Spring 2021

Well-Distributed Sequences: Number Theory, Optimal Transport, and Potential Theory

Louis Max Brown
Yale University Graduate School of Arts and Sciences, lmb353@cornell.edu

Abstract

Well-Distributed Sequences: Number Theory, Optimal Transport, and Potential Theory

Louis Brown

2021

The purpose of this dissertation will be to examine various ways of measuring how uniformly distributed a sequence of points on compact manifolds and finite combinatorial graphs can be, providing bounds and novel explicit algorithms to pick extremely uniform points, as well as connecting disparate branches of mathematics such as Number Theory and Optimal Transport. Chapter 1 sets the stage by introducing some of the fundamental ideas and results that will be used consistently throughout the thesis: we develop and establish Weyl's Theorem, the definition of discrepancy, LeVeque's Inequality, the Erdős-Turán Inequality, Koksma-Hlawka Inequality, and Schmidt's Theorem about Irregularities of Distribution. Chapter 2 introduces the Monge-Kantorovich transport problem with special emphasis on the Benamou-Brenier Formula (from 2000) and Peyre's inequality (from 2018). Chapter 3 explores Peyre's Inequality in further depth, considering how specific bounds on the Wasserstein distance between a point measure and the uniform measure may be obtained using it, in particular in terms of the Green's function of the Laplacian on a manifold. We also show how a smoothing procedure can be applied by propagating the heat equation on probability mass in order to get stronger bounds on transport distance using well-known properties of the heat equation. In Chapter 4, we turn to the primary question of the thesis: how to select points on a space which are as uniformly distributed as possible. We consider various diverse approaches one might attempt: an ergodic approach iterating functions with good mixing properties; a dyadic approach introduced in a 1975 theorem of Kakutani on proportional splittings on intervals; and a completely novel potential theoretic approach, assigning energy to point configurations and greedily minimizing the total potential arising from pair-wise point interactions. Such energy minimization questions are certainly not new, in the static setting—physicist Thomson posed the question of how to minimize the potential of electrons on a sphere as far back as 1904. However, a *greedy* approach to uniform distribution via energy minimization is novel, particularly through the lens of Wasserstein,

and yields provably Wasserstein-optimal point sequences using the Green's function of the Laplacian as our energy function on manifolds of dimension at least 3 (with dimension 2 losing at most a square root log factor from the optimal bound). We connect this to known results from Graham, Pausinger, and Proinov regarding best possible uniform bounds on the Wasserstein 2-distance of point sequences in the unit interval. We also present many open questions and conjectures on the optimal asymptotic bounds for total energy of point configurations and the growth of the total energy function as points are added, motivated by numerical investigations that display remarkably well-behaved qualities in the dynamical system induced by greedy minimization. In Chapter 5, we consider specific point sequences and bounds on the transport distance from the point measure they generate to the uniform measure. We provide provably optimal rates for the van der Corput sequence, the Kronecker sequence, regular grids and the measures induced by quadratic residues in a field of prime order. We also prove an upper bound for higher degree monomial residues in fields of prime order, and conjecture this to be optimal. In Chapter 6, we consider numerical integration error bounds over Lipschitz functions, asking how closely we can estimate the integral of a function by averaging its values at finitely many points. This is a rather classical question that was answered completely by Bakhalov in 1959 and has since become a standard example ('the easiest case which is perfectly understood'). Somewhat surprisingly perhaps, we show that the result is not sharp and improve it in two ways: by refining the function space and by proving that these results can be true uniformly along a subsequence. These bounds refine existing results that were widely considered to be optimal, and we show the intimate connection between transport distance and integration error. Our results are new even for the classical discrete grid. In Chapter 7, we study the case of finite graphs—we show that the fundamental question underlying this thesis can also be meaningfully posed on finite graphs where it leads to a fascinating combinatorial problem. We show that the philosophy introduced in Chapter 4 can be meaningfully adapted and obtain a potential-theoretic algorithm that produces such a sequence on graphs. We show that, using spectral techniques, we are able to obtain empirically strong bounds on the 1-Wasserstein distance between measures on subsets of vertices and the uniform measure, which for graphs of large diameter are much stronger than the trivial diameter bound.

Well-Distributed Sequences: Number Theory, Optimal Transport, and Potential Theory

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by
Louis Brown

Dissertation Director: Stefan Steinerberger

June 2021

Contents

Acknowledgements **x**

1 Introduction **1**

1.1 Equidistribution 1

1.1.1 Fejér's Theorem 5

1.2 Weyl's Criterion 7

1.3 LeVeque's Inequality 11

1.3.1 Discrepancy of the van der Corput Sequence 19

1.3.2 Proinov's Diaphony Bounds 20

1.4 Erdős-Turán Inequality 21

1.4.1 Erdős-Turán Inequality on the Kronecker Sequence 24

1.4.2 Tightness of the Erdős-Turán Inequality 25

1.5 Koksma's Inequality 26

2 A Brief Survey of Optimal Transport **29**

2.1 Monge and Kantorovich 29

2.2 The Wasserstein Distance 34

2.3 The Benamou-Brenier Formula 36

2.4 Peyre's Inequality 38

2.5 Wasserstein on Graphs 39

3 Using Peyre to get Bounds **41**

3.1 A Smoothing Procedure 41

3.2 Peyre with Green's 46

4 A Dynamical Approach **53**

4.1 Background 53

4.1.1 A Naive Approach: Maximize Distance 54

4.1.2 An Ergodic Approach: The Tent Map 55

4.1.3 A Dyadic Approach: Kakutani's Theorem 56

4.1.4 A Potential Theoretic Approach: Greedy Minimization 57

4.2 A Curious Phenomenon 59

4.2.1 Introduction 60

4.2.2 Connections To Other Problems. 63

4.2.3 Known results. 64

4.2.4 Riesz points 65

4.3 Results 66

4.3.1 Wasserstein distance. 66

4.3.2 Main Results. 67

4.3.3 Two Remarks. 70

4.3.4 Higher dimensions. 71

4.4 Proofs 73

4.4.1 Proof of Theorem 4.3.1, Corollary 2 and Corollary 3. 73

4.5 Proof of Theorem 4.3.2 76

4.5.1 Proof of the Proposition 77

4.5.2 Proof of Theorem 4.3.3 79

5 Explicit Sequences of Points and their Transport Distance **80**

5.1 Introduction 80

5.1.1 Introduction 80

5.1.2 Setup. 82

5.1.3 Existing Results in One Dimension. 82

5.1.4 Existing Results in Higher Dimensions 85

5.1.5 Proof of Theorem 5.1.3: Monomial Residues 86

5.2 Main Results . . . 88
5.2.1 A Random Walk. . . . 88
5.2.2 Kronecker sequences. . . . 89
5.2.3 Other manifolds. . . . 90
5.3 Proofs . . . 92
5.3.1 A recurring computation. . . . 92
5.3.2 Random Walks: Proof of Theorem 5.2.1. . . . 93
5.3.3 Kronecker sequences: Proof of Theorem 5.2.2 . . . 95
5.3.4 A General Manifold Result: Proof of Theorem 5.2.3 . . . 97

6 Numerical Integration and Error Bounds 100
6.1 Recent Results . . . 100
6.2 A New Result: Kronecker Integration . . . 102
6.3 The Case of the Regular Grid . . . 103
6.4 Error Bounds from Kantorovich-Rubinstein . . . 104
6.5 Numerical Integration: Proof of Theorems . . . 104
6.5.1 Integration Error of Kronecker Sequences: Proof of Theorem 6.2.1 . 104
6.5.2 Proof of Theorem 6.3.1 . . . 105

7 Wasserstein on Graphs 108
7.1 Introduction . . . 108
7.1.1 Introduction. . . . 108
7.1.2 Wasserstein Distance. . . . 109
7.2 The Algorithm . . . 112
7.2.1 Setup . . . 112
7.2.2 Description of the Algorithm . . . 113
7.2.3 A Theoretical Guarantee . . . 114
7.3 Spectral Bounds on Transport Distances . . . 115
7.3.1 Motivation . . . 115
7.3.2 Spectral Bounds on Transport . . . 116
7.3.3 Applying the Theorem to obtain Transport Bounds . . . 117

7.3.4 A Case Study: Cycle Graphs 118
7.3.5 A Case Study: Torus Grid Graphs 121
7.4 Numerics 122
7.4.1 Connectivity Graph of Level 2 Menger Sponge 123
7.4.2 Truncated Tetrahedral Graph 124
7.4.3 Frucht Graph 125
7.4.4 Faulkner-Younger Graph 125
7.4.5 Erdős-Rényi Random Graphs 126
7.4.6 Complete 3-ary Tree 127
7.5 Proofs 127
7.5.1 Proof of Theorem 7.2.1 127
7.5.2 Proof of Theorem 7.3.1 128
7.6 Connection to other Results 130
7.6.1 Low-discrepancy point sets. 130
7.6.2 Leja points 130

List of Figures

1.1 The first 7 terms of the Kronecker sequence with $\alpha = \sqrt{2}$ 4

1.2 The first 8 elements of the van der Corput sequence. 5

2.1 Portrait of Gaspard Monge 30

2.2 Transporting a Pile of Dirt Into a Hole 32

2.3 Photograph of Leonid Kantorovich 33

2.4 Transport distance on a graph 40

4.1 The discrepancy of the simple algorithm on $\{0, \pi - 3, 1\}$. 54

4.2 The first 7 elements of the naive algorithm on $\{0, \pi - 3, 1\}$. 54

4.3 The Tent map on $[0, 1]$ 55

4.4 The first 1000 terms of the sequence, iterating the tent map on $\pi - 3$ 56

4.5 The discrepancy of this sequence does not decay quickly 56

4.6 D_N obtained by greedily minimizing $\cos(2\pi x)$ 57

4.7 D_N obtained by greedily minimizing $\ln|2\sin(\pi x)| - \frac{1}{5}\cos(10\pi x)$ 57

4.8 250 points running the algorithm on $\{1/3, 4/5\}$ 61

4.9 A demonstration of the algorithm 61

4.10 Total energy functions stay remarkably small in L^∞ 62

5.1 The renormalized quadratic residues in $\mathbb{F}_{29}$ rescaled to the unit interval . . 80

5.2 A distribution with small Wasserstein transportation cost 85

7.1 A demonstration of the algorithm on the truncated tetrahedral graph . . . 108

7.2 The enumeration obtained by the algorithm on the Frucht Graph 109

7.3 The Nauru Graph on 24 vertices: algorithm starting at 1. 111
7.4 The difference between $L^{-1/2}(\delta_0)$ on $\mathbb{S}^1$ and the fractional Laplacian $L^{-1/2}(\delta_0)$ on C_{150} is small. 120
7.5 The Menger Sponge, whose 400 cubes form the vertices of a connectivity graph. 123
7.6 The tightness of the bound in Theorem 7.3.1 on the Level 2 Menger Sponge Connectivity Graph 123
7.7 The sequence of vertices picked by the algorithm on the Truncated Tetrahedral Graph. 124
7.8 The sequence of vertices picked by the algorithm on the Frucht Graph. . . . 124
7.9 The first ten vertices picked by the algorithm on the Faulkner-Younger Graph. 125
7.10 A sparse Erdős-Rényi Graph from $G(100, .06)$. 127
7.11 A dense Erdős-Rényi Graph from $G(100, .2)$. 127
7.12 The 3-ary tree of depth 4 127

List of Tables

7.1 $W_1(\mu, dx)$ for the Connectivity Graph of a Level 2 Menger Sponge 124
7.2 $W_1(\mu, dx)$ for the Truncated Tetrahedral Graph 125
7.3 $W_1(\mu, dx)$ for the Frucht Graph 125
7.4 $W_1(\mu, dx)$ for the Faulkner-Younger Graph 126
7.5 $W_1(\mu, dx)$ for Erdős-Rényi Random Graphs 126
7.6 $W_1(\mu, dx)$ for the 3-ary tree of depth 4 127

Acknowledgements

I am deeply grateful to everyone who has supported me in reaching this point. To Yvette Barnard and the administrative staff at the Yale Mathematics Department, for their always friendly support in logistical issues large and small throughout the past half-decade. To my advisor Stefan Steinerberger, for his infinite patience, enthusiasm, wisdom, and countless rounds of edits. To my close friend Alex Frederick for their insights in some very enlightening conversations about Gauss sums. To my parents and sister for their constant encouragement, even when I did not believe I could do it. And finally to my perpetual partner in crime and the love of my life Lily Lu-Lerner, who provides me the strength and purpose to push through.

Chapter 1

Introduction

1.1 Equidistribution

This chapter is comprised of a survey of results in the theory of uniform distribution, we refer to [10,44,47,78] as good points of reference. In this thesis we are interested in studying point sequences which are evenly distributed in Euclidean space and manifolds; we will start in the unit interval and then explore higher dimensional spaces and general manifolds, and we will conclude this study in the setting of finite graphs. If we are told to pick N points on the unit interval which are as evenly spread out as possible, this is very easy: simply choose $0/N, \ldots, (N-1)/N$. But if we do not know N in advance—if we are tasked with constructing an infinite sequence of points which, no matter how far into the sequence we go, the points are consistently spread out evenly—this turns out to be a very difficult challenge. Before we can discuss such sequences of points, we must specify what exactly being "evenly distributed" means. Here we summarize a variety of different ways to characterize regularity of point sequences on $[0,1]$ (all of which are mutually connected), after which we present two particular sequences which perform excellently by all of these notions: the Kronecker sequence and the van der Corput sequence.

- **Combinatorial.** For every $n \in \mathbb{N}$, the set $\{x_1, \ldots, x_n\}$ has the property that for every interval $J \subset [0,1]$, the number of elements in J is $|J| \cdot n$ with a very small error.

- **Analytical** (Erdős-Turán [49,50])**.** The sequence has the property that $\{x_1, \ldots, x_n\}$

satisfy favorable exponential sum estimates on expressions of the form

$$\sum_{k=1}^{n} \frac{1}{k} \left| \sum_{\ell=1}^{n} e^{2\pi i k x_\ell} \right| \qquad \text{and} \qquad \sum_{k=1}^{n} \frac{1}{k^2} \left| \sum_{\ell=1}^{n} e^{2\pi i k x_\ell} \right|^2.$$

The exponential sum $\sum_{\ell=1}^{n} \exp(2\pi i k x_\ell)$ is 'small' for 'small' values of k.

- **Numerical** (Koksma-Hlawka [69]). The set $\{x_1, \ldots, x_n\}$ is a good set for numerical integration: we have

$$\int_0^1 f(x)dx \sim \frac{1}{n} \sum_{k=1}^{n} f(x_k)$$

with a 'small' error for 'smooth' functions f.

- **Geometric** (Roth [116]). The two-dimensional set

$$\left\{ \left(\frac{i}{n}, x_i \right) : 1 \leq i \leq n \right\} \subset [0,1]^2$$

is regularly distributed in the unit square: every cartesian box $[a,b] \times [c,d]$ contains roughly $(b-a)(d-c)n$ elements with a small error (see Fig. 3).

The problem of picking sets and sequences of points which are well-behaved with respect to these metrics has been intensively studied for over a century starting with the seminal paper of Weyl [152]. We refer to the foundational results [1, 8, 18, 20, 49, 50, 69, 116, 121], the survey paper [17] and the textbooks [10, 36, 44, 47, 78] (also with regard to various different ways of interpreting the notion of 'small' and 'smooth' in the above statements and to which extent they are connected to one another). We will begin by considering the 'Combinatorial' notion: discrepancy. But first, it will be convenient to introduce the notion of *equidistribution*, which will be key to the main content of this thesis. As the name suggests, equidistribution refers to being evenly sampled throughout a measure space. More specifically, we will call a sequence of points $x_n \in [0,1)$ *equidistributed* or *uniformly distributed* over the unit interval $[0,1)$ if the following holds for all $\alpha \in [0,1)$:

$$\lim_{N \to \infty} \frac{\#\{1 \leq i \leq N : x_n \leq \alpha\}}{N} = \alpha.$$

In other words, the sequence should, in the limit, have the "correct" portion of its points located in each interval $[0, \alpha]$, where "correct" means "equal to the measure of that interval." Of course, this also guarantees the sequence has the correct portion of its points in each interval $[\alpha, \beta]$ as well. Equidistributed sequences are particularly useful for performing numerical integration. A classic instance of an equidistributed sequence, which we will use as a recurring example throughout this chapter, is the *Kronecker sequence* $x_n \equiv n\alpha \pmod 1$, where α is *badly approximable*. Bad approximability of α means that that there is some $c > 0$ such that, for all integers p, q, we have

$$\left|\alpha - \frac{p}{q}\right| > \frac{c}{q^2} \qquad \text{or, equivalently,} \qquad |q\alpha - p| > \frac{c}{q}.$$

α is called "badly approximable" because fractions with small denominator cannot get too close to α (or, equivalently, small integer multiples of α are not too close to integers). Note that we can also think of this definition in terms of the unit circle: the distance from $q\alpha$ to the nearest integer p is proportional to the angle between the point $q\alpha \pmod 1$ of the way around the circle and $(1, 0)$. Better yet, we can replace this with the Euclidean distance between the two points

$$\left|e^{2\pi i q\alpha} - 1\right| = 2\left|\sin(\pi q\alpha)\right|.$$

Since $\sin x$ satisfies the two-sided bound $2x/\pi \leq \sin x \leq x$ for $x \in [0, \pi/2]$, our definition is equivalent to the condition that $\left|e^{2\pi i q\alpha} - 1\right| > c/q$. By default we look at the Kronecker sequence for $\alpha = \sqrt{2}$, which we quickly prove here is badly approximable: Suppose

$$\left|q\sqrt{2} - p\right| = \varepsilon \leq \frac{1}{3}$$

for some positive $p, q \in \mathbb{N}$. Since $\varepsilon \leq 1/3$ we must have $q < p$, or else we would have

$$q\sqrt{2} - p \geq p(\sqrt{2} - 1) \geq \sqrt{2} - 1 > \frac{1}{3}.$$

Similarly, since $q\sqrt{2} - p \geq -1/3$ we must have $q > p/4$. Then

$$q\sqrt{2} + p < p(\sqrt{2} + 1) < 3p.$$

Thus, multiplying out,

$$\left|2q^2 - p^2\right| = \left|(q\sqrt{2} - p)(q\sqrt{2} + p)\right| < 3p\varepsilon.$$

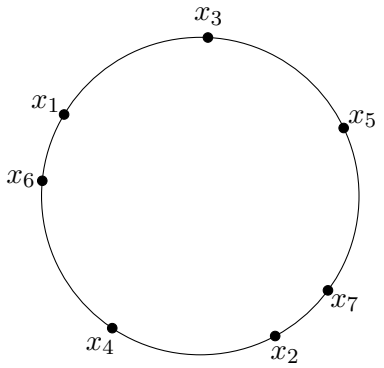

Figure 1.1: The first 7 terms of the Kronecker sequence with $\alpha = \sqrt{2}$

But the left-hand side is a positive integer and thus must be at least 1. Therefore ε is at least $(3p)^{-1} > (12q)^{-1}$ and $\sqrt{2}$ is badly approximable. There is a large body of work studying badly approximable numbers (see [102, 122, 123]). In particular, Perron explicitly constructs many badly approximable numbers and vectors, such as algebraic points on Veronese curves $(\alpha, \ldots, \alpha^n)$ for α algebraic of degree $n + 1$ and Schmidt's papers take an interesting game-theoretic approach to badly approximable numbers. The most well-studied class of badly approximable numbers are the quadratic irrationals—the argument above for $\sqrt{2}$ works just as well if 2 is replaced by any other natural number which is not a perfect square. In fact, there is a more general principle here: it is well known that a number is badly approximable precisely if the coefficients of its continued fraction expansion are bounded (see e.g. [120]). Of course, any periodic (or eventually periodic) sequence is bounded. It is also a classic theorem that a number has eventually periodic continued fraction coefficients if and only if it is a quadratic irrational—the forward direction is a result of Euler, and the converse of Lagrange. Thus, we immediately have that all quadratic irrationals are

badly approximable. It is known that the set of badly approximable numbers has Lebesgue measure 0. We refer to the seminal 1994 paper by Beck for much more on the matter [9]. While we could prove equidistribution of the Kronecker sequence directly, we will hold back until we have developed a particularly useful method for doing so in the next section. We introduce another classic example of an equidistributed sequence: the *van der Corput* sequence, where x_n is the rational number whose binary expansion, when the bimal point is removed and the bits reversed, is simply the binary expansion of n.

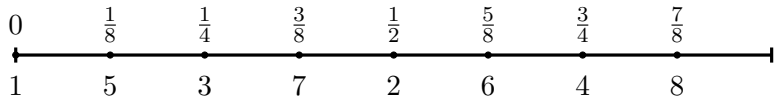

Figure 1.2: The first 8 elements of the van der Corput sequence.

Another way of thinking of the van der Corput sequence is that it greedily places a point at the midpoint of the largest gap so far. However, this interval is certainly not unique, so the question is ultimately which interval to fill in; it is not hard to see that, for poor choices of intervals, the sequence may not be equidistributed at all (see Chapter 4, §1.1). Thus, there is quite a bit of subtlety in the way van der Corput manages to pick these intervals. In fact, the van der Corput and Kronecker sequences are, in a precise sense, the *most* equidistributed sequences possible, in that they distribute uniformly at the optimal rate on all of the metrics listed at the beginning of the chapter (up to constants, see §3).

1.1.1 Fejér's Theorem

We pause here to recall a basic but important fact of Fourier analysis, which will be key to proving Weyl's Criterion in the following section: Fejér's Theorem, originally proved by Lipót Fejér in 1900. The *Fejér kernel* is defined as

$$\psi_m(x) = \sum_{|k|\leq m-1} \left(1 - \frac{|k|}{m}\right) e^{2\pi i k x}$$

For $x \notin \mathbb{Z}$, this summation can alternatively be written as

$$\psi_m(x) = \frac{1}{m}\left(\frac{\sin(m\pi x)}{\sin(\pi x)}\right)^2 = \frac{1}{m}\left(\frac{1-\cos(2\pi m x)}{1-\cos(2\pi x)}\right),$$

a fact which can be readily verified by expanding the right-hand expression(s) out in terms of trigonometric polynomials. (In fact, we may continuously extend the expression using l'Hôpital's Rule so that the equation holds for integers as well.) Notably, this means ψ_m is even and everywhere non-negative. Observe further that ψ_m has mean 1:

$$\int_0^1 \psi_m(x)dx = \widehat{\psi_m}(0) = 1.$$

We also recall the *convolution* of two integrable 1-periodic functions,

$$(f * g)(x) = \int_0^1 f(t)g(x-t)dt.$$

Convolution simply acts on the Fourier coefficients by point-wise multiplication,

$$\widehat{(f * g)}(k) = \widehat{f}(k)\widehat{g}(k).$$

In particular, convolving with the Fejér kernel yields

$$\widehat{(f * \psi_m)}(k) = \sum_{|k|\leq m-1} \left(1 - \frac{|k|}{m}\right) \widehat{f}(k)e^{2\pi ik},$$

which has a natural interpretation as the mean of the first m partial sums of the Fourier series of f. In fact, as $m \to \infty$ this converges to f.

Theorem (Fejér's Theorem [54])**.** *If $f : \mathbb{T} \to \mathbb{R}$ is continuous, then the sequence of convolved functions $f * \psi_m$ converges uniformly to f as $m \to \infty$.*

Proof: Summarized from [54]. Since f is continuous on a compact space, it is uniformly so. Thus, for any $\varepsilon > 0$, there exists $\delta \in (0, 1)$ such that, for all $x, y \in \mathbb{T}$ with $|x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$. We write

$$S_m(x) = (f * \psi_m)(x) = \frac{1}{m} \int_{\mathbb{T}} f(t) \left(\frac{1 - \cos(2\pi m(x-t))}{1 - \cos(2\pi(x-t))}\right) dt.$$

Then, for any $x \in \mathbb{T}$, we may break up the integral into two pieces: the region within δ of x, and the region that is not–call these A and B, respectively. Since f is continuous on a

compact space, it attains some maximum value $M = \|f\|_{L^\infty}$. Then we may bound

$$\left|\frac{1}{m}\int_B f(t)\left(\frac{1-\cos(2\pi m(x-t))}{1-\cos(2\pi(x-t))}\right)dt\right| \leq \frac{1}{m}\cdot\frac{2M}{1-\cos(2\pi\delta)}.$$

In particular, the right hand side does not depend on x and goes to 0 as $m \to \infty$. Thus, it suffices to consider region A. By construction, for all $y \in A$ we have $|f(x) - f(y)| < \varepsilon$. Thus, the integral over A can be bounded as

$$\left|\int_A f(t)\psi_m(x-t)dt - f(x)\int_A \psi_m(x-t)dt\right| \leq \varepsilon\int_A \psi_m(x-t)dt.$$

Finally, we bound $\int_A \psi_m(x-t)dt$. By the same argument as above, we see that

$$\lim_{m\to\infty}\int_B \psi_m(x-t)dt = 0,$$

and since ψ_m is even we can write

$$\int_A \psi_m(x-t) + \int_B \psi_m(x-t)dt = \int_{\mathbb{T}} \psi_m(x-t)dt = \int_{\mathbb{T}} \psi_m(t)dt = 1,$$

concluding the argument. □

1.2 Weyl's Criterion

Given a sequence of points $x_n \in [0,1)$, we define

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

i.e. μ_N is the distribution with point masses of equal weight $1/N$ placed at each of the first N terms of the sequence. Then we may take the Fourier transform of μ_N,

$$\widehat{\mu}_N(k) = \frac{1}{N}\sum_{n=1}^{N} e^{-2\pi i k x_n}.$$

Observe here that, since $\widehat{\mu}(k)$ is the average of numbers on the complex unit circle, we have $|\widehat{\mu}(k)| \leq 1$. We also point out that $\widehat{\mu}_N(-k) = \overline{\widehat{\mu}_N(k)}$, and consequently $|\widehat{\mu}(k)| = |\widehat{\mu}(-k)|$. Weyl's Criterion gives a beautifully simple necessary and sufficient condition for x_n to be uniformly distributed, in terms of the $\widehat{\mu}_N(k)$:

Theorem (Weyl's Criterion [152])**.** *The following are equivalent:*

1. *x_n is equidistributed.*

2. *For all properly Riemann-integrable f,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_0^1 f(x)dx.$$

3. *For all $k \neq 0$,*

$$\lim_{N\to\infty} \widehat{\mu}_N(k) = 0.$$

Proof: Summarized from [78]. **(1) $\Longrightarrow$ (2)**
Consider the characteristic function on the interval $[a, b]$

$$\mathbf{1}_{[a,b]}(x) = \begin{cases} 1 & x \in [a, b] \\ 0 & \text{else} \end{cases}.$$

We may rephrase our definition of equidistribution as follows: for all $a, b \in [0, 1)$,

$$\lim_{N\to\infty} \sum_{n=1}^{N} \frac{\mathbf{1}_{[a,b]}(x_n)}{N} = \int_0^1 \mathbf{1}_{[a,b]}(x)dx$$

since the integral is simply $b - a$. Thus, (2) holds for characteristic functions, and since the condition is linear in f it also holds for linear combinations of characteristic functions, i.e. step functions. Since every Riemann-integrable function can be arbitrarily well-approximated (in the L^1 norm) by step functions from both above and below, which in turn bound the sum on the left-hand side of (2), we obtain the desired result.

(2) $\Longrightarrow$ (3)

If we expand out condition (3) with the definition of the Fourier transform, we get

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{-2\pi i k x_n} = 0 = \int_0^1 e^{-2\pi i k x} dx$$

for all $k \neq 0$. This is precisely the equation in condition (2), with $f(x) = e^{-2\pi i k x}$.

(3) $\Longrightarrow$ (1)

By Fejér's Theorem, we can approximate any f continuous on $\mathbb{T}$ arbitrarily well (in the L^∞ norm) by trigonometric polynomials, linear combinations of the $e^{2\pi i k x}$. Then, let g be a trigonometric polynomial such that $\|f - g\|_{L^\infty} < \varepsilon$. By the triangle inequality we have

$$L = \left| \frac{1}{N} \sum_{n=1}^{N} (f(x_n) - g(x_n)) \right| \leq \frac{1}{N} \sum_{n=1}^{N} |f(x_n) - g(x_n)| < \varepsilon.$$

By (3), we have that condition (2) holds for g. So, applying another triangle inequality,

$$\begin{aligned} \left| \int_0^1 f(x)dx - \frac{1}{N} \sum_{n=1}^{N} f(x_n) \right| &\leq \left(\left| \int_0^1 (f(x) - g(x))dx \right| + \left| \int_0^1 g(x)dx - \frac{1}{N} \sum_{n=1}^{N} g(x_n) \right| + L \right) \\ &\leq \varepsilon + \left| \int_0^1 g(x)dx - \frac{1}{N} \sum_{n=1}^{N} g(x_n) \right| + \varepsilon. \end{aligned}$$

Applying the limit as $N \to \infty$ to both sides, the middle term on the right-hand side vanishes. Since we can pick g so that ε is arbitrarily small, we have that condition (2) holds for f. Finally, we may approximate the characteristic function of an interval arbitrarily well (in the L^1 norm) from both above and below by continuous functions, and so we conclude that condition (1) holds. □

Remarks.

- We may interpret condition (2) as confirmation that equidistributed sequences are good for numerical integration–if we simply average our function values on sufficiently many terms, we can get arbitrarily close to the true integral. In fact, this statement means that the μ_N *weakly converge* to dx. Note however that we have no guarantees on the speed of such a convergence (and thus, how many terms to take before we

can be satisfied with our numerical approximation of an integral). Investigating this matter will require more sophisticated machinery quantitatively measuring how far off the μ_N are from dx, which we will begin developing in the following section.

- This observation yields an intuitive interpretation of Weyl's Criterion as well: the Fourier coefficients $\widehat{dx}(k) = 0$ for all $k \neq 0$, and thus the same should be true of $\widehat{\mu}_N(k)$ in the limit. For any probability measure μ on the unit interval, we have $\widehat{\mu}(0) = 1$ by definition, so we may comfortably restrict ourselves to considering $k \neq 0$. More formally, applying Parseval's Theorem to the inner product of f with the measure μ,

$$\int_0^1 f d\mu = \sum_{k \in \mathbb{Z}} \widehat{f}(k)\widehat{\mu}(k),$$

so if we want μ to integrate arbitrary Riemann-integrable f correctly (i.e. the same way dx does,) we must have that all non-zero Fourier coefficients of μ are 0, because

$$\widehat{f}(0) = \int_0^1 f dx.$$

- Alternatively, we may view this geometrically:

$$\widehat{\mu}_N(k) = \frac{1}{N} \sum_{n=1}^{N} e^{-2\pi i k x_n}$$

is simply the average of the first N terms of the sequence, embedded on the complex unit circle by the map $x \mapsto e^{-2\pi i k x_n}$ that wraps the unit interval around it k times. For the sequence to be equidistributed, these averages must cancel nicely and converge to 0 for any fixed k.

We may now easily demonstrate the Kronecker sequence's equidistribution with nothing more than the geometric series formula and the triangle inequality:

$$|\widehat{\mu}_N(k)| = \left| \frac{1}{N} \sum_{n=1}^{N} e^{-2\pi i k n \alpha} \right| = \frac{1}{N} \left| \frac{1 - e^{-2\pi i k N \alpha}}{1 - e^{-2\pi i k \alpha}} \right| \leq \frac{2}{N} \left| \frac{1}{1 - e^{-2\pi i k \alpha}} \right|,$$

which certainly converges to 0 as $N \to \infty$ since α is fixed. Note that we have not actually

used α's bad approximability yet: all we needed here was that α is irrational, so that $1-e^{-2\pi ik\alpha} \neq 0$. Bad approximability tells us something much stronger: that $|1-e^{-2\pi ik\alpha}| > c/k$, which will be necessary to make strong quantitative bounds on *how equidistributed* the sequence is–but first, we must make precise what this means.

1.3 LeVeque's Inequality

We now introduce one particular way of measuring the regularity of a sequence. We let the *discrepancy* D_N of a sequence be

$$D_N = \sup_{\text{interval } J \subset [0,1)} \left| \frac{\#\{1 \leq n \leq N : x_i \in J\}}{N} - |J| \right|.$$

For convenience, we will slightly overload the terminology here and also refer to the term inside the supremum as the discrepancy of the *interval* J. This quantity is trivially bounded above by 1, so the supremum certainly exists. We could identify the points $0, 1$ and instead take the supremum over *arcs* of the circle, and this is in fact an equivalent definition. We can see this by noting that any arc containing 0 has as its complement an arc which does not, and thus corresponds to an interval in $[0, 1)$. Moreover, the discrepancy of an arc J and its complement J^c are equal, since

$$\left| \frac{\#\{1 \leq n \leq N : x_i \in J^c\}}{N} - |J^c| \right| = \left| \left(1 - \frac{\#\{1 \leq n \leq N : x_i \in J\}}{N}\right) - (1 - |J|) \right|.$$

As an immediate corollary of this, we have that D_N is *translation-invariant*, i.e. adding a constant α to a sequence on the unit interval (mod 1) preserves its discrepancy. With this alternate definition in mind, it is easy to see that, for any arc J, we can find an arc with discrepancy at least as large whose endpoints lie on terms in the sequence:

- If
$$\frac{\#\{1 \leq n \leq N : x_i \in J\}}{N} > |J|,$$
then we may shrink J to the largest closed arc inside it with endpoints on terms of the sequence. This arc contains the same number of terms as J, and has smaller length,

so the discrepancy will be greater.

- Similarly, if
$$\frac{\#\{1 \le n \le N : x_i \in J\}}{N} \le |J|,$$
then we may expand J to the smallest open arc containing it with endpoints on terms of the sequence. This arc contains the same number of terms as J, and has greater length, so the discrepancy will be greater.

Using the same complementing trick as before, this also tells us that, for any *interval* J, we can find an interval with discrepancy at least as large whose endpoints lie on terms of the sequence. Since there are only finitely many intervals whose endpoints are in the first N terms of the sequence, this supremum is achieved and can be simplified to a maximum. Specifically, we can compute the discrepancy as follows:

Theorem (Neiderreiter [95])**.** *For any ordered sequence of points $x_1 \le \cdots \le x_N$ in $[0, 1)$, their discrepancy is given by*

$$D_N = \frac{1}{N} + \max_{1 \le j \le N} \left(\frac{j}{N} - x_j \right) - \min_{1 \le i \le N} \left(\frac{i}{N} - x_i \right).$$

This gives us an easy way to compute the discrepancy of a sequence directly: simply add $1/N$ to the range of the function $i/N - x_i$. Further, the argmin and argmax of the function indicate which particular interval(s) achieve the discrepancy.

Proof. As noted above, the discrepancy of the sequence is always achieved as the discrepancy of a particular interval. For simplicity, we treat the unit interval as a circle, and thus can restrict ourselves to considering closed arcs between points. Then, depending on the values of i, j, we have the following two cases:

- For $i \le j$, we have
$$D_N([x_i, x_j]) = \left| \frac{j - i + 1}{N} - (x_j - x_i) \right|.$$

- For $i > j$, we have

$$\begin{aligned} D_N([x_i, x_j]) &= \left| \frac{(N+2) - (i-j+1)}{N} - (1 - (x_i - x_j)) \right| \\ &= \left| \frac{j-i+1}{N} - (x_j - x_i) \right|, \end{aligned}$$

since the arcs $[x_i, x_j]$ and $[x_j, x_i]$ together cover every point and only double count the endpoints $\{x_i, x_j\}$ (which are assumed distinct here).

Conveniently, the expression does not care about the order of i and j. So we have

$$\begin{aligned} D_N &= \max_{1 \le i,j \le N} \left| \frac{j-i+1}{N} - (x_j - x_i) \right| \\ &= \max_{1 \le i,j \le N} \left| \left(\frac{j}{N} - x_j \right) - \left(\frac{i}{N} - x_i \right) + \frac{1}{N} \right| \\ &= \frac{1}{N} + \max_{1 \le j \le N} \left(\frac{j}{N} - x_j \right) - \min_{1 \le i \le N} \left(\frac{i}{N} - x_i \right). \end{aligned}$$

□

This theorem makes it clear that the discrepancy is always at least $1/N$ (which may also be seen directly by taking an arbitrarily short interval around a point x_i), and so the question naturally arises: is it possible to construct a sequence for which the discrepancy never exceeds c/N for some constant c? The question was first answered in the negative by Tatyana van Aardenne-Ehrenfest in her 1945 paper "Proof of the Impossibility of a Just Distribution of an Infinite Sequence Over an Interval" [1]. van Aardenne-Ehrenfest proved that, for any sequence there are always infinitely many N for which

$$ND_N > c \frac{\log \log N}{\log \log \log N},$$

for some universal constant $c > 0$. Roth, in 1954 [116], strengthened this to say that, for infinitely many N and a universal constant $c > 0$, we have

$$ND_N > c\sqrt{\log N}.$$

It was not until 1972 that Wolfgang Schmidt [121] proved the sharpest form of the bound, with his seminal theorem bounding the asymptotic decay of discrepancy from below:

Theorem (Schmidt [121]). *For any sequence $\{x_n\}_{n=1}^{\infty}$ in the unit interval $[0,1]$ there are infinitely many integers N such that*

$$D_N \geq \frac{1}{100}\frac{\log N}{N}.$$

Recent work has been done exploring the optimal constant in the above theorem, we refer to [79,81]. Notably, both the van der Corput and Kronecker sequences attain this $\log N/N$ discrepancy (up to constants), and thus they are optimally equidistributed sequences in this sense. We will take a detour at the end of this section to survey some more precise, and quite fascinating, results regarding the discrepancy of the van der Corput sequence. Must any equidistributed sequence necessarily have discrepancy tending to 0 though? It is clear that if $D_N \to 0$ as $N \to \infty$ for a sequence x_n, then x_n is equidistributed; since the discrepancy is defined as the maximum discrepancy over all intervals, the maximum going to 0 also means that the discrepancy on each interval tends to 0. The converse is also true, though not nearly as immediate. In fact, it is *prima facie* a rather amazing property of equidistributed sequences: they are always uniformly so. That is, there is no way to construct a sequence such that every interval J has $|J| \cdot N + o(1)$ of the first N points in it without the supremum of the $o(1)$ errors across all intervals also being $o(1)$. Rather than proving it directly, we elect to crack this nut with a sledgehammer, and first prove a much stronger statement of which this is a direct consequence: LeVeque's inequality, which bounds D_N from above in terms of the $\widehat{\mu}(k)$. But first, we will need a lemma:

Lemma (See e.g. [78]). *Let x_n be a sequence in $[0,1)$. For $x \in [0,1]$, we define*

$$R_N(x) = \#\{1 \leq n \leq N : x_n \in [0,x)\} - Nx.$$

Then we have

$$\int_0^1 R_N(x)^2 dx = \left(\sum_{n=1}^{N}(x_n - 1/2)\right)^2 + \frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\left|N\widehat{\mu}_N(k)\right|^2.$$

Before presenting the proof, we remark that $R_N(x)/N$ is simply the signed discrepancy $\pm D_N$ of the interval $[0, x)$. It is useful to define the *star discrepancy* as

$$D_N^* = \sup_{x \in [0,1]} \left| \frac{\#\{1 \leq n \leq N : x_n < x\}}{N} - x \right| = \left\| \frac{R_N(x)}{N} \right\|_{L^\infty}.$$

By a similar calculation to the one in the Computation of Discrepancy Theorem,

$$D_N^* = \max_{1 \leq i \leq N} \max \left(\left| \frac{i}{N} - x_i \right|, \left| \frac{i-1}{N} - x_i \right| \right).$$

Certainly $D_N^* \leq D_N$, since it is a supremum over the subset of intervals anchored at 0 rather than all subintervals. In the other direction, we see that any interval can be expressed as the difference between two intervals anchored at 0, so by the triangle inequality we have $D_N \leq 2D_N^*$. Thus, while D_N^* is a less natural quantity, since it is not translation-invariant and anchoring at 0 is arbitrary, they only differ up to constants and are on the same order. As such, the L^2 computation in this lemma will be a key step in bounding D_N. This also recalls a crucial theme introduced by Weyl's Criterion: exponential sum estimates on the x_n are a good measurement of their uniformity.

Proof: Summarized from [78]. $R_N(x)$ is continuous at all but a finite set of points, and thus equals its Fourier series almost everywhere. We now compute its Fourier coefficients, starting with the case $k = 0$. Note that $\#\{1 \leq n \leq N : x_n \in [0, x)\}$ is the sum of characteristic functions of the $(x_i, 1]$, since $x_n \in [0, x) \iff x \in (x_n, 1]$. Then,

$$\widehat{R}_N(0) = \int_0^1 R_N(x)dx = \int_0^1 \sum_{n=1}^N \left(\mathbf{1}_{(x_n,1]} - x \right) dx = \sum_{n=1}^N (1/2 - x_n).$$

For $k \neq 0$, we can use integration by parts to compute

$$\begin{aligned} \widehat{R}_N(k) &= \int_0^1 R_N(x) e^{-2\pi i k x} dx = \left(\sum_{n=1}^N \int_{x_n}^1 e^{-2\pi i k x} dx \right) - N \int_0^1 x e^{-2\pi i k x} dx \\ &= \left(\sum_{n=1}^N \frac{1}{2\pi i k} (e^{-2\pi i k x_n} - 1) \right) + \frac{N}{2\pi i k} = \frac{N}{2\pi i k} \widehat{\mu}_N(k). \end{aligned}$$

Finally, by Parseval's Identity we have

$$\begin{aligned}\int_0^1 R_N(x)^2 dx &= \sum_{k\in\mathbb{Z}} \left|\widehat{R}_N(k)\right|^2 = \widehat{R}_N(0)^2 + \sum_{k\neq 0} \left|\frac{N}{2\pi i k}\widehat{\mu}_N(k)\right|^2 \\ &= \left(\sum_{n=1}^{N}(x_n - 1/2)\right)^2 + \frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\left|N\widehat{\mu}_N(k)\right|^2. \end{aligned}$$ □

With this Lemma in hand, we now state and prove LeVeque's inequality.

Theorem (LeVeque's Inequality [85])**.** *Let x_n be a sequence in $[0,1)$. Then*

$$D_N \leq \left(\frac{6}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\left|\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i k x_n}\right|^2\right)^{1/3}.$$

Proof: Summarized from [78]. Let

$$T_N(x) = \frac{R_N(x) - \widehat{R}_N(0)}{N}.$$

That is, shift R_N vertically to have mean 0, and scale down by N. Since R_N is piecewise linear with slope $-N$ and jumps of $+1$ at each x_n, T_N will be piecewise linear with slope -1 and jumps of $+1/N$ at each x_n. Since $R_N(0) = R_N(1) = 0$ (there are no x_n less than 0 and all x_n are less than 1), we also have $T_N(0) = T_N(1)$, so we may periodically extend the domain of T to all of $\mathbb{R}$, with period 1. Since T_N is mean 0 over $[0,1]$, we may fix α, β such that $T_N(\alpha) \geq 0$ and $T_N(\beta) \leq 0$, with $\alpha \in [0,1)$ and $\beta \in [\alpha, \alpha+1)$ (using periodicity of T_N). Then, since T_N only has positive jumps, it is positive on the interval $(\alpha, \alpha + T_N(\alpha))$ and in particular lies above the line $-x + T_N(\alpha) + \alpha$. Similarly, T_N is negative on the interval $(\beta + T_N(\beta), \beta)$ and lies below the line $-x + T_N(\beta) + \beta$. Thus, since the signs differ, the intervals $(\alpha, \alpha + T_N(\alpha))$ and $(\beta + T_N(\beta), \beta)$ are disjoint. So we have

$$\begin{aligned}\int_0^1 T_N(x)^2 dx &= \int_\alpha^{\alpha+1} T_N(x)^2 dx \geq \int_\alpha^{\alpha+T_N(\alpha)} T_N(x)^2 dx + \int_{\beta+T_N(\beta)}^{\beta} T_N(x)^2 dx \\ &\geq \int_\alpha^{\alpha+T_N(\alpha)} (-x + T_N(\alpha) + \alpha)^2 dx + \int_{\beta+T_N(\beta)}^{\beta} (-x + T_N(\beta) + \beta)^2 dx \\ &= \frac{T_N(\alpha)^3}{3} - \frac{T_N(\beta)^3}{3}.\end{aligned}$$

Since $x \mapsto x^3$ has positive second derivative $6x$ for positive x, its secant lines always lie above the curve (this is an application of Jensen's inequality). In particular,

$$\frac{x^3+y^3}{2} \geq \left(\frac{x+y}{2}\right)^3 \qquad \forall x, y \geq 0$$

and thus, setting $x = T_N(\alpha)$ and $y = -T_N(\beta)$ and multiplying both sides by $2/3$,

$$\frac{T_N(\alpha)^3 - T_N(\beta)^3}{3} \geq \frac{(T_N(\alpha) - T_N(\beta))^3}{12}.$$

Note that α and β can be chosen so that $T_N(\alpha) - T_N(\beta)$ is arbitrarily close to the supremum of $T_N(x) - T_N(y)$ without the sign restrictions on $T_N(x), T_N(y)$, so this inequality on $\int_0^1 T_N(x)^2 dx$ holds for all α, β. Further, for all $\alpha > \beta \in [0,1]$,

$$T_N(\alpha) - T_N(\beta) = \frac{R_N(\alpha) - R_N(\beta)}{N}$$

is simply the signed discrepancy of the interval $[\beta, \alpha)$. Clearly, swapping α and β will negate this expression, so we may apply an absolute value to the right hand side to receive the *unsigned* (positive) discrepancy on $[\beta, \alpha)$. Since this holds for all subintervals $[\beta, \alpha) \subset [0,1)$ we can take the supremum and conclude

$$\int_0^1 T_N(x)^2 dx \geq \frac{D_N^3}{12}.$$

We can also compute the left-hand side of the inequality directly, applying the Lemma:

$$\begin{aligned}
\int_0^1 T_N(x)^2 dx &= \int_0^1 \left(\frac{R_N(x) - \widehat{R}_N(0)}{N}\right)^2 dx \\
&= \frac{1}{N^2}\int_0^1 (R_N(x)^2 + \widehat{R}_N(0)^2 - 2R_N(x)\widehat{R}_N(0))dx \\
&= \frac{1}{N^2}\left(\int_0^1 R_N(x)^2 dx + \widehat{R}_N(0)^2 - 2\widehat{R}_N(0)^2\right) \\
&= \frac{1}{2\pi^2 N^2}\sum_{k=1}^{\infty}\frac{1}{k^2}\left|N\widehat{\mu}_N(k)\right|^2
\end{aligned}$$

Combining with the inequality from earlier immediately yields the desired result,

$$D_N \leq \left(\frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} \left| \widehat{\mu}_N(k) \right|^2 \right)^{1/3}. \qquad \square$$

The exponential sum expression appearing in this bound arises in many settings of interest, so it is worth naming. Let the *diaphony* [154] of a sequence of points x_n in the unit interval be defined as

$$F_N = \left(\sum_{k \neq 0} \frac{1}{k^2} \left| \hat{\mu}_N(k) \right|^2 \right)^{1/2}.$$

We note that diaphony has been studied in a variety of settings [35, 52, 63, 101, 153, 154]. F_N is always finite since, as observed before, $|\hat{\mu}_N(k)| \leq 1$, so the sum converges (and, in particular, we have $F_N \leq \pi/\sqrt{3}$). Alternatively, we may view this as the Sobolev $\dot{H}^{-1}$-norm. Diaphony does not have an immediate analogue in higher dimensional spaces: naively using the $\dot{H}^{-1}$ norm will not work since it diverges on point masses for dimension $d > 1$. One of the main contributions of this thesis is to interpret transport distance W_2 as a suitable generalization of diaphony, since it provides a lower bound in 1-dimension

$$W_2 \left(\frac{1}{N} \sum_{k=1}^{N} \delta_{x_k}, dx \right) \leq 2F_N$$

(this is Peyre's inequality) and is well-defined and finite in any dimension (see Chapter 2, §4). LeVeque's inequality gives us a quantitative bound on the discrepancy on the order of $F_N^{2/3}$, and in particular makes quite clear that discrepancy is bounded uniformly across intervals: since $|\widehat{\mu}_N(k)| \leq 1$, the bound on the right-hand side is at most 1. If x_n is equidistributed, then by Weyl's Criterion all the $\widehat{\mu}_N(k) \to 0$ as $N \to 0$, so, by the Dominated Convergence Theorem, F_N (and, in turn, D_N) goes to 0 as well. The traditional argument to directly show this fact uses compactness of the unit interval, the main advantage being that such an approach generalizes easily to higher dimensions whereas this one does not (as F_N will not even converge in dimension $d \geq 2$). If we apply LeVeque's inequality to the Kronecker

sequence, using our geometric series sum expression from earlier, we have

$$D_N \leq \left(\frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{N^2 k^2} \left| \frac{1 - e^{2\pi i k N \alpha}}{1 - e^{2\pi i k \alpha}} \right|^2 \right)^{1/3}.$$

One can do much better by using the continued fraction expansion of α (see [9, 72]).

1.3.1 Discrepancy of the van der Corput Sequence

Here we state, without proof, a variety of interesting results about the van der Corput sequence, beginning with the bound van der Corput himself gave in 1935 [146].

Theorem. *For $N \geq 1$, the van der Corput sequence satisfies*

$$ND_N^* \leq \log_2 N + 1.$$

Tijdeman then significantly improved the constant on this bound:

Theorem (Tijdeman, unpublished—see e.g. [53])**.** *For $N \geq 1$, we have*

$$ND_N^* \leq \frac{1}{3} \log_2 N + 1.$$

Béijan and Faure showed how far the discrepancy deviates from this bound:

Theorem (Béijan & Faure [11])**.** *For $N \geq 1$, $D_N = D_N^*$, and thus*

$$ND_N \leq \frac{1}{3} \log_2 N + 1.$$

Further, we have

$$\limsup_{N \to \infty} \left(ND_N - \frac{1}{3} \log_2 N \right) = \frac{4}{9} + \frac{1}{3} \log_2 3.$$

More recently (2005), Drmota-Larcher-Pillichshammer proved an amazing central limit theorem on the discrepancy of the van der Corput sequence:

Theorem (Drmota-Larcher-Pillichshammer [46])**.** *For all real y, we have*

$$\lim_{M \to \infty} \frac{1}{M} \left| \left\{ N < M : ND_N \leq \frac{1}{4} \log_2 N + y \frac{1}{4\sqrt{3}} \sqrt{\log_2 N} \right\} \right| = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-t^2/2} dt.$$

In 2017, Spiegelhofer proved a pair of theorems describing precisely how often the discrepancy of the van der Corput sequence can dip below Schmidt's $\log N/100N$ bound, improving known estimates. We combine these into a single theorem here:

Theorem (Spiegelhofer [125]). *For M sufficiently large, we have*

$$M^{.056} \leq \# \left\{ N < M : ND_N \leq \frac{\log N}{100} \right\} \leq M^{.183}.$$

It is not known what the optimal exponents in the above theorem are. We end this survey with an interesting fact regarding bit reversal. For a positive integer N, we define N^R to be the number whose binary expansion is the reverse of N's.

Theorem (Spiegelhofer [125]). *For all $N \in \mathbb{N}$, the van der Corput sequence satisfies*

$$ND_N = N^R D_{N^R}.$$

1.3.2 Proinov's Diaphony Bounds

In 1986, P. D. Proinov published a bound for the diaphony of sequences in the d-dimensional unit cube [108], but the proof was not released until a decade and a half later, in a Bulgarian-language monograph in 2000 [109]. An English translation of the proof was published by N. Kirk in 2020 [75], in a paper which both explains and expands upon Proinov's results. Proinov bounds diaphony using an L^2 variant of discrepancy, and Kirk extends these results to a dyadic variant of diaphony derived in turn from a Walsh variant of Fourier series, all of which are outside the scope of this thesis—we encourage the reader to reference [75] for these. For our sake, it will suffice to state the following lower bound on diaphony:

Theorem (Proinov [108]). *For any infinite sequence of points in the unit interval, we have for infinitely many N that*

$$F_N \geq \frac{\pi}{256\sqrt{\log 2}} \cdot \frac{\sqrt{\log N}}{N}.$$

The constant above is roughly .0147. Kirk improves this to .1619 using Proinov's techniques applied to more recent bounds on L^2 discrepancy. The optimal value of this constant (i.e. the supremum of all constants for which the theorem holds) is not known.

1.4 Erdős-Turán Inequality

We now consider another important inequality which, like LeVeque's, bounds the discrepancy of a sequence in terms of its exponential sums. In fact, we will see that LeVeque's inequality follows from the Erdős-Turán inequality up to constants, but we nonetheless believe it valuable to include the direct proof of the former as it is interesting in its own right. The proof we provide here is not the original, but a very elegant insight of Ganelius regarding Fourier series, published a decade after Erdős and Turán initially proved it [49,50].

Lemma (Ganelius [56]). *Let $V : \mathbb{R} \to \mathbb{R}$ be 1-periodic, bounded, and integrable, and set*

$$\omega(\delta) = \sup_{0 \le y-x \le \delta} (V(y) - V(x)).$$

Then, for any positive integer m, we have

$$\sup_{x \in \mathbb{R}} |V(x)| \le 4 \left(\omega(m^{-1}) + \sum_{k=0}^{m-1} \left(1 - \frac{k}{m}\right) \left|\widehat{V}(k)\right| \right).$$

There is a deliberate asymmetry in the definition of ω: it measures the maximum *increase* of V over intervals of length at most δ, not simply the maximum (absolute) *change* in value. (We will see the usefulness of this one-sidedness when applying the lemma to our setting.) Note that ω is monotonic and satisfies the triangle inequality

$$\omega(x) + \omega(y) \ge \omega(x + y),$$

and in particular, for any $n \in \mathbb{N}$, $n\omega(x) \ge \omega(nx)$. To prove the lemma, we recall Fejér's Theorem (proved above in §1.1).

Proof: Summarized from [56]. Let $\sigma_m = V * \psi_m$. Observe that, for all $\delta \in (0, 1/2)$, we have

$$\begin{aligned} \int_\delta^{1-\delta} \psi_m(x)dx &= 2\int_\delta^{1/2} \frac{1}{m} \left(\frac{\sin(m\pi x)}{\sin(\pi x)} \right)^2 dx \\ &\le \frac{2}{m} \int_\delta^{1/2} (\sin(\pi x))^{-2} = \frac{2}{m\pi} \cot(\pi\delta) \le \frac{2}{m\pi^2\delta}, \end{aligned}$$

where the last inequality follows from $\cot x \leq 1/x$ for $x \in (0, \pi)$, a consequence of the power series of cotangent. Then, since $\widehat{\psi}_m(0) = 1$, we have

$$\int_{-\delta}^{\delta} \psi_m(x)dx = 1 - \int_{\delta}^{1-\delta} \psi_m(x)dx \geq 1 - \frac{2}{m\pi^2\delta}.$$

Assume without loss of generality that $W = \sup_x |V(x)| = \sup_x V(x)$ (otherwise, simply run the following argument on $-V(-x)$: W, ω, and all the $|\widehat{V}(k)|$ are the same for $-V(-x)$). Then for any $\varepsilon > 0$, we may fix x_ε with $V(x_\varepsilon) > W - \varepsilon$, and thus, for any $x \in [x_\varepsilon - 2\delta, x_\varepsilon]$, we have, by definition of ω,

$$V(x) > V(x_\varepsilon) - \omega(2\delta) > W - \varepsilon - \omega(2\delta).$$

Then, by non-negativity of ψ_m, we may bound

$$\begin{aligned}\sigma_m(x_\varepsilon - \delta) &= \int_0^1 \psi_m(x)V(x_\varepsilon - \delta - x)dx \\ &\geq (W - \varepsilon - \omega(2\delta)) \int_{-\delta}^{\delta} \psi_m(x)dx - W \int_{\delta}^{1-\delta} \psi_m(x)dx\end{aligned}$$

We now choose $\delta = 8(m\pi^2)^{-1}$. If $m = 1$, the Lemma holds trivially, as $W \leq |\widehat{V}(0)| + \omega(1)$ (V cannot stray farther from its mean than its total range). So we may fix $m \geq 2$. Then $\delta \leq 4/\pi^2 < 1/2$, so our bounds on the integral of the Fejér kernel apply:

$$\int_{\delta}^{1-\delta} \psi_m(x)dx \leq \frac{2}{m\pi^2\delta} = \frac{1}{4} \qquad \text{and} \qquad \int_{-\delta}^{\delta} \psi_m(x)dx \geq 1 - \frac{2}{m\pi^2\delta} = \frac{3}{4}.$$

If $W < \omega(2\delta) \leq 2\omega(\delta) \leq 2\omega(m^{-1})$, then we are done immediately, so assume $W > \omega(2\delta)$ and choose $\varepsilon > 0$ with $W - \varepsilon - \omega(2\delta) \geq 0$. Then we may bound

$$\begin{aligned}\sigma_m(x_\varepsilon - \delta) &\geq (W - \varepsilon - \omega(2\delta)) \int_{-\delta}^{\delta} \psi_m(x)dx - W \int_{\delta}^{1-\delta} \psi_m(x)dx \\ &\geq \frac{3}{4}(W - \varepsilon - \omega(2\delta)) - \frac{1}{4}W = \frac{1}{2}W - \frac{3}{4}(\varepsilon + \omega(2\delta)).\end{aligned}$$

Since V is real, we have $\widehat{V}(-k) = \overline{\widehat{V}(k)}$, so

$$\begin{aligned}\sigma_m(x) &= \widehat{V}(0) + \sum_{k=1}^{m-1} (1-k/m)\left(\widehat{V}(k)e^{2\pi ikx} + \widehat{V}(-k)e^{-2\pi ikx}\right) \\ &= \widehat{V}(0) + \sum_{k=1}^{m-1} (1-k/m)\, 2\Re\left(\widehat{V}(k)e^{2\pi ikx}\right) \leq 2\sum_{k=0}^{m-1}(1-k/m)\left|\widehat{V}(k)\right|.\end{aligned}$$

Combining this with the inequality above yields

$$W \leq \frac{3}{2}(\varepsilon + \omega(2\delta)) + 4\sum_{k=0}^{m-1}(1-k/m)\left|\widehat{V}(k)\right|.$$

Since $16/\pi^2 < 2$, we have $\omega(2\delta) = \omega(16(m\pi^2)^{-1}) \leq 2\omega(m^{-1})$. Sending $\varepsilon \to 0$,

$$W \leq 3\omega(\delta) + 4\sum_{k=0}^{m-1}(1-k/m)\left|\widehat{V}(k)\right|. \qquad \square$$

The Erdős-Turán inequality [49, 50] now follows as an immediate corollary of Ganelius' Lemma: Setting $V(x) = -T_N(x)$, we see that V has a constant slope of $+1$ except at the terms of the sequence where it jumps by $-1/N$, so $\omega(x) \leq x$. Thus,

$$\sup_x |-T_N(x)| \leq 4\left(\omega(m^{-1}) + \sum_{k=0}^{m-1}(1-k/n)\left|\widehat{T}(k)\right|\right) \leq 4\left(\frac{1}{m} + \frac{1}{2\pi}\sum_{k=1}^{m-1}\frac{1}{k}\left|\widehat{\mu}_N(k)\right|\right),$$

recalling our computations of $\widehat{R}(k)$ from earlier (and in particular that $\widehat{R}(0) = 0$). As we saw in the proof of LeVeque's inequality,

$$D_N = \sup_{x,y}(T_N(y) - T_N(x)).$$

and thus, by the triangle inequality, we arrive at the Erdős-Turán inequality

$$D_N \leq 8\left(\frac{1}{m} + \frac{1}{2\pi}\sum_{k=1}^{m-1}\frac{1}{k}\left|\widehat{\mu}_N(k)\right|\right).$$

To get LeVeque's inequality (up to constants) from this, simply set $m = \lceil F_N^{-2/3} \rceil$:

$$D_N \lesssim F_N^{2/3} + \sum_{k=1}^{m-1} \frac{1}{k} |\widehat{\mu}_N(k)| \lesssim F_N^{2/3} + F_N \cdot F_N^{-1/3} \lesssim F_N^{2/3},$$

where the second inequality follows from applying the Cauchy-Schwartz inequality to the sequences $\frac{1}{k}|\widehat{\mu}_N(k)|$ and $\mathbf{1}_{[1,m-1]}$ (the sequence of $(m-1)$ 1s and 0s after that). Here and throughout the text we use $A \lesssim B$ to denote that $A \leq cB$ for some universal constant c–this is a useful shorthand that allows us to dispense of constants in favor of notational compactness, particularly when we are really interested in asymptotic scaling anyway. Indeed, we will often discard the 8 and $(2\pi)^{-1}$ shown above for a cleaner inequality, which is true up to constants. For an investigation of precisely which constants may replace the 8 and $(2\pi)^{-1}$ to keep the inequality strictly true, we refer to the Rivat-Tenenbaum paper on the matter [115]. Since LeVeque's inequality is, up to constants, merely a special case of the Erdős-Turán inequality, the latter is, again up to constants, the tighter bound of the two asymptotically. For instance, if we look at the Kronecker sequence on the golden ratio $\alpha = (1+\sqrt{5})/2$, the Erdős-Turán bound gives $D_N \lesssim (\log N)^2/N$ (see the following subsection) while LeVeque's inequality only yields $D_N \lesssim N^{-2/3}$. We refer to [93] for further exposition on these bounds as well as more examples displaying the relative weakness of LeVeque's inequality compared with the Erdős-Turán inequality.

1.4.1 Erdős-Turán Inequality on the Kronecker Sequence

For our running example of the Kronecker sequence, the Erdős-Turán inequality yields

$$D_N \leq 8\left(\frac{1}{m} + \frac{1}{2\pi}\sum_{k=1}^{m-1} \frac{|1-e^{2\pi i N k\alpha}|}{kN|1-e^{2\pi ik\alpha}|}\right).$$

By way of a standard dyadic decomposition trick, we see that, since α is badly approximable, $k|1-e^{2\pi ik\alpha}| \gtrsim 1$, and thus the set

$$\{|1-e^{2\pi ik\alpha}| : 2^\ell \leq k \leq 2^{\ell+1}\}$$

is $\sim 2^{-\ell}$ separated, and, for every element, we have $|1 - e^{2\pi ik\alpha}| \gtrsim 2^{-\ell}$. Thus,

$$D_N \lesssim \frac{1}{m} + \frac{1}{N} \sum_{\ell=0}^{\log m} \sum_{k=2^\ell}^{2^{\ell+1}} \frac{1}{k} \left|1 - e^{2\pi ik\alpha}\right|^{-1} \lesssim \frac{1}{m} + \frac{1}{N} \sum_{\ell=0}^{\log m} 2^{-\ell} \sum_{k=2^\ell}^{2^{\ell+1}} \left|1 - e^{2\pi ik\alpha}\right|^{-1}$$
$$\lesssim \frac{1}{m} + \frac{1}{N} \sum_{\ell=0}^{\log m} 2^{-\ell} \sum_{j=1}^{2^\ell} (j/2^{-\ell})^{-1} \lesssim \frac{1}{m} + \frac{1}{N} \sum_{\ell=0}^{\log m} \ell \sim \frac{1}{m} + \frac{(\log m)^2}{N}.$$

Setting $m = N$, we get a bound on the order of $(\log N)^2/N$. This is not the optimal rate however—Beck [9] has stronger bounds on the discrepancy of the Kronecker sequence.

1.4.2 Tightness of the Erdős-Turán Inequality

After considering the Erdős-Turán inequality bounding the size of discrepancy from above, it is natural to ask the question "How good of a bound is this?" That is, how close does the Erdős-Turán bound get to actual discrepancy? Imre Rusza proved that the discrepancy can never be significantly smaller than the Erdős-Turán bound.

Theorem (Rusza [117])**.** *For a sequence of points in the unit interval, let*

$$B = \inf_{m \in \mathbb{N}} \left(\frac{1}{m} + \sum_{k=1}^{m-1} \frac{1}{k} \left|\widehat{\mu}_N(k)\right| \right)$$

be the minimal bound given by the Erdős-Turán inequality. Then we have

$$D_N \gtrsim B^{3/2}.$$

Proof: Summarized from [117]. As in §3, we set

$$R_N(x) = \#\{1 \le n \le N : x_n \in [0, x)\} - Nx.$$

Then we have the following chain of inequalities:

$$D_N \ge D_N^* = \left\| \frac{R_N}{N} \right\|_{L^\infty} \ge \left\| \frac{R_N}{N} \right\|_{L^2}.$$

By the §3 Lemma, we have

$$\left\| \frac{R_N}{N} \right\|_{L^2}^2 \geq \frac{1}{2\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} |\widehat{\mu}_N(k)|^2 = \frac{1}{4\pi^2} F_N^2.$$

Combining with the above, we find

$$F_N^2 \leq 4\pi^2 D_N^2.$$

Then, through an application of Cauchy-Schwarz along the same lines as when we deduced LeVeque's inequality from the Erdős-Turán inequality, we have

$$S = \sum_{k=1}^{m-1} \frac{1}{k} |\widehat{\mu}_N(k)| \leq \pi \sqrt{2(m-1)} D_N.$$

This in turn proves that

$$B \leq \frac{1}{m} + S \leq \frac{1}{m} + \pi \sqrt{2(m-1)} D_N.$$

Setting $m = 1 + \lfloor (\pi \sqrt{2} D_N)^{-2/3} \rfloor$ yields the desired result:

$$B \leq 3 \cdot 2^{-1/3} (\pi D_N)^{2/3} < 6 D_N^{2/3}. \qquad \square$$

Indeed, Rusza constructs examples showing that this bound is, up to constants, optimal.

1.5 Koksma's Inequality

Weyl's Criterion tells us that we should only sample over equidistributed sequences for numerical integration, but of course there is massive variety within the class of equidistributed sequences. In practice it is not enough to know that a sequence converges if we have no idea how "quickly" it does so–maybe we need to take a trillion terms of the sequence before being less than 1000 away from the true integral. For our sequence to be useful in practice, we need pragmatic theoretical bounds which we can actually compute on how far these averages are from $\int_0^1 f dx$. Koksma's inequality [77] provides such an error bound for us.

But before we present Koksma's inequality, we need a lemma.

Lemma (See e.g. [78]). *For any sequence $0 = x_0 \leq x_1 \leq \cdots \leq x_N < x_{N+1} = 1$, and any function $f : [0,1] \to \mathbb{R}$ of bounded variation, we have*

$$\frac{1}{N}\sum_{i=1}^{N} f(x_i) - \int_0^1 f(x)dx = \sum_{i=0}^{N} \int_{x_i}^{x_{i+1}} \left(t - \frac{i}{N}\right) df(t),$$

where the integral on the right-hand side is the Riemann-Stieltjes integral.

Proof: Summarized from [78]. This follows from Abel summation and integration by parts:

$$\begin{aligned}
\sum_{i=0}^{N} \int_{x_i}^{x_{i+1}} \left(t - \frac{i}{N}\right) df(t) &= \int_0^1 t df(t) - \sum_{i=0}^{N} \frac{i}{N}(f(x_{i+1}) - f(x_i)) \\
&= \Big(tf(t)\Big)_0^1 - \int_0^1 f(t)dt + \frac{1}{N}\sum_{i=0}^{N-1} f(x_{i+1}) - f(1) \\
&= \frac{1}{N}\sum_{i=1}^{N} f(x_i) - \int_0^1 f(t)dt. \qquad \square
\end{aligned}$$

From this, we may easily arrive at Koksma's inequality:

Theorem (Koksma [77]). *For any sequence x_i in $[0,1)$, and any function $f : [0,1] \to \mathbb{R}$ with total variation $V(f) < \infty$, we have*

$$\left|\frac{1}{N}\sum_{i=1}^{N} f(x_i) - \int_0^1 f(t)dt\right| \leq V(f) D_N^*.$$

Koksma's inequality gives us a bound which is both useful and easy to compute: as long as we can bound the discrepancy of a sequence and the total variation of f, we have a guarantee on how close the average value of f over x_i is to the true integral. In particular, for a fixed f the bound is simply proportional to the discrepancy of the sequence.

Proof: Summarized from [78]. Without loss of generality, we may assume the first N x_i are ordered (as all our quantities are independent of the order of the x_i). Note that, for $t \in [x_i, x_{i+1}]$, we have

$$\left(t - \frac{i}{N}\right) \leq \max\left(\left|x_i - \frac{i}{N}\right|, \left|x_{i+1} - \frac{i}{N}\right|\right) \leq D_N^*,$$

since D_N^* is the maximum of the middle expression over all i (we can ignore the edge cases since $x_0 - 0/N = x_{N+1} - N/N = 0$). Then, since f has bounded variation, we may apply the lemma, and get

$$\begin{aligned}\left|\frac{1}{N}\sum_{i=1}^{N} f(x_i) - \int_0^1 f(t)dt\right| &= \sum_{i=0}^{N}\int_{x_i}^{x_{i+1}} \left(t - \frac{i}{N}\right) df(t) \\ &\leq \sum_{i=0}^{N}\int_{x_i}^{x_{i+1}} D_N^* df(t) \leq V(f)D_N^*. \end{aligned}$$ □

One particularly nice application of Koksma's inequality is an immediate bound on the size of the $\widehat{\mu}_N(k)$: we have

$$|\widehat{\mu}_N(k)| = \frac{1}{N}\sum_{i=1}^{N} e^{2\pi i(kx_i+\theta)} = \frac{1}{N}\sum_{i=1}^{N}\cos(2\pi(kx_i+\theta)),$$

where $\theta = \arg\widehat{\mu}_N(k)$ is chosen to rotate the exponential sum onto the real axis. Then, since $\cos(2\pi(kx+\theta))$ has total variation $V(f) = 4k$ (all k bumps go up and down by 2),

$$|\widehat{\mu}_N(k)| = \left|\widehat{\mu}_N(k) - \int_0^1 \cos(2\pi(kx+\theta))dx\right| \leq 4kD_N^*.$$

We can also apply our Erdős-Turán bound on D_N for the Kronecker sequence to bound the error of numerical integration on it:

$$\left|\frac{1}{N}\sum_{i=1}^{N} f(x_i) - \int_0^1 f(t)dt\right| \lesssim \frac{(\log N)^2}{N}.$$

We can interpret Koksma's inequality as a quantitative specification of Weyl's Criterion: if $D_N \to 0$, then numerical integrals over x_i approach true integrals, in particular so do the $\widehat{\mu}_N(k)$, and here is precisely how fast those limits converge in terms of D_N itself.

Chapter 2

A Brief Survey of Optimal Transport

2.1 Monge and Kantorovich

In Chapter 1, we have explored the use of discrepancy (the literal difference between a point measure and the Lebesgue on an interval) as a way to quantify the equidistribution of a sequence. In 1781, French geometer Gaspard Monge famously asked a more concrete question: if probability mass were physical mass (say, sand in piles) then how much work is needed to push one distribution of mass to another? Of course, there may be varying types of terrain which pose obstructions to travel in differing degrees, so for the problem to be well-specified we need a cost function $c(x, y)$ that tells us the amount of work needed to take one unit of mass from point x to point y (we will always assume the cost scales linearly with the mass: carrying two units of mass from x to y is twice as much work as carrying one between the same points). In his original formulation, Monge took the cost function to simply be the euclidean distance between x and y: transporting m mass over d distance costs $m \cdot d$. This special case is now referred to as the "Earth mover distance," recalling Monge's initial paper exploring the transfer of soil extracted from known sites in the ground to aboveground structures being built from it. Formally, if we have a starting measure μ which we would like to transport to a target measure ν (both on the same space

M), the plan to do so will be a *coupling* γ of μ and ν: a measure on $M \times M$ with marginals μ and ν, respectively. That is, for each pair of points in M, we indicate how much mass to move from one to the other, and the probability distribution induced by projecting γ onto the first coordinate should match μ while the probability distribution induced by projecting γ onto the second coordinate should match ν.

Figure 2.1: Portrait of Monge by François Delpech, Public domain, via Wikimedia Commons

We denote the collection of couplings on μ, ν as $\Gamma(\mu, \nu)$. Since we obviously want to minimize the amount of work done, the ultimate number we care about is the solution to the *Monge-Kantorovich minimization problem*

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{M \times M} c(x, y) d\gamma(x, y).$$

Monge's Problem was slightly narrower than this: he was interested specifically in *deterministic* transport plans, induced by functions $M \to M$ (where the mass of any point is entirely sent to another point, and no mass is "split"). It is understandable why this restriction might be imposed, from a physical perspective: if individuals are stationed at each point in tractors holding their respective piles of sand, each entire pile must be transported to the same place (it would cost much more time and effort to have to measure out the

appropriate portions to drop at each destination). Unfortunately, this is, in general, not possible to do, e.g. if μ has a single point mass, it can only be transported to another single point mass (though there are plenty of examples where deterministic transport is impossible despite both μ, ν being continuous). For our purposes, we will restrict to the case where the probability space is a Riemannian manifold (M, g). Then, we have the following theorem which will be immensely useful:

Theorem (Existence of Optimal Couplings, see e.g. [148])**.** *For any Riemannian manifold (M, g) and continuous cost function $c : M \times M \to \mathbb{R}^+$, and any pair μ, ν of probability distributions on M, there exists a coupling $\gamma \in \Gamma(\mu, \nu)$ which minimizes the transport cost. In particular, the infimum in the Monge-Kantorovich problem is attained as a minimum.*

The proof is a topological argument invoking the compactness of $\Gamma(\mu, \nu)$. Much stronger variants of this theorem are true, with fewer restrictions on the cost function and underlying space, but it will be sufficient for our purposes. (In general, the cost function need not even be non-negative, depending on the particular setting and application. Indeed, the transport itself need not be restricted to a transfer on a single space M.) While we can certainly talk about the transport of measures that are not probability distributions, the question only makes sense if the total mass of μ and ν is equal, so without loss of generality, as long as they are positive finite measures we may normalize them to be probability measures. In general, the problem of finding optimal couplings is quite hard, though there are settings in which it is not: for two distributions on $\mathbb{R}$ with cost function the standard euclidean distance, the solution is very intuitive. We may imagine the probability density function of the initial measure as aboveground piles of dirt and of the target measure as holes below ground, and it is clear that the optimal transport plan consists of simply bulldozing the dirt forward while allowing it to fall in the holes. Around two centuries after Monge formulated the problem, Kantorovich approached it from the perspective of the theory of linear programming (which he had invented, though it was later independently formulated by Koopmans, Dantzig, and others [41]). He thus proposed we consider instead the dual problem of finding

$$\sup \left\{ \int_M \phi(x) d\mu(x) + \int_M \psi(y) d\nu(y) : \phi(x) + \psi(y) \leq c(x, y) \right\},$$

where the supremum is taken over L^1 functions ϕ, ψ.

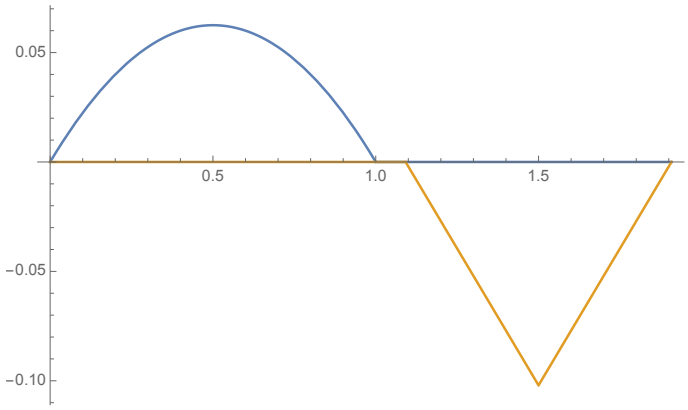

Figure 2.2: Here, it is easy to see how to optimally transport the lump of dirt to the hole

We can think of the dual problem with a real-life example as follows: Suppose we are discussing our soil transport question when an entrepreneurial stranger appears. They inform us that they overheard our conversation and they just so happen to run a soil transport company, and will offer their services with the following payment scheme: we pay some upon soil pickup and some upon soil drop-off, and they take care of everything in between. Specifically, they produce two functions: $\phi(x)$ is the price to pick up one unit of soil mass at location x, and $\psi(y)$ is the price to drop off one unit of soil mass at location y (picking up and dropping off are qualitatively different operations, they need not cost the same amount even at the same location). They, somehow, already know our cost function c and, while they do not (yet) know the precise distributions of soil we need transported, they guarantee us that ϕ and ψ are such that, for all $(x, y) \in M \times M$, $\phi(x) + \psi(y) \leq c(x, y)$, so the price we would pay them to pick up from x and drop off at y is no more than we would pay to move it ourselves. (They may have techniques, connections, and machinery that nonetheless make *their* cost function less than ours, so it could still be profitable for them as well.) Kantorovich's dual problem asks the maximum amount we would have to pay the entrepreneur. The answer: no less than it would cost us to move the soil ourselves!

Theorem (Kantorovich Duality [74])**.** *For any Riemannian manifold (M, g) and continuous cost function $c : M \times M \to \mathbb{R}^+$, and any pair μ, ν of probability distributions on M, the Monge-Kantorovich problem and the dual Kantorovich problem have the same solution.*

Moreover, the supremum in the dual problem is attained as a maximum. That is, there exists a pair of functions ϕ, ψ satisfying $\phi(x) + \psi(y) \leq c(x, y)$ and $\gamma \in \Gamma(\mu, \nu)$, with

$$\int_M \phi(x) d\mu(x) + \int_M \psi(y) d\nu(y) = \int_{M \times M} c(x, y) d\gamma(x, y).$$

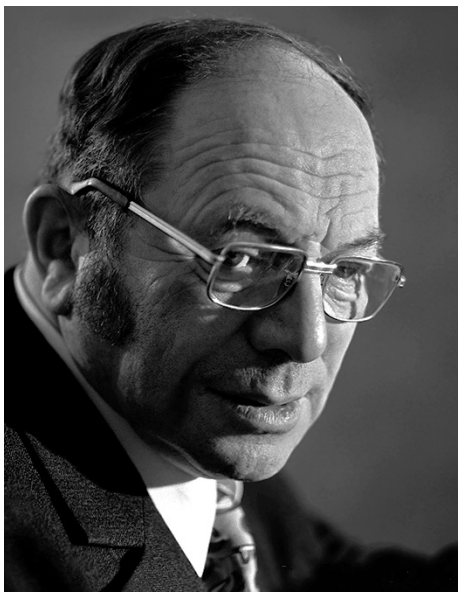

Figure 2.3: Kantorovich, courtesy Andrei Bogdanoff, CC BY 3.0, via Wikimedia Commons

In the special case when c is a metric, we can further make the statement that the optimal transport cost is

$$\sup \left\{ \int_M \phi d\mu - \int_M \phi d\nu \right\},$$

where ϕ is 1-Lipschitz (i.e., $\phi(x) - \phi(y) \leq d(x, y)$ for all $x, y \in M$). This equation is referred to as the Kantorovich-Rubinstein formula, and is easy to verify using Kantorovich duality: for any 1-Lipschitz ϕ simply set $\psi = -\phi$. Then, since ϕ is 1-Lipschitz by assumption,

$$\phi(x) + \psi(y) = \phi(x) - \phi(y) \leq d(x, y),$$

and thus ϕ, ψ are a permissible pair, and we have

$$\int_M \phi d\mu + \int_M \psi d\nu = \int_M \phi d\mu - \int_M \phi d\nu.$$

Conversely, given any permissible ϕ, ψ pair, we define

$$f(x) = \inf_y \left(d(x,y) - \psi(y) \right).$$

Note that f is 1-Lipschitz: we have

$$f(x) - f(x') = \inf_y (d(x,y) - \psi(y)) - \inf_y (d(x',y) - \psi(y)) \leq \sup_y (d(x,y) - d(x',y)) \leq d(x,x')$$

by the triangle inequality. By construction f satisfies

$$\phi(x) \leq f(x) \leq -\psi(x),$$

where the upper bound is realized by setting $y = x$. Thus,

$$\int_M f d\mu - \int_M f d\nu \geq \int_M \phi d\mu + \int_M \psi d\nu.$$

So the suprema of the two expressions is equal. In practice, this alternate form is often easier to use as it ostensibly only needs a single function ϕ to be optimized instead of a ϕ, ψ pair (though the proof provides a recipe to cook either up from the other).

2.2 The Wasserstein Distance

We now turn our attention to the Wasserstein distance, for which Santamborigo [119] or Villani [148] are great points of reference. The $p-$Wasserstein distance [147] between two measures μ and ν is defined as

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

where d is the metric on M and $p \in [1, \infty)$. That is, W_p^p is simply the optimal transport cost when $c = d^p$. A particularly natural case arises from $p = 1$, when we recover the Earth mover distance studied by Monge, also known as the Kantorovich-Rubinstein distance. As noted in the previous section, W_1 admits a particularly nice duality formula (though

Kantorovich duality exists for all p):

$$W_1(\mu,\nu) = \sup\left\{\int_M \phi d\mu - \int_M \phi d\nu\right\},$$

with the supremum taken over all 1-Lipschitz ϕ. The p-Wasserstein distance can be thought of analogously to the L^p-norm. One can easily verify that W_p satisfies the axioms for a metric, inheriting all the relevant properties from d. The argument for triangle inequality uses Minkowski's inequality and thus follows closely to the proof that L^p is a norm. Since we are working on compact spaces, we can always bound W_p from above by the diameter of the space (the maximum distance between two points). It is relatively easy to see that we have an (optimal) lower bound on the Wasserstein distance from a point distribution to the uniform distribution that is independent of the pointset $\{x_1,\ldots,x_N\} \subset M = \mathbb{T}^d$:

$$W_p\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \geq \frac{c_d}{N^{1/d}}.$$

The short argument is as follows: let $B_{\varepsilon N^{-1/d}}(x)$ be a ball of radius $\varepsilon N^{-1/d}$ centered around x. We note that the total volume of N such balls satisfies

$$\left|\bigcup_{i=1}^{N} B_{\varepsilon N^{-1/d}}(x_i)\right| \leq N\left|B_{\varepsilon N^{-1/d}}(0)\right| \leq \omega_d \varepsilon^d$$

for some universal constant ω_d depending only on the dimension. For ε sufficiently small (depending only on ω_d), this is much less than the volume of M and therefore most of the Lebesgue measure on M is at distance $> \varepsilon N^{-1/d}$ from the set $\{x_1,\ldots,x_N\}$. In the case where $N = n^d$, we can achieve optimal scaling on the constant c_d using a lattice configuration, picking the x_i to be the centers of d-dimensional hypercubes with side length $1/n$ which partition M. This is because, by Stirling's approximation, the volume of a d-dimensional unit ball is $\omega_d \sim d^{-(d+1)/2}$, so, to make the above argument work, we need

$$\varepsilon \lesssim d^{(d+1)/2d} \sim \sqrt{d},$$

and thus our lower bound becomes

$$\frac{c_d}{N^{1/d}} \sim \frac{\sqrt{d}}{n}.$$

On the other hand, if we transport the point measure on each x_i to the Lebesgue measure on its corresponding box, then, since the probability mass must travel at most $\sqrt{d}/n$ (the diameter of each box), we have

$$W_p\left(\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}, dx\right) \leq \frac{\sqrt{d}}{n},$$

as desired. The Wasserstein distance fully encodes the geometry of M: We can easily recover the metric on the space from W_p, for any p. Simply observe that $W_p(\delta_x, \delta_y) = d(x, y)$. From Hölder's inequality, we immediately deduce that $W_p \leq W_q$ for all $p \leq q$: Let $\gamma \in \Gamma(\mu, \nu)$. Then, applying Hölder's inequality on q/p and its Hölder conjugate k to d^p and 1, we have

$$\int_{M\times M} d^p \cdot 1 d\gamma \leq \left(\int_{M\times M} (d^p)^{q/p} d\gamma\right)^{p/q} \left(\int_{M\times M} 1^k d\gamma\right)^{1/k},$$

so, simplifying and taking the pth root of both sides,

$$\left(\int_{M\times M} d^p d\gamma\right)^{1/p} \leq \left(\int_{M\times M} d^q d\gamma\right)^{1/q},$$

and thus all couplings γ will have smaller total transport cost in W_p than W_q.

2.3 The Benamou-Brenier Formula

Just two decades ago, Jean-David Benamou and Yann Brenier approached Wasserstein distance from the perspective of computational fluid dynamics. In such context, it is natural to formulate the transport question continuously, asking not merely "How much mass should be sent from point x to point y?" but "How does it get there?" To this end, they ask not for a coupling of μ, ν, but a pair $\rho_t(x)$ and $v(t, x)$ of time-dependent density and velocity fields, respectively. This is particularly useful for applied settings, where we may be interested to

know what measure is "half-way between" μ and ν as they are being transported. With this viewpoint, they prove the following:

Theorem (Benamou-Brenier Formula [14])**.** *For any Riemannian manifold (M, g) and pair μ, μ' of positive measures on M, we have*

$$W_2(\mu, \mu') = \inf \left\{ \int_0^1 \int_M |v|^2 d\rho_t dt \right\},$$

where the infimum is taken over all ρ_t, v satisfying $\rho_0 = \mu, \rho_1 = \mu'$, and

$$\frac{\partial \rho_t}{\partial t} + \nabla \cdot (v\rho_t) = 0.$$

The first constraint merely says that ρ indeed transports from μ to μ', and the second is a continuity constraint stipulating that probability mass be locally conserved (i.e., no "teleporting" mass). The Benamou-Brenier Formula can also be phrased in terms of negative Sobolev norms, eliminating v from the picture altogether—we briefly introduce them here. The weighted Sobolev norm on signed measures is defined by

$$\|\nu\|_{\dot{H}^{-1}(\mu)} = \sup \left\{ \langle f, \nu \rangle \;\middle|\; \int_M |\nabla f|^2 \, d\mu \leq 1 \right\}.$$

Clearly, this expression is infinite if ν has non-zero total mass—we can simply set f to be an arbitrarily large constant function. In general, as long as ν has finite total mass we may shift it down by an appropriate multiple of the Lebesgue measure to attain zero mass. When μ is the Lebesgue measure, there is a nice spectral expression for this:

$$\|\nu\|^2_{\dot{H}^{-1}} = \sum_{k=1}^{\infty} \frac{\langle \nu, \phi_k \rangle^2}{\lambda_k},$$

where ϕ_k, λ_k are the eigenfunctions and eigenvalues of the Laplacian on M. Note that in the 1-dimensional setting this is the diaphony F_N, introduced in Chapter 1 §3. Using this notation, we may restate the Benamou-Brenier Formula as follows:

$$W_2(\mu, \mu') = \inf \left\{ \int_0^1 \|d\rho_t\|_{\dot{H}^{-1}(d\rho_t)} \right\}.$$

This infimum only needs the constraint that $\rho_0 = \mu, \rho_1 = \mu'$, since the negative Sobolev norm is itself an optimization. Thus, in this perspective, the Benamou-Brenier Formula can be understood to say that, asymptotically, W_2 agrees with $\dot{H}^{-1}$:

$$W_2(\mu + d\mu) = \|d\mu\|_{\dot{H}^{-1}(\mu)} + o(d\mu)$$

for a small measure $d\mu$ on M. It is then natural to ask for some non-asymptotic variant of this: can we bound W_2 by $\dot{H}^{-1}$?

2.4 Peyre's Inequality

Rémi Peyre, in 2018, answered this question positively:

Theorem (Peyre's Inequality [106]). *For any pair μ, ν of positive measures on a Riemannian manifold M, we have*

$$W_2(\mu, \nu) \leq 2\|\mu - \nu\|_{\dot{H}^{-1}(\mu)}.$$

In particular, when ν is the Lebesgue measure on M (as will be the case of interest for our setting, since we ultimately want to use this to consider transport cost between the uniform distribution and a point distribution), this is simply $W_2(\mu, dx) \leq 2\|\mu - dx\|_{\dot{H}^{-1}}$. We prove this by way of the following lemma:

Lemma (Peyre [106]). *For measures μ, μ' with $\mu' \geq \rho\mu$ for some $\rho > 0$, we have for all ν*

$$\|\nu\|_{\dot{H}^{-1}(\mu)} \geq \rho^{1/2}\|\nu\|_{\dot{H}^{-1}(\mu')}.$$

Proof: Summarized from [106]. Note that, since the integrand is everywhere non-negative,

$$\int_M |\nabla f|^2 d\mu' \geq \rho \int_M |\nabla f|^2 d\mu.$$

Thus, bringing ρ inside the integrand, for all f we have

$$\int_M |\nabla f|^2 d\mu' \leq 1 \implies \int_M |\nabla \rho^{1/2} f|^2 d\mu \leq 1,$$

and since $\langle \rho^{1/2} f, \nu \rangle = \rho^{1/2} \langle f, \nu \rangle$, the result follows. □

We may now prove Peyre's inequality using this and the Benamou-Brenier Formula:

Proof: Summarized from [106]. For $t \in [0,1]$, let

$$\mu_t = (1-t)\mu + t\nu,$$

so that $\mu_0 = \mu$ and $\mu_1 = \nu$. Then we have by the Benamou-Brenier Formula

$$W_2(\mu,\nu) \leq \int_0^1 \|\mu - \nu\|_{\dot{H}^{-1}(\mu_t)} dt$$

(since W_2 is the infimum of this expression over all such ρ_t interpolating μ, ν). Since $\nu \geq 0$, we have $\mu_t \geq (1-t)\mu$, and so, applying the lemma,

$$\|\mu - \nu\|_{\dot{H}^{-1}(\mu_t)} \leq (1-t)^{-1/2} \|\mu - \nu\|_{\dot{H}^{-1}(\mu)}.$$

Thus, we may integrate $\int_0^1 (1-t)^{-1/2} dt = \left[-2(1-t)^{1/2}\right]_0^1 = 2$, and so

$$W_2(\mu,\nu) \leq \int_0^1 (1-t)^{-1/2} \|\mu - \nu\|_{\dot{H}^{-1}(\mu)} dr = 2\|\mu - \nu\|_{\dot{H}^{-1}(\mu)}.$$ □

Ledoux provided the following generalization of Peyre's result not requiring $p = 2$:

Theorem (Ledoux [82])**.** *For any $1 \leq p < \infty$, if ν, μ satisfy $d\nu = f d\mu$ then we have*

$$W_p(\nu,\mu) \leq p\|f - 1\|_{H^{-1,p}(\mu)}.$$

2.5 Wasserstein on Graphs

The notion of Wasserstein distance can be immediately applied to finite graphs: we will work only in the simple case of unweighted graphs where all edges have the same weight and the distance between two vertices is given as the length of the shortest path connecting them; extensions to the weighted case (i.e. arbitrary metric graphs) are certainly conceivable. In

short, transporting ε L^1-mass over a single edge has a W_1 cost of ε. We will mostly be interested in looking at W_1 on the graph because of its particularly nice properties.

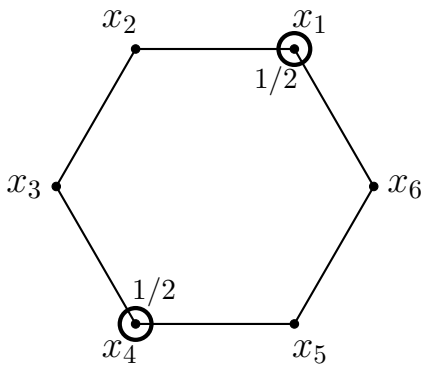

Figure 2.4: $W_1\left(\frac{1}{2}(\delta_{x_1}+\delta_{x_2}), dx\right) = 2/3$: We can transport $1/6$ units of mass from x_1 to each of x_2 and x_6, and similarly with x_4 to x_3 and x_5, incurring a total cost of $4 \times 1/6$.

Even though we are merely applying Wasserstein distance to a finite set of points, the fact that the metric is induced by a graph supplies an abundance of structure that allows us to make strong claims. For instance, we may apply Kantorovich-Rubinstein as follows:

Proposition 2.5.1 (Kantorovich-Rubinstein [105])**.** *Let $G = (V, E)$ be a finite, simple graph, let $f : V \to \mathbb{R}$ and let $W \subset V$ be a subset of vertices. Then*

$$\left| \frac{1}{|V|} \sum_{x \in V} f(x) - \frac{1}{|W|} \sum_{x \in W} f(x) \right| \leq W_1 \left(\frac{1}{|W|} \sum_{x \in W} \delta_x, dx \right) \max_{x_i \sim x_j} |f(x_i) - f(x_j)|.$$

Proof. The Kantorovich-Rubinstein formula tells us that, for any pair of measures μ_i,

$$W_1(\mu_1, \mu_2) = \max_{\text{non-constant } f} \left(\left| \int_V f(x) d\mu_1 - \int_V f(x) d\mu_2 \right| \left[\max_{x_i \sim x_j} |f(x_i) - f(x_j)| \right]^{-1} \right),$$

where the factor on the right simply scales down f to ensure it is 1-Lipschitz. Applying this to the case where $\mu_1 = dx$ and $\mu_2 = \frac{1}{|W|} \sum_{x \in W} \delta_x$, we arrive at the desired result. □

(In the context of graphs, we use dx to denote the uniform measure on the vertex set V.) This result can be interpreted as a bound on the error of numerical integration: averaging f over its values on W can only be so far off from the global average of f over V. We will look more closely at the graph setting in Chapter 7, particularly as it relates to sampling and numerical integration, and investigate the relationship between results on the graph and results on manifolds.

Chapter 3

Using Peyre to get Bounds

3.1 A Smoothing Procedure

This section collects some existing machinery and summarizes results from Steinerberger [127, 131], note Bobkov and Ledoux independently used very similar techniques in [22]. In particular it also contains a theorem formulated by Steinerberger and the author in [25]. Suppose we have a measure μ we would like to transport to the uniform measure dx—how can we get a good bound on the W_2 transport cost required to do this? The main idea of this section is to apply the heat equation to smooth the measure and then interpret the outcome ν of this smoothing process as the result of applying a particular transport plan moving μ to ν. We achieve this through a reinterpretation of parabolic second order differential equations, more specifically the heat equation

$$\left(\frac{\partial}{\partial t} - \Delta\right) u(t,x) = 0.$$

Observe that if $u(0,x) \geq 0$, we may interpret this process physically as representing the dynamics of particle density as particles diffuse over time (Fourier originally formulated it to model temperature flow, hence the name). Using the spectral expansion we see that high-frequency eigenfunctions, and linear combinations thereof, quickly decay under the heat equation: within a short amount of time, the solution $u(t,x)$ will be close to constant. On the other hand, if we run the equation for a sufficiently short time, then most of the

particles are only moving a little bit. Let (M, g) denote a compact, smooth Riemannian manifold without boundary. We denote the L^2-normalized eigenfunctions of the Laplacian as $(\phi_k)_{k=0}^{\infty}$ (where, as usual, $\phi_0 = 1$ is constant–for simplicity we normalize M to have volume 1). The argument is based on an explicit construction comprised of two steps: we first use the heat kernel as a way to organize transport to achieve a smoothened distribution of mass that is very close to flat, and then we apply Peyre's inequality

$$W_2(\mu, \nu) \leq 2\|\mu - \nu\|_{\dot{H}^{-1}(\mu)},$$

introduced in Chapter 2. The heat kernel $p(t, x, y) : \mathbb{R}_{\geq 0} \times M \times M \rightarrow \mathbb{R}_{\geq 0}$ satisfies

$$\int_M p(t, x, y) dy = 1.$$

In particular, it may be understood as a probability distribution. We re-interpret it as a transport plan telling us how to spread mass located at x. The result of this transport plan will be a new probability distribution given by

$$e^{t\Delta}\mu = \int_M p(t, x, y) d\mu(y) \quad \text{at } W_2\text{-cost} \quad \left(\int_M \int_M |x - y|^2 p(t, x, y) d\mu(y) dx \right)^{\frac{1}{p}}.$$

If we can bound the above cost, we may apply the triangle inequality to estimate

$$W_2(\mu, dx) \leq W_2(\mu, e^{t\Delta}\mu) + W_2(e^{t\Delta}\mu, dx).$$

Here, $e^{t\Delta}\mu$ is the solution of the heat equation after t units of time (since we work on a manifold without boundary, we do not need to specify boundary conditions). Fortunately, this transportation cost $W_2(\mu, e^{t\Delta}\mu)$ is easy to bound:

Lemma (See e.g. [57, 127, 131])**.** *Let μ be a probability measure on the compact manifold M. Then we have for all $t > 0$*

$$W_2(\mu, e^{t\Delta}\mu) \lesssim_M \sqrt{t},$$

where the implicit constant depends only on the manifold.

Note that there is nothing special about W_2 in this result–the same proof works just as well for all W_p, $p \in [1, \infty)$. We merely state is this way for simplicity because we are using it as part of a larger argument to bound W_2.

Proof: Summarized from [127]. We apply the heat equation for a short time to μ. We interpret the heat equation as convolution with the heat kernel and the heat kernel as a transport plan. The result of this transport plan will be a new mass distribution given by

$$\nu = \int_M e^{t\Delta}\delta_x(y)d\mu(y) \quad \text{at } W_2^2\text{-cost} \quad \int_M \int_M |x-y|^2 e^{t\Delta}\delta_x(y)d\mu(y)dx.$$

To bound this, we use a classical bound of Aronson [4, 86],

$$e^{t\Delta}\delta_x(y) \leq \frac{c_1}{t^{d/2}} \exp\left(-\frac{|x-y|^2}{c_2 t}\right),$$

where $d = \dim(M)$, and c_1, c_2 are constants depending only on (M, g), and obtain

$$\int_M \int_M |x-y|^2 e^{t\Delta}\delta_x(y)d\mu(y)dx \lesssim_{(M,g)} \int_M \int_M \frac{|x-y|^2}{t^{d/2}} \exp\left(-\frac{|x-y|^2}{c_2 t}\right) d\mu(y)dx.$$

However, it is easily seen that for some universal constants depending on the manifold

$$\int_M \frac{|x-y|^2}{t^{d/2}} \exp\left(-\frac{|x-y|^2}{c_2 t}\right) dx \lesssim_{c_2,(M,g)} t.$$

Altogether, since $c_2 \lesssim_{(M,g)} 1$, this implies the squared W_2 cost is bounded by

$$\int_M \int_M |x-y|^2 e^{t\Delta}\delta_x(y)d\mu(y)dx \lesssim_{(M,g)} t\|\mu\|_{L^1} = t. \qquad \square$$

With this smoothing trick and bound in hand, we return to our main problem of approximating the uniform measure with point measures, where we can now prove the following:

Theorem 3.1.1 (B. & Steinerberger [25]). *For any set of points* $\{x_1, \dots, x_k\}$ *on* $M = \mathbb{T}^d$,

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

Before explaining the proof, we note 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 the size of these exponential sums is available for many sequences.

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 useful generalization of Zinterhof's diaphony.

Recall that F_N becomes meaningless in dimensions $d \geq 2$ because Dirac deltas are no longer contained in the Sobolev space $\dot{H}^{-1}$ (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 [55] or Grabner, Klinger & Tichy [60]). In contrast, we can rewrite our inequality (even on general manifolds) as

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

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 Chapter 6.

Proof of Theorem 3.1.1: Summarized from [25]. We abbreviate

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

and first use the triangle inequality

$$W_2(\mu, dx) \leq W_2(\mu, e^{t\Delta}\mu) + W_2(e^{t\Delta}\mu, dx),$$

where $e^{t\Delta}$ is the solution of the heat equation at time t. We can define this operator spectrally via

$$e^{t\Delta}\delta_x = \sum_{k\in\mathbb{Z}^d} e^{-\|k\|^2 t} e^{2\pi i\langle k,x\rangle}\phi_k,$$

where $\phi_k(y) = e^{2\pi i\langle k,y\rangle}$ are the Laplacian eigenfunctions of M, with associated eigenvalue $\|k\|^2$. Then, applying Peyre's inequality, we have

$$W_2(e^{t\Delta}\mu, dx) \lesssim \left\|e^{t\Delta}\mu - dx\right\|_{\dot{H}^{-1}},$$

which we may expand spectrally, omitting the constant term $dx = \phi_0$ (which vanishes by orthogonality of eigenfunctions), and applying linearity of the heat operator $e^{t\Delta}$,

$$\|e^{t\Delta}\mu - dx\|^2_{\dot{H}^{-1}} = \sum_{\substack{k\in\mathbb{Z}^d\\ k\neq 0}} \frac{\left\langle e^{t\Delta}\mu, \phi_k\right\rangle^2}{\|k\|^2} = \sum_{\substack{k\in\mathbb{Z}^d\\ k\neq 0}} \frac{e^{-2\|k\|^2 t}}{\|k\|^2}\left|\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i\langle k,x_n\rangle}\right|^2.$$

Finally, note that, since

$$W_2(\mu, dx) \leq W_2(\mu, e^{t\Delta}\mu) + W_2(e^{t\Delta}\mu, dx),$$

we may square both sides to

$$\begin{aligned} W_2^2(\mu, dx) &\leq W_2^2(\mu, e^{t\Delta}\mu) + W_2^2(e^{t\Delta}\mu, dx) + 2W_2(\mu, e^{t\Delta}\mu)W_2(e^{t\Delta}\mu, dx) \\ &\lesssim W_2^2(\mu, dx) + W_2^2(\mu, e^{t\Delta}\mu), \end{aligned}$$

since the cross term is, up to constants, dominated by the others. Putting together all our transport bounds and rescaling t to absorb the 2 in $e^{-2\|k\|^2 t}$, we arrive at the result. □

3.2 Peyre with Green's

The Green's function of the Laplacian, defined spectrally as

$$G(x,y) = \sum_{k=1}^{\infty} \frac{\phi_k(x)\phi_k(y)}{\lambda_k},$$

is a kernel with mean value 0 and the property that

$$-\Delta_x \int_M G(x,y) f(y) dy = f(x),$$

i.e. it solves the equation $-\Delta u = f$. It scales approximately like $|x-y|^{2-d}$ in dimension $d \geq 3$ and $-\log|x-y|$ for $d = 2$. We refer to [5] for a good introduction to Green's function. Closed form expressions for Green's function are only known for certain very restricted instances. Nonetheless, there are circumstances where we either have an explicit description of it or define a point sequence using it (as in the next chapter), so it is useful to have an analogue of the inequality in the previous section in terms of the Green's function:

Theorem (Steinerberger [127]). *Let M be a smooth, compact $d-$dimensional manifold without boundary, $d \geq 3$, and let $G : M \times M \to \mathbb{R} \cup \{\infty\}$ denote the Green's function of the Laplacian. Then, for any set of n points $\{x_1, \dots, x_n\} \subset M$, we have*

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{x_k}, dx\right) \lesssim_M \frac{1}{n^{1/d}} + \frac{1}{n}\left|\sum_{k \neq \ell} G(x_k, x_\ell)\right|^{1/2}.$$

If the manifold is two-dimensional, $d = 2$, then we have

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{x_k}, dx\right) \lesssim_M \frac{\sqrt{\log n}}{n^{1/2}} + \frac{1}{n}\left|\sum_{k \neq \ell} G(x_k, x_\ell)\right|^{1/2}.$$

We will prove the case $d \geq 3$, taking note of what needs to be changed when $d = 2$.

Proof: Summarized from [127]. We abbreviate

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

and first use the triangle inequality

$$W_2(\mu, dx) \leq W_2(\mu, e^{t\Delta}\mu) + W_2(e^{t\Delta}\mu, dx),$$

where $e^{t\Delta}$ is the solution of the heat equation at time t. As before, we have

$$e^{t\Delta} f = \sum_{k=0}^{\infty} e^{-\lambda_k t} \langle f, \phi_k \rangle \phi_k,$$

where ϕ_k denotes the sequence of Laplacian eigenfunctions on the manifold, i.e. $-\Delta\phi_k = \lambda_k\phi_k$ (normalized to have $\|\phi_k\|_{L^2} = 1$). We have already seen in the Lemma above that $W_2(\mu, e^{t\Delta}\mu) \lesssim \sqrt{t}$. For the second term, we invoke Peyre's inequality

$$W_2(e^{t\Delta}\mu, dx) \lesssim \left\| e^{t\Delta}\mu - dx \right\|_{\dot{H}^{-1}},$$

where $\dot{H}^{-1}$ is a Sobolev space whose norm can be defined spectrally via

$$\|f\|^2_{\dot{H}^{-1}} = \sum_{k=1}^{\infty} \frac{\langle f, \phi_k \rangle^2}{\lambda_k}.$$

We note that the Green function is defined in a similar way, i.e. spectrally via

$$\int_M G(x,y) f(y) dy = \sum_{k=1}^{\infty} \frac{\langle f, \phi_k \rangle}{\lambda_k} \phi_k(x).$$

As a consequence, we have that

$$\int_{M\times M} G(x,y) f(x) f(y) dx dy = \left\langle \int_M G(x,y) f(y) dy, f(x) \right\rangle = \sum_{k=1}^{\infty} \frac{\langle f, \phi_k \rangle^2}{\lambda_k} = \|f\|^2_{\dot{H}^{-1}}.$$

We note that the heat equation and the Green function are both spectral multipliers and

thus, whenever $s_1 + t_1 = s_2 + t_2$ and all four numbers are positive,

$$\begin{aligned}\int_M \int_M G(x,y) e^{s_1 \Delta} f(x) e^{t_1 \Delta} g(y) dx dy &= \sum_{k=1}^{\infty} e^{-s_1 \lambda_k} \frac{\langle f, \phi_k \rangle \langle g, \phi_k \rangle}{\lambda_k} e^{-t_1 \lambda_k} \\ &= \sum_{k=1}^{\infty} e^{-s_2 \lambda_k} \frac{\langle f, \phi_k \rangle \langle g, \phi_k \rangle}{\lambda_k} e^{-t_2 \lambda_k} \\ &= \int_M \int_M G(x,y) e^{s_2 \Delta} f(x) e^{t_2 \Delta} g(y) dx dy.\end{aligned}$$

We can now write

$$\begin{aligned}\left\| e^{t\Delta}\mu - dx \right\|_{\dot{H}^{-1}}^2 &= \int_M \int_M G(x,y) e^{t\Delta}\mu(x) e^{t\Delta}\mu(y) dx dy \\ &= \frac{1}{N^2} \sum_{k,\ell} \int_M \int_M G(x,y) e^{t\Delta}\delta_{x_k}(x) e^{t\Delta}\delta_{x_\ell}(y) dx dy \\ &= \frac{1}{N^2} \sum_{k} \int_M \int_M G(x,y) e^{t\Delta}\delta_{x_k}(x) e^{t\Delta}\delta_{x_k}(y) dx dy \\ &+ \frac{1}{N^2} \sum_{k \neq \ell} \int_M \int_M G(x,y) e^{t\Delta}\delta_{x_k}(x) e^{t\Delta}\delta_{x_\ell}(y) dx dy.\end{aligned}$$

We use the self-adjointness of spectral multipliers (and both convolution with G as well as the heat kernel are spectral multipliers, moreover the heat kernel is a semigroup and $e^{t\Delta} e^{t\Delta} = e^{2t\Delta}$) and rewrite the first term as

$$\begin{aligned}\int_M \int_M G(x,y) e^{t\Delta}\delta_{x_k}(x) e^{t\Delta}\delta_{x_k}(y) dx dy &= \int_M \int_M G(x,y) \delta_{x_k}(x) e^{2t\Delta}\delta_{x_k}(y) dx dy \\ &= \int_M G(x_k,y) e^{2t\Delta}\delta_{x_k}(y) dy\end{aligned}$$

We have a very good understanding of the heat kernel since

$$e^{2t\Delta}\delta_{x_k}(y) \lesssim \begin{cases} t^{-d/2} & \text{if } d_g(x,y) \leq \sqrt{t} \\ \text{exponentially decaying} & \text{otherwise.} \end{cases}$$

More formally, we use a classical bound of Aronson [4, 86] for the heat kernel on manifolds

stating that for constants c_1, c_2 depending on the manifold

$$e^{t\Delta}\delta_x(y) \leq \frac{c_1}{t^{d/2}} \exp\left(-\frac{|x-y|^2}{c_2 t}\right),$$

where $|x-y| = d_g(x,y)$ denotes the geodesic distance (we will use this notation henceforth). We can couple this with the estimate (see e.g. Aubin [6])

$$G(x,y) \lesssim_M \frac{1}{|x-y|^{d-2}}$$

and obtain

$$\int_M G(x_k,y)e^{2t\Delta}\delta_{x_k}(y)dy \lesssim \int_{\mathbb{R}^n} \frac{e^{2t\Delta}\delta_0(x)}{|x|^{d-2}}dx \lesssim \int_0^\infty \frac{c_1}{t^{d/2}} \frac{\exp\left(-\frac{r^2}{c_2 t}\right)}{r^{n-2}} r^{n-1}dr \lesssim t^{1-\frac{d}{2}}.$$

This implies

$$\frac{1}{N^2}\sum_k \int_M \int_M G(x,y)e^{t\Delta}\delta_{x_k}(x)e^{t\Delta}\delta_{x_k}(y)dxdy \lesssim \frac{t^{1-\frac{d}{2}}}{N}.$$

It remains to bound the second term. We can again use the fact that Fourier multipliers commute to argue that

$$\begin{aligned}\int_M \int_M G(x,y)e^{t\Delta}\delta_{x_k}(x)e^{t\Delta}\delta_{x_\ell}(y)dxdy &= \int_M \int_M G(x,y)\delta_{x_k}(x)e^{2t\Delta}\delta_{x_\ell}(y)dxdy \\ &= \int_M G(x_k,y)e^{2t\Delta}\delta_{x_\ell}(y)dy.\end{aligned}$$

We understand this value for t very small since the Green function is integrable and

$$\lim_{t\to 0}\int_M G(x_k,y)e^{2t\Delta}\delta_{x_\ell}(y)dy = G(x_k,x_\ell)$$

and will now control the variation in time. We note that if $x_k = x_\ell$ for some $k \neq \ell$, then our upper bound is infinity/undefined and the entire statement is vacuously true. We can thus assume $x_k \neq x_\ell$ for $k \neq \ell$. The heat kernel solves the heat equation and thus

$$\frac{\partial}{\partial t}e^{t\Delta}\delta_{x_\ell}(y) = \Delta_y e^{t\Delta}\delta_{x_\ell}(y)$$

which we use in combination with

$$\Delta_y G(x,y) = 1 - \delta_x \qquad \text{and} \qquad \int_M e^{2t\Delta}\delta_{x_\ell}(y)dy = 1,$$

where the first identity is in the sense of distributions (and will be used paired against a smooth function). We write

$$\begin{aligned}\frac{1}{2}\frac{\partial}{\partial t}\int_M G(x_k,y)e^{2t\Delta}\delta_{x_\ell}(y)dy &= \int_M G(x_k,y)\Delta_y e^{2t\Delta}\delta_{x_\ell}(y)dy = \int_M \Delta_y G(x_k,y)e^{2t\Delta}\delta_{x_\ell}(y)dy \\ &= \int_M \left(1-\delta_{x_k}\right)e^{2t\Delta}\delta_{x_\ell}(y)dy = 1 - \left(e^{2t\Delta}\delta_{x_\ell}\right)(x_k) \leq 1\end{aligned}$$

This implies

$$\frac{1}{N^2}\sum_{k\neq \ell}\int_M\int_M G(x,y)e^{t\Delta}\delta_{x_k}(x)e^{t\Delta}\delta_{x_\ell}(y)dxdy \leq 2t + \frac{1}{N^2}\sum_{k\neq\ell} G(x_k,x_\ell).$$

Altogether, collecting all the estimates, we have

$$W_2(\mu,dx) \lesssim W_2(\mu,e^{t\Delta}\mu) + W_2(e^{t\Delta}\mu,dx) \lesssim \sqrt{t} + \left(\frac{t^{1-\frac{d}{2}}}{N} + 2t + \frac{1}{N^2}\sum_{k\neq\ell} G(x_k,x_\ell)\right)^{1/2}.$$

Setting $t = N^{-2/d}$ results in

$$W_2\left(\frac{1}{N}\sum_{k=1}^N \delta_{x_k}, dx\right) \lesssim_M \frac{1}{N^{1/d}} + \frac{1}{N}\left|\sum_{k\neq\ell} G(x_k,x_\ell)\right|^{1/2}.$$

We observe that the Green energy may actually be negative and, at first glance, it may look like one could get improved results. However, we will show in the proof of the corollary that

$$\frac{1}{N}\left|\sum_{k\neq\ell} G(x_k,x_\ell)\right|^{1/2} \gtrsim_M -N^{-1/d}$$

and thus our application of the triangle inequality in this form is not lossy. We quickly note

the necessary changes for the case $d = 2$. We observe that in that case

$$G(x, y) \lesssim |\log |x - y||$$

and thus, for the first term,

$$\begin{aligned}\int_M G(x_k, y) e^{2t\Delta} \delta_{x_k}(y) dy &\lesssim \int_{\mathbb{R}^n} |\log |x|| e^{2t\Delta} \delta_0(x) dx \\ &\lesssim \int_0^\infty \frac{c_1}{t} |\log r| \exp\left(-\frac{r^2}{c_2 t}\right) r dr \lesssim \log(1/t).\end{aligned}$$

The off-diagonal term behaves exactly as before and we obtain

$$W_2(\mu, dx) \lesssim_M \sqrt{t} + \left(\frac{\log(1/t)}{N} + t + \frac{1}{N^2} \sum_{k \neq \ell} G(x_k, x_\ell) \right)^{1/2}.$$

Setting $t = 1/N$ results in the desired statement. □

The result is sharp for $d \geq 3$ and sharp up to possibly the factor of $\sqrt{\log n}$ in $d = 2$. One way to see this is by computing asymptotics on the Green energy which follows as a byproduct from our approach (this Corollary can be interpreted as related to the work of Wagner [150] for Coulomb energy on the sphere).

Corollary 1 (Steinerberger [127])**.** *Let M be a smooth, compact $d-$dimensional manifold without boundary and $d \geq 3$, then*

$$\sum_{\substack{k,\ell=1 \\ k \neq \ell}}^{n} G(x_k, x_\ell) \gtrsim_M -n^{2-2/d}.$$

If the manifold is two-dimensional, then

$$\sum_{\substack{k,\ell=1 \\ k \neq \ell}}^{n} G(x_k, x_\ell) \gtrsim_M -n \log n.$$

The Theorem and the Corollary combined then show that for points minimizing the

Green energy for $d \geq 3$ we attain an optimal transport scaling of

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}, dx\right) \lesssim_M \frac{1}{n^{1/d}}.$$

This also refines [13] showing that minimizers of the Green energy equidistribute on the manifold: if we have an infinite sequence of points for which

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{\substack{k,\ell=1\\ k\neq\ell}}^{n} G(x_k, x_\ell) = 0,$$

then the sequence is asymptotically uniformly distributed (because the Wasserstein distance tends to 0). The Corollary follows from the argument developed in the Theorem. We again first deal with the case $d \geq 3$ and then discuss the necessary modifications for $d = 2$.

Proof: Summarized from [127]. We make use of the trivial identity

$$\left\|e^{t\Delta}\mu - dx\right\|_{\dot{H}^{-1}}^2 \geq 0.$$

At the same time, we can control its expansion in terms of Green's function and the bounds obtained in the theorem via

$$\begin{aligned}0 \leq \left\|e^{t\Delta}\mu - dx\right\|_{\dot{H}^{-1}}^2 \leq \frac{1}{N^2}\sum_{k}\int_M\int_M G(x,y)e^{t\Delta}\delta_{x_k}(x)e^{t\Delta}\delta_{x_k}(y)dxdy \\ + \frac{1}{N^2}\sum_{k\neq\ell}\int_M\int_M G(x,y)e^{t\Delta}\delta_{x_k}(x)e^{t\Delta}\delta_{x_\ell}(y)dxdy.\end{aligned}$$

We have the inequalities

$$\begin{aligned}\frac{1}{N^2}\sum_{k}\int_M\int_M G(x,y)e^{t\Delta}\delta_{x_k}(x)e^{t\Delta}\delta_{x_k}(y)dxdy &\lesssim \frac{t^{1-\frac{d}{2}}}{N} \\ \frac{1}{N^2}\sum_{k\neq\ell}\int_M\int_M G(x,y)e^{t\Delta}\delta_{x_k}(x)e^{t\Delta}\delta_{x_\ell}(y)dxdy &\leq \frac{t}{\mathrm{vol}(M)} + \frac{1}{N^2}\sum_{k\neq\ell}G(x_k,x_\ell)\end{aligned}$$

We may now set $t = N^{-2/d}$ to obtain the desired bound. Using the modified bounds yields the result for $d = 2$. □

Chapter 4

A Dynamical Approach

4.1 Background

In this section, we consider the problem of "How do we select points on a space which are as uniformly distributed as possible?" This is a natural question to ask, and one which has been studied in some form for a long time. If we know in advance how many points we are to place, we can exploit symmetries of the space to make nicely structured, well spread out designs, but in many real-life scenarios we do not. For instance, consider the problem of corporate strategists deciding where to place their next cafe in town to maximize convenience for consumers (and profit for themselves). Not only do they not know how many cafes they will ultimately build, but they did not necessarily have any say over the current placement: some short-term thinking former strategists made the call of where to place the first 100 cafes, and it is your job as the new hire to place the 101st, well aware that you (or a successor) will likely have to pick a location for the 102nd and 103rd cafes later. This is the problem of 'on-line selection.' The motivation is similar to that of Chapter 1, with the notable exception that when we are on a general manifold we measure the quality of the point set using Wasserstein transport cost as opposed to discrepancy. We will consider some known results and approaches to the problem, first looking at the problem on the unit interval / torus, and broadening our scope to more general manifolds later. It is not known what the optimal rate of discrepancy is for sequences in $[0,1]^d$ with $d > 1$—finding good constructions in 1 dimension could pave the way for higher-dimensional generalizations.

4.1.1 A Naive Approach: Maximize Distance

On first inspection, the problem may seem trivial: why not simply place each point as far from the existing points as possible? More formally, we may set

$$x_{n+1} = \arg\max_{x \in [0,1)} \left(\min_{1 \le i \le n} |x - x_i| \right),$$

breaking ties by taking the smallest such x. Below we plot the discrepancy of the sequence obtained by this algorithm with initial point set $\{0, \pi - 3, 1\}$.

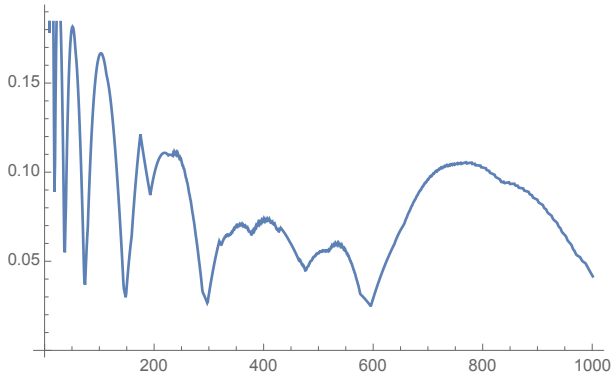

Figure 4.1: The discrepancy of the simple algorithm on $\{0, \pi - 3, 1\}$.

Why does this seemingly decent algorithm perform so horrendously in terms of discrepancy? Quite simply, the sequence it generates is not even equidistributed. To understand why, consider the first few points it picks.

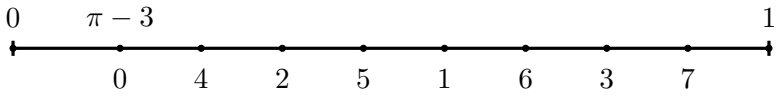

Figure 4.2: The first 7 elements of the naive algorithm on $\{0, \pi - 3, 1\}$.

Notice that each time the interval between two points in the sequence is bisected, two subintervals of the same length appear: we start with a subinterval of length $\pi - 3$ and another of length $4 - \pi$, and then turn the latter into two subintervals of length $(4 - \pi)/2$, then 4 of length $(4 - \pi)/4$, etc. Notably, the intervals of the same length are all adjacent and form a single block. Thus, the algorithm will spend 2^k steps filling in all the intervals in such a block, while completely ignoring intervals elsewhere. This explains the pattern of

increasingly long bumps in the discrepancy plot: while it is initially good to fill in longest intervals, it eventually becomes quite bad to keep picking points in the same small cluster of intervals. This also showcases the subtlety of the van der Corput sequence: while it is true that the van der Corput sequence bisects a maximal length interval at each step, it is the *order* in which these intervals are filled which makes it such an equidistributed sequence.

4.1.2 An Ergodic Approach: The Tent Map

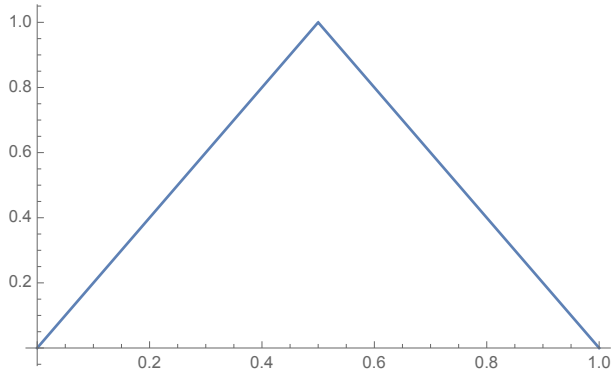

Figure 4.3: The Tent map on $[0, 1]$

One way we could approach this problem is to pick a particularly chaotic map with nice mixing properties–say, the Tent map

$$f(x) = \begin{cases} 2x & x < .5 \\ 2(1-x) & x \geq .5 \end{cases},$$

and simply iterate it. That is, arbitrarily pick some x_1 to start, then recursively set $x_{i+1} = f(x_i)$ to get the next point. This is just operating on the binary expansion of x: if the first bit after the bimal point is 0, it is removed, and if it is 1 then it is removed and the remaining bits are all flipped (swapping all 0s and 1s). Thus, if we pick a starting value x_1 with random bits, we should get a random sequence of points. However, random sequences of points are not particularly well-distributed in general. The star discrepancy of a random sequence of points on the unit interval is precisely the Kolmogorov-Smirnov statistic between the cumulative distribution function of the empirical distribution and the

uniform one. Then by Kolmogorov's Theorem, we have that, as $N \to \infty$, $\sqrt{N}D_N^*$ converges to the Kolmogorov distribution [76]. In particular, this tells us that we almost surely have infinitely many values of N for which $D_N^* > N^{-1/2}$, and thus the discrepancy does not decay particularly quickly (recall that the optimal rate is on the order of $\log N/N$).

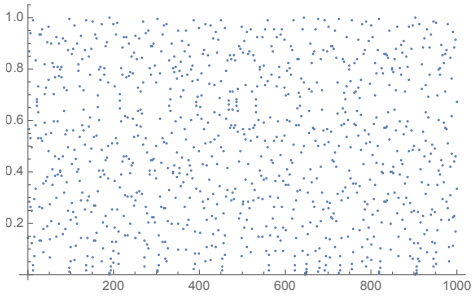

Figure 4.4: The first 1000 terms of the sequence, iterating the tent map on $\pi - 3$

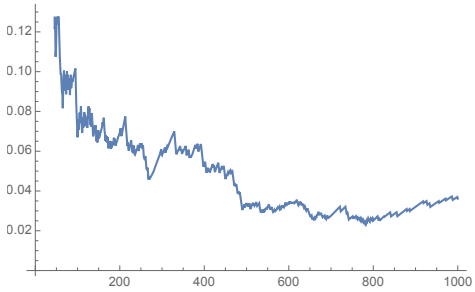

Figure 4.5: The discrepancy of this sequence does not decay quickly

4.1.3 A Dyadic Approach: Kakutani's Theorem

In the 1975 Conference on Measure Theory in Oberwolfach, Shizuo Kakutani (then a Yale faculty) gave a talk on this question. His framing was slightly different: he was not considering sequences of points but sequences of *partitions*. He proved the following theorem:

Theorem (Kakutani [73]). *Let $S \subset [0,1]$ be any finite set of points. Then, for any $\alpha \in (0,1)$, the following algorithm produces an equidistributed sequence: at each step, identify all the longest empty sub-intervals (x,y), and place new points at all the $\alpha x + (1-\alpha y)$.*

Note that there is a bit of "cheating" happening here: Kakutani is not adding a single point in each step, but a slew of points (one for each maximal length interval). The question for our sake is then "What order do we place these points in one at a time?" and for that we are essentially back at square 1. It is not hard to see that poor choices of ordering lead to the sequence failing to be equidistributed at all (see §1.1 above for an example of this). Since Kakutani's result, many authors have expanded, generalized, and refined the procedure, see [30, 31, 45, 96, 111, 149]. In particular, Volčič [149] presented a generalization of Kakutani's procedure called ρ-refinement: rather than fixing a ratio α to dyadically split intervals, ρ-refinement begins with a finite set of points and partitions the longest

empty interval(s) homothetically to the current partition—that is, it scales down the current partition to decompose the longest empty interval(s). Volčič proved the following:

Theorem (Volčič [149]). *If $\{\pi_n\}$ is a uniformly distributed sequence of partitions on $[0,1]$, then we may consider an associated point sequence, randomly ordering all endpoints which appear in the same step. With probability 1, this sequence is uniformly distributed.*

4.1.4 A Potential Theoretic Approach: Greedy Minimization

Another way we might attempt to choose points, motivated this time by physical considerations is by picking some function $f : [0,1) \to \mathbb{R}$, where

$$\sum_{i<j} f(x_i - x_j) \qquad \text{represents the energy}$$

of the system with particles placed at the points x_i. (We want f to be even, i.e. symmetric about $1/2$, so that the potential energy created from the interaction between x_i, x_j does not depend on the order in which we subtract them.) Then we set

$$x_k = \arg\min_{x\in[0,1)} \sum_{i<k} f(x - x_i),$$

greedily throwing in the point which contributes the lowest energy to the system and breaking ties arbitrarily. Here, we examine the outcome for two particular choices of f: $f_1(x) = \cos(2\pi x)$, and $f_2(x) = -\ln|2\sin(\pi x)| - \frac{1}{5}\cos(10\pi x)$.

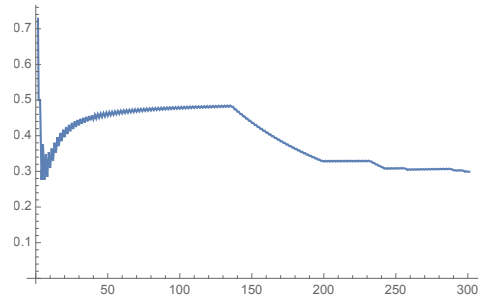

Figure 4.6: D_N obtained by greedily minimizing $\cos(2\pi x)$

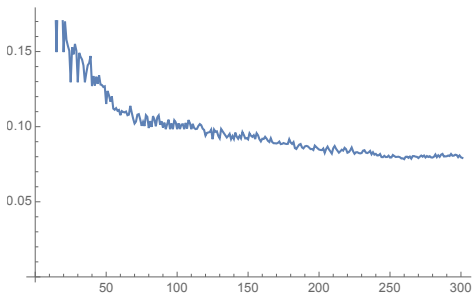

Figure 4.7: D_N obtained by greedily minimizing $\ln|2\sin(\pi x)| - \frac{1}{5}\cos(10\pi x)$

It is clear that we will need to further constrain the potential function if we are to get

sequences with good discrepancy. If the function satisfies a concavity property, Florian Pausinger proved that we retrieve a sequence of van der Corput type (if we start with a singleton set), as well as the following theoretical bound on the discrepancy:

Theorem (Pausinger [100])**.** *Let $f : [0,1] \to \mathbb{R}$ be bounded and symmetric about $1/2$. Further, assume $\widehat{f}(k) > c|k|^{-2}$ for some $c > 0$ and all $k \neq 0$. Then all sequences defined via the greedy algorithm on an arbitrary initial set satisfy*

$$D_N \leq \frac{\tilde{c}}{N^{1/3}},$$

where $\tilde{c} > 0$ depends on the initial set.

This answered a question posed by Steinerberger in [130]. We can interpret the condition on the Fourier coefficients as a quantitative positive definiteness, requiring the Fourier coefficients not to decay too quickly.

Proof: Summarized from [100]. Assume without loss of generality that f is mean 0 (otherwise, we may simply shift it down without impacting the algorithm, since the argmin of $\sum f(x - x_k)$ is independent of constant shifts). Then

$$\begin{aligned}\sum_{m,\ell=1}^{n} f(x_m - x_\ell) &= nf(0) + 2\sum_{\substack{m,\ell=1 \\ m<\ell}}^{n} f(x_m - x_\ell) \\ &= nf(0) + 2\sum_{\ell=2}^{n}\sum_{m=1}^{\ell-1} f(x_\ell - x_m) \leq nf(0)\end{aligned}$$

since, by definition of the greedy algorithm,

$$\sum_{m=1}^{\ell-1} f(x_\ell - x_m) = \min_{x\in[0,1)} \sum_{m=1}^{\ell-1} f(x - x_m) \leq \int_{\mathbb{T}} \sum_{m=1}^{\ell-1} f(x - x_m)dx = 0.$$

Thus

$$\sum_{m,\ell=1}^{n} f(x_m - x_\ell) \leq nf(0). \qquad (\diamond)$$

On the other hand, we have

$$\sum_{m,\ell=1}^{n} f(x_m - x_\ell) = \sum_{k\in\mathbb{Z}} \widehat{f}(k) \sum_{m,\ell=1}^{n} e^{2\pi i k(x_m - x_\ell)} = \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left(\sum_{m=1}^{n} e^{2\pi i k x_m}\right)\left(\sum_{m=1}^{n} e^{2\pi i k(-x_m)}\right)$$
$$= \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left(\sum_{m=1}^{n} e^{2\pi i k x_m}\right)\overline{\left(\sum_{m=1}^{n} e^{2\pi i k x_m}\right)} = \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left|\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2.$$

Thus, combining with $\diamond$,

$$\sum_{k\in\mathbb{Z}} \widehat{f}(k) \left|\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2 \leq n f(0).$$

Now, we use the fact that $\widehat{f}(k) \geq c|k|^{-2}$:

$$n f(0) \geq \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left|\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2 \geq n^2 \sum_{k\neq 0} \frac{c}{k^2} \left|\frac{1}{n}\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2 = c n^2 \|\mu_n\|_{\dot{H}^{-1}}^2,$$

so $\|\mu_n\|_{\dot{H}^{-1}} \lesssim n^{-1/2}$. Finally, by LeVeque's inequality, we can bound the discrepancy as

$$D_n \lesssim \|\mu_n\|_{\dot{H}^{-1}}^{2/3} \lesssim n^{-1/3}$$

and we have the desired result. □

4.2 A Curious Phenomenon

In this section we will further explore the dynamical system from §1.4—with the right conditions on the potential function, this curious system results in sequences of points with seemingly remarkable properties. As in Pausinger's result, let the even function $f : \mathbb{T} \to \mathbb{R}$ satisfy $\widehat{f}(k) \geq c|k|^{-2}$—recall this is a quantitative form of positive definiteness. One example of such a function is the second Bernoulli polynomial which, identifying $\mathbb{T} \sim [0, 1]$, is

$$f(x) = x^2 - x + \frac{1}{6}.$$

Starting with an arbitrary initial set of points, we define a sequence greedily via

$$x_n = \arg\min_{x \in \mathbb{T}} \sum_{k=1}^{n-1} f(x - x_k).$$

As before, when we take the argmin, we may pick arbitrarily from any of the points where the minimum is attained if there is not a unique point. Such sequences $(x_n)_{n=1}^{\infty}$ seem to be astonishingly regularly distributed, with quickly decaying discrepancy and W_2 transportation cost to the uniform measure. However, these observations are mainly empirical at this stage. We refer to the papers [126, 130] for some numerical experiments (see also below). We summarize the existing results and derive some new ones; however, the overall phenomenon is largely unexplained. We prove

$$W_2\left(\mu, \nu\right) \leq \frac{c}{\sqrt{n}}, \quad \text{where } \mu = \frac{1}{n}\sum_{k=1}^{n} \delta_{x_k}$$

is the empirical distribution and $\nu = dx$ is the Lebesgue measure. Much stronger results seem to be true and it is an interesting problem to understand this dynamical system better. We obtain optimal results in dimension $d \geq 3$: using $G(x, y)$ to denote the Green's function of the Laplacian on a compact manifold, we show that

$$x_n = \arg\min_{x \in M} \sum_{k=1}^{n-1} G(x, x_k) \quad \text{satisfies} \quad W_2\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{x_k}, dx\right) \lesssim \frac{1}{n^{1/d}}.$$

4.2.1 Introduction

We begin by listing some open problems related to the algorithm above:

1. **Open Problem 1.** Is it true that

$$\sum_{k,\ell=1}^{n} f(x_k - x_\ell) \lesssim \log n?$$

2. **Open Problem 2.** Is it true that

$$\left\| \sum_{k=1}^{n} f(x - x_k) \right\|_{L^\infty} \lesssim \log n?$$

Before embarking on a discussion of these problems, we quickly illustrate Open Problem 2 with a simple example. Indeed, Open Problem 2 can be stated in very simple terms.

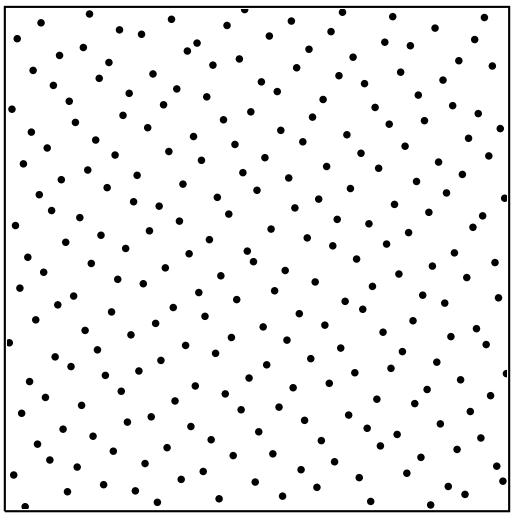

Figure 4.8: 250 points created starting with $\{1/3, 4/5\}$ and using the second Bernoulli polynomial for f. We display the points $(n/250, x_n) \in [0,1]^2$ for $1 \leq n \leq 250$. Why is this distribution so regular?

We fix again $f(x) = x^2 - x + 1/6$ and obtain a sequence by starting with $x_1 = 0.3, x_2 = 0.8$ and using the greedy algorithm to obtain all subsequent elements of the sequence. We set

$$f_n(x) = \sum_{k=1}^{n} f(x - x_k).$$

As seen in Figure 4.3, the function f_n does not seem to be very large: this is only possible if

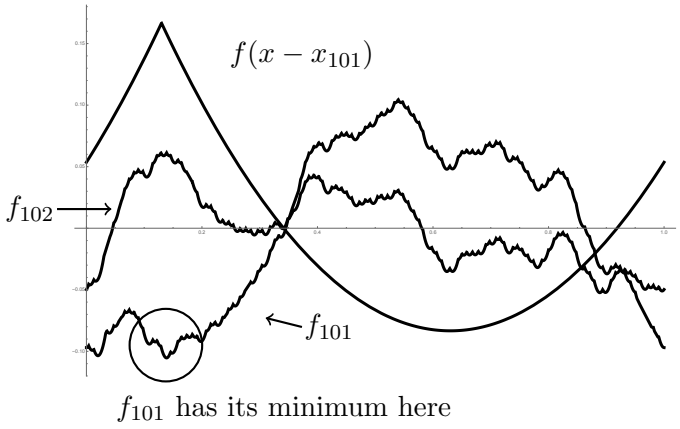

Figure 4.9: We obtain f_{102} by finding the point x_{101} at which f_{101} assumes its minimum and $f_{102}(x) = f_{101}(x) + f(x - x_{101})$.

the sequence elements are so regular that the sum over $f(x-x_k)$ leads to good cancellation properties. This is one instance of the 'curious' phenomenon alluded to in the section title: why is $\|f_n\|_{L^\infty}$ so remarkably small in n? The inequalities posed in Open Problems 1 and

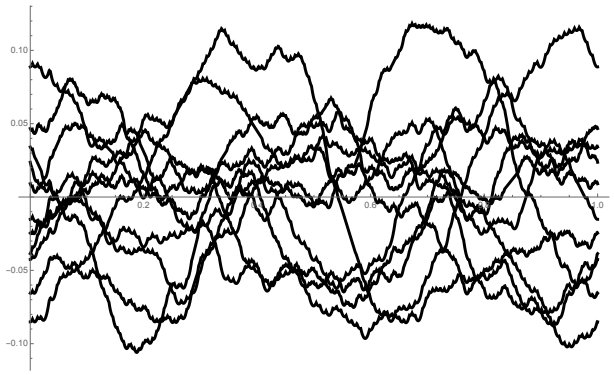

Figure 4.10: The functions $f_{100}, f_{110}, f_{120} \ldots, f_{200}$. We observe that they are quite different from one another and have an interesting behavior. Most importantly, they all seem to be quite small with $\|f_n\|_{L^\infty}$ barely exceeding $\|f\|_{L^\infty}$.

2 above, if true, would indicate that the sequence $(x_n)_{n=1}^{\infty}$ satisfies very good distribution properties. What is remarkable is that, in a certain sense, these properties would be close to optimal. We will derive that

$$\sum_{k,\ell=1}^{n} f(x_k - x_\ell) \lesssim n \qquad \text{and} \qquad \left\| \sum_{k=1}^{n} f(x-x_k) \right\|_{L^\infty} \lesssim n^{1/2}$$

but these estimates seem to be very far from sharp. Any improvement of these estimates would immediately yield improvements of our other results via the arguments outlined below. In the converse direction, it is an interesting question whether the following is true: if $f : \mathbb{T} \rightarrow \mathbb{R}$ is an even, continuous function with mean 0 such that $\widehat{f}(k) > c|k|^{-2}$ for all $k \neq 0$, is it true that for any sequence $(x_k)_{k=1}^{\infty}$

$$\left\| \sum_{k=1}^{n} f(x-x_k) \right\|_{L^\infty} \qquad \text{is unbounded in } n?$$

For the 'double' sum, this is a known result of Proinov [108], translated into English by Kirk [75] (see Chapter 1, §3.2), which shows that, for some constant $c_f > 0$ depending only

on the function and any sequence $(x_n)_{n=1}^{\infty}$,

$$\sum_{k,\ell=1}^{n} f(x_k - x_\ell) \geq c_f \log n \qquad \text{for infinitely many } n.$$

In particular, this shows that the bounds conjectured in Open Problem 1 would be optimal. It seems reasonable to assume that the condition $\widehat{f}(k) \geq c|k|^{-2}$, or some condition like it, is necessary for this phenomenon to occur; it is certainly necessary for our proof that the sequence $(x_n)_{n=1}^{\infty}$ is uniformly distributed. It may be of interest to study the dynamical system when f is a trigonometric polynomial: it seems that in this case the sequence $(x_n)_{n=1}^{\infty}$ will not even be uniformly distributed.

4.2.2 Connections To Other Problems.

We start with a simple example. Let us define, as above, $f(x) = x^2 - x + 1/6$ and consider the sequence obtained via

$$x_n = \arg\min_{x \in \mathbb{T}} \sum_{k=1}^{n-1} f(x - x_k)$$

when starting with $\{1/3, 4/5\}$. The sequence is easy to compute and starts

$$\frac{1}{3}, \frac{4}{5}, 0.066, 0.566, 0.941, 0.441, 0.191, 0.691, \ldots$$

Empirically, this sequence (and seemingly any sequence obtained in this way) seems to have remarkable regularity properties. We now state our question as follows.

> **Open Problem 3.** Are these greedy sequences, up to constants, comparable to the behavior of the best Kronecker sequence or the van der Corput sequence in all the ways outlined in Chapter 1, §1?

If this were indeed the case, it could have very interesting consequences. Recall, both the Kronecker sequence and the van der Corput sequence are known to be optimal in the one-dimensional setting (a result of Schmidt [121], see also [80, 81] for an improved constant). However, nobody knows what sequences are optimal in even $d = 2$ dimensions (we refer to the excellent survey of Bilyk [17]). So, if there was a greedy-type construction with

optimal behavior in $d = 1$ dimension, it might suggest sequences of similar quality in higher dimensions as well – this would be interesting because the greedy sequence seems to be unlike any that has been studied; in particular, if it enjoys good distribution properties, this seems like it would have to be because of a different underlying mechanism.

4.2.3 Known results.

This type of construction was first proposed by the Steinerberger in [130]. There it was shown that if the function is

$$f(x) = -\log\left(2\sin\left(\pi|x|\right)\right),$$

then the arising sequence satisfies the discrepancy bound

$$D_N \leq c\frac{\log N}{\sqrt{N}}.$$

The arguments are based on the explicit structure of the Fourier series of $f(x)$ and do not generalize to other functions. It is already discussed in [130] that much stronger results seem to be true and that the sequence arising from this function f seems, numerically, as well behaved with regard to all these aspects as the Kronecker or the van der Corput sequence. This particular choice of f has a natural geometric interpretation as $-\log$ of the chord length from 1 to $e^{2\pi i x}$ on the complex unit circle, and thus minimizing $\sum_{k=1}^{n-1} f(x - x_k)$ is precisely maximizing the product of chord lengths to points on the unit circle. Steinerberger proved stronger results about this particular f in [129], contextualizing the problem as one about polynomials with zeroes on the unit circle and considering an L^1 measure of discrepancy (as opposed to the L^∞ discrepancy considered in this thesis). The same idea, interpreted differently, has also led to a numerical scheme that seems to be effective at regularizing point sets [126]. It was also noted in [130] that if one starts with a single element $\{x_1\}$, then the arising sequence seems to be related to the van der Corput sequence—this is indeed the case and was subsequently proven by Pausinger [100]. Pausinger's theorem (see §1.3) holds for the much larger family of strictly convex functions $f : [0, 1] \to \mathbb{R}$ that are symmetric around

$x = 1/2$. There it is also shown that for functions of this type satisfying $\widehat{f}(k) \geq c|k|^{-2}$ for all $k \neq 0$, the arising construction results in a sequence with

$$D_N \leq \frac{\tilde{c}}{N^{1/3}}.$$

The bound stated in Open Problem 1 would improve this estimate to $\lesssim \log N/N^{2/3}$ (which, however, is still not at the $\log N/N$ level that we observe numerically). Other types of greedy constructions of sequences have been considered in the literature, we refer to work of Kakutani [73] and Temylakov: [142,143] and §6.11 in [140]. Temlyakov has since used this type of sequence to establish an endpoint result for a result in Numerical Integration [141].

4.2.4 Riesz points

Riesz points refer, at great level of generality, to point sets minimizing energy expressions of the form

$$\arg \min_{x \in M^n} \sum_{\substack{k,\ell=1 \\ k \neq \ell}}^{n} \frac{1}{\|x_k - x_\ell\|^s}.$$

The problem was first stated on $\mathbb{S}^2$ with $s = 1$ by Thomson [144] in 1904 and has since inspired a large body of work, we refer to [24, 40, 64, 118] and references therein. We make a connection with two contributions in particular. The first is due to Beltran, Corral and Criado del Rey [13]: they show that if we consider sets of n points on a compact manifold chosen so that they minimize

$$\sum_{\substack{k,\ell=1 \\ k \neq \ell}}^{n} G(x_k, x_\ell), \quad \text{where } G \text{ is the Green's function}$$

of the Laplacian on M, then the sequence of point measures on the first n terms converges weakly to the uniform measure

$$\frac{1}{n} \sum_{k=1}^{n} \delta_{x_k} \rightharpoonup dx.$$

This can be considered the static analogue (since one finds the minimal arrangement for all n points) of our problem (keeping the previous $n-1$ points fixed and then greedily adding the point which will minimize total energy). The second contribution that we highlight is

very recent and due to Marzo & Mas [92]. They studied the specific problem of minimizing the $s-$Riesz energy

$$E_s = \sum_{\substack{k,\ell=1 \\ k \neq \ell}}^{n} \frac{1}{\|x_k - x_\ell\|^s} \qquad \text{on } \mathbb{S}^d$$

and estimating the spherical cap discrepancy of the minimizing point set: in short, if the points are uniformly distributed, then we would expect the number of points in each spherical cap to be proportional to the volume of the cap; the largest discrepancy is known as spherical cap discrepancy. They use ideas dating back to Wolff: the Riesz energy E_s is comparable to a negative Sobolev norm and, more precisely, for all $f \in L^2(\mathbb{S}^d)$,

$$\|f\|^2_{H^{(s-d)/2}} \lesssim_{s,d} \int_{\mathbb{S}^d \times \mathbb{S}^d} \frac{f(x)f(y)}{\|x-y\|^s} dxdy \lesssim_{s,d} \|f\|^2_{H^{(s-d)/2}}.$$

This, while not directly related to our approach, is at least philosophically connected: we will estimate the Wasserstein distance in negative Sobolev spaces and use the underlying L^2-structure. We take a similar approach to combinatorial graphs in Chapter 7.

4.3 Results

4.3.1 Wasserstein distance.

The main purpose of this section is to (1) describe the phenomenon and its connections in a concise way and (2) to point out that we can obtain slightly improved regularity results by switching to the Wasserstein distance, introduced in Chapter 2 §2. We will only discuss the case where one measure is the empirical distribution

$$\mu = \frac{1}{n}\sum_{k=1}^{n} \delta_{x_k} \qquad \text{and the other measure is} \qquad \nu = dx.$$

In our setting, we trivially have $0 \leq W_p(\mu, \nu) \leq 1$. Pausinger's inequality [100]

$$D_N \leq \frac{c}{N^{1/3}}$$

can be coupled with the Monge-Kantorovich formula to obtain

$$W_1\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}, dx\right) \lesssim \frac{1}{n^{1/3}}.$$

4.3.2 Main Results.

Our main result is an improvement for the W_2-distance. Hölder's inequality shows that $W_1(\mu,\nu) \leq W_2(\mu,\nu)$, so the result also implies improved bounds for the W_1 distance.

Theorem 4.3.1 (B. & Steinerberger [26]). *Let the even function $f : \mathbb{T} \to \mathbb{R}$ satisfy $\widehat{f}(k) \geq c|k|^{-2}$ for some fixed constant $c > 0$ and all $k \neq 0$. Define a sequence via*

$$x_n = \arg\min_{x\in\mathbb{T}} \sum_{k=1}^{n-1} f(x - x_k),$$

then this sequence satisfies

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}, dx\right) \lesssim \frac{1}{n^{1/2}},$$

where the implicit constant depends only on the initial set, $f(0)$ and c.

One way of interpreting the Theorem is as follows: given $\{x_1, \ldots, x_n\}$ we can interpret these points as Dirac measures with weight $1/n$. It is then possible to 'break' these points up and move their L^1-mass a distance of, on average, not more than $\sim n^{-1/2}$ to recreate the uniform distribution. The result seems to be far from the truth, which we believe to be at scale n^{-1} up to logarithmic factors (see below). We also obtain the following corollary (which was suggested to us together with its proof by Igor Shparlinski).

Corollary 2 (Shparlinski). *Suppose $f : \mathbb{T} \to \mathbb{R}$ is even, has mean value 0 and satisfies both*

$$\widehat{f}(k) > 0 \qquad \textit{and} \qquad \sum_{k\in\mathbb{Z}} \widehat{f}(k) < \infty.$$

Then, for any sequence (x_n) arising from the algorithm outlined above,

$$\left\| \sum_{k=1}^{n} f(x - x_k) \right\|_{L^\infty} \lesssim \sqrt{n}.$$

Again, we believe this to be far from optimal and expect the quantity to grow not much faster than (at most) logarithmically. We have a refinement of this statement in the case where $\widehat{f}(k) \sim |k|^{-2}$:

Theorem 4.3.2 (B. & Steinerberger [26]). *Let $f : \mathbb{T} \rightarrow \mathbb{R}$ be an even function with mean 0 satisfying $c_1|k|^{-2} \leq \widehat{f}(k) \leq c_2|k|^{-2}$ for all $k \neq 0$ for some universal $c_1, c_2 > 0$. Then*

$$\left\| \sum_{k=1}^{n} f(x - x_k) \right\|_{L^\infty} \lesssim n^{1/3} \qquad \textit{for infinitely many } n.$$

The argument is slightly finer than this: we will prove that

$$\left\| \sum_{k=1}^{n} f(x - x_k) \right\|_{L^\infty} \lesssim n^{1/3} \left\| \sum_{k=1}^{n} f(x - x_k) \right\|_{L^1}^{1/3}$$

and then prove that the L^1-term has to be $\lesssim 1$ infinitely many times. We note that this result is below the $n^{1/2}-$threshold that we would expect from randomly chosen points. Again, as mentioned above, we expect the error rate to actually be much smaller than this. We will now discuss why Wasserstein distance is a very canonical way of capturing problems of this type. We state this formally in the following estimate.

Corollary 3 (B. & Steinerberger [26]). *Suppose $f : \mathbb{T} \rightarrow \mathbb{R}$ is even, has mean value 0 and satisfies $\widehat{f}(k) \geq c|k|^{-2}$ for $k \neq 0$. Then, for any set $\{x_1, \dots, x_n\} \subset \mathbb{T}$, we have*

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{x_k}, dx\right) \lesssim \frac{1}{n}\left(\sum_{k,\ell=1}^{n} f(x_k - x_\ell)\right)^{1/2}.$$

Fix now a function such that $\widehat{f}(k) \sim |k|^{-2}$ for $k \neq 0$ (in the sense of having corresponding upper and lower bounds). Open Problem 1 asks whether

$$\sum_{k,\ell=1}^{n} f(x_k - x_\ell) \lesssim \log n \qquad \text{might hold}$$

and, conversely, which kind of lower bounds exist. Corollary 2 shows that any such estimate

would imply

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

This connects to yet another problem, that of *irregularities of distribution*. The results of van Aardenne-Ehrenfest, Roth, and Schmidt described in Chapter 1 §3 show that, for discrepancy, irregularities of distribution are unavoidable. A natural question now is the following: does a similar phenomenon exist for the Wasserstein distance? This was answered by Cole Graham [61] who proved the following result.

Theorem (Graham [61]). *For any sequence $(x_n)_{n=1}^{\infty}$ in $[0,1]$, we have*

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}, dx\right) \gtrsim \frac{\sqrt{\log n}}{n} \quad \textit{for infinitely many } n.$$

Steinerberger has already remarked in [131] that this is sharp for the Kronecker sequence $x_n = \{n\alpha\}$ for any badly approximable α. An implication of Graham's result coupled with our Corollary above is the following result that was first established by Proinov.

Theorem (Proinov, [108]). *Let $f : \mathbb{T} \to \mathbb{R}$ be a function with mean value 0 satisfying $\widehat{f}(k) \geq c|k|^{-2}$ for $k \neq 0$. Then, for any sequence $(x_n)_{n=1}^{\infty}$, we have*

$$\sum_{k,\ell=1}^{n} f(x_k - x_\ell) \gtrsim \log n \qquad \textit{for infinitely many } n.$$

Using again the Kronecker sequence, we can show that there are sequences for which this notion of energy does indeed grow very slowly; this result is folklore, we include it for the convenience of the reader. The same result is also known for the van der Corput sequence, we refer to Proinov & Grozdanov [110].

Proposition 4.3.1 (B. & Steinerberger [26]). *Let $f : \mathbb{T} \to \mathbb{R}$ have mean value 0 satisfying $\widehat{f}(k) \leq c|k|^{-2}$ for $k \neq 0$. Then, for any badly approximable α, the sequence $x_n = \{n\alpha\}$ has*

$$\sum_{k,\ell=1}^{n} f(x_k - x_\ell) \lesssim \log n.$$

The proof of the proposition makes explicit use of a rather delicate property of the

sequence $\{n\alpha\}$. It is thus even more striking that, possibly, the greedy sequence

$$x_n = \arg\min_{x \in \mathbb{T}} \sum_{k=1}^{n-1} f(x - x_k)$$

might conceivably behave in a similar manner. Naturally, this falls into the realm of Approximation Theory and, more specifically, the Greedy Algorithm [43, 139, 140] and its use in Approximation Theory. Indeed, we can interpret this greedy sequence as a way to approximate the constant function 0 by means of translates $f(x - x_k)$. The Greedy Algorithm is well understood to yield reasonable estimates for a broad class of functions—what is of special interest here is that in our case the greedy algorithm seems to perform much better than one would usually expect from a greedy algorithm; moreover, it seems to be comparable in efficiency to subtle constructions in Number Theory that make use of delicate notions such as badly approximable numbers.

4.3.3 Two Remarks.

All our estimates are based on the inequality

$$\sum_{k,\ell=1}^{n} f(x_k - x_\ell) \leq n f(0).$$

It is not difficult to see (see below) that this is indeed satisfied for our greedy construction. However, the inequality (and therefore our main Theorem) is also valid if x_n is chosen in such a way that

$$\sum_{k=1}^{n-1} f(x_n - x_k) \leq 0.$$

We observe that f has mean value 0 and thus

$$\int_{\mathbb{T}} \sum_{k=1}^{n-1} f(x - x_k) dx = 0$$

and it is always possible to choose a new element x_n with this property (and, usually, there are many of those). However, presumably these elements can be chosen in rather terrible ways and there is no reason to expect these sequences $(x_n)_{n=1}^{\infty}$ to have particularly good

distribution properties; it would seem our Theorem is close to optimal for these types of sequences though we do not know how to show this. It also shows the bottleneck in our current approach: we do not know how to make use of the fact that the algorithm chooses the minimal value and not merely a value not exceeding the expected value. The second remark concerns uniform distribution of the sequence $(x_n)_{n=1}^{\infty}$. We have the following fact.

Corollary 4 (B. & Steinerberger [26]). *If $\widehat{f}(k) > 0$ for all $k \neq 0$, then the sequence x_n defined via*

$$x_n = \arg\min_{x \in \mathbb{T}} \sum_{k=1}^{n-1} f(x - x_k).$$

is uniformly distributed on $\mathbb{T}$.

The argument is so short that we can give it right here.

Proof: Summarized from [26]. We have

$$nf(0) \geq \sum_{m,\ell=1}^{n} f(x_m - x_\ell) = \sum_{\substack{k \in \mathbb{Z} \\ k \neq 0}} \widehat{f}(k) \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2 \geq \widehat{f}(k) \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2$$

from which we obtain

$$\frac{1}{n} \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right| \leq \sqrt{\frac{f(0)}{\widehat{f}(k)}} \frac{1}{\sqrt{n}}.$$

This tends to 0 from which we obtain uniform distribution from Weyl's theorem. □

We emphasize that the argument also shows that the size of $\widehat{f}(k)$ will play a role in the quality of the distribution: if it decays rapidly, the convergence rate might be quite slow.

4.3.4 Higher dimensions.

The same phenomenon exists in higher dimensions and it does so at a great level of generality. Indeed, the scaling in higher dimensions is fundamentally different and this allows us to obtain optimal results. Let (M, g) be a smooth compact manifold without boundary. We use ϕ_k to denote the L^2-normalized eigenfunctions of the Laplace operator

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

We will now define admissible kernels $K : M \times M \to \mathbb{R}$ to be functions of the form

$$K(x,y) = \sum_{k=1}^{\infty} a_k \frac{\phi_k(x)\phi_k(y)}{\lambda_k}$$

where the coefficient a_k is assumed to satisfy a two-sided bound:

$$c_1 < a_k < c_2 \qquad \text{for all } k \geq 1$$

and some positive constants c_1, c_2. We note that the sum starts at $k = 1$ and thus excludes the trivial (constant) eigenfunction ϕ_0. In particular, all these kernels have mean value 0. This definition is an extension of our assumption $\widehat{f}(k) \geq c|k|^{-2}$ in the one-dimensional setting. A particularly natural kernel arises from setting $a_k = 1$ in which case we obtain the Green's function of the Laplacian $G(x,y)$. This function has the property that

$$-\Delta_x \int_M G(x,y) f(y) dy = f(x),$$

i.e. it solves the equation $-\Delta u = f$. We will now consider sequences of the form

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

Theorem 4.3.3 (B. & Steinerberger [26])**.** *Let x_n be a sequence obtained in such a way on a $d-$dimensional compact manifold. Then*

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

We note that this result is optimal for $d \geq 3$. We do not know whether the logarithmic factor is necessary for $d = 2$. The main ingredient is a favorable estimate of the Wasserstein distance that was recently obtained by Steinerberger [127] that allows for a greedy formulation. We note that while the static case, the structure of point sets minimizing the Green energy, has been an active field of study [12, 13, 15, 34, 39, 57, 84, 92, 127], we are not aware

of results in the dynamic setting. This theorem and its proof are explored further in the following chapter.

Corollary 5 (B. & Steinerberger [26])**.** *If $d \geq 3$, then there exists a sequence of points $(x_n)_{n=1}^{\infty}$ on $\mathbb{T}^d$ such that*

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}, dx\right) \lesssim_d \frac{1}{n^{1/d}} \qquad \textit{uniformly in } n.$$

This Corollary seems to be new: it gives a constructive proof that Wasserstein distance does not have an irregularities of distribution phenomenon in dimensions $d \geq 3$. We have the same result up to a factor of $\sqrt{\log n}$ in two dimensions. By Graham's result [61], the loss of a factor of $\sqrt{\log n}$ is indeed necessary in $d = 1$.

4.4 Proofs

4.4.1 Proof of Theorem 4.3.1, Corollary 2 and Corollary 3.

Proof: Summarized from [26]. The proof decomposes into two parts. In the first part we argue exactly as in [100]. We can assume w.l.o.g. that f has mean value 0. We first observe

$$\sum_{m,\ell=1}^{n} f(x_m - x_\ell) \leq nf(0). \qquad (\diamond)$$

which follows from the identity

$$\sum_{m,\ell=1}^{n} f(x_m - x_\ell) = nf(0) + 2\sum_{\substack{m,\ell=1 \\ m<\ell}}^{n} f(x_m - x_\ell),$$

the reformulation

$$\sum_{\substack{m,\ell=1 \\ m<\ell}}^{n} f(x_m - x_\ell) = \sum_{\ell=2}^{n}\sum_{m=1}^{\ell-1} f(x_\ell - x_m)$$

and the greedy algorithm: by definition of x_ℓ, we have

$$\sum_{m=1}^{\ell-1} f(x_\ell - x_m) = \min_x \sum_{m=1}^{\ell-1} f(x - x_m) \leq \int_{\mathbb{T}} \sum_{m=1}^{\ell-1} f(x - x_m)dx = 0.$$

Rewriting quantities in terms of Fourier Analysis then shows that

$$\sum_{m,\ell=1}^{n} f(x_m - x_\ell) = \sum_{k\in\mathbb{Z}} \widehat{f}(k) \sum_{m,\ell=1}^{n} e^{2\pi i k(x_m - x_\ell)} = \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left(\sum_{m=1}^{n} e^{2\pi i k x_m}\right)\left(\sum_{m=1}^{n} e^{2\pi i k(-x_m)}\right)$$
$$= \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left(\sum_{m=1}^{n} e^{2\pi i k x_m}\right)\overline{\left(\sum_{m=1}^{n} e^{2\pi i k x_m}\right)} = \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left|\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2.$$

We first use this fact to establish the statement of the Corollary 2. This corollary was suggested to the author and Steinerberger by Igor Shparlinski, and we are grateful to be able to incorporate it here. Note that

$$\sum_{\ell=1}^{n} f(x - x_\ell) = \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left(\sum_{\ell=1}^{n} e^{-2\pi i k x_\ell}\right) e^{2\pi i k x}$$

and thus, using the Cauchy-Schwarz inequality,

$$\left\|\sum_{\ell=1}^{n} f(x - x_\ell)\right\|_{L^\infty} \leq \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left|\sum_{\ell=1}^{n} e^{-2\pi i k x_\ell}\right| = \sum_{k\in\mathbb{Z}} \widehat{f}(k)^{1/2} \widehat{f}(k)^{1/2} \left|\sum_{\ell=1}^{n} e^{-2\pi i k x_\ell}\right|$$
$$\leq \left(\sum_{k\in\mathbb{Z}} \widehat{f}(k)\right)^{1/2} \left(\sum_{k\in\mathbb{Z}} \widehat{f}(k) \left|\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2\right)^{1/2}$$
$$= \sqrt{f(0)} \left(\sum_{m,\ell=1}^{n} f(x_m - x_\ell)\right)^{1/2}.$$

Coupled with the inequality $(\diamond)$ above, we obtain

$$\left\|\sum_{\ell=1}^{n} f(x - x_\ell)\right\|_{L^\infty} \leq f(0)\sqrt{n} \lesssim \sqrt{n},$$

which was the desired statement. To prove Theorem 4.3.1 and Corollary 3, we may further assume $\widehat{f}(k) \geq c|k|^{-2}$ for all $k \neq 0$, and thus

$$n^2 \sum_{\substack{k\in\mathbb{Z}\\ k\neq 0}} \frac{c}{k^2} \left|\frac{1}{n}\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2 \leq \sum_{k\in\mathbb{Z}} \widehat{f}(k) \left|\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2 = \sum_{m,\ell=1}^{n} f(x_m - x_\ell) \leq n f(0),$$

so we have

$$\sum_{\substack{k\in\mathbb{Z}\\k\neq 0}} \frac{c}{k^2}\left|\frac{1}{n}\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2 \leq \frac{f(0)}{n}.$$

Reformulating,

$$\left(\sum_{\substack{k\in\mathbb{Z}\\k\neq 0}} \frac{1}{k^2}\left|\frac{1}{n}\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2\right)^{1/2} \lesssim_{c,f(0)} \frac{1}{\sqrt{n}}.$$

We note that this last argument has been previously stated in the literature in a very different context (integration error of periodic functions in terms of Zinterhof's diaphony [154]) in a paper of Zinterhof & Stegbuchner [153]. It remains to prove the second Corollary and Theorem 4.3.1. For that we use Peyré's inequality [106] (see Chapter 2 §4): this estimate states that, for any measure μ on $\mathbb{T}$

$$W_2(\mu, dx) \lesssim \|\mu\|_{\dot{H}^{-1}} = \left(\sum_{k\neq 0} \frac{|\widehat{\mu}(k)|^2}{k^2}\right)^{1/2}.$$

We apply this estimate to the measure

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

to obtain

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}, dx\right) \lesssim \left(\sum_{k\neq 0}\frac{1}{k^2}\left|\frac{1}{n}\sum_{\ell=1}^{n} e^{2\pi i k x_\ell}\right|^2\right)^{1/2} \lesssim \frac{1}{\sqrt{n}}.$$

This establishes Theorem 4.3.1. Corollary 2 follows from remarking that, as seen above,

$$\frac{1}{n^2}\sum_{m,\ell=1}^{n} f(x_m - x_\ell) = \sum_{k\in\mathbb{Z}} \widehat{f}(k)\left|\frac{1}{n}\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2.$$

Moreover, since $\widehat{f}(k) \geq c|k|^{-2}$ for $k \neq 0$ and $\widehat{f}(0) = 0$, we can bound this quantity by

$$\sum_{k\in\mathbb{Z}} \widehat{f}(k)\left|\frac{1}{n}\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2 \gtrsim \sum_{k\neq 0}\frac{1}{|k|^2}\left|\frac{1}{n}\sum_{m=1}^{n} e^{2\pi i k x_m}\right|^2 = \|\mu\|_{\dot{H}^{-1}}.$$

Finally, by another application of Peyré's inequality, we have

$$\sum_{k\neq 0} \frac{1}{|k|^2} \left| \frac{1}{n} \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2 \gtrsim W_2(\mu, dx). \qquad \square$$

4.5 Proof of Theorem 4.3.2

Lemma (B. & Steinerberger [26]). *We have*

$$\left\| \sum_{k=1}^{n} f(x - x_k) \right\|_{L^1} \leq 2f(0) \qquad \text{for infinitely many } n.$$

Proof: Summarized from [26]. Suppose that the inequality fails for some fixed n. Since

$$\int_{\mathbb{T}} \sum_{k=1}^{n} f(x - x_k) dx = 0,$$

we have that the positive mass and the negative mass cancel and thus, by pigeonholing,

$$\min_{x\in\mathbb{T}} \sum_{k=1}^{n} f(x - x_k) \leq -\frac{1}{2} \left\| \sum_{k=1}^{n} f(x - x_k) \right\|_{L^1} \leq -f(0).$$

This, in turn, then implies that

$$\sum_{k,\ell=1}^{n+1} f(x_k - x_\ell) = \sum_{k,\ell=1}^{n} f(x_k - x_\ell) + f(0) + 2\sum_{k=1}^{n} f(x_{n+1} - x_k) \leq \sum_{k,\ell=1}^{n} f(x_k - x_\ell) - f(0)$$

and we see that the quantity is decaying since $f(0) > 0$. However, the quantity cannot decay indefinitely since

$$\sum_{k,\ell=1}^{n} f(x_k - x_\ell) = \sum_{\substack{k\in\mathbb{Z} \\ k\neq 0}} \widehat{f}(k) \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2 \geq 0.$$

This means that the desired inequality has to eventually be true. The argument shows something slightly stronger than this: since

$$\sum_{k,\ell=1}^{n} f(x_k - x_\ell) \leq f(0)n,$$

we can infer that if the L^1-norm is bigger than $2f(0)$ for some fixed n, then it holds true for some $m \leq 2n$. However, we will not need this refined information. $\square$

Proof of Theorem 4.3.2: Summarized from [26]. We now fix such a value of n where the L^1-norm is smaller than $2f(0)$. We argue that

$$\begin{aligned}\left\| \frac{d}{dx} \sum_{m=1}^{n} f(x-x_m) \right\|_{L^2} &= \left\| \frac{d}{dx} \sum_{k \in \mathbb{Z}} \widehat{f}(k) \sum_{m=1}^{n} e^{2\pi i k(x-x_m)} \right\|_{L^2} \lesssim \left\| \frac{d}{dx} \sum_{k \in \mathbb{Z}} \frac{1}{k^2} \sum_{m=1}^{n} e^{2\pi i k(x-x_m)} \right\|_{L^2} \\ &\lesssim \left\| \sum_{k \in \mathbb{Z}} \frac{1}{k} \sum_{m=1}^{n} e^{2\pi i k(x-x_m)} \right\|_{L^2} = \left(\sum_{k \neq 0} \frac{1}{k^2} \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2 \right)^{1/2} \\ &\lesssim \left(\sum_{k,\ell=1}^{n} f(x_k - x_\ell) \right)^{1/2} \lesssim n^{1/2}.\end{aligned}$$

The final ingredient in our argument is a Gagliardo-Nirenberg inequality: for differentiable $g : \mathbb{T} \rightarrow \mathbb{R}$ with mean value 0, we have

$$\|g\|_{L^\infty(\mathbb{T})} \lesssim \left\| \frac{d}{dx} g \right\|_{L^2(\mathbb{T})}^{2/3} \|g\|_{L^1(\mathbb{T})}^{1/3},$$

which establishes the desired result. $\square$

4.5.1 Proof of the Proposition

Proof: Summarized from [26]. We have

$$\sum_{m,\ell=1}^{n} f(x_m - x_\ell) = \sum_{\substack{k \in \mathbb{Z} \\ k \neq 0}} \widehat{f}(k) \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2 \lesssim \sum_{k=1}^{\infty} \frac{1}{k^2} \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2.$$

This quantity was estimated in [131], we summarize the argument here. We observe that we trivially have the inequality

$$\sum_{k=n^2}^{\infty} \frac{1}{k^2} \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2 \leq n^2 \sum_{k=n^2}^{\infty} \frac{1}{k^2} \lesssim 1.$$

It thus remains to estimate the first n^2 sums. We split into dyadic pieces and estimate

$$\sum_{2^\ell \leq k \leq 2^{\ell+1}} \frac{1}{k^2} \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2 \lesssim \frac{1}{2^{2\ell}} \sum_{2^\ell \leq k \leq 2^{\ell+1}} \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2.$$

We recall the geometric series and use it to estimate

$$\left| \sum_{m=1}^{n} e^{2\pi i k m \alpha} \right| = \left| \frac{e^{2\pi i k n \alpha} - 1}{e^{2\pi i k \alpha} - 1} \right| \leq \frac{2}{|e^{2\pi i k \alpha} - 1|} \lesssim \frac{1}{\|k\alpha\|},$$

where $\|x\| = \min(x - \lfloor x \rfloor, \lceil x \rceil - x)$ is the distance to the nearest integer. Since α is assumed to be badly approximable, i.e.

$$\left| \alpha - \frac{p}{q} \right| \geq \frac{c_\alpha}{q^2},$$

we have that, for any $2^\ell \leq k_1 < k_2 \leq 2^{\ell+1}$,

$$|\|k_1\alpha\| - \|k_2\alpha\|| \geq \frac{c_\alpha}{2^{\ell+1}}.$$

Moreover, we also have

$$\frac{c_\alpha}{2^{\ell+1}} \leq \|k_1\alpha\|, \|k_2\alpha\| \leq 1 - \frac{c_\alpha}{2^{\ell+1}}.$$

This shows that the sum

$$\sum_{2^\ell \leq k \leq 2^{\ell+1}} \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2 \lesssim \sum_{2^\ell \leq k \leq 2^{\ell+1}} \frac{1}{\|k\alpha\|^2}$$

can be estimated from above by

$$\sum_{2^\ell \leq k \leq 2^{\ell+1}} \frac{1}{\|k\alpha\|^2} \lesssim_\alpha \sum_{k=1}^{2^\ell} \frac{1}{(k/2^\ell)^2} \lesssim 2^{2\ell}.$$

Altogether this shows that over every dyadic block

$$\sum_{2^\ell \leq k \leq 2^{\ell+1}} \frac{1}{k^2} \left| \sum_{m=1}^{n} e^{2\pi i k x_m} \right|^2 \lesssim \frac{1}{2^{2\ell}} \lesssim 1$$

and thus the sum simplifies to the number of dyadic blocks up to n^2 which is $\sim \log n$. $\square$

4.5.2 Proof of Theorem 4.3.3

Proof: Summarized from [26]. We can see that K is positive-definite and equivalent to the Green's function G. Thus, it suffices to prove the desired result for the Green's function G instead. The proof follows by induction from the main result of [127]. Fixing a $d-$dimensional manifold (M, g) with $d \geq 3$, we have for any set of n points $\{x_1, \ldots, x_n\} \subset M$ that

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}, dx\right) \lesssim_M \frac{1}{n^{1/d}} + \frac{1}{n}\left|\sum_{k\neq\ell} G(x_k, x_\ell)\right|^{1/2}.$$

If the manifold is two-dimensional, $d = 2$, then we have

$$W_2\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{x_k}, dx\right) \lesssim_M \frac{\sqrt{\log n}}{n^{1/2}} + \frac{1}{n}\left|\sum_{k\neq\ell} G(x_k, x_\ell)\right|^{1/2}.$$

We emphasize that $G(\cdot, y)$ has mean value 0 and thus, by the usual argument, we obtain

$$\min_{x\in M}\sum_{k=1}^{n-1} G(x, x_k) \leq 0$$

and thus we may bound

$$\sum_{k\neq\ell} G(x_k, x_\ell) \leq 0.$$

We recall the Corollary from [127] which implies that for *any* set of points

$$\sum_{\substack{k,\ell=1\\k\neq\ell}}^{n} G(x_k, x_\ell) \gtrsim_M -n^{2-2/d}$$

for $d \geq 3$. If the manifold is two-dimensional, then we have the estimate

$$\sum_{\substack{k,\ell=1\\k\neq\ell}}^{n} G(x_k, x_\ell) \gtrsim_M -n\log n.$$

These two results combined imply the desired statement. □

Chapter 5

Explicit Sequences of Points and their Transport Distance

5.1 Introduction

5.1.1 Introduction

Here we study the problem of measuring the regularity of point sets $\{x_1, \ldots, x_N\} \subset \mathbb{T}^d$ as well as infinite sequences. There are many classical notions of regularity (discussed at the beginning of Chapter 1) as well as good constructions of sets minimizing these notions that have been proposed. Typically, the regularity of *sequences* is considered using discrepancy, whereas we would look at the regularity of *measures* using transport distance, though of course we can always interpret sequences of points as sequences of measures, by placing a $\frac{1}{N}\delta_{x_k}$ Dirac measure at each of the first N points. The results of Chapter 4 also suggest that a potential theoretic approach to regularity is useful, interpreting the points as particles with energy interactions. To get the full story, we will need to use multiple approaches in conjunction. This chapter and Chapter 6 follow [25] closely.

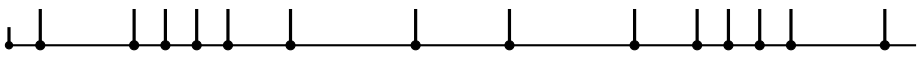

Figure 5.1: The renormalized quadratic residues in $\mathbb{F}_{29}$ rescaled to $[0, 1]$. Every dot except the one at zero represents two quadratic residues corresponding to two Dirac delta measures. How costly is it to move this point measure to the uniform distribution on $[0, 1]$?

The classical theory has developed a useful machinery in terms of exponential sums that exploits 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 $\mathbb{T}^d$? Here, the 'cost' of transporting δ units of measure across distance d is understood to be the W_1 cost, $\delta \cdot d$. An even more practical example is the following: suppose we have people evenly distributed over $\mathbb{T}^d$ and N supermarkets placed in $\{x_1, \dots, x_N\} \subset \mathbb{T}^d$. Demand and supply are exactly matched: how far would the trucks have to drive to distribute the goods from the supermarkets evenly? This is Monge's transportation problem discussed in Chapter 2. It is easy to see that

$$W_p\left(\frac{1}{N}\sum_{k=1}^{N} \delta_{x_k}, dx\right) \geq \frac{c_d}{N^{1/d}},$$

where c_d is a universal constant depending only on the dimension (see Chapter 2 §2 for the argument). This scaling is, for example, assumed by a rescaling of $\mathbb{Z}^d$ intersected with $\mathbb{T}^d \cong [0,1]^d$. This chapter is motivated by the following questions:

1. Do the classical constructions of regular sequences in $\mathbb{T}^d$ from [36, 44, 47, 78] 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 Optimal Transport, where 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 μ to μ or random

points to random points [2, 3, 23, 71, 136–138, 145]. As far as we know, special structures arising from Number Theory or Combinatorics have not been considered before (however, there are some interesting precursors in [21, 25, 65, 127, 131–134]).

5.1.2 Setup.

We recall the $p-$Wasserstein distance discussed in Chapter 2. As throughout the thesis, our two measures under consideration here are

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

where dx refers to the normalized volume measure. As mentioned above, we have an (optimal) lower bound that is independent of the set $\{x_1, \dots, x_N\} \subset \mathbb{T}^d$

$$W_1\left(\frac{1}{N}\sum_{k=1}^{N} \delta_{x_k}, dx\right) \geq \frac{c_d}{N^{1/d}}.$$

5.1.3 Existing Results in One Dimension.

There are several recent results in the one-dimensional setting. Given a finite set on the one-dimensional torus $\{x_1, \dots, x_N\} \subset \mathbb{T}$, we associate to it the measure

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

Then we may bound the transport cost

$$W_1(\mu, dx) \lesssim D_N(\mu)$$

using Kantorovich-Rubinstein duality (this is carried out in greater detail in [21] or [130]). We recall another notion of regularity introduced in Chapter 1 §3: Zinterhof's diaphony

[47, 154], which can be defined as

$$F_N(\mu) = \left(\sum_{\substack{k \in \mathbb{Z} \\ k \neq 0}} \frac{|\widehat{\mu}(k)|^2}{k^2} \right)^{1/2}.$$

One of the key points of this Chapter is that we are able to generalize Zinterhof's diaphony to higher dimensions. Recall Peyre's inequality [106] (see Chapter 2 §4), which says that

$$W_2(\mu, dx) \lesssim F_N(\mu).$$

Summarizing, we have two inequalities and Holder's inequality

$$W_1(\mu, dx) \lesssim D_N(\mu) \qquad \text{and} \qquad W_1(\mu, dx) \leq W_2(\mu, dx) \lesssim F_N(\mu)$$

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. A simple example is given by the van der Corput sequence in base $r \in \mathbb{N}$ (see e.g. [44]). The element x_n is given by writing n in base r, inverting the digits at the comma and then reinterpreting this as a real number; the van der Corput sequence in base 2 starts with $0.5, 0.25, 0.75, 0.125, 0.625$ and so on. It is known to satisfy $D_N \lesssim_r N^{-1} \log N$. Using an existing result of Proinov & Grozdanov [110], we can obtain the following improved estimate on the transport distance.

Theorem 5.1.1 (Proinov & Grozdanov [110])**.** *Let $(x_n)_{n=1}^{\infty}$ denote the van der Corput sequence in base r. Then, uniformly in N,*

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

A recent result of Graham [61] (see Chapter 4 §3.2) shows that this is the optimal rate. Peyre's inequality $F_N \gtrsim W_2(\mu, dx)$ [106], implies the same result for the Zinterhof diaphony which recovers a result of Proinov [108]. A natural question is whether this rate

of growth is attained by other sequences as well. Steinerberger recently remarked [131] 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 [28]).

Theorem 5.1.2 (Kronecker Sequence and Quadratic Residues [131])**.** *Let α be badly approximable and $x_n = \{n\alpha\}$, 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 $x_k = \left\{k^2/p\right\}$ for $1 \leq k \leq p$. Then

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

The bound on the Kronecker sequence follows from applying Peyre's inequality to the diaphony estimate shown in the proof of the Chapter 4 Proposition (see §5.1). Recall from Chapter 1 that the Kronecker sequence has optimal discrepancy on the order of

$$D_N \lesssim \frac{\log N}{N}.$$

Notice this loses a factor of $\sqrt{\log N}$ from our W_2 bound. While D_N is giving the worst case discrepancy over all intervals, analogous to an L^∞ norm, the W_2 transport distance is more similar to an L^2 norm, giving the average distance points must be transported. Thus, it is notable that Kronecker's points are substantially more well-distributed 'on average' than in the worst-case interval. As for the quadratic residues, we will see that the argument from [131] generalizes to any monomial residues, the main difference being that we can complete the square and push this argument through for *any* quadratic polynomial, whereas it is not so simple for higher degrees. That is, for an odd prime p, $a \not\equiv 0 \pmod{p}$, we have

$$ax^2 + bx + c \equiv a(x + (2a)^{-1}b)^2 + c - (4a)^{-1}b^2 \pmod{p},$$

so the residues of any quadratic polynomial are simply the shifted residues of a quadratic

monomial. The matter of whether the result below can be extended to higher degree non-monomials is an interesting open question.

Theorem 5.1.3. *Let p be a prime, n a positive integer and $m \in \mathbb{Z}\backslash p\mathbb{Z}$, and let $x_k = \{mk^n/p\}$ for $1 \leq k \leq p$. Then*

$$W_2\left(\frac{1}{p}\sum_{k=1}^{p}\delta_{x_k}, dx\right) \leq \frac{2\pi}{p\sqrt{3}}\left(1 + (\gcd(n, p-1) - 1)^2\left(p - \frac{1}{p}\right)\right)^{1/2} \lesssim_n \frac{1}{\sqrt{p}}.$$

We present the proof at the end of this section. The above 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 [131]. For that reason, we believe that there are many interesting results in $d = 1$ that are within reach.

5.1.4 Existing Results in Higher Dimensions

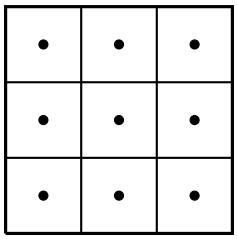

Figure 5.2: The regular grid distribution with small Wasserstein transportation cost – however, these constructions are not uniform in N.

We first present some recent results on randomly selected points in dimension 2. In 2016, Ambrosio-Stra-Trevisan proved the following:

Theorem (Ambrosio-Stra-Trevisan [3])**.** *Suppose $D = [0,1]^2$ or D is a 2-dimensional compact manifold with volume 1, and points x_i are picked uniformly at random over D. Then*

$$\lim_{n\to\infty}\frac{n}{\log n}\mathbb{E}\left[W_2^2\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}, dx\right)\right] = \frac{1}{4\pi}.$$

That is, on average the squared W_2 distance between randomly selected point masses and the uniform distribution tends to $\log n/4\pi n$. Following this result, Ledoux provided a bound on the plane $\mathbb{R}^2$ using similar techniques:

Theorem (Ledoux [82])**.** *Suppose points x_i are picked from the standard Gaussian distribution μ on $\mathbb{R}^2$. Then, for some universal constant $C > 0$ and all $n \geq 2$, we have*

$$\frac{\log n}{Cn} \leq \mathbb{E}\left[W_2^2\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}, \mu\right)\right] \leq \frac{C(\log n)^2}{n}.$$

Ledoux further conjectured that the left-hand side is of the correct order, which is supported by simulations and heuristics. We now consider higher-dimensional manifolds: It is easy to see that for any fixed set of points $\{x_1, \ldots, x_N\} \subset \mathbb{T}^d$, the lattice construction (see Fig. 5.2) 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 5.2: where would one put the next point and the point after that? A general result has recently been obtained by the author and Steinerberger [26] 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 (see Theorem 4.3.3 in Chapter 4 §3.4).

5.1.5 Proof of Theorem 5.1.3: Monomial Residues

Proof. Consider a monomial $f(x) = mx^n$, where n is a positive integer and $m \in \mathbb{Z}$, and let

$$\mu_p = \frac{1}{p}\sum_{k=1}^{p}\delta_{\{f(k)/p\}},$$

where $p \nmid m$ is an odd prime. That is, place a point mass of weight $1/p$ at each of the residues of f in $\mathbb{F}_p$ (with multiplicity). Then the Fourier coefficients of μ_p given by

$$\widehat{\mu}_p(j) = \int_0^1 e^{-2\pi i j x} d\mu_p = \frac{1}{p}\sum_{k=1}^{p} e^{-2\pi i j f(k)/p}$$

are Gauss sums. If $p \mid j$ then clearly $\widehat{\mu}_p(j) = 1$, so we restrict to the case $p \nmid j$. Then, we may express these Fourier coefficients in terms of the non-trivial Dirichlet characters

$\chi : (\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{S}^1$ of order dividing n:

$$\widehat{\mu}_p(j) = \frac{1}{p} \sum_{\substack{\chi^n=1 \\ \chi \neq 1}} \tau_j(\chi) \qquad \text{where} \qquad \tau_j(\chi) = \sum_{k=1}^{p-1} \chi(k) e^{-2\pi i k j m/p}.$$

We can see this as follows: first, recall that $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic—fix a generator g. Then the Dirichlet characters of order dividing n are determined by their value at g. We may, without loss of generality, assume $n \mid p-1$; otherwise, we can simply replace n by $\gcd(n, p-1)$ and receive the same set of residues. Letting $\zeta = e^{2\pi i/n}$, the Dirichlet characters are precisely those given by $\chi_q(g) = \zeta^q$. Note

$$\sum_{k=1}^{p} e^{-2\pi i k j m/p} = 0,$$

as this is summing over all p of the pth roots of unity. Thus, we have

$$\sum_{\substack{\chi^n=1 \\ \chi \neq 1}} \tau_j(\chi) = \sum_{\substack{\chi^n=1 \\ \chi \neq 1}} \sum_{k=1}^{p-1} \chi(k) e^{-2\pi i k j m/p} + \sum_{k=1}^{p} e^{-2\pi i k j m/p}.$$

We may switch the order of summation: if we fix $k = g^x$ and consider the total contribution of $e^{-2\pi i k j m/p}$ to the sum, we see that this will be the sum over all χ of order dividing n (*including* $\chi = 1$ now) of $\chi(k) e^{-2\pi i k j m/p}$. But, the $\chi(k)$ are simply running over each of the $n/\gcd(n, x)$ roots of unity $\gcd(n, x)$ times. This is always 0, unless $n = \gcd(n, x)$—that is, unless $n \mid x$, in which case it is adding n copies of 1. $n \mid x$ precisely when k is an nth power residue mod p. Thus, the summation does exactly what we want: it eliminates the contribution of all non-residues, and extracts the correct contribution from each residue, which necessarily appears as an nth power of n elements in $\mathbb{Z}/p\mathbb{Z}$ (except if $k = 0$, where it appears as an nth power of precisely one element—0). Since p is prime and χ is non-trivial, it is automatically primitive, and thus we have $|\tau_j(\chi)| = \sqrt{p}$ (this can be seen by an elementary argument: simply expand the sum $\tau_j(\chi)\overline{\tau_j(\chi)}$). Since there are precisely $n-1$ non-trivial characters of order dividing n, we have by the triangle inequality, for $j \notin p\mathbb{Z}$,

$$|\widehat{\mu}_p(j)| \leq \frac{1}{p}(n-1)\sqrt{p} = \frac{n-1}{\sqrt{p}} \lesssim_n \frac{1}{\sqrt{p}}.$$

Note that we may bound the discrepancy with the Erdős-Turán inequality:

$$D_N \lesssim \frac{1}{q} + \sum_{k=1}^{q} \frac{|\widehat{\mu}_p(k)|}{k} \lesssim_n \frac{1}{q} + \sum_{\substack{1 \leq k \leq q \\ p|k}} \frac{1}{k} + \frac{1}{\sqrt{p}} \sum_{k=1}^{q} \frac{1}{k} \lesssim \frac{1}{q} + \frac{1}{p} \log\left(\frac{q}{p}\right) + \frac{\log q}{\sqrt{p}}.$$

Setting $q = p$ yields $D_N \lesssim_n \log p/\sqrt{p}$. To bound W_2, we use Peyre's inequality:

$$\begin{aligned} W_2(\mu_p, dx) &\leq 2 \left(\sum_{\substack{k \in \mathbb{Z} \\ k \neq 0}} \frac{|\widehat{\mu}_p(k)|^2}{k^2} \right)^{1/2} \leq 2 \left(\sum_{\substack{k \in p\mathbb{Z} \\ k \neq 0}} \frac{1}{k^2} + \sum_{k \notin p\mathbb{Z}} \frac{(n-1)^2}{pk^2} \right)^{1/2} \\ &= \frac{2\pi}{p\sqrt{3}} \left(1 + (n-1)^2 \left(p - \frac{1}{p} \right) \right)^{1/2} \lesssim_n \frac{1}{\sqrt{p}}. \end{aligned}$$

□

As with the Kronecker sequence, it is worth noting that the Wasserstein bound is better than the discrepancy bound, here by a factor of $\log p$ (though the true values are unknown).

5.2 Main Results

5.2.1 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 of points but a sequence of probability measures. Let μ_k be the measure that arises from an unbiased random walk on $\mathbb{T} \cong [0, 1]$ where each step is $\pm\alpha$ (independently and with likelihood 1/2 each) and α is a quadratic irrational. This model was studied by Su [132] (see also Hensley & Su [65] and Su [134]). The main result in [132] showed that the measure arising after k random steps satisfies

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

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 5.2.1 (B. & Steinerberger [25])**.** *We have*

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

We emphasize that the framework discussed in this chapter enables us to reduce Theorem 5.2.1 to standard estimates. This is presumably not optimal and stronger results should be true. Hensley & Hu [65] 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.

5.2.2 Kronecker sequences.

We now consider a natural higher-dimensional generalization of Kronecker sequences (irrational rotations) on $\mathbb{T}$. We say that a vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d) \in \mathbb{R}^d$ is badly approximable if, for all $p \in \mathbb{Z}^d$ and $q \neq 0$, we have

$$\|qv - p\| \geq \frac{c_\alpha}{q^{1/d}}.$$

Since p does not appear on the right-hand side, we may alternatively write

$$\min_{p \in \mathbb{Z}} \|qv - p\| \geq \frac{c_\alpha}{q^{1/d}}.$$

Further, since all norms are equivalent in finite-dimensional vector spaces, we may replace the ℓ^2 norm with an ℓ^∞ norm. This yields

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

where $\|\cdot\|$ is the distance to the nearest integer. By Dirichlet's approximation theorem, this is the optimal scaling: for any $\alpha \in \mathbb{R}^d$ there is always some c_α such that, for infinitely many $q \in \mathbb{Z}\backslash\{0\}$, we have

$$\max_{1 \leq j \leq d} \|\alpha_j q\| \leq \frac{a_\alpha}{q^{1/d}}.$$

The existence of badly approximable vectors follows from continued fraction expansion when $d = 1$. The first examples in higher dimensions are due to Perron [102], Davenport [42] showed that there are uncountably many such vectors for $d = 2$ and Schmidt [123] extended

this result to $d \geq 3$. The Kronecker sequence is then defined via

$$x_n = (n\alpha_1, n\alpha_2, \dots, n\alpha_d) \mod 1,$$

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 5.2.2 (B. & Steinerberger [25]). *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 this 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 α being badly approximable is necessary; however, in light of results in $d = 1$, this is quite conceivable.

5.2.3 Other manifolds.

Nothing about our approach is particularly tied to the torus $\mathbb{T}^d$. Indeed, Theorem 3.1.1

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

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

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

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 [131]) assumes the form

$$W_2\left(\frac{1}{N}\sum_{k=1}^{N} \delta_{x_k}, dx\right) \lesssim_M \inf_{t>0} \quad \sqrt{t} + \left(\sum_{k=1}^{\infty} \frac{e^{-2\lambda_k t}}{\lambda_k} \left|\frac{1}{N}\sum_{n=1}^{N} \phi_k(x_n)\right|^2\right)^{1/2}.$$

For most manifolds, we do not have an explicit expression for the eigenfunctions ϕ_k and the inequality is thus of limited use. Chapter 3 §2 recalls a substitute inequality in cases where the Green's function $G(x,y)$ or good estimates for it are known [127]. 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. [55, 58–60, 94]). 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 5.2.3 (B. & Steinerberger [25]). *Let (M,g) be a compact manifold without boundary, normalized to have volume 1, and let $f : \mathbb{T} \rightarrow \mathbb{R}$ be differentiable. Then, for some constant $c_M > 0$ depending only on the manifold, we have*

$$\left| \int_{\mathbb{M}} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \leq c_M \inf_{t>0} \left[\sqrt{t}\|\nabla f\|_{L^\infty} + \left\| e^{t\Delta} \sum_{k=1}^{N} \delta_{x_k} \right\|_{\dot{H}^{-1}} \|\nabla f\|_{L^2} \right]$$

Alternatively, rewriting the Sobolev norm in terms of the spectral expansion, we could also write the upper bound on the integration error as

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

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

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

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 the lemma from below)

$$\frac{1}{N}\sum_{k=1}^{\infty} \frac{e^{-2\lambda_k t}}{\lambda_k} \lesssim_M \frac{1}{N}\sum_{k=1}^{\infty} \frac{e^{-k^{2/d} t}}{k^{2/d}} \lesssim_d \frac{1}{N} t^{-\frac{d-2}{2}},$$

for $0 < t < 1/2$. 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

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

However, this is inferior to classical Monte-Carlo and thus perhaps not useful.

5.3 Proofs

5.3.1 A recurring computation.

We collect a simple Lemma that will reappear in several different arguments.

Lemma (B. & Steinerberger [25])**.** *We have, for $m + d \geq 1$, the estimate*

$$\sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \|k\|^m \lesssim_{m,d} t^{-\frac{m+d}{2}}.$$

If $m + d = 0$, then we have, for $0 < t < 1/2$,

$$\sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \|k\|^m \lesssim_{m,d} \log\left(\frac{1}{t}\right).$$

Proof: Summarized from [25]. 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

$$\sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \|k\|^m \lesssim_d \sum_{\substack{k \in \mathbb{Z} \\ k \neq 0}} e^{-|k|^2 t} |k|^{m+d-1}.$$

If $m + d = 0$, then we can easily bound the sum via

$$\sum_{\substack{k\in\mathbb{Z}\\ k\neq 0}} e^{-|k|^2 t}|k|^{-1} \lesssim \int_1^\infty \frac{e^{-x^2 t}}{x} dx$$

This integral is the complete gamma function and can be rewritten in terms of the exponential integral via

$$\int_1^\infty \frac{e^{-x^2 t}}{x} dx = -\frac{1}{2}\mathrm{Ei}(-t) = \frac{1}{2}\int_t^\infty \frac{e^{-x}}{x} dx.$$

It is easy to see that

$$\int_t^\infty \frac{e^{-x}}{x} dx \lesssim \int_t^1 \frac{1}{x} dx + \int_1^\infty \frac{e^{-x}}{x} dx \lesssim \log\left(\frac{1}{t}\right).$$

It remains to deal with the case in which $m + d \geq 1$, where we estimate the sum via a different integral. Note that

$$\sum_{\substack{k\in\mathbb{Z}\\ k\neq 0}} e^{-|k|^2 t}|k|^{m+d-1} \lesssim \int_0^\infty e^{-x^2 t} x^{m+d-1} dx = c_{m+d} t^{-\frac{m+d}{2}}. \qquad \square$$

5.3.2 Random Walks: Proof of Theorem 5.2.1.

Proof: Summarized from [25]. We have that the measure μ_k describing the distribution of the random walk after k steps is given by

$$\mu_k = \mu_{k-1} * \mu \qquad \text{where } * \text{ denotes convolution, and} \qquad \mu = \frac{1}{2}\delta_\alpha + \frac{1}{2}\delta_{-\alpha}.$$

Therefore

$$|\widehat{\mu_k}(\ell)| = |\widehat{\mu}(\ell)|^k = |\cos(2\pi\ell\alpha)|^k.$$

Using Peyre's estimate, we reduce the problem to estimating the sum

$$W_2(\mu_k, dx) \leq \left(\sum_{\substack{\ell\in\mathbb{Z}\\ \ell\neq 0}} \frac{|\cos(2\pi\ell\alpha)|^{2k}}{\ell^2}\right)^{1/2}.$$

We use, as we often do, that $\ell\alpha$ cannot be close to an integer for many values of ℓ. 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 α is badly approximable, we have that two distinct elements $\ell_1, \ell_2 \in I_j$ satisfy $|\ell_1 - \ell_2| \gtrsim_\alpha k$. We can now write

$$\sum_{\substack{\ell \in \mathbb{Z} \\ \ell \neq 0}} \frac{|\cos(2\pi\ell\alpha)|^{2k}}{\ell^2} = \sum_{j=0}^{k-1} \sum_{\ell \in I_j} \frac{|\cos(2\pi\ell\alpha)|^{2k}}{\ell^2}.$$

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 $\gtrsim_\alpha k/(j+1)$ and any two consecutive elements are $\gtrsim_\alpha k$ separated implying that

$$\sum_{\ell \in I_j} \frac{1}{\ell^2} \lesssim_\alpha \sum_{h=0}^{\infty} \frac{1}{(k/(j+1) + hk)^2} \lesssim_\alpha \frac{(j+1)^2}{k^2}.$$

However, we also have

$$\max_{x \in I_j} |\cos(2\pi x)|^{2k} \leq \left(1 - \left(\frac{\min\{j, k-j\}}{k}\right)^2\right)^{2k}.$$

By symmetry, it suffices to sum j up to $k/2$. We then obtain

$$\begin{aligned} \sum_{0 \leq j \leq k/2} \max_{x \in I_j} |\cos(2\pi x)|^{2k} \sum_{\ell \in I_j} \frac{1}{\ell^2} &\lesssim \sum_{0 \leq j \leq k/2} \left(1 - \frac{j^2}{k^2}\right)^{2k} \frac{(j+1)^2}{k^2} \\ &\lesssim k \int_0^1 \left(1 - x^2\right)^{2k} x^2 dx \lesssim \frac{1}{\sqrt{k}}. \end{aligned}$$

□

5.3.3 Kronecker sequences: Proof of Theorem 5.2.2

Proof: Summarized from [25]. Let us consider the Kronecker sequence

$$x_n = (n\alpha_1, n\alpha_2, n\alpha_3, \ldots, n\alpha_d) \bmod 1.$$

We assume that α 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 [120]) states that α is badly approximable if and only if the linear form induced by α 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_α 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,

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

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

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

We split frequencies into dyadic scales and first estimate

$$\sum_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{1}{\|\langle k, \alpha \rangle\|^2}.$$

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

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

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

$$\sum_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{1}{\|\langle k, \alpha \rangle\|^2} \lesssim_\alpha \sum_{h=1}^{2^{\ell \cdot d}} \frac{1}{(h 2^{-\ell \cdot d})^2} \lesssim 2^{2\ell \cdot d}.$$

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

$$\begin{aligned}
\sum_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{e^{-\|k\|^2 t}}{\|k\|^2} \frac{1}{\|\langle k, \alpha \rangle\|^2} &\leq \left(\max_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{e^{-\|k\|^2 t}}{\|k\|^2} \right) \sum_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{1}{\|\langle k, \alpha \rangle\|^2} \\
&\lesssim_\alpha \left(\max_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{e^{-\|k\|^2 t}}{\|k\|^2} \right) 2^{2\ell \cdot d} \\
&\lesssim_d \left(\max_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{e^{-\|k\|^2 t}}{\|k\|^2} \right) \sum_{2^\ell \le \|k\| \le 2^{\ell+1}} 2^{\ell \cdot d} \\
&\leq \sum_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{e^{-\|k\|^2 (t/2)}}{(\|k\|/2)^2} 2^{\ell \cdot d}.
\end{aligned}$$

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

$$\sum_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{e^{-\|k\|^2 (t/2)}}{(\|k\|/2)^2} 2^{\ell \cdot d} \lesssim_d \sum_{2^\ell \le \|k\| \le 2^{\ell+1}} \frac{e^{-\|k\|^2 (t/2)}}{\|k\|^2} \|k\|^d$$

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

$$\sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} \frac{e^{-\|k\|^2 t}}{\|k\|^2} \frac{1}{\|\langle k, \alpha \rangle\|^2} \lesssim \sum_{\substack{k \in \mathbb{Z}^d \\ k \neq 0}} e^{-\|k\|^2 t} \|k\|^{d-2} \lesssim \frac{1}{t^{d-1}}.$$

Therefore

$$W_2(\mu, dx) \lesssim_\alpha \sqrt{t} + \frac{2}{N}\frac{1}{t^{\frac{d-1}{2}}}$$

which implies, for the choice $t = N^{-2/d}$ that

$$W_2(\mu, dx) \lesssim_\alpha \frac{1}{N^{1/d}}. \qquad \square$$

5.3.4 A General Manifold Result: Proof of Theorem 5.2.3

Proof: Summarized from [25]. The proof combines two estimates. We first replace the point measure on the x_k

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

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 equality, the Kantorovich-Rubinstein duality formula, in one dimension, see Chapter 2 §1). However, it is known that (see Chapter 3 §1 Lemma)

$$W_1(\mu, e^{t\Delta}\mu) \lesssim_M \sqrt{t}$$

and we obtain 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 δ_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

$$\begin{aligned}
\left|\int_M f(x)d\mu - \int_M f(x)e^{t\Delta}\mu dx\right| &= \left|\int_M \frac{1}{N}\sum_{k=1}^N (p_t(x_k,y)f(y) - f(x_k))dy\right| \\
&\leq \frac{1}{N}\sum_{k=1}^N \left|\int_M p_t(x_k,y)f(y) - f(x_k)dy\right| \\
&\leq \frac{1}{N}\sum_{k=1}^N \left|\int_M p_t(x_k,y)f(y) - p_t(x_k,y)f(x_k)dy\right| \\
&\leq \frac{1}{N}\sum_{k=1}^N \int_M \|\nabla f\|_{L^\infty} p_t(x_k,y)|x_k - y|dy \\
&\leq \|\nabla f\|_{L^\infty} \max_{x\in M}\int_M p_t(x,y)|x-y|dy.
\end{aligned}$$

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. [131]) 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

$$\left|\int_M f(x)dx - \int_M f(x)e^{t\Delta}\mu\ dx\right| \qquad \text{from above.}$$

We interpret this as an inner product

$$\left|\int_M f(x)dx - \int_M f(x)e^{t\Delta}\mu\ dx\right| = \left|\left\langle f, e^{t\Delta}\mu - 1\right\rangle\right|.$$

A duality argument now shows that

$$\left|\left\langle f, e^{t\Delta}\mu - 1\right\rangle\right| \leq \|f\|_{\dot{H}^1} \|e^{t\Delta}\mu\|_{\dot{H}^{-1}}$$

which is the desired result. One could also avoid the language of functional analysis and estimate, after noticing that $e^{t\Delta}\mu - 1$ and ϕ_k both have mean value 0 for $k \geq 1$,

$$\begin{aligned}
\left|\left\langle f, e^{t\Delta}\mu - 1\right\rangle\right| &= \left|\sum_{k=0}^{\infty} \left\langle f, \phi_k\right\rangle \left\langle e^{t\Delta}\mu - 1, \phi_k\right\rangle\right| = \left|\sum_{k=1}^{\infty} \lambda_k^{1/2} \left\langle f, \phi_k\right\rangle \lambda_k^{-1/2} \left\langle e^{t\Delta}\mu - 1, \phi_k\right\rangle\right| \\
&\leq \left(\sum_{k=1}^{\infty} \lambda_k \left\langle f, \phi_k\right\rangle^2\right)^{1/2} \left(\sum_{k=1}^{\infty} \lambda_k^{-1} \left\langle e^{t\Delta}\mu - 1, \phi_k\right\rangle^2\right)^{1/2} \\
&= \left(\sum_{k=1}^{\infty} \lambda_k \left\langle f, \phi_k\right\rangle^2\right)^{1/2} \left(\sum_{k=1}^{\infty} \lambda_k^{-1} \left\langle e^{t\Delta}\mu, \phi_k\right\rangle^2\right)^{1/2}.
\end{aligned}$$

As for the first term, we observe that

$$\sum_{k=1}^{\infty} \lambda_k \left\langle f, \phi_k\right\rangle^2 = \int_M (-\Delta f) f dx = \int_M |\nabla f|^2 dx = \|\nabla f\|_{L^2}^2.$$

As for the second sum, we observe that, using the self-adjointness of the heat propagator and the fact that ϕ_k is an eigenfunction of the Laplacian

$$\left\langle e^{t\Delta}\mu, \phi_k\right\rangle = \left\langle \mu, e^{t\Delta}\phi_k\right\rangle = e^{-\lambda_k t} \left\langle \mu, \phi_k\right\rangle.$$

This then results in

$$\left(\sum_{k=1}^{\infty} \lambda_k^{-1} \left\langle e^{t\Delta}\mu, \phi_k\right\rangle^2\right)^{1/2} = \left(\sum_{k=1}^{\infty} \frac{e^{-2\lambda_k t}}{\lambda_k} \left\langle \mu, \phi_k\right\rangle^2\right)^{1/2}$$

and concludes the desired result. □

Chapter 6

Numerical Integration and Error Bounds

6.1 Recent Results

Let us consider the problem of numerically integrating a function $f : [0,1]^d \to \mathbb{R}$ which we assume to be Lipschitz. It is a classic 1959 result of Bakhvalov [7] (see also Novak [97]) that there are sets of points $(x_k)_{k=1}^N$ such that for all differentiable functions $f : [0,1]^d \to \mathbb{R}$ with Lipschitz constant $\|\nabla f\|_{L^\infty}$, we have

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

and that this result is optimal in the power of N and its dependence on the Lipschitz constant $\|\nabla f\|_{L^\infty}$: there are functions f for which the error is at that scale (up to constants). We can easily construct such functions: consider the Lipschitz function

$$f(x) = \min_{1 \leq i \leq n} \|x - x_i\|.$$

Then the sum $\frac{1}{N} \sum_{k=1}^{N} f(x_k) = 0$, and the true value of the integral will be on the order of $N^{-1/d}$, the average distance to points in the set (see Chapter 2 §2 for the argument in the context of transport distance). In fact, the traditional problem formulation presumes as a

foregone conclusion a linear dependence on the Lipschitz constant: rather than writing the inequality as above, authors simply restrict their attention to the unit ball of of Lipschitz functions in F_d^{lip}, with norm

$$\|f\| = \max\left(\|f\|_{L^\infty}, \sup_{x,y \in [0,1]^d} \frac{|f(x) - f(y)|}{\|x - y\|_\infty} \right) \leq 1$$

and write instead

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

Recently, Hinrichs, Novak, Ullrich and Wozniakowski [68] (see also [66, 67]) 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 $N = m^d$. Bakhalov's proof shows that C_d can in fact be taken to be C^d for some universal constant $C > 1$ [98]. The theorem also shows conversely that this is sharp, in the sense that the following lower bound exists as well: for any set of points $(x_k)_{k=1}^N$, there is a function f with $\|f\| = 1$ such that

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

where c_d again can be taken to be c^d for some universal constant $0 < c < 1$. The proof in fact does not require points to be uniformly weighted, nor even linearly aggregated at all: it applies to the worst case error of *any* algorithm which takes the values at points x_k as inputs and outputs an integral estimate. While Bakhalov's result is sharp for worst case functions, we may nonetheless asymptotically improve it in the general case by considering a different norm on f.

6.2 A New Result: Kronecker Integration

Theorem 6.2.1 (B. & Steinerberger [25]). *Let $d \geq 2$ and let $\alpha \in \mathbb{R}^d$ be a badly approximable vector. Then, for some $c_\alpha > 0$ and all differentiable $f : \mathbb{T}^d \rightarrow \mathbb{R}$*

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

The main novelties are that:

1. The result holds uniformly in N along a sequence. In the classical theory, we are told in advance how many points need to be picked and get to distribute them to minimize integration error, but it is a harder problem to generate an infinite sequence of points which integrates well no matter how many terms we take. Theorem 6.2.1 shows that we can recover the classical estimate and even do slightly better in this setting.

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: given *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. Indeed, the result is actually true on general $d-$dimensional manifolds, we refer to Theorem

5.2.3. 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

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

6.3 The Case of the Regular Grid

Let us return to the case of the regular grid, with $N = m^d$ 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 sampling Lipschitz functions). The classic estimate for a regular grid $(x_n)_{n=1}^N$ is

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

Sukharev [135] (see also [98]) proved the sharp constant is $d/(2d+2)$. It is known that 'the result cannot be significantly improved for uniformly continuous functions' (Dick & Pillichshammer [44, §1.3]). Indeed, there is a corresponding result of Larcher (unpublished, see [44, §1.3]) that shows that the estimate is optimal with regards to modulus of continuity. However, there is an explicit improvement in terms of L^p-spaces that seems to be new.

Theorem 6.3.1 (B. & Steinerberger [25])**.** *We have, for some explicit constant c_d depending on the dimension, for all differentiable $f : [0,1]^d \to \mathbb{R}$ sampled on the regular grid $(x_k)_{k=1}^N$*

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

We observe that this is a slightly better estimate than Theorem 5.2.4 (an $L^1(\mathbb{T}^d)$ norm instead of the larger $L^2(\mathbb{T}^d)$ norm); this is maybe to be expected since one would assume that stronger estimates become available for the regular grid. We will also show that this is the best possible bound in terms of these L^p-spaces. It is an interesting question whether this bound (L^1 instead of L^2) is also true for the Kronecker sequence (Theorem 6.2.1).

More generally, one could ask whether there is a sequence $(x_n)_{n=1}^{\infty}$ that uniformly attains the same error estimate as Theorem 6.3.1.

6.4 Error Bounds from Kantorovich-Rubinstein

Recall from Chapter 2 §1, the Kantorovich-Rubinstein formula: the optimal transport cost between two measures μ, ν is given by

$$W_1(\mu, \nu) = \sup \left\{ \int_M \phi d\mu - \int_M \phi d\nu \right\},$$

where the supremum is taken over all 1-Lipschitz functions ϕ. Taking $\mu = dx$ and $\nu = \frac{1}{N}\sum_{k=1}^{N} \delta_{x_k}$ reveals that numerical integration error and optimal transport are essentially asking the same question! More precisely, the worst case integration error over all 1-Lipschitz functions sampling over a given set $(x_k)_{k=1}^{N}$ is precisely the W_1 transport cost between the point measure $\frac{1}{N}\sum_{k=1}^{N} \delta_{x_k}$ and dx. This allows us to refine the rate of growth on the constants with respect to dimension in the regular grid case: by the argument provided in Chapter 2 §2, both c_d and C_d are $\sim \sqrt{d}$. See [151] for more results using this approach to bounding numerical integration error through Kantorovich-Rubinstein duality.

6.5 Numerical Integration: Proof of Theorems

6.5.1 Integration Error of Kronecker Sequences: Proof of Theorem 6.2.1

Proof: Summarized from [25]. Having proven Theorem 5.2.3 in the previous chapter, we can now outline a proof of Theorem 6.2.1 which follows quite easily by combining several of our existing arguments. We will make use of the inequality

$$\left| \int_{\mathbb{M}} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \leq c_M \inf_{t>0} \left[\sqrt{t}\|\nabla f\|_{L^\infty} + \left\| e^{t\Delta} \sum_{k=1}^{N} \delta_{x_k} \right\|_{\dot{H}^{-1}} \|\nabla f\|_{L^2} \right]$$

in the special case where the manifold is given by $M = \mathbb{T}^d$ and the set of points is given by

$$x_n = (n\alpha_1, \ldots, n\alpha_d) \bmod 1$$

where α is badly approximable. The only quantity that requires computation is the $\dot{H}^{-1}$ norm. This, however, was already done in the proof of Theorem 5.2.2 where we used that

$$W_2(\mu, dx) \lesssim_d \sqrt{t} + W_2(e^{t\Delta}\mu, dx)$$

and then estimated that

$$W_2(e^{t\Delta}\mu, dx) \lesssim_{\alpha,d} \frac{1}{N}\frac{1}{t^{\frac{d-1}{2}}}.$$

This results in

$$\left| \int_{\mathbb{M}} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \lesssim_{d,\alpha} \inf_{t>0} \left[\sqrt{t}\|\nabla f\|_{L^\infty} + \frac{1}{N}\frac{1}{t^{\frac{d-1}{2}}}\|\nabla f\|_{L^2} \right].$$

We set

$$t = \frac{1}{N^{2/d}} \frac{\|\nabla f\|_{L^2}^{2/d}}{\|\nabla f\|_{L^\infty}^{2/d}}.$$

This results in

$$\left| \int_{\mathbb{M}} f(x)dx - \frac{1}{N}\sum_{k=1}^{N} f(x_k) \right| \lesssim_{d,\alpha} \frac{1}{N^{1/d}} \|\nabla f\|_{L^2}^{\frac{1}{d}} \|\nabla f\|_{L^\infty}^{\frac{d-1}{d}}. \qquad \square$$

6.5.2 Proof of Theorem 6.3.1

We now turn to Theorem 6.3.1, the case of the regular grid. The proof is based on a simple Poincaré-type inequality for Lipschitz functions vanishing at a fixed point.

Lemma (See e.g. [51], notes below)**.** *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

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

The inequality is not new and follows from combining two known results. The inequality

$$\left| \int_{[0,1]^d} f(x)dx \right| \lesssim \int_{[0,1]^d} \frac{|\nabla f|}{|x|^{d-1}} dx$$

is used as a first step in the proof of Morrey's inequality (see Evans [51, §5.6.2]). This is now combined with an interpolation estimate: it is easy to see that the function $g(x) = |x|^{1-d}$ is contained in the Lorentz space $L^{\frac{d}{d-1},\infty}$. Thus, by the Hölder inequality in Lorentz spaces due to O'Neil [99], we have

$$\int_{[0,1]^d} \frac{|\nabla f|}{|x|^{d-1}} dx \lesssim \|f\|_{L^{d,1}}.$$

We recall the definition of the $L^{d,1}$ norm and use the Hölder inequality to obtain

$$\begin{aligned}
\|f\|_{L^{d,1}} &= d \cdot \|\lambda \cdot |\{|f| > \lambda\}|^{1/d}\|_{L^1(\frac{d\lambda}{\lambda})} = d \cdot \int_0^\infty |\{|f| > \lambda\}|^{1/d} d\lambda \\
&= d \cdot \int_0^{\|f\|_{L^\infty}} |\{|f| > \lambda\}|^{1/d} d\lambda \\
&\leq d \cdot \|f\|_{L^\infty}^{\frac{d-1}{d}} \cdot \left(\int_0^\infty |\{|f| > \lambda\}| \, d\lambda \right)^{\frac{1}{d}} = d \cdot \|f\|_{L^\infty}^{\frac{d-1}{d}} \|f\|_{L^1}^{\frac{1}{d}}.
\end{aligned}$$

Using this simple statement, we can now prove Theorem 6.3.1.

Proof: Summarized from [25]. The proof of Theorem 6.3.1 follows easily from the Lemma which we apply, in isolation, to each fundamental cell of size $N^{-1/d}$. Rescaling the inequality in the Lemma then shows that for any such box $B = [0, N^{-1/d}]^d$, we have

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

Summing over all boxes leads to

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

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

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

Optimality. We quickly construct an example showing that our result is optimal. Let us consider the Lipschitz 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

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

while also observing that

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

By letting $\varepsilon \to 0$, we see that our estimate has the optimal exponents.

Chapter 7

Wasserstein on Graphs

7.1 Introduction

7.1.1 Introduction.

This section follows [27] closely. The purpose of this section is to explore a basic problem on metric spaces, already discussed throughout the earlier chapters on manifolds, in the setting of a finite graph $G = (V, E)$. Given a graph, how does one construct a sequence $x_1, x_2, \ldots$ of vertices such that their distribution is *uniformly good*—by this we mean that if one takes the first k vertices $\{x_1, \ldots, x_k\}$, then this set is very nearly as evenly distributed on the set as any set of k vertices would be. As noted earlier, the precise notion of 'well-distributed' under inspection will depend on the actual setting; the question is frequently interesting for several different such notions.

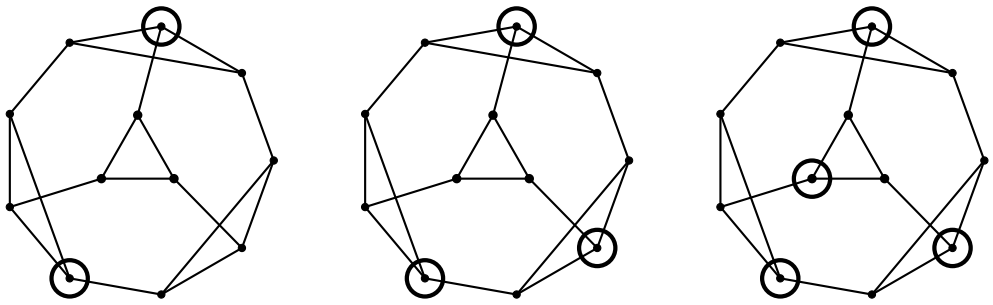

Figure 7.1: The first k (here $k = 2, 3, 4$) elements of the sequence are nearly as evenly distributed as any set of k vertices could be.

There are many different reasons why one could be interested in such sequences: they

are natural sampling points for functions (especially for on-line selection and in cases where one does not know in advance how many points one can sample) but there is also an obvious combinatorial question ('How well distributed can sequences of vertices on graphs be? What is the unavoidable degree of irregularity?'). We will now state one informal version of the main problem before stating a more precise version further below.

> **Main Problem** (informal version)**.** Given a finite graph $G = (V, E)$, how would one select a sequence of vertices that are *uniformly good*? In what metric would one measure the 'goodness' of such a sequence?

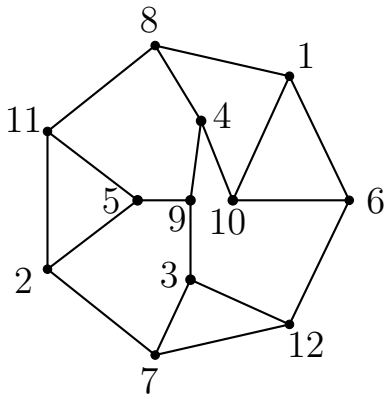

Figure 7.2: The Frucht Graph and the enumeration obtained by the algorithm when starting with $x_1 = 1$.

Fig. 7.2 contains a simple such example: taking the Truncated Tetrahedral Graph, in which order should one select the vertices so as to obtain a sequence that is uniformly evenly distributed? Even without making the notion of quality precise, we can get some intuition from this simple example. The enumeration of the vertices was automatically generated by the algorithm discussed below.

7.1.2 Wasserstein Distance.

The Wasserstein distance W_p, as well as the appropriate formulation on graphs, were introduced in Chapter 2 §2 and §5, respectively. As a reminder, on an abstract metric space X

equipped with a metric d, we define

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

where $\Gamma(\mu,\nu)$ denotes the collection of all measures on $X \times X$ with marginals μ and ν, respectively (also called the set of all couplings of μ and ν). We will, throughout this section, work exclusively with the Earth Mover Distance W_1 (although extensions to more general W_p are certainly conceivable). The Earth Mover's Distance is particularly nice to work with: by Kantorovich-Rubinstein duality (see Chapter 2)

$$W_1(\mu,\nu) = \sup \left\{ \int_X f d\mu - \int_X f d\nu : f \text{ is 1-Lipschitz} \right\}.$$

Because of this, W_1 is translation invariant—for positive measures μ, ν, μ', we have

$$W_1(\mu,\nu) = W_1(\mu+\mu', \nu+\mu').$$

Thus, if we have a positive measure μ decomposed into non-positive measures μ_1, μ_2 as $\mu = \mu_1 + \mu_2$, we may use this translation invariance to write

$$W_1(\mu,\nu) = W_1(\mu_1^+ + \mu_2^+, \nu + \mu_1^- + \mu_2^-),$$

where $\mu_i^+ = \max\{\mu_i, 0\}$ is the positive part of μ_i and $\mu_i^- = \max\{-\mu_i, 0\}$ is the negative part. One natural way of making the question precise is thus as follows.

> **Main Problem** (formal version)**.** Given a finite graph $G = (V, E)$, how would one select a sequence of vertices such that
>
> $$W_1\left(\frac{1}{k} \sum_{j=1}^{k} \delta_{x_j}, dx \right) \qquad \text{is small for all } k,$$
>
> where dx is the normalized counting measure with weight $|V|^{-1}$ on each vertex.

How small one could expect this quantity to be will depend on the particular geometry of the graph. If $M = \mathbb{T}^d$ is the $d-$dimensional torus, then, as shown in Chapter 2,

$$W_1\left(\frac{1}{n}\sum_{j=1}^{n}\delta_{x_j}, dx\right) \geq c_d n^{-1/d}$$

for all sets of points $\{x_1, \ldots, x_n\} \subset \mathbb{T}^d$. This clearly shows that the geometry (here: the dimension d) plays a role in what we can expect. We also emphasize that it is almost surely the case that our main question (as asked in §1.1) is of interest also for many other ways of making the notion of even distribution quantitative and Wasserstein distance may be one of many (though certainly a rather canonical one).

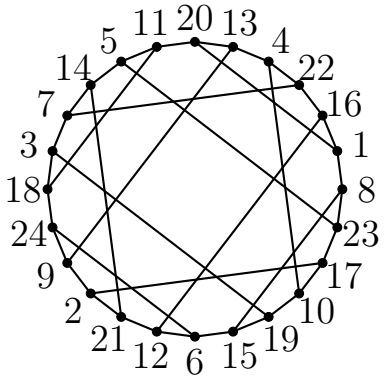

Figure 7.3: The Nauru Graph on 24 vertices: algorithm starting at 1.

We conclude our short introduction to the problem at hand by stating the formulation of Kantorovich-Rubinstein duality when X is a finite graph.

Proposition 7.1.1 (Kantorovich-Rubinstein, see e.g., [74,105])**.** *Let $G = (V, E)$ be a finite, simple graph, let $f : V \to \mathbb{R}$ and let $W \subset V$ be a subset of vertices. Then*

$$\left|\frac{1}{|V|}\sum_{x\in V} f(x) - \frac{1}{|W|}\sum_{x\in W} f(x)\right| \leq W_1\left(\frac{1}{|W|}\sum_{x\in W}\delta_x, dx\right) \max_{x_i\sim x_j} |f(x_i) - f(x_j)|.$$

This shows that our notion of uniform distribution of a subset of vertices has a natural connection to the question of sampling on graphs (i.e., reconstructing the average value of a 'smooth' function by sampling in a subset of the vertices). The theory of sampling on graphs is in its infancy but rapidly developing, we refer to [70, 89, 103, 104, 124].

7.2 The Algorithm

7.2.1 Setup

We recall that for a finite graph $G = (V, E)$, we can define the adjacency matrix

$$A = (a_{ij})_{i,j=1}^{|V|} \qquad \text{where} \qquad a_{ij} = \begin{cases} 1 & \text{if } x_i \sim_E x_j \\ 0 & \text{otherwise} \end{cases}$$

as well as the degree matrix

$$D = (d_{ij})_{i,j=1}^{|V|} \qquad \text{where} \qquad d_{ij} = \begin{cases} \deg(x_i) & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$$

With these definitions, we can define a notion of a Laplacian via

$$L = D - A.$$

We denote the eigenvectors of L by ϕ_i with corresponding eigenvalues λ_i, i.e., $L\phi_i = \lambda_i \phi_i$. Note that L has all its eigenvalues in $[0, 2\max_{v \in V} \deg(v)]$. Since L's columns sum to 0, we have for all measures μ that $L\mu$ has no net mass, i.e., is orthogonal to the constant vector, or has mean 0. Since L is symmetric, it is diagonalizable and all pairs of eigenvectors with distinct eigenvalues are orthogonal. The induced ordering of eigenvectors is analogous to the continuous case: small eigenvalue means slow oscillation frequency and the oscillation increases with the eigenvalue—the larger the eigenvalue, the more oscillation there is. In particular, ϕ_1 is constant with $\lambda_1 = 0$. Since we assume G is connected, there is only one instance of the trivial eigenvalue. As L is diagonalizable, we can also take arbitrary powers and define the fractional Laplacian L^α for $\alpha > 0$. Setting $n = |V|$,

$$L^\alpha v = \sum_{i=1}^{n} \langle v, \phi_i \rangle \, \lambda_i^\alpha \phi_i.$$

To define $L^{-\alpha}$, we need to adjust this definition slightly: since $\lambda_1 = 0$, we first shift v down by its mean to avoid dividing by 0. (In other words, we simply ignore the component of v in the direction of the constant eigenvector ϕ_1.) That is,

$$L^{-\alpha} v = \sum_{i=2}^{n} \langle v, \phi_i \rangle \lambda_i^{-\alpha} \phi_i.$$

Of course, multiplying a vector by a matrix can be interpreted as applying an operator to a function, since vectors indexed by vertices are simply functions $V \to \mathbb{R}$. Note that $AD^{-1} = I - LD^{-1}$ is a diffusion operator: each vertex splits its mass uniformly among its neighbors, and hence mass is preserved. This operator has eigenvalues in $[-1, 1]$. If we instead apply the transpose $D^{-1}A$, this corresponds to each vertex *taking* an equal portion of each of its neighbors masses, which, in general, is not mass-preserving. If G is k-regular (each vertex has equal degree k), then $D = kI$ is scalar and these two notions coincide and equal $\frac{1}{k}A$. We refer to [38,62] for a good introduction to these notions and many references.

7.2.2 Description of the Algorithm

We present an algorithm, parametrized by $0 < \alpha < 1$, for greedily picking well-distributed vertices x_k on graphs. (If $\alpha \gg 1$ the algorithm degenerates and repeatedly selects the same small subset of distant vertices over and over.) First, x_1 is chosen arbitrarily. Then, vertices are picked recursively according to the following:

$$x_{k+1} = \arg\min_{x \in V} \left(L^{-\alpha} \sum_{j=1}^{k} \delta_{x_j} \right)(x),$$

breaking ties arbitrarily (Figures 7.2, 3, 5, and 9-11 display applications of this algorithm to various graphs). If we write out the algorithm explicitly in terms of the spectrum of the Laplacian operator $L = D - A$, it becomes

$$x_{k+1} = \arg\min_{x \in V} \sum_{i=2}^{n} \sum_{j=1}^{k} \frac{\phi_i(x_j)}{\lambda_i^{\alpha}} \phi_i(x),$$

where the ϕ_i are normalized with $\|\phi_i\|_{L^2} = 1$, and we skip the constant eigenvector ϕ_1 since it has eigenvalue 0 (note that this choice, while seemingly arbitrary, has no impact on the algorithm as any contribution from the constant vector can be ignored when computing $\arg\min$—the algorithm is independent of choice of right inverse of the Laplacian). That is, we add up the projections of the indicator vector of the current vertex set, $\sum_{j=1}^{k} \delta_{x_j}$, onto each eigenvector, scaling down by the α power of the respective eigenvalue. Consider the case of a cycle graph: if the number of vertices is sufficiently large, this is well approximated by a torus. Setting $\alpha = 1/2$ and identifying the torus with $[0, 2\pi)/\sim := \mathbb{T}$, we have a simple explicit formula for the inverse Laplacian of a point mass (see §3.4 for the derivation):

$$L^{-1/2}(\delta_{x_k}) = -\frac{1}{\pi} \ln |2 \sin((x - x_k)/2)|.$$

Note that $2\sin((x - x_k)/2)$ is precisely the Euclidean distance between the points at angles x_i and x on the unit circle (i.e., $|e^{2\pi i x} - e^{2\pi i x_k}|$). Thus, the algorithm is simply maximizing the product of distances between points on the circle, by setting

$$x_{k+1} = \arg\min_{x \in \mathbb{T}} \left(-\frac{1}{\pi} \sum_{j=1}^{k} \ln |2 \sin((x - x_j)/2)| \right).$$

The arising sequence appears to behave on par with provably optimally regular sequences, and Steinerberger recently proved strong results on the regularity of such a sequence in [129], using techniques which are specific to this setting and unlikely to generalize to other graphs. (We explore this example in more detail in §3.4.) Nonetheless, these remarkable results on the torus and cycle graph give us hope that the algorithm may work comparably well on graphs more generally.

7.2.3 A Theoretical Guarantee

We prove a theoretical guarantee, for any finite graph G, that these sequences do exhibit at least a certain degree of regularity.

Theorem 7.2.1. *Let G be a simple connected graph and let*

$$\mu_k = \frac{1}{k}\sum_{j=1}^{k} \delta_{x_j},$$

where vertices x_j are selected as above. Then, for all $1 \leq k \leq n$,

$$\sum_{i=2}^{n} \frac{|\langle \mu_k, \phi_i \rangle|^2}{\lambda_i^{2\alpha}} \leq \left(\max_{j \leq k} \left\| L^{-2\alpha} \left(\delta_{x_j} \right) \right\|_{\ell^2}^2 \right) k^{-1}.$$

Remark 1. Observe for the sake of comparison that

$$\sum_{i=2}^{n} |\langle \mu_k, \phi_i \rangle|^2 = \|\mu_k\|_{\ell^2}^2 - |\langle \mu_k, \phi_1 \rangle|^2 = \frac{1}{k} - \frac{1}{n}.$$

Thus, it is natural that the bound in the Theorem should be $\sim k^{-1}$. However, λ_i (and thus $\lambda_i^{2\alpha}$) may be arbitrarily close to 0, scaling up the terms in the sum substantially. The only way to prevent this is for μ_k to be almost orthogonal to low-frequency eigenfunctions (which is cf. Erdős-Turán [49, 50] a natural way of defining regularity, as it means that μ_k is concentrated at high frequencies).

Remark 2. Note that

$$\max_{j \leq k} \left\| L^{-2\alpha} \left(\delta_{x_j} \right) \right\|_{\ell^2}^2 \leq \max_{x \in V} \left\| L^{-2\alpha} \left(\delta_x \right) \right\|_{\ell^2}^2.$$

The term on the right side is an interesting quantity in itself, and there may be good bounds for it in terms of the geometry of the graph. As can be seen from the expansion into eigenfunctions, this quantity measures, implicitly, how much low-frequency eigenfunctions concentrate in a particular vertex. In vertex-transitive graphs like cycle graphs and torus grid graphs we see that the quantity is actually independent of the vertex x.

7.3 Spectral Bounds on Transport Distances

7.3.1 Motivation

The motivation behind the algorithm is two-fold:

1. The greedy algorithm tries to minimize a Sobolev norm

$$\left\| \frac{1}{k} \sum_{j=1}^{k} \delta_{x_j} - dx \right\|_{\dot{H}^{-1}}.$$

2. Peyre's inequality [106] (see Chapter 2 §4) shows that, in the continuous setting,

$$W_2(\mu, dx) \lesssim \|\mu - dx\|_{\dot{H}^{-1}}.$$

The purpose of this section is to establish a connection between problems of optimal transport and spectral properties of the Laplacian. This is known to hold in the continuous case, we recall the following bound:

Theorem (Carroll, Massaneda, Ortega-Cerda [32])**.** *Let (M, g) be a compact Riemannian manifold with normalized volume measure dx and $\partial M = \emptyset$. If $-\Delta_g \phi = \lambda \phi$ on M, then, for some constant $C > 0$ depending only on (M, g),*

$$W_1\left(\phi^+ dx, \phi^- dx\right) \leq \frac{C}{\sqrt{\lambda}} \|\phi\|_{L^1(M)}.$$

This inequality is sharp. We recall the basic intuition that a Laplacian eigenfunction may, at scale $\lambda^{-1/2}$ (the wavelength), be understood as a random wave. This suggests that one has to move mass at least a distance comparable to the wavelength and examples on the torus $\mathbb{T}^d$ or the sphere $\mathbb{S}^d$ show that this is indeed the case.

7.3.2 Spectral Bounds on Transport

The purpose of this section is to show that a variation of this result exists on finite graphs; we will prove this for the Earth Mover's Distance $p = 1$.

Theorem 7.3.1. *Let $M = I - AD^{-1}$ and let $M\phi_k = \lambda_k \phi_k$. Then $0 \leq \lambda_k \leq 2$ and*

$$W_1(\phi_k^+, \phi_k^-) \leq \frac{1}{1 - |1 - \lambda_k|} \|\phi_k\|_{\ell^1}.$$

Note that, since $\phi = \phi_k^+ - \phi_k^-$ has mean 0, the measures ϕ_k^+ and ϕ_k^- have the same mass,

and thus one can be transported to the other. When we consider the asymptotic behavior of $W_1(\phi_k^+, \phi_k^-)$ on cycle graphs of increasing size, we see that this bound is a natural analog of Peyre's result [106] to graphs—it scales sharply with respect to $\dot{H}^{-1}(\phi_k)$ (see §3.4). We observe that this bound degenerates if $|\lambda_k - 1|$ is close to 1 and this a consequence of the proof. We also note that we always have the trivial transport inequality

$$W_1(\phi_k^+, \phi_k^-) \leq \operatorname{diam}(G) \left\|\phi_k^+\right\|_{\ell^1} = \frac{\operatorname{diam}(G)}{2} \|\phi_k\|_{\ell^1},$$

and thus the bound in the Theorem is preferable to the trivial bound only when

$$|1 - \lambda_k| < 1 - \frac{2}{\operatorname{diam}(G)}.$$

In many of the interesting cases for applications (graphs with good mixing properties), we can expect a spectral gap that quantitatively bounds $|\lambda_2 - 1| < 1$.

7.3.3 Applying the Theorem to obtain Transport Bounds

For an arbitrary distribution μ, we may use this bound to measure the Wasserstein distance to the uniform distribution on a graph with n vertices. The observation above motivates splitting μ into mid-range and extreme-frequency components,

$$\overline{\mu} = \sum_{|1-\lambda_k|<1-2/\operatorname{diam}(G)} \langle \mu, \phi_k \rangle \, \phi_k$$

and

$$\underline{\mu} = \sum_{|1-\lambda_k|\geq 1-2/\operatorname{diam}(G)} \langle \mu, \phi_k \rangle \, \phi_k.$$

We then transport $\overline{\mu}$ by propagating infinitely, and bound $\underline{\mu}$ with a diameter bound:

$$\begin{aligned} W_1(\mu, dx) &= W_1(\overline{\mu}^+ + \underline{\mu}^+, \overline{\mu}^- + \underline{\mu}^- + dx) \\ &\leq W_1\left(\overline{\mu}^+, \overline{\mu}^-\right) + W_1\left(\underline{\mu}^+, \underline{\mu}^- + dx\right) \\ &\leq \sum_{i=0}^{\infty} \|A^i D^{-i} \overline{\mu}\|_{\ell^1} + \frac{\operatorname{diam}(G)}{2} \|\underline{\mu} - dx\|_{\ell^1} \end{aligned}$$

While the above spectral bound is not guaranteed to be smaller than the diameter bound, empirical evidence suggests that it is in general a stronger bound, particularly for graphs with large diameter. We can test the quality of this bound by using linear programming [105] to compute exact Wasserstein distances. In §4, we display the results of doing so with

$$\mu = \frac{1}{k}\sum_{j=1}^{k} \delta_{x_j}$$

using the algorithm to pick the x_j:

$$x_{k+1} = \arg\min_{x \in V} \left(L^{-1/2} \sum_{j=1}^{k} \delta_{x_j} \right)(x),$$

compared against picking vertices x_k uniformly at random (without repetition) and averaging over 1000 Wasserstein distances obtained in this manner. This approach yields promising computational results across a number of large graphs.

7.3.4 A Case Study: Cycle Graphs

In this subsection, we will look carefully at the behavior of the cycle graphs C_n in the context of the above result. We recall the eigenvectors of $M = \frac{1}{2}L$ on C_n are

$$\phi_k(x) = (n/2)^{-1/2} \cos\left(\frac{2\pi k x}{n}\right) \qquad \text{and} \qquad \phi_{n-k}(x) = (n/2)^{-1/2} \sin\left(\frac{2\pi k x}{n}\right),$$

for $0 < k < n/2$, with $\phi_0(x) \equiv n^{-1/2}$ and, if n is even, $\phi_{n/2}(x) = n^{-1/2}(-1)^x$ and corresponding eigenvalues

$$\lambda_k = 1 - 1\cos\left(\frac{2\pi k}{n}\right).$$

Note that ϕ_0 is the constant eigenvector here and, since cosine is even, $\lambda_k = \lambda_{n-k}$ for all $k \neq 0$. We have elected to use the real eigenvectors in order to apply our arguments, though it is worth noting that we can change basis for the dimension 2 eigenspaces and simply write

$$\phi_k(x) = n^{-1/2} \exp\left(\frac{2\pi i k x}{n}\right),$$

where $0 \le k < n$ with all λ_k the same as above. Then,

$$\frac{1}{1-|1-\lambda_k|} = \frac{1}{1-\cos\left(\frac{2\pi k}{n}\right)} \approx 2\left(\frac{n}{2\pi k}\right)^2$$

for small k, by Taylor expansion. Further,

$$\|\phi_k\|_{\ell^1} \approx \|\phi_{n-k}\|_{\ell^1} \approx (n/2)^{-1/2} \int_0^n \left|\cos\left(\frac{2\pi kx}{n}\right)\right| dx = \frac{\sqrt{8n}}{\pi}.$$

On the other hand, $W_1(\phi_k^+, \phi_k^-) \approx W_1(\phi_{n-k}^+, \phi_{n-k}^-)$ can be approximated by the continuous analogue, where it is clear from symmetry that the optimal way to transport the sine wave is sending all mass to the nearest zero, where the positive and negative mass will cancel. This endures a cost of

$$(n/2)^{-1/2} 4k \int_0^{n/4k} x \sin\left(\frac{2\pi kx}{n}\right) dx = (n/2)^{-1/2} 4k \left(\frac{n}{2\pi k}\right)^2,$$

integrating by parts. Putting it all together, this yields

$$W_1(\phi_k^+, \phi_k^-) \approx \frac{\pi k}{n} \cdot \frac{1}{1-|1-\lambda_k|} \|\phi_k\|_{\ell^1}.$$

We may let $k \le n/100$ so that k is small enough for the Taylor expansion to be good, but nonetheless on the order of n: then we see the bound in Theorem 7.3.1 is sharp up to constants. Observe that applying $L^{-1/2}$ to the point mass δ_0 yields

$$\begin{aligned} L^{-1/2}(\delta_0) = \sum_{k=1}^{n-1} \frac{\phi_k(0)}{\lambda_k^{1/2}} \phi_k &= \sum_{k=1}^{n-1} \frac{1}{\left(2-2\cos\left(\frac{2\pi k}{n}\right)\right)^{1/2} n} \exp\left(\frac{2\pi ikx}{n}\right) \\ &= 2 \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{\left(2-2\cos\left(\frac{2\pi k}{n}\right)\right)^{1/2} n} \cos\left(\frac{2\pi kx}{n}\right) \\ &\approx \frac{1}{\pi} \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{k} \cos\left(\frac{2\pi kx}{n}\right), \end{aligned}$$

for all odd n, again approximating with Taylor expansion. (If n is even, we will get an extra $(-1)^x/n\sqrt{2}$ term corresponding to $\phi_{n/2}$, but this will vanish in the limit we are about to take.) We caution the reader that the λ_k above are the eigenvalues of $L = 2M$, and are

thus double the eigenvalues of M referred to in the preceding computations. Rescaling with $x = n\theta/2\pi$, we have

$$L^{-1/2}(\delta_0) \approx \frac{1}{\pi} \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{k} \cos(k\theta).$$

Fixing θ and letting $n \to \infty$, this is simply the Fourier series for

$$\lim_{n\to\infty} L^{-1/2}(\delta_0) = -\frac{1}{\pi} \ln|2\sin(\theta/2)|.$$

Taking this limit and rescaling really is just transitioning us to the continuous setting: the eigenfunctions of the Laplacian on the unit circle $\mathbb{S}^1$ identified with $[0, 2\pi)/\sim$ are $(2\pi)^{-1/2}\exp(ikx)$, with eigenvalue k^2, for $k \in \mathbb{Z}$. So the fractional inverse Laplacian $L^{-1/2}$ of a point mass on the circle is

$$\begin{aligned} L^{-1/2}(\delta_0) &= \sum_{k\neq 0} \frac{\phi_k(0)}{\lambda_k^{1/2}} \phi_k = \sum_{k\neq 0} \frac{1}{2\pi|k|} \exp(ikx) \\ &= \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{1}{k} \cos(kx) = -\frac{1}{\pi} \ln|2\sin(x/2)|, \end{aligned}$$

precisely our function in the discrete case.

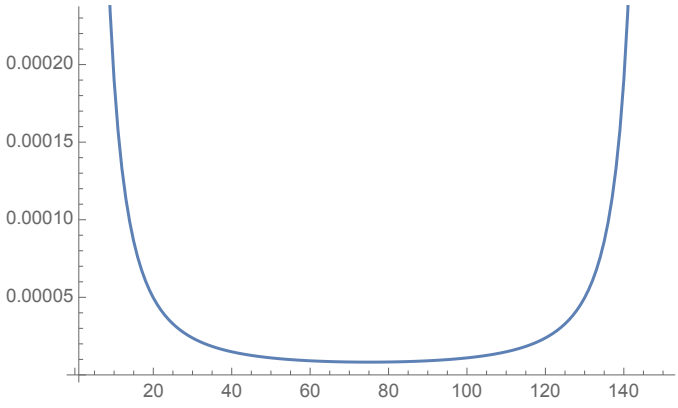

Figure 7.4: The difference between $L^{-1/2}(\delta_0)$ on $\mathbb{S}^1$ and the fractional Laplacian $L^{-1/2}(\delta_0)$ on C_{150} is small.

In Figure 7.6 we show the difference between the inverse fractional Laplacian on a point mass $L^{1/2}(\delta_0)$ in the continuous and discrete settings. The two outputs are almost identical and thus their difference is quite small (this holds even for *very* small values of n). It is worth

recalling the recent Theorem of Pausinger, which proves that this algorithm belongs to a large class which produces the van der Corput sequence on the torus (and thus achieves optimal discrepancy, up to constants) [100]. Further, Steinerberger's recent result [129] indicates that this algorithm performs extremely well on the torus, and by extension large cycle graphs, and that the sequence of points rapidly becomes very evenly distributed. See Chapter 4, §2 for much more on such algorithms in the continuous setting.

7.3.5 A Case Study: Torus Grid Graphs

In this subsection, we examine another class of graphs: torus grid graphs. The $m \times n$ torus grid graph $T_{m,n}$ is the Cartesian product of C_m and C_n, so we can apply many of our computations from the previous subsection here. In particular, the eigenvectors for $T_{m,n}$ are the Kronecker products of pairs of eigenvectors (ϕ_j, ϕ_k) from C_m and C_n, respectively, with corresponding eigenvalues under $M = \frac{1}{4}L$ of

$$\lambda_{j,k} = \frac{\lambda_j + \lambda_k}{2} = 1 - \frac{1}{2}\left(\cos\left(\frac{2\pi j}{m}\right) + \cos\left(\frac{2\pi k}{n}\right)\right).$$

(The factor of $1/2$ appears because cycle graphs are 2-regular while torus grid graphs are 4-regular, and vanishes if we instead use L here.) Then we have

$$\frac{1}{1 - |1 - \lambda_{j,k}|} = \frac{1}{1 - \frac{1}{2}\left(\cos\left(\frac{2\pi j}{m}\right) + \cos\left(\frac{2\pi k}{n}\right)\right)} \approx \frac{4}{\left(\frac{2\pi j}{m}\right)^2 + \left(\frac{2\pi k}{n}\right)^2}$$

for small j, k, by Taylor expansion. Further,

$$\|\phi_{j,k}\|_{\ell^1} = \|\phi_j\|_{\ell^1}\|\phi_k\|_{\ell^1} \approx \frac{8\sqrt{nm}}{\pi^2}.$$

To find $W_1(\phi_{j,k}^+, \phi_{j,k}^-)$, note that the support of $\phi_{j,k}^+$ consists of checkerboarded rectangles, and the most efficient way to transport the positive mass to the negative mass will be along the higher frequency direction—that is, horizontally if $m/j < n/k$, and vertically otherwise.

Without loss of generality, we suppose we are in the former case. Then this incurs a cost of

$$\sum_{i=0}^{n-1} |\phi_k(i)| W_1(\phi_j^+, \phi_j^-) = \|\phi_k\|_{\ell^1} W_1(\phi_j^+, \phi_j^-) \approx \frac{\sqrt{8n}}{\pi} \cdot (m/2)^{-1/2} 4j \left(\frac{m}{2\pi j}\right)^2.$$

Applying our assumption that $m/j < n/k$ to our earlier estimate, we see

$$\frac{1}{1-|1-\lambda_{j,k}|} \approx \frac{4}{\left(\frac{2\pi j}{m}\right)^2 + \left(\frac{2\pi k}{n}\right)^2} \leq \frac{4}{\left(\frac{2\pi k}{n}\right)^2 + \left(\frac{2\pi k}{n}\right)^2} = 2\left(\frac{n}{2\pi k}\right)^2.$$

Putting it all together,

$$\begin{aligned} \frac{\sqrt{8n}}{\pi}(m/2)^{-1/2} 4j \left(\frac{m}{2\pi j}\right)^2 &\approx W_1(\phi_{j,k}^+, \phi_{j,k}^-) \\ &\leq \frac{1}{1-|1-\lambda_{j,k}|} \|\phi_{j,k}\|_{\ell^1} \leq 2\left(\frac{n}{2\pi k}\right)^2 \cdot \frac{8\sqrt{nm}}{\pi^2}, \end{aligned}$$

and thus, taking the quotient of the two sides in the above inequality our bound is off by (at most) a factor of $\pi k^2 m (n^2 j)^{-1}$. Note that when $k/n = j/m$ (i.e., when the horizontal and vertical components of $\phi_{j,k}$ have the same frequency), this simplifies to $\pi k/n$, precisely our result on cycle graphs.

7.4 Numerics

Below we provide numerics on a variety of graphs demonstrating the performance of the algorithm and the bound from Theorem 7.3.1. In particular, we compare the performance of vertices selected according to our algorithm against that of randomly selected vertices. The differences between our vertex sequences and random vertex sequences may seem marginal, but this is partly due to the fact that the diameter of some of these graphs is quite small. For instance, the Truncated Tetrahedral graph, with diameter 3, only has 12 vertices, so we will hardly be able to distinguish the performance of the algorithm's vertices from randomly selected vertices on such a small set—it is impressive that we see a difference at all. We see that for the Faulkner-Younger Graph and the Level 2 Menger Sponge the difference becomes significantly more drastic. Many of the graphs are quite well connected, which

makes the transport problem easier than on sparse graphs (e.g., on complete graphs it makes no difference at all which vertices are selected, the transport cost only depends on the number of vertices). In all the tables in this section, x_j were computed directly using the recursive definition for the algorithm ($\alpha = .5$) given in Section 2 with any ties broken randomly, and the exact Wasserstein distances

$$W_1\left(\frac{1}{k}\sum_{j=1}^{k}\delta_{x_j}, dx\right)$$

were subsequently computed using the dual linear program in [105] in the "Algorithm" row. For graphs which are not vertex-transitive, the performance of the algorithm depends upon the arbitrary initial vertex chosen, and thus all choices of initial vertex were attempted and the transport costs averaged. In the "Random" row, 1000 uniformly randomly selected sets of k distinct vertices were taken, and the corresponding Wasserstein distances were averaged.

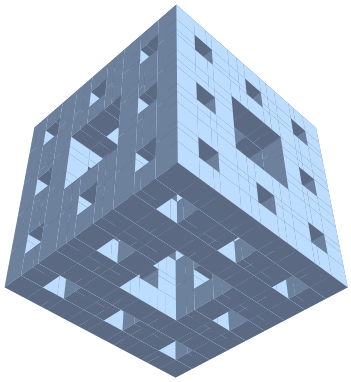

Figure 7.5: The Menger Sponge, whose 400 cubes form the vertices of a connectivity graph.

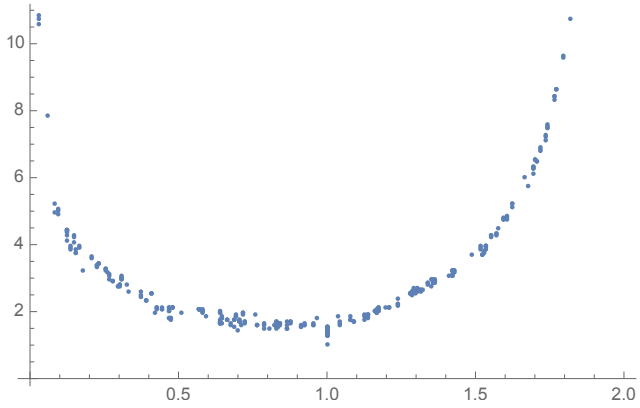

Figure 7.6: The tightness of the bound in Theorem 7.3.1, applied to the eigenfunctions of the Level 2 Menger Sponge Connectivity Graph (computed as a quotient of the right and left sides).

7.4.1 Connectivity Graph of Level 2 Menger Sponge

The Level 2 Menger sponge is the object obtained beginning with a cube and drilling out the middle square of each face (viewed as a three by three grid of squares), and then iterating this process one more time on the smaller cubes (see Fig. 7.7). We can then generate a

connectivity graph of the remaining 400 smaller cubes (each one ninth the side length of the original cube). Note that this is not a regular graph. In Figure 7.8, we see the that, on the Level 2 Menger Sponge Connectivity Graph, the Theorem 7.3.1 bound is tightest for mid-range eigenvalues. This is to be expected, due to the blow-up of the $1/(1-|1-\lambda|)$ term at the extremes, where a diameter bound is tighter (see §3.2). But we see here that, even for very small eigenvalues, the bound is fairly tight.

No. of vertices	1	3	5	10	15	20	25	30
Algorithm	9.94	6.48	5.01	3.54	2.92	2.55	2.31	2.11
Random	9.94	6.73	5.52	4.25	3.63	3.20	2.90	2.69

Table 7.1: $W_1(\mu, dx)$ for the Connectivity Graph of a Level 2 Menger Sponge

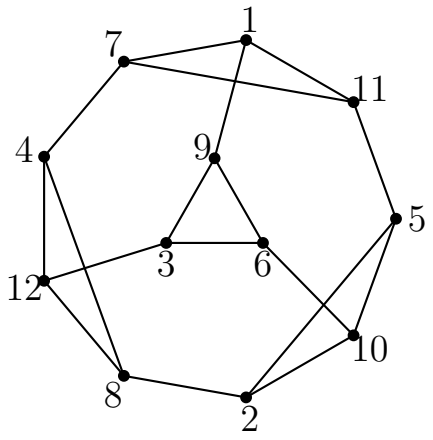

Figure 7.7: The sequence of vertices picked by the algorithm on the Truncated Tetrahedral Graph.

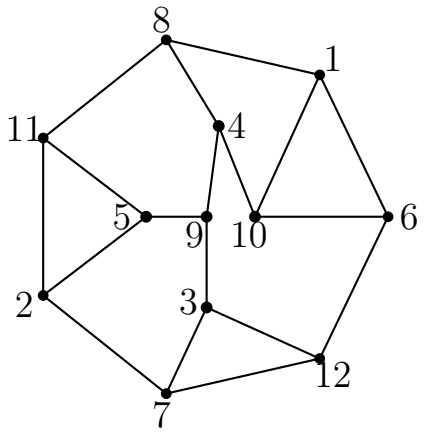

Figure 7.8: The sequence of vertices picked by the algorithm on the Frucht Graph.

7.4.2 Truncated Tetrahedral Graph

The Truncated Tetrahedral Graph (see Fig. 7.9) is a 3-regular, vertex-transitive graph on 12 vertices. It is the 1-skeleton of the Archimedean solid formed by truncating each vertex of a tetrahedron.

No. of vertices	1	2	3	4	5	6	7	8	9	10
Algorithm	1.92	1.17	0.83	0.67	0.58	0.50	0.42	0.33	0.28	0.23
Random	1.92	1.35	1.01	0.84	0.72	0.58	0.52	0.43	0.34	0.27

Table 7.2: $W_1(\mu, dx)$ for the Truncated Tetrahedral Graph

7.4.3 Frucht Graph

The Frucht Graph (see Fig. 7.10) is a 3-regular graph on 12 vertices, and has trivial automorphism group despite being degree-regular.

No. of vertices	1	2	3	4	5	6	7	8	9	10
Algorithm	1.93	1.17	0.86	0.67	0.59	0.50	0.42	0.34	0.29	0.23
Random	1.93	1.34	1.04	0.85	0.73	0.59	0.52	0.43	0.35	0.27

Table 7.3: $W_1(\mu, dx)$ for the Frucht Graph

7.4.4 Faulkner-Younger Graph

The Faulkner-Younger Graph on 44 vertices (see Fig. 7.11) is a 3-regular non-Hamiltonian graph (that is, there is no path along its edges that traverses every vertex exactly once).

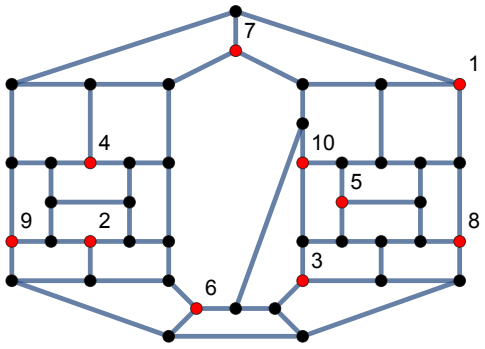

Figure 7.9: The first ten vertices picked by the algorithm on the Faulkner-Younger Graph. Each label is above and to the right of the corresponding vertex.

No. of vertices	1	2	3	4	5	6	7	8	9	10
Algorithm	4.17	2.67	2.05	1.71	1.52	1.34	1.23	1.15	1.05	0.97
Random	4.17	3.08	2.57	2.24	2.01	1.83	1.68	1.56	1.46	1.37

Table 7.4: $W_1(\mu, dx)$ for the Faulkner-Younger Graph

7.4.5 Erdős-Rényi Random Graphs

The Erdős-Rényi model for random graphs $G(n, p)$ is given by including an edge between each pair of the n vertices independently with probability p. Here we display the performance of the algorithm on two such graphs, one taken from $G(100, .06)$ (a sparse graph, see Fig. 7.12) and another from $G(100, .2)$ (a dense graph, see Fig. 7.13).

No. of vertices	1	3	5	10	15	20	25	30
Algorithm, Sparse Graph	2.72	2.61	2.13	1.52	1.22	1.02	0.85	0.74
Random, Sparse Graph	2.72	2.11	1.82	1.44	1.22	1.06	0.93	0.83
Algorithm, Dense Graph	1.79	1.53	1.31	1.01	0.86	0.80	0.75	0.70
Random, Dense Graph	1.79	1.46	1.26	0.99	0.88	0.81	0.75	0.70

Table 7.5: $W_1(\mu, dx)$ for Erdős-Rényi Random Graphs

It is no surprise that the dense graph exhibits little variation between the transport cost of random vertices and of the algorithm's—after all, any pair of vertices has many short paths between them. In fact, this particular graph has diameter 3. Thus, for sufficiently dense graphs it is largely irrelevant which vertices are selected: the transport cost will be low. The sparse graph displayed has diameter 6 and is thus more interesting: while random vertices initially outperform the algorithm, by 15 vertices selected they are matched, after which the algorithm surpasses the random vertices. That is, even in highly irregular graphs such as this one where random vertices perform well at first, the algorithm nonetheless manages to catch up even with a relatively small number of vertices.

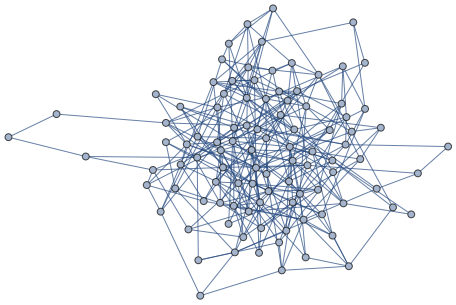

Figure 7.10: A sparse Erdős-Rényi Graph from $G(100, .06)$.

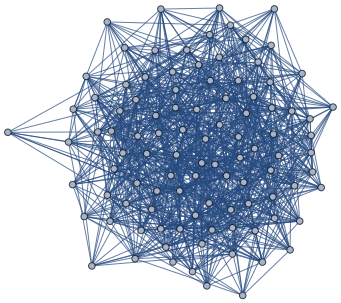

Figure 7.11: A dense Erdős-Rényi Graph from $G(100, .2)$.

7.4.6 Complete 3-ary Tree

The complete 3-ary tree of depth 4 is a rooted tree, where each vertex has 3 children, except for the fourth generation of vertices which all have no children, yielding a total of 40 vertices (see Fig. 7.14).

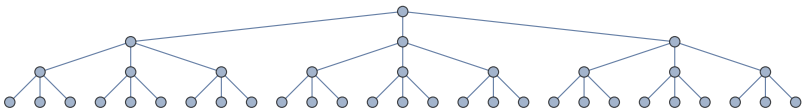

Figure 7.12: The 3-ary tree of depth 4

No. of vertices	1	2	3	4	5	6	7	8	9	10
Algorithm	4.25	3.58	2.75	2.83	2.55	2.07	1.98	1.73	1.34	1.37
Random	4.25	3.62	3.12	2.93	2.74	2.48	2.33	2.17	1.99	1.86

Table 7.6: $W_1(\mu, dx)$ for the 3-ary tree of depth 4

7.5 Proofs

7.5.1 Proof of Theorem 7.2.1

Proof. Note first that, since, for all $v \in \mathbb{R}^n$, $L^{-2\alpha}v$ has mean 0 (being spanned by ϕ_i, $i > 1$, and thus orthogonal to the constant ϕ_1), we will always have

$$\min_{x \in V} \left(L^{-2\alpha} \sum_{j=1}^{k} \delta_{x_j} \right)(x) < 0.$$

Using the ℓ^2 norm

$$\|v\|_{\ell^2}^2 = \sum_{i=1}^{n} v_i^2,$$

we observe

$$\begin{aligned}
\left\|L^{-\alpha}(k\mu_k)\right\|_{\ell^2}^2 &= \left\|L^{-\alpha}\left((k-1)\mu_{k-1}\right)\right\|_{\ell^2}^2 + \left\|L^{-\alpha}\left(\delta_{x_k}\right)\right\|_{\ell^2}^2 \\
&\quad + 2\left\langle L^{-\alpha}((k-1)\mu_{k-1}), L^{-\alpha}\left(\delta_{x_k}\right)\right\rangle \\
&= \left\|L^{-\alpha}\left((k-1)\mu_{k-1}\right)\right\|_{\ell^2}^2 + \left\|L^{-\alpha}\left(\delta_{x_k}\right)\right\|_{\ell^2}^2 \\
&\quad + 2\left\langle L^{-2\alpha}((k-1)\mu_{k-1}), \delta_{x_k}\right\rangle,
\end{aligned}$$

since $L^{-\alpha}$ is self-adjoint. Rewriting the inner product term,

$$\left\langle L^{-2\alpha}((k-1)\mu_{k-1}), \delta_{x_k}\right\rangle = \left(L^{-2\alpha}\sum_{j=1}^{k-1}\delta_{x_j}\right)(x_k).$$

But x_k was chosen by the algorithm specifically to minimize that quantity—thus, it is certainly less than the average value of 0, and so

$$\left\|L^{-\alpha}(k\mu_k)\right\|_{\ell^2}^2 \leq \left\|L^{-\alpha}\left((k-1)\mu_{k-1}\right)\right\|_{\ell^2}^2 + \left\|L^{-\alpha}\left(\delta_{x_k}\right)\right\|_{\ell^2}^2.$$

Then, by induction, we obtain the desired inequality:

$$\sum_{i=2}^{n}\frac{|\langle\mu_k,\phi_i\rangle|^2}{\lambda_i^\alpha} = \left\|L^{-\alpha}(\mu_k)\right\|_{\ell^2}^2 \leq \left(\max_{j\leq k}\left\|L^{-\alpha}\left(\delta_{x_j}\right)\right\|_{\ell^2}^2\right)k^{-1}. \qquad \square$$

7.5.2 Proof of Theorem 7.3.1

Proof. We recall that AD^{-1} can be interpreted as the propagator of the random walk on the graph $G = (V, E)$. Moreover, we have

$$AD^{-1}\phi_k = (1-\lambda_k)\phi_k$$

and observe that $|1-\lambda_k| \leq 1$. We proceed in a similar manner to [131] and interpret diffusion on the graph as one of many ways to transport mass. In particular, we will apply

the elementary estimate

$$W_1(AD^{-1}v, v) \leq \|v\|_{\ell^1} = \sum_{j=1}^{n} |v_j|$$

to $v = ((AD^{-1})^i \phi_k)^{\pm}$,

$$W_1(AD^{-1}((AD^{-1})^i\phi_k)^{\pm}), ((AD^{-1})^i\phi_k)^{\pm}) \leq \|((AD^{-1})^i\phi_k)^{\pm}\|_{\ell^1} = \frac{|1-\lambda_k|^i}{2}\|\phi_k\|_{\ell^1}.$$

In particular, we will transport ϕ_k to $(AD^{-1})^m\phi_k$ through its positive and negative parts after each diffusion. For large m, this measure almost vanishes since

$$(AD^{-1})^m\phi_k = (1-\lambda_k)^m \phi_k.$$

We perform this operation until some arbitrary m and then use the trivial bound on the remaining measure. This shows that the total transport can be bounded by

$$\begin{aligned} W_1(\phi_k^+, \phi_k^-) &\leq \frac{|1-\lambda_k|^m \operatorname{diam}(G)}{2}\|\phi_k\|_{\ell^1} + \sum_{i=0}^{m-1} |1-\lambda_k|^i \|\phi_k\|_{\ell^1} \\ &= \left(\frac{|1-\lambda_k|^m \operatorname{diam}(G)}{2} + \frac{1-|1-\lambda_k|^m}{1-|1-\lambda_k|}\right)\|\phi_k\|_{\ell^1} \\ &= \left(|1-\lambda_k|^m\left[\frac{\operatorname{diam}(G)}{2} - \frac{1}{1-|1-\lambda_k|}\right] + \frac{1}{1-|1-\lambda_k|}\right)\|\phi_k\|_{\ell^1}. \end{aligned}$$

We observe that this bound is monotonic in m, with direction depending on the sign of the bracketed expression. If

$$|1-\lambda_k| \geq 1 - \frac{2}{\operatorname{diam}(G)},$$

the bound is monotonically increasing and we set $m = 0$, recovering the initial diameter bound. If

$$|1-\lambda_k| < 1 - \frac{2}{\operatorname{diam}(G)},$$

we have a monotonically decreasing bound, and take the limit as $m \to \infty$, yielding the desired bound of

$$W_1(\phi_k^+, \phi_k^-) \leq \frac{1}{1-|1-\lambda_k|}\|\phi_k\|_{\ell^1}. \qquad \square$$

7.6 Connection to other Results

7.6.1 Low-discrepancy point sets.

A classical problem in the study of *irregularities of distribution* is to construct sequences $(x_n)_{n=1}^{\infty}$ on the unit interval $[0,1]$ such that $\{x_1, \dots, x_n\}$ is fairly evenly distributed over the unit interval for all $n \in \mathbb{N}$. The problem has now been solved completely: as discussed in Chapter 1, Schmidt [121] proved that for any sequence on $[0,1]$, there exist infinitely many $N \in \mathbb{N}$ such that the discrepancy satisfies

$$D_N \geq \frac{1}{100} \cdot \frac{\log N}{N}.$$

Recall Steinerberger's result [130] from Chapter 4 §2.3: greedy sequences defined via

$$x_{n+1} = \arg\min_{x \in \mathbb{T}} \left((-\Delta)^{-1/2} \sum_{k=1}^{n} \delta_{x_k} \right)(x)$$

satisfy, for all $N \in \mathbb{N}$,

$$D_N \lesssim \frac{\log N}{\sqrt{N}},$$

This result is conjectured to be far from optimal, and numerical examples show that the arising sequences seem to be remarkably close to the best possible bound $N^{-1} \log N$ (down to the level of the constant). The argument is somewhat different and uses the Koksma-Hlawka inequality and classical Fourier Analysis. In particular, this result is stronger than what is guaranteed by Theorem 7.2.1.

7.6.2 Leja points

Leja points can be defined, in the utmost level of generality, for any symmetric kernel $k : X \times X \to \mathbb{R} \cup \{\infty\}$ on a compact Hausdorff space. We remain on smooth compact manifolds M, a natural example for the kernel is

$$k(x,y) = \frac{1}{d_g(x,y)^s} \qquad \text{where } d_g(x,y) \text{ is the geodesic distance and } s > 0.$$

We can then define, in an iterative fashion, for a given initial point $x_1 \in M$, a sequence $(x_k)_{k=1}^{\infty}$, in such a way that

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

Put differently, we greedily add a new point x_n in such a way that the total energy

$$\sum_{\substack{k,\ell=1 \\ k \neq \ell}}^{n} k(x_k, x_\ell) \qquad \text{is as small as possible.}$$

These sets were introduced by Edrei [48] and intensively studied by Leja [83] after whom they are named. The most commonly used kernel is $k(x, y) = -\log|x - y|$ (such that minimizing the sum is the same as maximizing the product of the distances). Leja points have a number of applications in numerical analysis [16,29,91,113,114]. Pausinger [100] recently gave a very precise description of Leja sequences on $\mathbb{T}$ for fairly general kernel functions and established a connection to binary digit expansion. For the Riesz kernel $k(x, y) = |x - y|^{-s}$, it is known that Leja sequences are asymptotically uniformly distributed [90]. We are not aware of any study of Leja *vertices* on graphs; while one could take existing kernels, for example $k(x, y) = |x - y|^{-s}$, and consider them on graphs, there is little reason to assume that such vertices will have many special properties: Graphs are simply too flexible. We can summarize the approach in this chapter as stating that

> there is a very good reason to believe (see the Figures in this chapter) that considering $k(x, y)$ to be the Green's function of the inverse Laplacian leads to well-distributed sets of vertices.

Moreover, we are able to analyze the continuous limit of manifolds and are able to obtain a quantitative bound showing that the bounds are more regularly distributed than simply exhibiting uniform distribution.

Bibliography

[1] T. van Aardenne-Ehrenfest, Proof of the Impossibility of a Just Distribution of an Infinite Sequence Over an Interval, Proc. Kon. Ned. Akad. Wetensch. 48, 3–8 (1945).

[2] M. Ajtai, J. Komlos, G. Tusnady, On optimal matchings, Combinatorica 4, 259–264 (1984).

[3] L. Ambrosio, F. Stra, D. Trevisan, A PDE approach to a 2-dimensional matching problem, Probab. Theory Related Fields 173, 433–477 (2019).

[4] D. Aronson, Non-negative solutions of linear parabolic equations, Ann. Sci. Norm. Sup. 22, 607–694 (1968).

[5] T. Aubin, Green's Function for Riemannian Manifolds, in: Nonlinear Analysis on Manifolds. Monge-Ampère Equations, Grundlehren der mathematischen Wissenschaften (A Series of Comprehensive Studies in Mathematics) 252, 101–114 (1982).

[6] T. Aubin, Some Nonlinear Problems in Riemannian Geometry, Springer Monographs in Mathematics, Springer Berlin Heidelberg (1998).

[7] N.S. Bakhvalov, On the approximate calculation of multiple integrals, J. Complexity 31, 502–516 (2015) [English translation; the original appeared in Vestnik MGU, Ser. Math. Mech. Astron. Phys. Chem 4, 3–18 (1959)].

[8] J. Beck, A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution, Compositio Math. 72 3, 269–339 (1989).

[9] J. Beck, Probabilistic Diophantine approximation. I. Kronecker sequences, Ann. of Math. 140, no. 2, 449–502 (1994).

[10] J. Beck, W. Chen, Irregularities of Distribution, Cambridge Tracts in Mathematics 89, Cambridge University Press (1987).

[11] R. Béjian, H. Faure, Discrépance de la suite de Van der Corput, Séminaire Delange-Pisot-Poitou. Théorie des nombres 19.1, 1–14 (1977–1978).

[12] C. Beltran, A facility location formulation for stable polynomials and elliptic Fekete points, Found. Comput. Math. 15, 125–157 (2015).

[13] C. Beltran, N. Corral, J. Criado del Rey, Discrete and Continuous Green Energy on Compact Manifolds, Journal of Approximation Theory 237, 160–185 (2019).

[14] J.D. Benamou, Y. Brenier, A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem, Numer. Math. 84, 375–393 (2000).

[15] L. Betermin, E. Sandier, Renormalized energy and asymptotic expansion of optimal logarithmic energy on the sphere, Constructive Approximation 47, 39–74 (2018).

[16] L. Bialas-Ciez, J.-P. Calvi, Pseudo Leja sequences, Annali di Matematica 191, 53–75 (2012).

[17] D. Bilyk, Roth's Orthogonal Function Method in Discrepancy Theory and Some New Connections in the book "Panorama of Discrepancy Theory", Lecture Notes in Math 2107 Springer Verlag, 71–158 (2014).

[18] D. Bilyk, M. Lacey, On the small ball Inequality in three dimensions, Duke Math. J. 143, no. 1, 81–115 (2008).

[19] D. Bilyk, F. Dai, S. Steinerberger, General and Refined Montgomery Lemmata, Math. Ann. 373, 1283–1297 (2019).

[20] D. Bilyk, M. Lacey, A. Vagharshakyan, On the small ball inequality in all dimensions, J. Funct. Anal. 254, no. 9, 2470–2502 (2008).

[21] M. Blumlinger, Asymptotic distribution and weak convergence on compact Riemannian manifolds, Monatsh. Math.110, no. 3-4, 177–188 (1990).

[22] S. Bobkov, M. Ledoux, A simple Fourier analytic proof of the AKT optimal matching theorem, arXiv: 1909.06193 (2019).

[23] F. Bolley, A. Guillin, C. Villani, Quantitative concentration inequalities for empirical measures on non-compact spaces, Probab. Theory Related Fields 137, no. 3-4, 541–593 (2007).

[24] J. Brauchart, Optimal logarithmic energy points on the unit sphere, Math. Comp. 77, 1599-1613 (2008).

[25] L. Brown, S. Steinerberger, On the Wasserstein Distance between Classical Sequences and the Lebesgue Measure, Trans. Amer. Math. Soc. 373, 8943–8962 (2020).

[26] L. Brown, S. Steinerberger, Positive-definite Functions, Exponential Sums and the Greedy Algorithm: a Curious Phenomenon, Journal of Complexity 61 (2020).

[27] L. Brown, Sequences of well-distributed vertices on graphs and spectral bounds on optimal transport, arXiv: 2008.11296.

[28] D. A. Burgess, The distribution of quadratic residues and non-residues, Mathematika 4, 106–112 (1957).

[29] D. Calvetti, L. Reichel, D. Sorensen, An implicitly restarted Lanczos method for large symmetric eigenvalue problems, Electronic Transactions on Numerical Analysis 2, 1–21 (1994).

[30] I. Carbone, Discrepancy of LS-sequences of partitions and points, Annali di Matematica 191, 819–844 (2012).

[31] I. Carbone, M. Iacò, A. Volčič, A dynamical system approach to the Kakutani-Fibonacci sequence, Ergodic Theory and Dynamical Systems 34(6), 1794–1806 (2014).

[32] T. Carroll, X. Massaneda, J. Ortega-Cerda, An enhanced uncertainty principle for the Vaserstein distance, Bull. London Math. Soc. 52, 1158–1173 (2020).

[33] J. Cassels, Simultaneous diophantine approximation. II, Proc. Lond. Math. Soc. (3) 5, 435–448 (1955).

[34] D. Chafaï, A. Hardy, M. Maïda, Concentration for Coulomb gases and Coulomb transport inequalities, Journal of Functional Analysis 275, 1447–1483 (2018).

[35] H. Chaix, H. Faure, Discrépance et diaphonie en dimension un. Acta Arith. 63, 103–141 (1993).

[36] B. Chazelle, The discrepancy method. Randomness and complexity, Cambridge University Press (2000).

[37] X. Cheng, G. Mishne, S. Steinerberger, The Geometry of Nodal Sets and Outlier Detection, Journal of Number Theory 185, 48–64 (2018).

[38] F. Chung, Spectral Graph Theory, CBMS Regional Conference Series in Mathematics 92, American Mathematical Society (1997).

[39] J. Criado del Rey, On the separation distance of minimal Green energy points on compact Riemannian manifolds, arXiv:1901.00779 (2019).

[40] B. Dahlberg, Regularity Properties of Riesz Potentials, Indiana University Mathematics Journal 28, 257–268 (1979).

[41] G.B. Dantzig, Reminiscences About the Origins of Linear Programming, In: Bachem A., Korte B., Grötschel M. (Eds.) Mathematical Programming The State of the Art., Springer-Verlag Berlin Heidelberg, 78–86 (1983).

[42] H. Davenport, Simultaneous Diophantine approximation, Mathematika 1, 51–72 (1954).

[43] R. DeVore, V. Temlyakov, Some remarks on greedy algorithms, Adv. Comput. Math. 5, no. 2-3, 173–187 (1996).

[44] J. Dick, F. Pillichshammer, Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration, Cambridge University Press (2010).

[45] M. Drmota, M. Infusino, On the discrepancy of some generalized Kakutani's sequences of partitions, In: Uniform Distribution Theory, Mathematical Institute of the Slovak Academy of Sciences, 7(1), 75–104 (2012).

[46] M. Drmota, G. Larcher, and F. Pillichshammer, Precise distribution properties of the van der Corput sequence and related sequences, Manuscripta Math. 118, 11–41 (2005).

[47] M. Drmota, R. Tichy, Sequences, discrepancies and applications, Lecture Notes in Mathematics 1651, Springer-Verlag Berlin (1997).

[48] A. Edrei, Sur les déterminants récurrents et les singularités dune fonction donnée par son développement de Taylor, Compositio Mathematica 7, 20–88 (1940).

[49] P. Erdős and P. Turán, On a problem in the theory of uniform distribution. I, Nederl. Akad. Wetensch. 51, 1146–1154 (1948).

[50] P. Erdős and P. Turán, On a problem in the theory of uniform distribution. II. Nederl. Akad. Wetensch. 51, 1262–1269 (1948).

[51] L. Evans, Partial differential equations. Second edition, Graduate Studies in Mathematics 19, American Mathematical Society (2010).

[52] H. Faure, Discrepancy and diaphony of digital (0, 1)-sequences in prime bases, Acta Arith. 117, 125–148 (2005).

[53] H. Faure, P. Kritzer, F. Pillichshammer, From van der Corput to modern constructions of sequences for quasi-Monte Carlo rules, Indagationes Mathematicae 26.5, 760–822 (2015).

[54] L Fejér, Sur les fonctions intégrables et bornées, C.R. Acad. Sci. Paris, 984–987 (1900).

[55] W. Freeden, On Integral Formulas of the (Unit) Sphere and Their Applications to Numerical Computations of Integrals, Computing, 25, 131–146 (1980).

[56] T. Ganelius, Some applications of a lemma on Fourier series, Acad. Serbe Sci. Publ. Inst. Math. 11, 9–18 (1957).

[57] D. García-Zelada, Concentration for Coulomb Gases on compact manifolds, Electron. Commun. Probab. 24, no. 12, 1–18 (2019)

[58] P. Grabner, Erdős-Turan type discrepancy bounds, Monatsh. Math. 111, no. 2, 127–135 (1991).

[59] P. Grabner, R. Tichy, Spherical designs, discrepancy and numerical integration, Math. Comp. 60, no. 201, 327–336 (1993).

[60] P. Grabner, B. Klinger, R. Tichy, Discrepancies of point sequences on the sphere and numerical integration, Multivariate approximation (Witten-Bommerholz, 1996), Math. Res., 101, Akademie Verlag, Berlin, 95–112 (1997).

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

[62] A. Grigor'yan, Introduction to Analysis on Graphs, University Lecture Series 71, American Mathematical Society (2018).

[63] V. S. Grozdanov, On the diaphony of one class of one-dimensional sequences, Int. J. Math and Math. Sci. 19.1, 115–124 (1996).

[64] D. Hardin, E. Saff, Discretizing manifolds via minimum energy points, Not. Amer. Math. Soc. (2004).

[65] D. Hensley, F. Su, Random walks with badly approximable numbers. Unusual applications of number theory, DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 64, Amer. Math. Soc., 95–101 (2004).

[66] A. Hinrichs, E. Novak, M. Ullrich, H. Wozniakowski, The curse of dimensionality for numerical integration of smooth functions, Math. Comp. 83, no. 290, 2853–2863 (2014).

[67] A. Hinrichs, E. Novak, M. Ullrich, H. Wozniakowski, The curse of dimensionality for numerical integration of smooth functions II, J. Complexity 30, no. 2, 117–143 (2014).

[68] A. Hinrichs, E. Novak, M. Ullrich, H. Wozniakowski, Product rules are optimal for numerical integration in classical smoothness spaces, J. Complexity 38, 39–49 (2017).

[69] E. Hlawka, Funktionen von beschränkter Variation in der Theorie der Gleichverteilung, Ann. Mat. Pura Appl. (4) 54, 325–333 (1961).

[70] P. Hu, W. C. Lau, A Survey and Taxonomy of Graph Sampling, arXiv:1308.5865 (2013).

[71] M. Huesmann, K.-T. Sturm, Optimal transport from Lebesgue to Poisson, Ann. Probab. 41, no. 4, 2426–2478 (2013).

[72] M. Iacò, Low discrepancy sequences: Theory and Applications, arXiv:1502.04897 (2015).

[73] S. Kakutani, A problem of equidistribution on the unit interval $[0, 1]$, Measure Theory: Proceedings of the Conference held at Oberwolfach 15–21 June 1975, 369–375 (1975).

[74] L.V. Kantorovich, On the Translocation of Masses, J Math Sci 133, 1381–1382 (2006).

[75] N. Kirk, On Proinov's Lower Bound for the Diaphony, Uniform distribution theory 15.2, 39–72 (2020).

[76] A. Kolmogorov, Sulla determinazione empirica di una legge di distribuzione, G. Ist. Ital. Attuari. 4, 83–91 (1933).

[77] J. Koksma, Een algemeene stelling uit de theorie der gelijkmatige verdeeling modulo 1, Mathematica B (Zutphen) 11, 7–11 (1942–1943).

[78] L. Kuipers, H. Niederreiter, Uniform distribution of sequences, Pure and Applied Mathematics. Wiley-Interscience, New York-London-Sydney (1974).

[79] G. Larcher, Discrepancy estimates for sequences: new results and open problems, in: Uniform Distribution and Quasi-Monte Carlo Methods: Discrepancy, Integration and Applications, Radon Series on Computational and Applied Mathematics 15 (2014).

[80] G. Larcher, On the star discrepancy of sequences in the unit interval, Journal of Complexity 31, 474–485 (2015).

[81] G. Larcher, F. Puchhammer, An improved bound for the star discrepancy of sequences in the unit interval, Uniform Distribution Theory 11, no. 1, 1–14 (2016).

[82] M. Ledoux, On Optimal Matching of Gaussian Samples, J. Math. Sci. 238, 495–522 (2019).

[83] F. Leja, Sur certaines suites liées aux ensembles plans et leur application à la représentation conforme, Annales Polonici Mathematici 4, 8–13 (1957).

[84] N. Lev, J. Ortega-Cerdà, Equidistribution estimates for Fekete points on complex manifolds, Journal of the European Mathematical Society 18, no. 2, 425–464 (2016).

[85] W. J. LeVeque, An inequality connected with Weyl's criterion for uniform distribution, Proc. Sympos. Pure Math. Vol. VIII, Amer. Math. Soc. 22–30 (1965).

[86] P. Li, S.-T. Yau, On the parabolic kernel of the Schrodinger operator, ActaMath. 156, 153–201 (1986).

[87] E. Lieb, On the lowest eigenvalue of the Laplacian for the intersection of two domains, Invent. Math. 74, 441–448 (1983).

[88] J. Lierl, S. Steinerberger, A Local Faber-Krahn inequality and Applications to Schrodinger's Equation, Comm. PDE 43, 66–81 (2018).

[89] G. Linderman, S. Steinerberger, Numerical Integration on Graphs: where to sample and how to weigh, Math. Comp. 89 (2020).

[90] G. A. López, E. B. Saff, Asymptotics of greedy energy points, Math. Comp. 79 (272), 2287–2316 (2010).

[91] S. De Marchi, G. Elefante, Quasi-Monte Carlo integration on manifolds with mapped low-discrepancy points and greedy minimal Riesz s-energy points, Applied Numerical Mathematics 127, 110–124 (2018).

[92] J. Marzo and A. Mas, Discrepancy of Minimal Riesz Energy Points, to appear in Constr. Approx.

[93] H. L. Montgomery, Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis, CBMS 84 (1990).

[94] F. Narcowich, X. Sun, J. Ward, Z. Wu, LeVeque type inequalities and discrepancy estimates for minimal energy configurations on spheres, J. Approx. Theory 162, no. 6, 1256–1278 (2010).

[95] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, Society for Industrial and Applied Mathematics (1992).

[96] S. Ninomiya, Constructing a new class of low-discrepancy sequences by using the β-adic transformation, Mathematics and Computers in Simulation (MATCOM), Elsevier, vol. 47(2), 403–418 (1998).

[97] E. Novak, Some results on the complexity of numerical integration, Monte Carlo and quasi-Monte Carlo methods, Springer Proc. Math. Stat. 163, Springer, 161–183 (2016).

[98] E. Novak, H. Wozniakowski, Tractability of multivariate problems. Vol. 1: Linear information, EMS Tracts in Mathematics 6. European Mathematical Society (EMS) (2008).

[99] R. O'Neil, Convolution operators and $L(p, q)$ spaces, Duke Mathematical Journal 30, 129–142 (1963).

[100] F. Pausinger, Greedy energy minimization can count in binary: point charges and the van der Corput sequence, Annali di Matematica 200, 165–186 (2021).

[101] F. Pausinger, W. Ch. Schmid, A good permutation for one-dimensional diaphony, Monte Carlo Methods Appl. 16, 307–322 (2010).

[102] O. Perron, Über diophantische Approximationen, Math. Ann. 83, 77–84 (1921).

[103] I. Pesenson, Sampling in Paley-Wiener spaces on combinatorial graphs, Trans. Amer. Math. Soc. 360, 5603–5627 (2010).

[104] I. Pesenson and M. Pesenson, Sampling, filtering and sparse approximations on combinatorial graphs, J. Fourier Anal. Appl. 16, 921–942 (2010).

[105] G. Peyré and M. Cuturi, Computational optimal transport, Foundations and Trends in Machine Learning 11, no. 5-6, 355–607 (2019).

[106] R. Peyre, Comparison between W_2 distance and $\dot{H}^{-1}$ norm, and Localization of Wasserstein distance, ESAIM: COCV 24, 1489–1501 (2018).

[107] G. Pólya and G. Szegő, Isoperimetric Inequalities in Mathematical Physics. Annals of Mathematics Studies, no. 27, Princeton University Press (1951).

[108] P. D. Proinov, On irregularities of distribution, C. R. Acad. Bulgare Sci. 39. No. 9, 31–34 (1986).

[109] P. D. Proinov, Quantitative Theory of Uniform Distribution and Integral Approximation, University of Plovdiv, Bulgaria (2000).

[110] P. D. Proinov, V. Grozdanov, On the diaphony of the van der Corput-Halton sequence, J. Number Theory 30, no. 1, 94–104 (1988).

[111] R. Pyke, W. R. van Zwet, Weak convergence results for the Kakutani interval splitting procedure, Ann. Probab. 32, no. 1A, 380–423 (2004).

[112] M. Rachh, S. Steinerberger, On the location of maxima of solutions of Schrödinger's equation, Comm. Pure. Appl. Math. 71, 1109–1122 (2018).

[113] L. Reichel, The Application of Leja Points to Richardson Iteration and Polynomial Preconditioning, Linear Algebra and its Applications 154–156, 389–414 (1991).

[114] L. Reichel, Newton Interpolation at Leja points, BIT 30, 332–346 (1990).

[115] J. Rivat, G. Tenenbaum, Constantes d'Erdős-Turán, Ramanujan Journal, 111–121, (2005).

[116] K. F. Roth, On irregularities of distribution, Mathematika 1, 73–79 (1954).

[117] I. Z. Rusza, On an Inequality of Erdős and Turán Concerning Uniform Distribution Modulo One, II, J. Number Theory 49, no. 1, 84–88 (1994).

[118] E. Saff, V. Totik, Logarithmic potentials with external fields, Springer (2013).

[119] F. Santambrogio, Optimal transport for applied mathematicians, Calculus of variations, PDEs, and modeling, Progress in Nonlinear Differential Equations and their Applications 87, Birkhauser/Springer, Cham (2015).

[120] W. M. Schmidt, Diophantine approximation, Lecture Notes in Mathematics 785, Springer, Berlin (1980).

[121] W. M. Schmidt, Irregularities of distribution. VII, Acta Arith. 21, 45–50 (1972).

[122] W. M. Schmidt, Badly approximable systems of linear forms, J. Number Theory 1, 139–154 (1969).

[123] W. Schmidt, On badly approximable numbers and certain games, Trans. Am. Math. Soc. 123, 178–199 (1966).

[124] D. Shuman, S. Narang, P. Frossard, A. Ortega, P. Vandergheynst, The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains, IEEE Signal Process. Mag. 30(3), 83–98 (2013).

[125] L. Spiegelhofer, Discrepancy Results for The Van Der Corput Sequence, Uniform distribution theory 13.2, 57–69 (2018).

[126] S. Steinerberger, A Nonlocal Functional promoting Low-Discrepancy Point Sets, Journal of Complexity 54 (2019).

[127] S. Steinerberger, A Wasserstein Inequality and Minimal Green Energy on Compact Manifolds, arXiv:1907.09023 (2019).

[128] S. Steinerberger, Directional Poincare Inequalities along Mixing Flows, Arkiv for Matematik 54, 555–569 (2016).

[129] S. Steinerberger, Polynomials with Zeros on the Unit Circle: Regularity of Leja Sequences, arXiv:2006.10708, accepted to Mathematika (2020).

[130] S. Steinerberger, Dynamically Defined Sequences with Small Discrepancy, Monatshefte Math. 191, 639–655 (2020).

[131] S. Steinerberger, Wasserstein Distance, Fourier Series and Applications, Monatshefte Math. 194, 305–338 (2021).

[132] F. Su, Convergence of random walks on the circle generated by an irrational rotation, Trans. Amer. Math. Soc. 350, no. 9, 3717–3741 (1998).

[133] F. Su, A LeVeque-type lower bound for discrepancy, Monte Carlo and quasi-Monte Carlo methods 1998, Springer, Berlin, 448–458 (2000).

[134] F. Su, Discrepancy convergence for the drunkard's walk on the sphere, Electron. J. Probab. 6 , no. 2, 1–20 (2001).

[135] A. G. Sukharev, Optimal numerical integration formulas for some classes of functions, Sov. Math. Dokl. 20, 472–475 (1979).

[136] M. Talagrand, Matching theorems and empirical discrepancy computations using majorizing measures, J. Amer. Math. Soc. 7, 455–537 (1994).

[137] M. Talagrand, The Transportation Cost from the Uniform Measure to the Empirical Measure in Dimension ≥ 3, The Annals of Probability, Vol. 22, 919–959 (1994).

[138] M. Talagrand, J. E. Yukich, The integrability of the square exponential transportation cost, Ann. Appl. Probab. 3, no. 4, 1100–1111 (1993).

[139] V. Temlyakov, Greedy approximation, Acta Numer. 17, 235–409 (2008).

[140] V. Temlyakov, Greedy approximation, Cambridge Monographs on Applied and Computational Mathematics 20, Cambridge University Press (2011).

[141] V. Temlyakov, Numerical integration without smoothness assumption, arXiv:2003.14331 (2020).

[142] V. Temlyakov, Cubature formulas, discrepancy, and nonlinear approximation, J. Complexity 19, 352–391 (2003).

[143] V. Temlyakov, Connections between numerical integration, discrepancy, dispersion, and universal discretization, SIAM Journal of Computational Mathematics, Volume S5, 185–209 (2019).

[144] J. J. Thomson, On the Structure of the Atom: an Investigation of the Stability and Periods of Oscillation of a number of Corpuscles arranged at equal intervals around the Circumference of a Circle; with Application of the Results to the Theory of Atomic Structure, Philosophical Magazine Series 6, Volume 7, Number 39, 237–265 (1904).

[145] N. Garcia Trillos, D. Slepcev, On the rate of convergence of empirical measures in ∞-transportation distance, Canad. J. Math. 67, no. 6, 1358–1383 (2015).

[146] J.G. van der Corput, Verteilungsfunktionen I-II, Proc. Akad. Amsterdam 38: 813—821, 1058—1066 (1935).

[147] L. N. Vasershtein, Markov processes on a countable product space describing large systems of automata, Problemy Peredavci Informacii, 3, 64–72 (1969).

[148] C. Villani, Topics in Optimal Transportation, Graduate Studies in Mathematics, American Mathematical Society (2003).

[149] A. Volčič. A generalization of Kakutani's splitting procedure. Annali di Mat. Pura e Appl. 190 (no. 1), 45–54 (2011).

[150] G. Wagner, On the means of distances on the surface of a sphere (lower bounds), Pacific J. Math. 144, 389–398 (1990).

[151] J. Weed, F. Bach, Sharp asymptotic and finite-sample rates of convergence of empirical measures in Wasserstein distance, Bernoulli 25(4A), 2620–2648 (2019).

[152] H. Weyl, Ueber die Gleichverteilung von Zahlen mod. Eins, Math. Ann. 77 (3), 313–352 (1916).

[153] P. Zinterhof, H. Stegbuchner, Trigonometrische Approximation mit Gleichverteilungsmethoden, Stud. Sci. Math. Hung. 13, 273–289 (1978).

[154] P. Zinterhof, Über einige Abschätzungen bei der Approximation von Funktionen mit Gleichverteilungsmethoden, Österreich. Akad. Wiss. Math.-Naturwiss. Kl. S.-B. II 185, no. 1-3, 121–132 (1976).