ON THE WASSERSTEIN DISTANCE BETWEEN CLASSICAL SEQUENCES AND THE LEBESGUE MEASURE

LOUIS BROWN AND STEFAN STEINERBERGER

Abstract. We discuss the classical problem of measuring the regularity of distribution of sets of \(N\) points in \(\mathbb{T}^d\). A recent line of investigation is to study the cost (= mass \(\times\) distance) necessary to move Dirac measures placed in these points to the uniform distribution. We show that Kronecker sequences satisfy optimal transport distance in \(d \geq 3\) dimensions. This shows that for differentiable \(f : \mathbb{T}^d \rightarrow \mathbb{R}\) and badly approximable vectors \(\alpha \in \mathbb{R}^d\), we have

\[
\left| \int_{\mathbb{T}^d} f(x) \, dx - \frac{1}{N} \sum_{k=1}^{N} f(k\alpha) \right| \leq c_\alpha \frac{\|\nabla f\|_{L^\infty}^{(d-1)/d} \|\nabla f\|_{L^2}^{1/d}}{N^{1/d}}.
\]

We note that the result is uniformly true for a sequence instead of a set. Simultaneously, it refines the classical integration error for Lipschitz functions, \(\|\nabla f\|_{L^\infty} \approx N^{-1/d}\). We obtain a similar improvement for numerical integration with respect to the regular grid. The main ingredient is an estimate involving Fourier coefficients of a measure; this allows for existing estimates to be conviently ‘recycled’. We present several open problems.

1. Introduction

1.1. Introduction. We study the problem of measuring the regularity of points set \(\{x_1, \ldots, x_N\} \subset \mathbb{T}^d\) as well as infinite sequences. There are many classical notions of regularity as well as good constructions of sets minimizing these notions that have been proposed; we refer to the classical textbooks \([13, 15, 16, 32]\). The classical theory has developed a useful machinery in terms of exponential sums that exploit

\[
\text{Figure 1. Three Dirac measures with weight } N^{-1}: \text{ we want to transport the measure so that the resulting measure is the Lebesgue measure; one way is to associate nearby regions in space of equal volume } N^{-1} \text{ and measure the cost of transporting the weight.}
\]
regularities of number-theoretic constructions. We will not, initially, pursue this path and instead ask a different question: consider the measure

$$\mu = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}.$$ 

How would we go about distributing this measure in such a way that the end result is the Lebesgue measure on $T^d$? Here, the ‘cost’ of transporting $\delta$ units of measure across distance $d$ is understood to be $\delta \cdot d$. An even more practical example is the following: suppose we have people evenly distributed over $T^d$ and $N$ storage units placed in $\{x_1, \ldots, x_N\}$. Demand and supply are exactly matched: how would the trucks have to drive to distribute the goods from the supermarket evenly? This is Monge’s transportation problem from 1781. It is fairly easy to see that

$$\text{transportation cost} \geq c_d N^{-1/d},$$

where $c_d$ is a universal constant depending only on the dimension. This scaling is, for example, assumed for a rescaling of $\mathbb{Z}^d$ intersected with $T^d \cong [0,1]^d$. Our paper was motivated by the following questions

1. Do the classical constructions of regular sequences in $T^d$ from $[13,15,16,32]$ have an optimal transportation cost? Do they have it uniformly in $N$?
2. How does one go about proving such results?
3. Does this perspective lead to new results?

We emphasize that these types of problems, estimating transport cost from one measure to another, have been actively investigated in the field of Optimal Transport $[42,57]$. Here, the emphasis is usually on existence and uniqueness of optimal transport maps as well as fine qualitative and quantitative properties. Many special cases have been actively investigated in probability theory, we emphasize the problems of estimating the transport of random points to the Lebesgue measure, more generally, random points drawn from a measure $\mu$ to $\mu$ or random points to random points $[1,2,9,21,31,55,54,56]$. As far as we know, special structures arising from Number Theory or Combinatorics have not been considered before (with some precursors in $[7,8,26,45,48,50,51,52]$).

1.2. Setup. We introduce the $p-$Wasserstein distance (the example above had $p = 1$ and is also known as the ‘Earth Mover distance’) between two measures $\mu$ and $\nu$ as

$$W_p(\mu, \nu) = \left( \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{M \times M} |x - y|^p d\gamma(x, y) \right)^{1/p},$$

where $| \cdot |$ is the usual distance on the torus and $\Gamma(\mu, \nu)$ denotes the collection of all measures on $M \times M$ with marginals $\mu$ and $\nu$, respectively (also called the set of all couplings of $\mu$ and $\nu$). Our two measures under consideration are

$$\mu = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k} \quad \text{and} \quad \nu = dx,$$

where $dx$ refers to the normalized volume measure. It is relatively easy to see that, we have an (optimal) lower bound that is independent of the set $\{x_1, \ldots, x_N\} \subset T^d$

$$W_1 \left( \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, dx \right) \geq \frac{c_d}{N^{1/d}}.$$
We refer to Santambrogio \cite{42} or Villani \cite{57} for nice introductions to Optimal Transport and the Wasserstein distance.

1.3. Existing Results in One Dimension. There are series of recent results in the one-dimensional setting. Given a finite set on the one-dimensional torus \( \{ x_1, \ldots, x_N \} \subset \mathbb{T} \), we associate to it the measure

\[
\mu = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}.
\]

A natural quantity that is frequently studied (see e.g. \cite{15,16,32}) is the discrepancy

\[
D_N(\mu) = \sup_{J \text{ interval}} |\mu(J) - |J||.
\]

It is easy to see that

\[
N^{-1} \leq D_N \leq 1.
\]

The inequality

\[
W_1(\mu, dx) \lesssim D_N(\mu)
\]

follows from Monge-Kantorovich duality (this is carried out in greater detail in \cite{7} or \cite{46}). Here and in what follows, we use \( A \lesssim B \) to denote the existence of a universal constant \( c > 0 \) such that \( A \leq cB \). Another notion of regularity is Zinterhof’s diaphony \cite{16,58} and can be defined as

\[
F_N(\mu) = \left( \sum_{k \in \mathbb{Z}, k \neq 0} |\hat{\mu}(k)|^2 \right)^{1/2}.
\]

A recent inequality of Peyre \cite{38} can be reinterpreted as saying (see \cite{45,48}) that

\[
W_2(\mu, dx) \lesssim F_N(\mu).
\]

Summarizing, we have two inequalities and Holder’s inequality

\[
W_1(\mu, dx) \lesssim D_N(\mu)
\]

\[
W_2(\mu, dx) \lesssim F_N(\mu)
\]

\[
W_1(\mu, dx) \leq W_2(\mu, dx).
\]

For classical one-dimensional constructions in Number Theory, the notions \( D_N \) and \( F_N \) have been studied intensively. This connection immediately implies a series of results for the Wasserstein distance: the upper bounds that we obtain for the \( W_2 \) distance are better, by a factor of \( (\log N)^{1/2} \) than the estimate on \( D_N \). We rephrase an existing result of Proinov & Grozdanov \cite{41} in this manner.

**Theorem 1** (Proinov & Grozdanov \cite{41}). Let \((x_n)_{n=1}^\infty\) denote the van der Corput sequence in base \( r \). Then

\[
W_2 \left( \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, dx \right) \lesssim_r \frac{(\log N)^{1/2}}{N}.
\]

It seems reasonable to conjecture that this is the smallest possible growth rate. At least by using the bound \( W_2 \lesssim F_N \), it is not possible to obtain a better result: Proinov \cite{40} showed that for any sequence \((x_n)_{n=1}^\infty\), we necessarily have

\[
F_N \gtrsim \frac{(\log N)^{1/2}}{N} \quad \text{for infinitely many values of } N.
\]
It seems exceedingly likely that this is indeed the best one can do also for the $W_2$ distance but we are not aware of any results in this direction. The second author recently remarked [45] that the $(n\alpha)$--sequence satisfies a similar growth. Moreover, quadratic residues of a finite field, suitably rescaled, behave better than one would obtain using the Polya-Vinogradov estimate (see for example [10]).

**Theorem 2** (na Sequence and Quadratic Residues, [45]). Let $\alpha$ be badly approximable and $x_n = \{na\}$, then

$$W_2 \left( \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, dx \right) \lesssim_{\alpha} \frac{(\log N)^{1/2}}{N}.$$ 

Moreover, let $p$ be a prime and let $x_k = \{k^2/p\}$ for $1 \leq k \leq p$. Then

$$W_2 \left( \frac{1}{p} \sum_{k=1}^{p} \delta_{x_k}, dx \right) \lesssim \frac{1}{\sqrt{p}}.$$ 

This connection between the Wasserstein distance, diaphony, the Sobolev space $\dot{H}^{-1}$ and the corresponding exponential sum estimate does not seem to have been noticed before the paper [45]. For that reason, we believe that there are interesting results in $d = 1$ that are not yet known but possibly within reach.

1.4. **Existing Results in Higher Dimensions.** It is easy to see that for any fixed set of points $\{x_1, \ldots, x_N\} \subset T^d$, the lattice construction (see Fig. 2) is optimal up to constants. However, if one were to construct an infinite sequence $(x_n)_{n=1}^{\infty}$ with estimates that are uniformly good, a lattice construction does not seem to be particularly useful; see Fig 2: where would one put the next point and the point after that?

![Figure 2](image-url)

**Figure 2.** A distribution with small Wasserstein transportation cost – however, these constructions are not uniform in $N$.

A general result has recently been obtained by the authors [8] on general compact manifolds. If $(M, g)$ is a compact manifold without boundary and $G(\cdot, \cdot)$ denotes the Green’s function of the Laplacian $-\Delta_g$, then the greedy construction

$$x_n = \arg \min_{x \in M} \sum_{k=1}^{n-1} G(x, x_k).$$

has good distribution properties.
Theorem 3 (B & S [8]). Let $x_n$ be a sequence obtained in this way on a $d$-dimensional compact manifold. Then

$$W_2 \left( \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, dx \right) \lesssim_M \begin{cases} \frac{N^{-1/2} (\log N)^{1/2}}{N^{-1/d}} & \text{if } d = 2 \\ N^{-1/d} & \text{if } d \geq 3. \end{cases}$$

It is not clear to us whether the $(\log N)^{1/2}$ factor is necessary. The crucial ingredient to these types of results is a smoothing procedure introduced in [45] coupled with Peyre’s estimate [38]. This allows us to obtain a general bound which is particularly simple on $T_d$ and reads as follows:

$$W_2 \left( \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, dx \right) \lesssim \inf_{t > 0} \left[ \sqrt{t} + \sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \| \sum_{n=1}^{N} e^{2\pi i (k, x_n)} \|^2 \right]$$

This inequality has a series of remarkable features:

(1) it is phrased exclusively in terms of exponential sums that have been well studied for a variety of sequences; in particular, information about these exponential sums is available.

(2) The quantity on the right-hand side reduces to the notion of diaphony $F_N$ in the one-dimensional case $d = 1$ and $t = 0$.

(3) However, in contrast to classical diaphony, the quantity is finite for any set of points and any dimension $d \in \mathbb{N}$ for all $t > 0$. It can thus be regarded as a generalization of diaphony.

We note that diaphony has been studied in a variety of settings [12, 19, 25, 37, 58, 59]. We interpret it, in one dimension, as the quantity

$$F_N = \left\| \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k} \right\|_{H^{-1}},$$

where $H^{-1}$ is the Sobolev space. We observe that this quantity becomes meaningless in dimensions $d \geq 2$ because Dirac deltas are no longer contained in the Sobolev space $H^{-1}$ (or, put differently, the infinite sums do not converge). This has been a persistent issue in trying to define notions of discrepancy in higher dimensions on other geometries (see e.g. Freeden [20] or Grabner, Klinger & Tichy [24]). In contrast, we can rewrite our inequality (even on general manifolds) as

$$W_2 \left( \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, dx \right) \lesssim \inf_{t > 0} \left[ \sqrt{t} + e^{t\Delta} \left\| \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k} \right\|_{H^{-1}} \right],$$

where $e^{t\Delta}$ is the heat propagator, i.e. the forward evolution of the heat equation. This quantity is always finite for any $t > 0$. We believe this to be an insight that might be useful in discrepancy theory as a suitable generalization of diaphony to higher dimensions. We also note that this notion is intimately tied to the integration error for Lipschitz functions, see §2.6. below.

2. Main Results

2.1. A Random Walk. We have already mentioned a series of results for $d = 1$. We add another one to the list: here, we do not consider a sequence insomuch as an actual probability measure. Let $\mu_k$ be the measure that arises from an unbiased
random walk on $T \cong [0, 1]$ where each step is $\pm \alpha$ (independently and with likelihood $1/2$ each) and $\alpha$ is a quadratic irrational. This model was studied by Su [50] (see also Hensley & Su [26] and Su [52]). The main result in [50] showed that the measure arising after $k$ random steps satisfies

$$D_N(\mu_k) \lesssim \alpha^{-1/2}.$$  

We note that this result immediately implies $W_1(\mu_k, dx) \lesssim k^{-1/2}$. Here, we show that for this model we can obtain a (worse) bound for the (larger) $W_2$-distance.

**Theorem 4.** We have

$$W_2(\mu_k, dx) \lesssim \alpha^{-1/4}.$$  

We emphasize that the arising computations follow from standard estimates. This estimate is presumably not optimal and stronger results should be true. Hensley & Hu [26] discuss their result and put it in direct relation to the Wasserstein distance. We hope that our approach will be a useful technique for these types of problems.

### 2.2. Kronecker sequences.

Kronecker sequences are the natural generalization of \{n\alpha\} sequences (irrational rotations) on $T$. We say that a vector $(\alpha_1, \alpha_2, \ldots, \alpha_d) \in \mathbb{R}^d$ is badly approximable if, for all positive integers $q \neq 0$, we have

$$\max_{1 \leq j \leq d} \|\alpha_jq\| \geq \frac{c_{\alpha}}{q^{1/d}},$$  

where $\| \cdot \|$ is the distance to the nearest integer. By Dirichlet’s approximation theorem, this is the optimal scaling. The existence of such a vector follows from continued fraction expansion when $d = 1$. The first examples in higher dimensions are due to Perron [39], Davenport [14] showed that there are uncountably many such vectors for $d = 2$ and Schmidt [43] extended this result to $d \geq 3$. The Kronecker sequence is then defined via

$$x_n = n\alpha \mod 1,$$  

where mod 1 is to be interpreted component-wise. We now establish that these sequences have uniformly good transport properties to the uniform measure.

**Theorem 5.** Let $d \geq 2$ and let $\alpha \in \mathbb{R}^d$ be badly approximable. Then the Kronecker sequence satisfies

$$W_2\left(\frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, dx\right) \lesssim c_{\alpha, d} N^{-1/d}$$  

We emphasize that the result is best possible (up to constants) as well as uniform in $N$. It is not at all clear to us whether the condition of $\alpha$ being badly approximable is necessary; it is clear that for irrational $\alpha$ that can be very well approximated by rational numbers, the rate at which the Wasserstein distance goes to 0 can be arbitrarily slow and being badly approximable is at the other end of the spectrum.

### 2.3. Integration.

Let us consider the problem of numerically integrating a function $f : T^d \to \mathbb{R}$ which we assume to be Lipschitz. It is a classic 1959 result of Bakhvalov [4] (see also Novak [35]) that there are points $(x_k)_{k=1}^{N}$ such that for all differentiable functions $f : T^d \to \mathbb{R}$ with Lipschitz constant $\|\nabla f\|_{L^\infty}$

$$\left| \int_{T^d} f(x) dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq c_d \|\nabla f\|_{L^\infty} N^{-1/d}$$  

where

$$c_d = \frac{d}{4d + 4}.$$  

This result is best possible (up to constants) and uniform in $N$. It is not at all clear to us whether the condition of $f$ being Lipschitz is necessary; it is clear that for $f$ that can be very well approximated by polynomials of degree $k$, the rate at which the approximation error goes to 0 can be arbitrarily slow and being Lipschitz is at the other end of the spectrum.
and that this result is optimal in the power of \( N \) and the its dependance on the Lipschitz constant \( \| \nabla f \|_{L^\infty} \): there are functions \( f \) for which the error is at that scale (up to constants). Recently, Hinrichs, Novak, Ullrich and Woźniakowski \[29\] (see also \[27, 28\]) established rather precise estimates on the constant \( c_d \) and showed that product rules (regular grid structures) are a good choice whenever the number of points \( N \) is of the form \( N = m^d \). We improve this result in the following manner.

**Theorem 6.** Let \( d \geq 2 \) and let \( \alpha \) be a badly approximable vector \( \alpha \in \mathbb{R}^d \). Then, for some universal \( c_\alpha > 0 \) and all differentiable \( f : \mathbb{T}^d \rightarrow \mathbb{R} \)

\[
\left| \int_{\mathbb{T}^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(k\alpha) \right| \leq c_\alpha \| \nabla f \|_{L^\infty(T^d)}^{(d-1)/d} \| \nabla f \|_{L^1(T^d)}^{1/d} N^{-1/d}.
\]

The main novelties are that

1. the result holds uniformly in \( N \) along a sequence and
2. the error estimate is actually smaller than the classically assumed dependence on the Lipschitz constant. We note that, trivially

\[
\| \nabla f \|_{L^2} \leq \| \nabla f \|_{L^\infty}
\]

which recovers the traditional estimate. At first, this seems like a contradiction to the fact that the dependence on the Lipschitz constant is optimal – however, it merely implies that extremal functions for the estimate have to have \( \| \nabla f \|_{L^2} \sim \| \nabla f \|_{L^\infty} \) which is perhaps not surprising (one would expect them to grow at maximal speed away from the points, so \( |\nabla f| \) should be fairly constant).

3. The result is an explicit improvement in the case where the function \( f \) has a large derivative in a small region.

We also emphasize that there is nothing particularly special about the Kronecker sequence: any sequence for which we can establish optimal Wasserstein bounds along the lines outlined above, we will also obtain a version of the integration result, the proof is identical (see §2.6. below). Indeed, the result is actually true on general \( d \)-dimensional manifolds, we refer to Theorem 8 below. We also note that the result has similar flavor and scaling as a result that was previously obtained by the second author \[49\] in a different context. If \( \alpha \in \mathbb{R}^d \) is badly approximable, then it is possible to obtain directional Poincaré inequalities without loss on \( \mathbb{T}^d \): for all \( f \in C^\infty(\mathbb{T}^d) \) with mean value 0, we have

\[
\| \nabla f \|_{L^2}^{(d-1)/d} \| \langle \nabla f, \alpha \rangle \|_{L^2}^{1/d} \geq c_\alpha \| f \|_{L^2}.
\]

### 2.4. The Case of the Regular Grid.

Let us return to the case of the regular grid (refering to \( \mathbb{T}^d \) or \([0, 1]^d \) and having \( N = m^d \) points that are arranged as a regular grid). Since we have just improved the classic integration error for the Kronecker sequence, we would expect a similar improvement to hold for the regular grid (which is well understood to be, in a sense, an optimal set for the sampling for Lipschitz functions). The classic estimate for a regular grid \( (x_n)_{n=1}^N \) is

\[
\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N} \sum_{k=1}^{N} f(x_k) \right| \leq c_d \| \nabla f \|_{L^\infty} N^{-1/d}.
\]

Sukharev \[53\] (see also \[36\]) has determined the sharp constant. It is known that ‘the result cannot be significantly improved for uniformly continuous functions’
Theorem 7. We have, for some explicit constant $c_d$ depending only on the dimension, for all differentiable $f : [0, 1]^d \to \mathbb{R}$ sampled on the regular grid $(x_k)_{k=1}^N$

$$\left| \int_{[0,1]^d} f(x) dx - \frac{1}{N} \sum_{k=1}^N f(x_k) \right| \leq c_d \| \nabla f \|_{L^\infty_{(T^d)}} \| \nabla f \|_{L^1(T^d)} N^{-1/d}.$$ 

We observe that this is even a slightly better estimate than Theorem 6 (an $L^1(T^d)$ norm instead of the larger $L^2(T^d)$ norm); this is maybe to be expected, we would assume the strongest estimates become available for the regular grid. We will also show that this is the best possible bound in terms of these $L^1$-spaces. It is an interesting question whether this bound ($L^1$ instead of $L^2$) is also true for the Kronecker sequence (Theorem 6). More generally, one could ask whether there is a sequence $(x_n)_{n=1}^\infty$ that uniformly attains the same error estimate as Theorem 7.

Our argument is based on a type of Poincaré inequality that we could not find in the literature and that may be of independent interest: let $B \subset \mathbb{R}^d$ denote the unit ball centered at the origin and let $f : B \to \mathbb{R}$ be Lipschitz and satisfy $f(0) = 0$, then

$$\left| \int_B f(x) dx \right| \lesssim_d \int_B \frac{\| \nabla f \|_{d-1}}{\| x \|^{d-1}} dx \lesssim_d \| \nabla f \|_{L^\infty_{B}} \| \nabla f \|_{L^1_{B}} ^{\frac{d}{2}}.$$ 

This inequality may have other applications.

2.5. Other manifolds. Nothing about our approach is particularly tied to the torus $T^d$. Indeed, the main inequality

$$W_2 \left( \frac{1}{N} \sum_{k=1}^N \delta_{x_k}, dx \right) \lesssim \inf_{t > 0} \sqrt{t} + \left( \sum_{k \in \mathbb{Z}^d \setminus \{0\}} e^{-\|k\|^2 t} \left| \frac{1}{N} \sum_{n=1}^N e^{2\pi i (k, x_n)} \right|^2 \right)^{1/2}$$ 

can be generalized just as easily to other manifolds. Let us fix a manifold $(M, g)$ and use $\phi_k$ denote the sequence of Laplacian eigenfunction

$$-\Delta \phi_k = \lambda_k \phi_k.$$ 

We assume that $\phi_0 = 1$ is the trivial (constant) eigenfunction and that they are normalized to $\| \phi_k \|_{L^2} = 1$. Then the inequality (see [45]) assumes the form

$$W_2 \left( \frac{1}{N} \sum_{k=1}^N \delta_{x_k}, dx \right) \lesssim M \inf_{t > 0} \sqrt{t} + \left( \sum_{k \in \mathbb{Z}^d \setminus \{0\}} e^{-2\lambda_k t} \left| \frac{1}{N} \sum_{n=1}^N \phi_k(x_n) \right|^2 \right)^{1/2}.$$ 

For most manifolds, we do not have an explicit expression of the eigenfunctions $\phi_k$ and the inequality is thus of limited use. There exists a substitute inequality in cases where the Green’s function $G(x, y)$ or good estimates for it are known [45]. However, the Laplacian eigenfunctions are completely explicit on the sphere and are simply the classical spherical harmonics that have already been frequently used to define notions of discrepancy on the sphere (see e.g. [20, 22, 23, 24, 34]). We
believe that our notion can be a useful addition. As an example of its usefulness, we give the general version of the result above.

**Theorem 8.** Let \((M, g)\) be a compact manifold without boundary normalized to have volume 1 and let \(f : T \to \mathbb{R}\) be differentiable. Then, for some constant \(c_M > 0\) depending only on the manifold, we have

\[
\left| \int_M f(x)dx - \frac{1}{N} \sum_{k=1}^N f(x_k) \right| \leq c_M \inf_{t > 0} \left[ \sqrt{t} \| \nabla f \|_{L^\infty} + \left( \sum_{k=1}^\infty \frac{e^{-2\lambda_k t}}{\lambda_k} \right) \sqrt{t} \left( \sum_{n=1}^N \phi_k(x_n) \right)^2 \right]^{1/2} \| \nabla f \|_{L^2} \]

Alternatively, rewriting the Sobolev in terms of the spectral expansion, we could also write the integration error, equivalently,

\[
\inf_{t > 0} \left[ \sqrt{t} \| \nabla f \|_{L^\infty} + \left( \sum_{k=1}^\infty \frac{e^{-2\lambda_k t}}{\lambda_k} \right) \sqrt{t} \left( \sum_{n=1}^N \phi_k(x_n) \right)^2 \right]^{1/2} \| \nabla f \|_{L^2}.
\]

One possible application is to estimate the error of points chosen randomly with respect to the volume measure \(dx\). We observe, from \(L^2\)-orthogonality of the Laplacian eigenfunctions, that if \((x_n)_{n=1}^N\) are chosen independently at random, then

\[
\mathbb{E} \sum_{k=1}^\infty \frac{e^{-2\lambda_k t}}{\lambda_k} \left( \frac{1}{N} \sum_{n=1}^N \phi_k(x_n) \right)^2 = \sum_{k=1}^\infty \frac{e^{-2\lambda_k t}}{\lambda_k} \frac{1}{N^2} \sum_{n=1}^N \sum_{\ell=1}^N \mathbb{E} \phi_k(x_n) \phi_k(x_\ell) = \frac{1}{N} \sum_{k=1}^\infty \frac{e^{-2\lambda_k t}}{\lambda_k}
\]

Weyl’s Theorem implies that, on a compact \(d\)-dimensional manifold, \(\lambda_k \sim k^{2/d}\). For example, on \(d\)-dimensional manifolds with \(d \geq 3\), we have (using Lemma 1 from below), for \(0 < t < 1/2\),

\[
\frac{1}{N} \sum_{k \in \mathbb{Z}} \frac{e^{-2\lambda_k t}}{\lambda_k} \lesssim_M \frac{1}{N} \sum_{k=1}^\infty \frac{e^{-k^{2/d} t}}{k^{2/d}} \lesssim d^{1/2} t^{-d/2}.
\]

Minimizing in \(t\) suggests the value

\[
t^{1/2} = \frac{1}{N^{1/d}} \left( \frac{\| \nabla f \|_{L^2}}{\| \nabla f \|_{L^\infty}} \right)^{2/d}
\]

resulting in the ‘typical bound’ for random points

\[
\left| \int_M f(x)dx - \frac{1}{N} \sum_{k=1}^N f(x_k) \right| \lesssim \| \nabla f \|_{L^\infty} \| \nabla f \|_{L^2}^2 N^{-1/d}.
\]

However, this is inferior to classical Monte-Carlo and thus perhaps not very useful in applications. In some cases, when the function \(f\) is close to being constant, this estimate may be better.

3. Proofs

3.1. A recurring computation. We collect a simple Lemma that will reappear in several different arguments.
Lemma 1. We have, for $m + d \geq 1$, the estimate
\[
\sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \|k\|^m \lesssim_{m,d} t^{-m+d}.
\]
If $m + d = 0$, then we have, for $0 < t < 1/2$,
\[
\sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \|k\|^m \lesssim_{m,d} t^{-m+d} \log \left( \frac{1}{t} \right).
\]

Proof. By moving to polar coordinates noting that, for all $\ell \geq 1$,
\[
\# \left\{ k \in \mathbb{Z}^d \setminus \{0\} : \ell \leq \|k\| < \ell + 1 \right\} \leq c_d \ell^{d-1},
\]
we can reduce the sum to a one-dimensional quantity
\[
\sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \|k\|^m \lesssim_d \sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \|k\|^{m+d-1}.
\]
If $m + d = 0$, then we can easily bound the sum via
\[
\sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \|k\|^{-1} \lesssim \int_{1}^{\infty} e^{-x^2 t} \frac{dx}{x}
\]
This integral is the complete gamma function and can be rewritten in terms of the exponential integral via
\[
\int_{1}^{\infty} e^{-x^2 t} \frac{dx}{x} = -\frac{1}{2} \text{Ei}(-t) = \frac{1}{2} \int_{t}^{\infty} e^{-x} \frac{dx}{x}.
\]
It is easy to see that
\[
\int_{t}^{\infty} e^{-x} \frac{dx}{x} \lesssim \int_{t}^{1} \frac{dx}{x} + \int_{1}^{\infty} \frac{e^{-x}}{x} \frac{dx}{x} \lesssim \left( \frac{1}{t} \right).
\]
It remains to deal with the case $m + d \geq 1$ where we estimate the sum via a different integral. Note that
\[
\sum_{k \in \mathbb{Z}^d} e^{-\|k\|^2 t} \|k\|^{m+d-1} \lesssim \int_{0}^{\infty} e^{-x^2 t} x^{m+d-1} dx = c_{m+d} t^{-\frac{m+d}{2}}.
\]

3.2. Random Walks: Proof of Theorem 4.

Proof. We have that the measure $\mu_k$ describing the distribution of the random walk after $k$ steps is given by
\[
\mu_k = \mu_{k-1} \ast \mu \quad \text{where } \ast \text{ denotes convolution}
\]
and
\[
\mu = \frac{1}{2} \delta_\alpha + \frac{1}{2} \delta_{-\alpha}.
\]
Therefore
\[
|\hat{\mu}(\ell)| = |\hat{\mu}(\ell)|^k = |\cos (2\pi \ell \alpha)|^k.
\]
Using Peyre’s estimate, we reduce the problem to estimating the sum
\[ W_2(\mu_k, dx) \leq \left( \sum_{\ell \in \mathbb{Z}} \frac{|\cos(2\pi \ell \alpha)|^{2k}}{\ell^2} \right)^{1/2}. \]

We use, as we often do, that \( \ell \alpha \) cannot be close to an integer for many values of \( \ell \). More precisely, we define the \( k \) sets
\[ I_j = \left\{ \ell \in \mathbb{Z} \setminus \{0\} : \frac{j}{k} \leq \{\ell \alpha\} \leq \frac{j+1}{k} \right\} \quad \text{for} \quad 0 \leq j \leq k-1. \]

Since \( \alpha \) is badly approximable, we have that two distinct elements \( \ell_1, \ell_2 \in I_j \) satisfy \( |\ell_1 - \ell_2| \gtrsim k \).

We can now write
\[ \sum_{\ell \not\in 0} \frac{|\cos(2\pi \ell \alpha)|^{2k}}{\ell^2} = \sum_{j=0}^{k-1} \sum_{\ell \in I_j} \frac{|\cos(2\pi \ell \alpha)|^{2k}}{\ell^2}. \]

We have
\[ \sum_{\ell \in I_j} \frac{|\cos(2\pi \ell \alpha)|^{2k}}{\ell^2} \lesssim \max_{x \in I_j} |\cos(2\pi x)|^{2k} \sum_{\ell \in I_j} \frac{1}{\ell^2}. \]

However, the smallest element in \( I_j \) is \( \gtrsim \alpha \frac{k}{(j+1)} \) and any two consecutive elements are \( \gtrsim \alpha \frac{k}{j} \) separated implying that
\[ \sum_{\ell \in I_j} \frac{1}{\ell^2} \lesssim \sum_{h=0}^{\infty} \frac{1}{(k/(j+1) + hk)^2} \lesssim \left( \frac{j+1}{k} \right)^2. \]

However, we also have
\[ \max_{x \in I_j} |\cos(2\pi x)|^{2k} \leq \left( 1 - \left( \min \left\{ j, k-j \right\} \right) \right)^{2k}. \]

By symmetry, it suffices to sum \( j \) up to \( k/2 \). We then obtain
\[ \sum_{0 \leq j \leq k/2} \max_{x \in I_j} |\cos(2\pi x)|^{2k} \sum_{\ell \in I_j} \frac{1}{\ell^2} \lesssim \sum_{0 \leq j \leq k/2} \left( 1 - \frac{j^2}{k^2} \right)^{2k} \frac{(j+1)^2}{k^2} \lesssim k \int_0^1 \left( 1 - x^2 \right)^{2k} dx \lesssim \frac{1}{\sqrt{k}}. \]

\[ \square \]

3.3. Lattices. We quickly discuss the case where \( N = m^d \) with \( m \) prime and the points are arranged as a regular grid. Abbreviating
\[ \mu = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, \]
we see that the Cartesian product structure allows us to decouple different dimensions when computing the Fourier coefficients and we obtain
\[ \mu(\ell) = \frac{1}{N} \sum_{k=1}^{N} e^{2\pi i (x_k, \ell)} = \begin{cases} 1 & \text{if } m \text{ divides } \ell_i \text{ for } 1 \leq i \leq d \\ 0 & \text{otherwise.} \end{cases} \]
This shows that
\[ \mu(\ell) = \begin{cases} 
1 & \text{if } \ell \in m\mathbb{Z}^d \\
0 & \text{otherwise.} 
\end{cases} \]

While it is easy to see without any computation that
\[ W_2(\mu, dx) \lesssim d N^{-1/d}, \]
we carry out the computations to demonstrate our method. We have
\[
W_2(\mu, dx) \lesssim \inf_{t > 0} \left[ \sqrt{t} + \left( \sum_{k \in \mathbb{Z}^d \setminus \{0\}} e^{-\|k\|^2 t} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i (k, x_n)} \right|^2 \right)^{1/2} \right].
\]

The exponential sum simplifies, for \( d \geq 3 \), to
\[
\sum_{k \in \mathbb{Z}^d \setminus \{0\}} \frac{e^{-\|k\|^2 t}}{\|k\|^2} \left| \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i (k, x_n)} \right|^2 = \sum_{k \in m \mathbb{Z}^d \setminus \{0\}} \frac{e^{-\|k\|^2 (m^2 t)}}{\|k\|^2} \lesssim \frac{1}{m^2} \sum_{k \in m \mathbb{Z}^d \setminus \{0\}} \frac{e^{-\|k\|^2 (m^2 t)}}{\|k\|^2} \lesssim \frac{1}{m^2} \frac{1}{N \sqrt{t} d}.
\]

Thus
\[
W_2(\mu, dx) \lesssim \inf_{t > 0} \left[ \sqrt{t} + \frac{1}{\sqrt{N \sqrt{t}} \left( \frac{1}{d} - \frac{1}{2} \right)^{1/2}} \right] \lesssim d \frac{1}{N^{1/d}}.
\]

We observe that the same argument has a logarithmic loss for \( d = 2 \) dimensions. Since we know that no such loss is present, we see it as an indication that the \((\log N)^{1/2}\) factor present in many of our two-dimensional results might be an artefact of the method.

### 3.4. Kronecker sequences.

**Proof.** Let us consider the Kronecker sequence
\[ x_n = n\alpha \mod 1. \]

We assume that \( \alpha \) is badly approximable, which means that, for some universal constant \( c_{\alpha} > 0 \) and all integers \( q \neq 0 \), we have
\[
\max_{1 \leq j \leq d} \|\alpha_j q\| \geq \frac{c_{\alpha}}{q^{1/d}},
\]
where \( \|\cdot\| \) is the distance to the nearest integer. Khintchine’s transference principle (see, for example, the textbook of Schmidt [44]) states that \( \alpha \) is badly approximable if and only if the linear form induced by \( \alpha \) is badly approximable, i.e. if for all \( 0 \neq k \in \mathbb{Z}^d \)
\[
\|\langle k, \alpha \rangle\| \geq \frac{c_{\alpha}}{\|k\|^d},
\]
where \( \|\cdot\| \) is the distance to the nearest integer and \( c_{\alpha} \) is a universal constant. This is the property we are going to use. Observe that, abbreviating
\[
\mu = \frac{1}{N} \sum_{k=1}^{N} \delta_{x_k},
\]
then, arguing via the geometric series,

\[
|\hat{\mu}(k)| = \frac{1}{N} \left| \sum_{\ell=1}^{N} e^{2\pi i (k, x_\ell)} \right| = \frac{1}{N} \left| \sum_{\ell=1}^{N} e^{2\pi i \ell (k, \alpha)} \right| \leq \frac{2}{N} \|k\| \|\langle k, \alpha \rangle\|,
\]

where \(\|\langle k, \alpha \rangle\|\) is the distance to the nearest integer. We are left with estimating

\[
W_2(\mu, dx) \leq \inf_{t>0} \left[ \sqrt{t} + \frac{2}{N} \left( \sum_{k \neq 0} e^{-\|k\|^2 t} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^2} \right)^{1/2} \right].
\]

We split frequencies into dyadic scales and first estimate

\[
\sum_{2^\ell \leq \|k\| \leq 2^{\ell+1}} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^2}.
\]

Clearly, for any \(k_1 \neq k_2\) in this dyadic scale, we have

\[
\|\langle k_1 - k_2, \alpha \rangle\| \geq \alpha \|k_1 - k_2\| \geq 2^{-\ell d}.
\]

This means that these \(\sim 2^{\ell d}\) terms are roughly evenly spread and we have

\[
\sum_{2^\ell \leq \|k\| \leq 2^{\ell+1}} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^2} \lesssim \alpha \sum_{h=1}^{2^{d-1}} \frac{1}{(h2^{-\ell d})^2} \lesssim 2^{2\ell d}.
\]

This shows that the typical size of such a term (of which there are \(2^{\ell d}\)) is \(2^{\ell d}\) and thus we can estimate a dyadic block by increasing the multiplier as in

\[
\sum_{2^\ell \leq \|k\| \leq 2^{\ell+1}} e^{-\|k\|^2 t} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^2} \lesssim \left( \max_{2^\ell \leq \|k\| \leq 2^{\ell+1}} e^{-\|k\|^2 t} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^2} \right) \sum_{2^\ell \leq \|k\| \leq 2^{\ell+1}} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^2} \lesssim \alpha \left( \max_{2^\ell \leq \|k\| \leq 2^{\ell+1}} e^{-\|k\|^2 t} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^2} \right) \lesssim \alpha (2^{\ell d}) \sum_{2^\ell \leq \|k\| \leq 2^{\ell+1}} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^2} \lesssim 2^{\ell d}.
\]

This, in turn, can be rewritten as the kind of sum already studied above since

\[
\sum_{2^\ell \leq \|k\| \leq 2^{\ell+1}} e^{-\|k\|^2 (t/2)} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^2} \lesssim 2^{\ell d} \sum_{2^\ell \leq \|k\| \leq 2^{\ell+1}} e^{-\|k\|^2 (t/2)} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^d} \lesssim 2^{\ell d}.
\]

Altogether, using the Lemma above as well as \(d \geq 2\),

\[
\sum_{k \neq 0, k \neq 0} e^{-\|k\|^2 t} \frac{1}{\|k\|^2 \|\langle k, \alpha \rangle\|^d} \lesssim \sum_{k \neq 0, k \neq 0} e^{-\|k\|^2 t} \|k\|^{-d} \lesssim \frac{1}{t\pi_t d}.
\]

Therefore

\[
W_2(\mu, dx) \lesssim \sqrt{t} + \frac{2}{N} \frac{1}{t\pi_t d}.
\]
which implies, for the choice $t = N^{-2/d}$ that
\[ W_2(\mu, dx) \lesssim \frac{1}{N^{1/d}}. \]

3.5. **Proof of Theorem 7.** The proof is based on the following Lemma which may be of independent interest; it is a Poincaré-type statement for Lipschitz functions vanishing in a fixed point (as opposed to more classical statements where the function vanishes on the boundary or is forced to have mean value 0).

**Lemma 2.** Let $f : [0, 1]^d \to \mathbb{R}$ be differentiable and assume that $f(1/2, 1/2, \ldots, 1/2) = 0$.

Then we have the estimate
\[ \left| \int_{[0,1]^d} f(x) dx \right| \leq c_d \| \nabla f \|_{L^\infty} \frac{2^{d-1} \cdot d}{d-1} \| \nabla f \|_{L^1}, \]

**Proof.** We think of $\| \nabla f \|_{L^1}$ as a fixed quantity (the ‘budget’) and of $\| \nabla f \|_{L^\infty}$ as a constraint. Instead of integrating over the unit cube, we integrate over the unit ball $B$ centered at 0 which is also where we assume the function to vanish (for simplicity of exposition). For applications to the unit cube, it then requires us to rescale the ball by a factor only depending on the dimension and this factor only affects the constant $c_d$; henceforth, we fix $B$ to be the standard unit ball. We integrate
\[ f(x) - f(0) = \int_0^{\| x \|} \left< \nabla f \left( t \frac{x}{\| x \|} \right), \frac{x}{\| x \|} \right> dt. \]

Applying the Cauchy-Schwarz inequality and integrating over $x$, we obtain
\[ \int_B f(x) dx \lesssim \int_B \frac{| \nabla f |}{\| x \|^{d-1}} dx. \]

We maximize this upper bound instead. Clearly, it is maximized by putting the maximum weight as close to the origin as possible while preserving the $L^1$-mass. This means that, subject to our constraints, the quantity is maximized when
\[ | \nabla f | = \begin{cases} \| \nabla f \|_{L^\infty} & \text{if } \| x \| \leq \varepsilon, \\ 0 & \text{otherwise.} \end{cases} \quad \text{where } \varepsilon^{d-1} \| \nabla f \|_{L^\infty} = \| f \|_{L^1}. \]

This shows
\[ \int_B \frac{| \nabla f |}{\| x \|^{d-1}} dx \lesssim \| f \|_{L^\infty} \int_{\| x \| \leq \varepsilon} \frac{1}{\| x \|^{d-1}} dx \]
\[ \lesssim \| f \|_{L^\infty} \cdot \varepsilon \lesssim \| \nabla f \|_{L^\infty} \frac{d-1}{d-1} \| \nabla f \|_{L^1}^{\frac{d-1}{d-1}}. \]

We believe this inequality
\[ \int_B f(x) dx \lesssim \int_B \frac{| \nabla f |}{\| x \|^{d-1}} dx \]
to be quite interesting. In particular, an application of Holder’s inequality shows that, for Lipschitz functions on the $d$—dimensional ball $f : \mathbb{B} \to \mathbb{R}$ that vanish in the origin and all $p > d$

$$\int_{\mathbb{B}} f(x)dx \lesssim_{d,p} \|\nabla f\|_{L^p}.$$ 

**Proof of Theorem 7.** The proof of Theorem 7 follows easily from the Lemma which we apply, in isolation, to each fundamental cell of size $N^{-1/d}$. Rescaling the inequality in Lemma 2 then shows that for any such box $B = [0, N^{-1/d}]^d$, we have

$$\left| \int_B f(x)dx \right| \leq \frac{Cd}{N} \|\nabla f\|_{L^\infty(B)}^{d-1} \|\nabla f\|_{L^1(B)}^{d-1}.$$ 

Summing over all boxes leads to

$$\left| \int_{[0,1]^d} f(x)dx - \frac{1}{N} \sum_{k=1}^N f(x_k) \right| \lesssim \frac{\|\nabla f\|_{L^\infty}^{d-1}}{N} \sum_B \|\nabla f\|_{L^1} \left( \sum_B \|\nabla f\|_{L^1} \right)^{1/d} \left( \sum_B \frac{d-1}{d} \right)^{d-1} \frac{1}{N} \|\nabla f\|_{L^1} \|\nabla f\|_{L^1}^{1/d} N^{-1/d}.$$ 

We emphasize that the argument by itself actually yields a slightly stronger result in terms of local $L^1$—norms over $N^{-1/d}$—boxes

$$\int_{[0,1]^d} f(x)dx - \frac{1}{N} \sum_{k=1}^N f(x_k) \lesssim \frac{\|\nabla f\|_{L^\infty}^{d-1}}{N} \sum_B \|\nabla f\|_{L^1}.$$ 

**Optimality.** We quickly construct an example showing that our result is optimal. Let us consider (a slightly smoothed version of) the function

$$f(x) = \min \left\{ \varepsilon, \min_{1 \leq i \leq N} \|x - x_i\| \right\}.$$ 

We only work in the regime where $\varepsilon \ll N^{-1/d}$ in which case we see that

$$\int_{[0,1]^d} f(x)dx - \frac{1}{N} \sum_{k=1}^N f(x_k) = \int_{[0,1]^d} f(x)dx \geq \varepsilon \left( 1 - c_d N \varepsilon^d \right)$$

while also observing that

$$\|\nabla f\|_{L^\infty} = 1 \quad \text{and} \quad \|\nabla f\|_{L^1} \sim N \varepsilon^d.$$ 

By letting $\varepsilon \to 0$, we see that our estimate has the optimal exponents.
3.6. Proof of Theorem 8.

Proof. The proof combines two estimates. We first replace the discrete measure
\[ \mu = \sum_{k=1}^{N} \delta_{x_k} \]
by the smoothed measure \( e^{t \Delta} \mu \). The second step of the argument is merely a duality estimate (or, alternatively, an application of the Cauchy-Schwarz inequality). The first step is comprised of the estimate
\[ \left| \int_M f d\mu - \int_M f e^{t \Delta} \mu dx \right| \lesssim \sqrt{t} \|\nabla f\|_{L^\infty} , \]
which can be understood in at least two different ways. We describe both of them. The first case is physical: we interpret the heat equation as a process that transports a Dirac measure to a nearby neighborhood. The physical scaling is that within \( t \) units of time, the mass is transported roughly distance \( \sqrt{t} \). However, the effect of transporting mass is naturally aligned to the setting of a Lipschitz function since
\[ \left| \int_M f d\mu - \int_M f d\nu \right| \leq \|\nabla f\|_{L^\infty} W_1(\mu, \nu). \]
(This inequality becomes an equation in one dimension and is known as Monge-Kantorovich duality in that setting, see e.g. [57]). However, it is known that (see e.g. [45])
\[ W_1(\mu, e^{t \Delta} \mu) \lesssim \sqrt{t} \|\mu\| \]
and since \( \mu \) is normalized to be a probability measure, we obtain \( \|\mu\| = 1 \) and the desired estimate. The second step is more explicit. We introduce the heat kernel \( p_t(x, y) \) as the solution of the heat equation started with the measure \( \delta_x \) and run up to time \( t \) and then evaluated in \( y \). Then it follows from conservation of mass that
\[ \int_M p_t(x, y) dy = 1 \]
and the mean-value theorem implies
\[ \left| \int_M f(x) d\mu - \int_M f(x) e^{t \Delta} \mu dx \right| = \left| \int_M \frac{1}{N} \sum_{k=1}^{N} (p_t(x_k, y) f(y) - f(x_k)) dy \right| \]
\[ \leq \frac{1}{N} \sum_{k=1}^{N} \left| \int_M p_t(x_k, y) f(y) - f(x_k) dy \right| \]
\[ \leq \frac{1}{N} \sum_{k=1}^{N} \left| \int_M p_t(x_k, y) f(y) - p_t(x_k, y) f(x_k) dy \right| \]
\[ \leq \frac{1}{N} \sum_{k=1}^{N} \left| \int_M \|\nabla f\|_{L^\infty} p_t(x_k, y) |x_k - y| dy \right| \]
\[ \leq \|\nabla f\|_{L^\infty} \max_{x \in M} \int_M p_t(x_k, y) |x_k - y| dy. \]
However, the last term can be controlled using Aronson’s estimate
\[ p_t(x, y) \leq \frac{c_1}{t^{n/2}} \exp \left( - \frac{|x - y|^2}{c_2 t} \right) , \quad \forall t > 0, x, y \in M, \]
where the constant $c_1, c_2$ depend only on the manifold. A simple computation then shows (see e.g. [45]) that
\[
\int_M p_t(x_k, y)|x_k - y|dy \lesssim_M \sqrt{t}.
\]
We now come to the final part of the argument. It remains to estimate the error
\[
\left| \int_M f(x)dx - \int_M f(x)e^{t\Delta}dx \right| \quad \text{from above.}
\]
We interpret this as an inner product
\[
\left| \int_M f(x)dx - \int_M f(x)e^{t\Delta}dx \right| = \left| \langle f, e^{t\Delta} - 1 \rangle \right|.
\]
A duality argument now shows that
\[
\left| \langle f, e^{t\Delta} - 1 \rangle \right| \leq \|f\|_H \|e^{t\Delta} - 1\|_{H^{-1}}
\]
which is the desired result. One could also avoid the language of functional analysis and estimate, after noticing that $e^{t\Delta} - 1$ has mean value 0 and that $\phi_k$ has mean value 0 for $k \geq 1$,}
\[
\left| \langle f, e^{t\Delta} - 1 \rangle \right| = \left| \sum_{k=0}^{\infty} \langle f, \phi_k \rangle \langle e^{t\Delta} - 1, \phi_k \rangle \right|
\]
\[
= \sum_{k=1}^{\infty} \lambda_k^{1/2} \langle f, \phi_k \rangle \lambda_k^{-1/2} \langle e^{t\Delta} - 1, \phi_k \rangle
\]
\[
\leq \left( \sum_{k=1}^{\infty} \lambda_k \langle f, \phi_k \rangle^2 \right)^{1/2} \left( \sum_{k=1}^{\infty} \lambda_k^{-1} \langle e^{t\Delta} - 1, \phi_k \rangle^2 \right)^{1/2}
\]
\[
= \left( \sum_{k=1}^{\infty} \lambda_k \langle f, \phi_k \rangle^2 \right)^{1/2} \left( \sum_{k=1}^{\infty} \lambda_k^{-1} \langle e^{t\Delta} - 1, \phi_k \rangle^2 \right)^{1/2}.
\]
As for the first term, we observe that
\[
\sum_{k=1}^{\infty} \lambda_k \langle f, \phi_k \rangle^2 = \int_M (\Delta f)f dx = \int_M |\nabla f|^2 dx = \|\nabla f\|_{L^2}^2.
\]
As for the second sum, we observe that, using the self-adjointness of the heat propagator and the fact that $\phi_k$ is an eigenfunction of the Laplacian
\[
\langle e^{t\Delta} \mu, \phi_k \rangle = \langle e^{t\Delta} \phi_k, \mu \rangle = e^{-\lambda_k t} \langle \mu, \phi_k \rangle.
\]
This then results in
\[
\left( \sum_{k=1}^{\infty} \lambda_k^{-1} \langle e^{t\Delta} \mu, \phi_k \rangle^2 \right)^{1/2} = \left( \sum_{k=1}^{\infty} e^{-2\lambda_k t} \langle \mu, \phi_k \rangle^2 \right)^{1/2}
\]
and concludes the desired result. \qed
References

[1] M. Ajtai, J. Komlos, and G. Tusnady, On optimal matchings, Combinatorica 4 (1984), 259–264.
[2] L. Ambrosio, F. Stra and D. Trevisan, A PDE approach to a 2-dimensional matching problem. Probab. Theory Related Fields 173 (2019), no. 1-2, 433–477.
[3] D. Aronson, Aronson, Don, Non-negative solutions of linear parabolic equations, Ann. Sci. Norm. Sup. 22, p. 607–694, (1968).
[4] N.S. Bakhvalov, On the approximate calculation of multiple integrals, J. Complexity 31 (2015) 502516 [English translation; the original appeared in Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem. 4, 3–18, 1959].
[5] J. Beck, A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution. Compositio Math. 72 3, 269–339 (1989).
[6] J. Beck and W. Chen, Irregularities of Distribution, Cambridge Tracts in Mathematics (No. 89), Cambridge University Press, 1987.
[7] M. Blumlinger, Asymptotic distribution and weak convergence on compact Riemannian manifolds. Monatsh. Math. 110 (1990), no. 3-4, 177–188.
[8] L. Brown and S. Steinerberger, Positive-definite Functions, Exponential Sums and the Greedy Algorithm: a curious Phenomenon, arXiv:1908.11228.
[9] F. Bolton, A. Guillin and C. Villani, Quantitative concentration inequalities for empirical measures on non-compact spaces. Probab. Theory Related Fields 137 (2007), no. 3-4, 541–593.
[10] D. A. Burgess, The distribution of quadratic residues and non-residues. Mathematika 4 (1957), 106–112.
[11] J. Cassels, Simultaneous diophantine approximation. II. Proc. Lond. Math. Soc. (3) 5, 435–448 (1955).
[12] H. Chaix and H. Faure, Discrpeance et diaphonie en dimension un. Acta Arith. 63: 103–141, 1993.
[13] B. Chazelle, The discrepancy method. Randomness and complexity. Cambridge University Press, Cambridge, 2000.
[14] H. Davenport, Simultaneous Diophantine approximation. Mathematika 1, 51–72.
[15] J. Dick and F. Pillichshammer, Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration. Cambridge University Press, Cambridge, 2010.
[16] M. Drmota, R. Tichy, Sequences, discrepancies and applications. Lecture Notes in Mathematics, 1651. Springer-Verlag, Berlin, 1997.
[17] P. Erd˝os and P. Tur´an, On a problem in the theory of uniform distribution. I. Nederl. Akad. Wetensch. 51: 1146–1154, (1948).
[18] P. Erd˝os and P. Tur´an, On a problem in the theory of uniform distribution. II. Nederl. Akad. Wetensch. 51: 1262–1269, (1948).
[19] H. Faure, Discrepancy and diaphony of digital (0, 1)-sequences in prime bases. Acta Arith. 117: 125–148, 2005.
[20] W. Freeden. On Integral Formulas of the (Unit) Sphere and Their Applications to Numerical Computations of Integrals. Computing, 25:131146, 1980.
[21] N. Garcı Trillos, D. Slepcev, On the rate of convergence of empirical measures in \(c\)-transportation distance. Canad. J. Math. 67 (2015), no. 6, 1358–1383.
[22] P. Grabner, Erd˝os-Turan type discrepancy bounds. Monatsh. Math. 111 (1991), no. 2, 127–135.
[23] P. Grabner and R. Tichy, Spherical designs, discrepancy and numerical integration. Math. Comp. 60 (1993), no. 201, 327–336.
[24] P. Grabner, B. Klinger and R. Tichy, Discrepancies of point sequences on the sphere and numerical integration. Multivariate approximation (Witten-Bommerholz, 1996), 95–112, Math. Res., 101, Akademie Verlag, Berlin, 1997.
[25] V. S. Grozdanov, On the diaphony of one class of one-dimensional sequences. Int. J. Math and Math. Sci. 19.1: 115–124,1996.
[26] D. Hensley and F. Su, Random walks with badly approximable numbers. (English summary) Unusual applications of number theory, 95–101, DIMACS Ser. Discrete Math. Theoret. Comput. Sci., 64, Amer. Math. Soc., Providence, RI, 2004.
[27] A. Hinrichs, E. Novak, M. Ullrich and H. Wozniakowski, The curse of dimensionality for numerical integration of smooth functions. Math. Comp. 83 (2014), no. 290, 2853–2863.
[28] A. Hinrichs, E. Novak, M. Ullrich and H. Wozniakowski, The curse of dimensionality for numerical integration of smooth functions II. J. Complexity 30 (2014), no. 2, 117–143.
[29] A. Hinrichs, E. Novak, M. Ullrich and H. Wozniakowski, Product rules are optimal for numerical integration in classical smoothness spaces. J. Complexity 38 (2017), 39–49.
[30] E. Hlawka, Funktionen von beschränkter Variation in der Theorie der Gleichverteilung, Ann. Mat. Pura Appl. (4) 54 (1961), 325–333.
[31] M. Huesmann, K.-T. Sturm, Optimal transport from Lebesgue to Poisson. Ann. Probab. 41 (2013), no. 4, 2426–2478.
[32] L. Kuipers and H. Niederreiter, Uniform distribution of sequences. Pure and Applied Mathematics. Wiley-Interscience, New York-London-Sydney, 1974.
[33] W. J. Leveque, An inequality connected with Weyl’s criterion for uniform distribution. 1965 Proc. Sympos. Pure Math., Vol. VIII pp. 22–30 Amer. Math. Soc., Providence, R.I.
[34] F. Narcowich, X. Sun, J. Ward and Z. Wu, Z. LeVeque type inequalities and discrepancy estimates for minimal energy configurations on spheres. J. Approx. Theory 162 (2010), no. 6, 1256–1278.
[35] E. Novak, Some results on the complexity of numerical integration. (English summary) Monte Carlo and quasi-Monte Carlo methods, 161–183, Springer Proc. Math. Stat., 163, Springer, 2016.
[36] E. Novak and H. Wozniakowski, Tractability of multivariate problems. Vol. 1: Linear information. EMS Tracts in Mathematics, 6. European Mathematical Society (EMS), Zürich, 2008.
[37] F. Pausinger and W. Ch. Schmid, A good permutation for one-dimensional diaphony. Monte Carlo Methods Appl. 16: 307–322, 2010.
[38] R. Peyre, Comparison between $W_2$ distance and $H^{-1}$ norm, and Localization of Wasserstein distance, to appear in ESAIM: COCV 24, p. 1489 – 1501, 2018.
[39] O. Perron, ber diophantische Approximationen. Math. Ann. 83(12), 77–84 (1921)
[40] P. Proinov, On irregularities of distribution. C. R. Acad. Bulgare Sci. 39 (1986), no. 9, 31–34.
[41] P. Proinov and V. Grozdanov, On the diaphony of the van der Corput-Halton sequence. J. Number Theory 30 (1988), no. 1, 94–104.
[42] F. Santambrogio, Optimal transport for applied mathematicians. Calculus of variations, PDEs, and modeling. Progress in Nonlinear Differential Equations and their Applications, 87. Birkhäuser/Springer, Cham, 2015.
[43] W. Schmidt, On badly approximable numbers and certain games. Trans. Am. Math. Soc. 123, 178–199 (1966)
[44] W. Schmidt, Diophantine approximation. Lecture Notes in Mathematics, 785. Springer, Berlin, 1980.
[45] S. Steinerberger, Wasserstein Distance, Fourier Series and Applications. [arXiv:1809.08011]
[46] S. Steinerberger, Dynamically Defined Sequences with Small Discrepancy. [arXiv:1902.03269]
[47] S. Steinerberger, A Nonlocal Functional promoting Low-Discrepancy Point Sets, Journal of Complexity, accepted.
[48] S. Steinerberger, A Wasserstein Inequality and Minimal Green Energy on Compact Manifolds, [arXiv:1907.09023]
[49] S. Steinerberger, Directional Poincare Inequalities along Mixing Flows, Arkiv fr Matematik 54 (2016), p. 555–569.
[50] F. Su, Convergence of random walks on the circle generated by an irrational rotation. Trans. Amer. Math. Soc. 350 (1998), no. 9, 3717–3741.
[51] F. Su, A LeVeque-type lower bound for discrepancy. (English summary) Monte Carlo and quasi-Monte Carlo methods 1998 (Claremont, CA), 448–458, Springer, Berlin, 2000.
[52] F. Su, Discrepancy convergence for the drunkard’s walk on the sphere. Electron. J. Probab. 6 (2001), no. 2, 20 pp.
[53] A. G. Sukharev, Optimal numerical integration formulas for some classes of functions, Sov. Math. Dokl. 20, 472–475, 1979.
[54] M. Talagrand and J. E. Yukich, The integrability of the square exponential transportation cost. Ann. Appl. Probab. 3 (1993), no. 4, 1100–1111.
[55] M. Talagrand, Matching theorems and empirical discrepancy computations using majorizing measures, J. Amer. Math. Soc. 7 (1994), 455–537.
[56] M. Talagrand, The Transportation Cost from the Uniform Measure to the Empirical Measure in Dimension $\geq 3$, The Annals of Probability, Vol. 22, No. 2 (Apr., 1994), pp. 919–959.
[57] C. Villani, Topics in Optimal Transportation, Graduate Studies in Mathematics, American Mathematical Society, 2003.
[58] P. Zinterhof, Über einige Abschätzungen bei der Approximation von Funktionen mit Gleichverteilungsmethoden. Österreich. Akad. Wiss. Math.-Naturwiss. Kl. S.-B. II 185 (1976), no. 1-3, 121–132.
[59] P. Zinterhof and H. Stegbuchner, Trigonometrische Approximation mit Gleichverteilungsmethoden, Stud. Sci. Math. Hung. 13, p. 273–289 (1978).

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, NEW HAVEN, CT 06511, USA
E-mail address: louis.brown@yale.edu

DEPARTMENT OF MATHEMATICS, YALE UNIVERSITY, NEW HAVEN, CT 06511, USA
E-mail address: stefan.steinerberger@yale.edu