimal distance $\ll N^{-1/d}$ from the uniform distribution, see (4). For instance, the classical explicit construction of Lubotzky, Phillips and Sarnak \cite{27, 28} is a symmetric point set of size $N$ with equal weights on $SU(2)$ and $SO(3)$ for which $q(\nu) = 2\sqrt{N-1}/N$, a value which is in fact smallest possible among all symmetric sets of the same size. Note that a symmetric set with equal weights corresponds to a self-adjoint Markov operator $T_\nu$. Observe also that (14) and (4) together imply that any $\nu \in \mathcal{P}(G)$ supported on at most $N$ points (not necessarily symmetric or with equal weights) satisfies $q(\nu) \gg N^{1/2}$, in accordance with general results of Kesten on spectral properties of random walks \cite{25}.

We now return to random walks, and prove a nonasymptotic upper estimate for the distance of the empirical measure from the uniform distribution under a spectral gap condition. Let $Y_1, Y_2, \ldots, Y_N$ be independent $G$-valued random variables with distribution $\nu_1, \nu_2, \ldots, \nu_N \in \mathcal{P}(G)$, respectively. Let $S_n = Y_1Y_2\cdots Y_n$ be the corresponding random walk, and let $\mu_n = N^{-1} \sum_{n=1}^{N} \delta_{S_n}$ be its empirical measure.

**Theorem 8.** Assume that \( \sum_{1 \leq m < n \leq N} q(\nu_{m+1} \ast \nu_{m+2} \ast \cdots \ast \nu_n) \leq BN \) with some constant $B \geq 0$.

(i) If $G = \mathbb{R}/\mathbb{Z}$, then

$$\sqrt{\mathbb{E}W_2^2(\mu_n, \Vol)} \leq \sqrt{\frac{1+2B}{3N}}.$$

(ii) If $G = \mathbb{R}^2/\mathbb{Z}^2$, then

$$\sqrt{\mathbb{E}W_2^2(\mu_n, \Vol)} \leq \sqrt{\frac{1+2B}{\pi}} \cdot \sqrt{\log N} + \frac{3\sqrt{1+2B}}{\sqrt{N}}.$$

(iii) If $d \geq 3$, then

$$\sqrt{\mathbb{E}W_2^2(\mu_n, \Vol)} \leq \left( \frac{8+16B}{d(d-2)} \right)^{1/d} \frac{\sqrt{d}}{\sqrt{\pi}N^{1/d}} + \frac{C(G)(1+B)^{3/(2d)}}{N^{3/(2d)}},$$

with some constant $C(G) > 0$ depending only on $G$.

For instance, if $\sup_{2 \leq n \leq N} \|T_{\nu_n}\|_{\text{op}} \leq p < 1$, then Theorem 8 applies with $B = p/(1-p)$. Random walks with i.i.d. steps having a spectral gap attain optimal rate, even without assuming normality of the Markov operator.

We conclude this section with the proof of Theorems 7 and 8.

**Proof of Theorem 7.** Observe that the $d_2^{2d}$-dimensional subspace in $L^2_0(G, \Vol)$ spanned by the orthonormal system \{ $\sqrt{\pi_{mn}} : 1 \leq m, n \leq d_\pi$ \} is invariant under $T_\nu$, and that the restriction of $T_\nu$ to this subspace has Hilbert–Schmidt norm square

$$\sum_{m,n=1}^{d_\pi} \|T_\nu \sqrt{\pi_{mn}}\|_{L^2_0(G, \Vol)}^2 = d_\pi \|\hat{\nu}(\pi)\|_{\text{HS}}^2.$$
Therefore $d_\pi \| \mathcal{P}(\pi) \|_{HS}^2 \leq d_\pi^2 q(\nu)^2$, and the Berry–Esseen inequality (13) implies that for all real $t > 0$,

$$W_2(\nu, \text{Vol}) \leq (dt)^{1/2} + 2q(\nu) \left( \sum_{x \in G \atop \pi \neq \pi_0} \frac{e^{-\lambda_\pi x^t}}{\lambda_\pi^2} \right)^{1/2} = (dt)^{1/2} + 2q(\nu) \left( \sum_{k=1}^\infty \frac{e^{-\lambda_k t}}{\lambda_k} \right)^{1/2}. \tag{15}$$

For the sake of simplicity, in the second step we expressed the infinite series in terms of the spectrum of $G$ as a manifold. Assume first, that $G = \mathbb{R}/\mathbb{Z}$. Then (15) yields

$$W_2(\nu, \text{Vol}) \leq t^{1/2} + 2q(\nu) \left( \sum_{k \in \mathbb{Z} \atop k \neq 0} \frac{e^{-4\pi^2 k^2 t}}{4\pi^2 k^2} \right)^{1/2} \leq t^{1/2} + 2q(\nu) \left( \sum_{k \in \mathbb{Z} \atop k \neq 0} \frac{1}{4\pi^2 k^2} \right)^{1/2} = t^{1/2} + 2q(\nu) \left( \frac{1}{12} \right)^{1/2},$$

and claim (i) follows from letting $t \to 0$.

Next, assume that $G = \mathbb{R}^2/\mathbb{Z}^2$. Then (15) yields

$$W_2(\nu, \text{Vol}) \leq (2t)^{1/2} + 2q(\nu) \left( \sum_{k \in \mathbb{Z}^2 \atop k \neq 0} \frac{e^{-4\pi^2 |k|^2 t}}{4\pi^2 |k|^2} \right)^{1/2}. \tag{16}$$

To estimate the infinite series, let $N(x) = |\{k \in \mathbb{Z}^2 : 0 < |k|^2 < x\}|$ denote the number of lattice points other than the origin in the open disk centered at the origin of radius $\sqrt{x}$. Estimating $N(x)$ is the famous Gauss circle problem, but we shall only need the trivial upper bound

$$N(x) \leq \pi \left( \sqrt{x} + \frac{1}{\sqrt{2}} \right)^2 - 1 \leq \pi x + \left( \pi \sqrt{2} + \frac{\pi}{2} - 1 \right) \sqrt{x} \quad (x \geq 1).$$

This follows from drawing unit squares centered at lattice points $k$ with $0 < |k|^2 < x$, and noting that the union of these pairwise disjoint unit squares is a subset of the disk centered at the origin of radius $\sqrt{x} + 1/\sqrt{2}$. Note that $N(x) = 0$ for $0 \leq x \leq 1$. Elementary calculations show that for all $0 < t \leq 1$,

$$\sum_{k \in \mathbb{Z}^2 \atop k \neq 0} \frac{e^{-4\pi^2 |k|^2 t}}{4\pi^2 |k|^2} = \int_0^\infty \frac{e^{-4\pi^2 tx}}{4\pi^2 x} \, dN(x) = - \int_0^\infty N(x) \, d\left( \frac{e^{-4\pi^2 tx}}{4\pi^2 x} \right)$$

$$\leq \int_1^\infty \left( \pi x + \left( \pi \sqrt{2} + \frac{\pi}{2} - 1 \right) \sqrt{x} \right) \left( \frac{t}{x} + \frac{1}{4\pi^2 x^3} \right) e^{-4\pi^2 tx} \, dx$$

$$\leq \int_1^1 \frac{1}{4\pi x} e^{-4\pi^2 tx} \, dx + \int_1^\infty \frac{3\pi \sqrt{2} + \frac{5}{2} \pi - 3}{4\pi^2} e^{-4\pi^2 tx} \, dx$$

$$\leq \int_1^1 \frac{1}{4\pi x} \, dx + \int_1^\infty \frac{t}{4\pi} e^{-4\pi^2 tx} \, dx + \frac{3\pi \sqrt{2} + \frac{5}{2} \pi - 3}{4\pi^2} e^{-4\pi^2 tx} \, dx$$

$$= \frac{1}{4\pi} \log \frac{1}{t} + \frac{e^{-4\pi^2 t}}{16\pi^3} + \frac{3\pi \sqrt{2} + \frac{5}{2} \pi - 3}{4\pi^2}.$$

Letting $\tau$ denote the constant term in the last line of the previous formula, (16) thus gives

$$W_2(\nu, \text{Vol}) \leq (2t)^{1/2} + 2q(\nu) \left( \frac{1}{4\pi} \log \frac{1}{t} + \tau \right)^{1/2}.$$

Choose $t = q(\nu)^2 \leq 1$, and note that $2^{1/2} + 2\tau^{1/2} \approx 2.77$. This proves claim (ii).
Finally, assume that \( d \geq 3 \). Similarly to the proof of Theorem 2, the Minakshisundaram–Pleijel asymptotic expansion of the heat kernel now gives

\[
\sum_{k=0}^{\infty} e^{-\lambda_k t} = \frac{1}{(4\pi t)^{d/2}} (1 + O(t)) \quad \text{as } t \to 0^+.
\]

Consequently for any \( 0 < t \leq 1 \),

\[
\sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} = \int_0^1 \sum_{k=1}^{\infty} e^{-\lambda_k u} du + \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} \leq \int_t^1 \left( \frac{1}{(4\pi u)^{d/2}} + \frac{C(G)}{u^{d/2-1}} \right) du + C(G)
\]

\[
\leq \frac{1}{(4\pi)^{d/2}} \frac{1}{(d/2-1)t^{d/2-1}} + C(G) R_d(t),
\]

where

\[
R_d(t) = \begin{cases} 
1 & \text{if } d = 3, \\
1 + \log \frac{1}{t^{d/2-2}} & \text{if } d = 4, \\
\frac{1}{t^{d/2-2}} & \text{if } d \geq 5.
\end{cases}
\]

Hence (15) yields

\[
W_2(\nu, \text{Vol}) \leq (dt)^{1/2} + 2q(\nu) \left( \frac{1}{(4\pi)^{d/2}(d/2-1)t^{d/2-1}} + C(G) R_d(t) \right)^{1/2}.
\]

The optimal choice is

\[
t^{d/2} = \min \left\{ 1, \frac{4q(\nu)^2}{(4\pi)^{d/2}d(d/2-1)} \right\}.
\]

If \( t = 1 \), then claim (iii) follows from the trivial estimate \( W_2(\nu, \text{Vol}) \leq C(G) \). If \( t < 1 \), then \( q(\nu) R_d(t)^{1/2} \leq C(G)q(\nu)^{3/d} \), and claim (iii) follows once again.

**Proof of Theorem 8.** Applying the Berry–Esseen inequality (13) and the triangle inequality for the \( L^2 \)-norm gives that for any real \( t > 0 \),

\[
\sqrt{\mathbb{E} W_2^2(\mu_N, \text{Vol})} \leq (dt)^{1/2} + 2 \left( \sum_{\pi \in G} \frac{e^{-\lambda_\pi t}}{\lambda_\pi} d_\pi \mathbb{E} ||\hat{\mu}_N(\pi)||_{HS}^2 \right)^{1/2}.
\]

Here

\[
\mathbb{E} ||\hat{\mu}_N(\pi)||_{HS}^2 = \mathbb{E} \left\| \frac{1}{N} \sum_{n=1}^{N} \pi(S_n) \right\|_{HS}^2 = \frac{1}{N^2} \sum_{m,n=1}^{N} \mathbb{E} \text{tr} (\pi(S_m)^* \pi(S_n)).
\]

Since \( \pi(x) \) is a \( d_\pi \times d_\pi \) unitary matrix, the total contribution of the diagonal terms \( m = n \) is \( d_\pi / N \). If \( m < n \), then

\[
|\mathbb{E} \text{tr}(\pi(S_m)^* \pi(S_n))| = |\mathbb{E} \text{tr}(\pi(Y_m)^* \pi(Y_{m+1})^* \cdots \pi(Y_1)^* \pi(Y_1) \cdots \pi(Y_{n-1}) \pi(Y_n))|
\]

\[
= |\mathbb{E} \text{tr}(\pi(Y_{m+1}Y_{m+2} \cdots Y_n))|
\]

\[
= |\text{tr}(\nu_{m+1} \nu_{m+2} \cdots \nu_n)^*(\pi)|
\]

\[
\leq \sqrt{d_\pi} ||(\nu_{m+1} \nu_{m+2} \cdots \nu_n)^*(\pi)||_{HS}
\]

\[
\leq d_\pi q(\nu_{m+1} \nu_{m+2} \cdots \nu_n).
\]
In the last two steps we used the Cauchy–Schwarz inequality, and an observation from the proof of Theorem 7. Estimating the terms \( m > n \) is entirely analogous, and we obtain
\[
\frac{1}{N^2} \sum_{m,n=1}^{N} \mathbb{E} \text{tr} \left( \pi(S_m)^* \pi(S_n) \right) \leq \frac{d_\pi}{N} + 2d_\pi \frac{1}{N^2} \sum_{1 \leq m < n \leq N} q(\nu_{m+1} \ast \nu_{m+2} \ast \cdots \ast \nu_n) \leq \frac{d_\pi}{N}(1 + 2B).
\]

Hence for any real \( t > 0 \),
\[
\sqrt{\mathbb{E} W_2^2(\mu_N, \text{Vol})} \leq (dt)^{1/2} + \frac{2(1 + 2B)^{1/2}}{N^{1/2}} \left( \sum_{\pi \in \hat{G}} \frac{e^{-\lambda_\pi t}}{\lambda_\pi} d_\pi^2 \right)^{1/2}.
\]

The rest of the proof is identical to that of Theorem 7 with \( q \) replaced by \((1 + 2B)^{1/2}/N^{1/2}\).

\[\square\]

5.2. Nonuniform spectral gaps

Let \( G \) be a semisimple, compact, connected Lie group. The proof of Theorem 4 is based on a result of Varjú [39], who proved that for any \( \nu \in \mathcal{P}(G) \) and any \( r > 0 \),
\[
1 - \max_{\pi \in \hat{G}} \|\widehat{\nu}(\pi)\|_{\text{op}} \geq c_0 \left( 1 - \max_{\pi \in \hat{G}} \|\widehat{\nu}(\pi)\|_{\text{op}} \right) \frac{1}{\log^2 (r + 2)},
\]
(17)

where \( c_0, r_0 > 0 \) are constants depending only on \( G \). Varjú calls this a nonuniform spectral gap estimate, since the lower bound to the gap depends on the size \( r \) of the Laplace eigenvalues. In fact, he proved (17) with a factor \( 1/\log^\gamma (r + 2) \), where \( 1 \leq \gamma \leq 2 \) is a constant depending explicitly on the group; for the sake of simplicity, we formulated Theorem 4 in the worst case \( \gamma = 2 \). The spectral gap conjecture basically states that the same holds with \( \gamma = 0 \).

**Proof of Theorem 4.** Let \( c_0, r_0 > 0 \) be as in (17), and let \( \nu \in \mathcal{P}(G) \) be such that \( \nu^{*n} \to \text{Vol} \) weakly. For any \( \pi \in \hat{G} \), \( \pi \neq \pi_0 \) we then have \( \nu^{*n}(\pi) = \widehat{\nu}(\pi)^n \to 0 \). Consequently \( \text{SRad}(\widehat{\nu}(\pi)) < 1 \), and there exists an integer \( m_0 = m_0(\nu) \) such that \( \|\widehat{\nu}(\pi)^{m_0}\|_{\text{op}} < 1 \) for all of the finitely many \( \pi \in \hat{G} \) with \( 0 < \lambda_\pi \leq r_0 \). We now apply (17) to \( \nu^{*m_0} \), and obtain that for any integer \( n \geq 1 \) and any \( r > 0 \),
\[
\max_{\pi \in \hat{G}} \|\widehat{\nu}(\pi)^n\|_{\text{op}} \leq \left( \max_{\pi \in \hat{G}} \|\widehat{\nu}(\pi)^{m_0}\|_{\text{op}} \right)^{1/n/m_0} \leq 2 \exp \left( -\frac{bn}{\log^2 (r + 2)} \right)
\]
(18)

with the constant
\[
b = b(\nu) := \frac{c_0}{m_0} \left( 1 - \max_{\pi \in \hat{G}} \|\widehat{\nu}(\pi)^{m_0}\|_{\text{op}} \right) > 0.
\]

First, we estimate \( W_2(\nu^{*n}, \text{Vol}) \). The nonuniform spectral gap estimate (18) yields
\[
d_\pi \|\widehat{\nu}(\pi)^n\|_{\text{HS}} \leq d_\pi^2 \|\widehat{\nu}(\pi)^n\|_{\text{op}}^2 \leq 4d_\pi^2 \exp \left( -\frac{bn}{\log^2 (\lambda_\pi + 2)} \right).
\]
Hence the Berry–Esseen inequality (13) gives that for any real $t > 0$,

$$W_2(\nu^n, \text{Vol}) \ll t^{1/2} + \left( \sum_{\pi \in \hat{G}} \sum_{x \neq \pi_0} \frac{e^{-\lambda_x t}}{\lambda_x} d_\pi^2 \exp \left( -\frac{bn}{\log^2(\lambda_x + 2)} \right) \right)^{1/2}$$

$$= t^{1/2} + \left( \sum_{k=1}^\infty \frac{e^{-\lambda_k t}}{\lambda_k} \exp \left( -\frac{bn}{\log^2(\lambda_k + 2)} \right) \right)^{1/2}.$$

For the sake of simplicity, in the second step we expressed the infinite series in terms of the spectrum of $G$ as a manifold. Using Weyl’s law $|\{k \in \mathbb{N} : \lambda_k \leq x\}| \sim \kappa_d x^{d/2}$ as $x \to \infty$ with some (explicit) constant $\kappa_d > 0$, it is straightforward to check that for any $x > 0$ with $xt$ large enough, $\sum_{\lambda_k > x} e^{-\lambda_k t} \ll x^{d/2} e^{-xt}$. Estimating the terms $\lambda_k \leq x$ and $\lambda_k > x$ separately, we thus deduce

$$\sum_{k=1}^\infty \frac{e^{-\lambda_k t}}{\lambda_k} \exp \left( -\frac{bn}{\log^2(\lambda_k + 2)} \right) \ll \sum_{\lambda_k \leq x} \exp \left( -\frac{bn}{\log^2(x + 2)} \right) + \sum_{\lambda_k > x} e^{-\lambda_k t} \ll x^{d/2} \exp \left( -\frac{bn}{\log^2(x + 2)} \right) + x^{d/2} e^{-xt}.$$  

One readily checks that the optimal choice is $x = \exp(a_0 n^{1/3})$ and $t = n^{1/3} \exp(-a_0 n^{1/3})$ with a suitably small constant $a_0 = a_0(\nu) > 0$, in which case we obtain

$$W_2(\nu^n, \text{Vol}) \ll t^{1/2} + \exp(-a_0 n^{1/3}) \ll \exp\left( -(a_0/4) n^{1/3} \right),$$

as claimed.

Next, we estimate $\sqrt{\mathbb{E}W_2^2(\mu_N, \text{Vol})}$. The Berry–Esseen inequality (13) and the triangle inequality for the $L^2$-norm give that for any real $t > 0$,

$$\sqrt{\mathbb{E}W_2^2(\mu_N, \text{Vol})} \leq (dt)^{1/2} + 2 \left( \sum_{\pi \in \hat{G}} \sum_{\pi \neq \pi_0} \frac{e^{-\lambda_x t}}{\lambda_x} d_\pi \mathbb{E}||\tilde{\mu}_N(\pi)||_{HS}^2 \right)^{1/2}.$$

Following the steps in the proof of Theorem 8, here

$$\mathbb{E}||\tilde{\mu}_N(\pi)||_{HS}^2 \leq \frac{d_\pi}{N} + \frac{2}{N^2} \sum_{1 \leq m < n \leq N} |\text{tr}(\tilde{\mu}(\pi)^{n-m})| \leq \frac{d_\pi}{N} + \frac{2d_\pi}{N^2} \sum_{1 \leq m < n \leq N} ||\tilde{\mu}(\pi)^{n-m}||_\text{op}.$$  

The nonuniform spectral gap estimate (18) shows that

$$\sum_{1 \leq m < n \leq N} ||\tilde{\mu}(\pi)^{n-m}||_\text{op} \leq 2 \sum_{1 \leq m < n \leq N} \exp\left( -\frac{b(n-m)}{\log^2(\lambda_\pi + 2)} \right) \ll N \log^2(\lambda_\pi + 2),$$

therefore

$$\sqrt{\mathbb{E}W_2^2(\mu_N, \text{Vol})} \ll t^{1/2} + 2 \left( \sum_{\pi \in \hat{G}} \sum_{\pi \neq \pi_0} \frac{e^{-\lambda_x t}}{\lambda_x} d_\pi \log^2(\lambda_\pi + 2) \right)^{1/2}$$

$$= t^{1/2} + \frac{1}{N^{1/2}} \left( \sum_{k=1}^\infty \frac{e^{-\lambda_k t}}{\lambda_k} \log^2(\lambda_k + 2) \right)^{1/2}$$

with an implied constant depending only on $\nu$ and $G$. From Weyl’s law we deduce

$$\sum_{k=1}^\infty \frac{e^{-\lambda_k t}}{\lambda_k} \log^2(\lambda_k + 2) \leq \sum_{\lambda_k \ll 1/t} \frac{\log^2(\lambda_k + 2)}{\lambda_k} + \sum_{\lambda_k \gg 1/t} e^{-\lambda_k t} \log^2 \frac{1}{t} \ll t^{1-d/2} \log^2 \frac{1}{t},$$
hence
\[ \sqrt{\mathbb{E} W_2^2(\mu_N, \text{Vol})} \ll t^{1/2} + \frac{1}{N^{1/2}} t^{1/2 - d/4} \log \frac{1}{t}. \]

The optimal choice is \( t = N^{-2/d}(\log N)^{4/d} \), and the claim follows. \( \square \)

**Funding**

The author is supported by the Austrian Science Fund (FWF) projects F-5510 and Y-901. I would like to thank the two anonymous referees for useful comments and suggestions.

**References**


