

ON COMBINATORIAL PROPERTIES OF GREEDY WASSERSTEIN MINIMIZATION

STEFAN STEINERBERGER

ABSTRACT. We discuss a phenomenon where Optimal Transport leads to a *remarkable* amount of combinatorial regularity. Consider infinite sequences $(x_k)_{k=1}^{\infty}$ in $[0, 1]$ constructed in a greedy manner: given x_1, \dots, x_n , the new point x_{n+1} is chosen so as to minimize the Wasserstein distance W_2 between the empirical measure of the $n+1$ points and the Lebesgue measure,

$$x_{n+1} = \arg \min_x W_2 \left(\frac{1}{n+1} \sum_{k=1}^n \delta_{x_k} + \frac{\delta_x}{n+1}, dx \right).$$

This leads to fascinating sequences (for example: $x_{n+1} = (2k+1)/(2n+2)$ for some $k \in \mathbb{Z}$) which coincide with sequences recently introduced by Ralph Kritzinger in a different setting. Numerically, the regularity of these sequences rival the best known constructions from Combinatorics or Number Theory. We prove a regularity result below the square root barrier.

1. INTRODUCTION AND MAIN RESULTS

1.1. Motivation. Given any finite set of points $x_1, \dots, x_n \in [0, 1]$, we propose adding a new point in such a way that the Wasserstein W_2 –distance between the new empirical measure and the Lebesgue measure on $[0, 1]$ is minimized. Formally, we consider a notion of energy $E : [0, 1] \rightarrow \mathbb{R}_{\geq 0}$ via

$$E(x) = W_2 \left(\frac{1}{n+1} \sum_{k=1}^n \delta_{x_k} + \frac{\delta_x}{n+1}, dx \right)$$

and then choose x_{n+1} to be any number in which E assumes a global minimum. Throughout this paper, we will only consider the W_2 distance between a discrete measure and the uniform measure on $[0, 1]$ for which the following convenient formula exists: assuming $0 \leq a_1 \leq a_2 \leq \dots \leq a_n \leq 1$, then

$$W_2 \left(\frac{1}{n} \sum_{k=1}^n \delta_{a_k}, dx \right)^2 = \sum_{i=1}^n \int_{(i-1)/n}^{i/n} (x - a_i)^2 dx.$$

This procedure is perhaps best illustrated with an example. Consider the initial set of points $\{1/\pi, 1/e, 1/\sqrt{2}\}$ on $[0, 1]$. A computation shows that the point x_4 that one needs to add to minimize the W_2 –transport cost is $x_4 = 7/8$. Repeating the computation after having added x_4 , the optimal point in the next step is $x_5 = 1/10$. Continuing the computation, one arrives at the sequence

$$\frac{1}{\pi}, \frac{1}{e}, \frac{1}{\sqrt{2}}, \frac{7}{8}, \frac{1}{10}, \frac{7}{12}, \frac{7}{14}, \frac{13}{16}, \frac{3}{18}, \dots$$

Key words and phrases. Wasserstein distance, Greedy Sequences, Irregularities of Distribution, Kronecker, van der Corput, Kritzinger, L^2 –discrepancy.

S.S. is supported by the NSF (DMS-2123224) and the Alfred P. Sloan Foundation.

which has a number of remarkable properties. The elements are rational (with numerator odd and denominator being $2n$, see [18] or §2.1). They are uniformly distributed in the unit interval in a highly regular manner that is illustrated in Fig. 1: newly added points avoid existing points and fill in existing gaps – considering the construction, this is not too surprising. However, they seem to do so at the optimal rate which is very difficult to do (see §1.3).

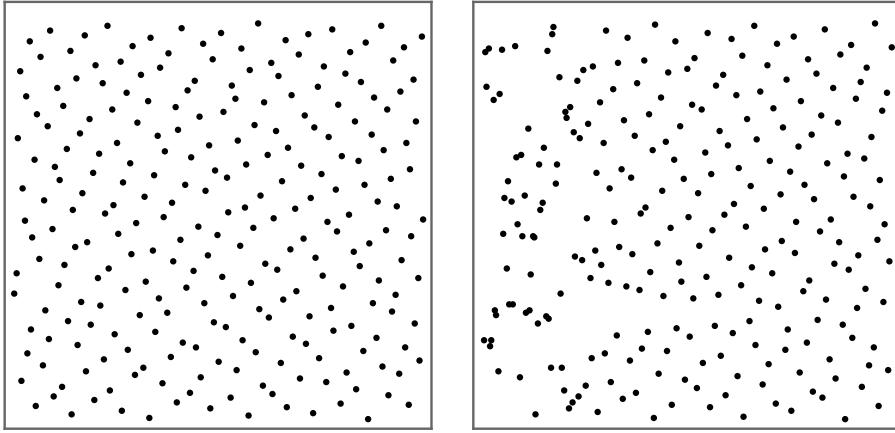


FIGURE 1. An illustration of the first n elements of a sequence: the points $(k/n, x_k)$ for $1 \leq k \leq n$ (here $n = 250$). Left: starting with $1/\pi, 1/e, 1/\sqrt{2}$. Right: starting with 50 random elements.

We consider another example (see Fig. 1, right) where we start with 50 randomly sampled points in $[0, 1]$. The greedy construction is eager to avoid existing clusters while filling in existing gaps – what is totally remarkable is that after the greedy construction has added another 50 elements, it seems to have completely repaired the existing defects and continues in a most regular fashion.

1.2. Kritzinger sequences. This coincides with another type of construction that was recently given by Ralph Kritzinger [18] using a seemingly different metric.

Theorem 1. *Greedy W_2 minimization produces a Kritzinger sequence.*

Kritzinger was motivated by a different notion of regularity which is a classical object in the irregularities of distribution: the L^2 –discrepancy. Formally, Kritzinger proposes, given x_1, \dots, x_n to pick x_{n+1} so that it minimizes

$$F(x) = -2 \sum_{k=1}^n \max \{x_k, x\} + (n+1)^2 x^2 - x.$$

This expression is at first glance perhaps a bit difficult to interpret while the Wasserstein W_2 interpretation allows access to the theory of Optimal Transport. Theorem 1 is only true on the unit interval (due to the rigidity of Optimal Transport in one dimension) and fails on domains in $d \geq 2$ dimensions where W_2 minimization and Kritzinger sequences are different. We first collect some of the results from [18].

Theorem (Kritzinger [18]). *If x_n is constructed using the greedy method, then it is different from x_1, \dots, x_{n-1} and $x_n = (2k+1)/(2n)$ for some $k \in \mathbb{N}_{\geq 0}$. Moreover, any Kritzinger sequence satisfies, for all $n \in \mathbb{N}$,*

$$\int_0^1 |\#\{1 \leq k \leq n : x_k \leq x\} - nx|^2 dx \leq \frac{n}{3} + c,$$

where the constant c only depends on the initial set of points.

We give a new proof of this result in our framework before going on to further refining the estimate. The estimate essentially says that for *most* points $0 < x < 1$

$$\#\{1 \leq k \leq n : 0 \leq x_k \leq x\} = n \cdot x \pm \mathcal{O}(\sqrt{n}).$$

This is the typical square root law that one would also expect from random variables: it is a common theme in the construction of greedy sequences that one can usually establish a $\mathcal{O}(\sqrt{n})$ error bound using an averaging trick. Simultaneously, it seems difficult to go past it in most cases.

1.3. How regular are these sequences really? Suppose now that $(x_k)_{k=1}^{\infty}$ is an arbitrary sequence in $[0, 1]$. In 1935 van der Corput [10, 11] asked whether there exists an infinite sequence so that for some universal constant C and all $n \in \mathbb{N}$

$$\max_{0 \leq x \leq 1} |\#\{1 \leq k \leq n : x_k \leq x\} - nx| \leq C.$$

This was shown to be impossible by van Aardenne-Ehrenfest [1] and started the theory of *irregularities of distribution*. After seminal contributions by K. F. Roth [21], it was then proven by W. M. Schmidt [22] in 1972 that for *every* sequence there are infinitely many integers $n \in \mathbb{N}$ such that

$$\max_{0 \leq x \leq 1} |\#\{1 \leq k \leq n : x_k \leq x\} - nx| \geq c \log n.$$

The optimal constant is now known to satisfy $0.065 \leq c \leq 0.2223$ where the lower bound is due to Larcher & Puchhammer [16] and the upper bound is due to Ostromoukhov [19]. There are two classical types of sequences with very good behavior: the Kronecker sequences, also known as irrational rotations on the torus, $x_n = n\alpha \bmod 1$. The other classical example is the van der Corput sequence in base 2 defined via digit expansions in base 2 (see [9, 12, 13, 17] for details).

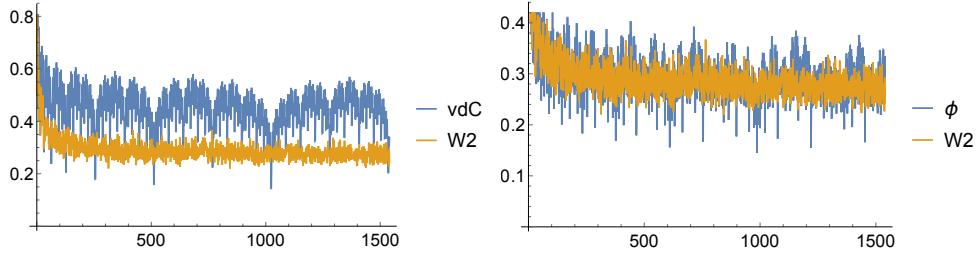


FIGURE 2. Comparison to van der Corput (left) and $n\phi \bmod 1$ (right).

We compared the maximum deviation (multiplied with $1/\log n$ to rescale the leading asymptotic order) between the Kritzinger sequence started with $x_1 = 1/2$, the van der Corput sequence in base 2 and the Kronecker sequence with $\alpha = (1+\sqrt{5})/2$. The Kritzinger sequence is somewhat superior to the van der Corput sequence and very comparable to the golden ratio Kronecker sequence (albeit it seems to have less variance). We emphasize that these two sequences are among the most regular sequences known and their behavior is within a small factor from the theoretical limit: indeed, irrational rotations on the torus are among the most well studied object in many different fields of mathematics while the van der Corput sequence was specifically designed to be as regular as possible. The Kritzinger sequence appears to draw its regularity from the geometry of the Wasserstein distance.

1.4. Improved Regularity. Our main result is a new regularity statement for Kritzinger sequences. Such a sequence is assumed to be initialized with an arbitrary finite set of real numbers and then extended in the greedy fashion.

Theorem 2. *Every Kritzinger sequence $(x_k)_{k=1}^\infty$ has infinitely many $n \in \mathbb{N}$ with*

$$\max_{0 \leq x \leq 1} \left| \int_0^x \#\{1 \leq k \leq n : x_k \leq y\} - ny \, dy \right| \leq 2 \cdot n^{1/3}.$$

In light of Fig. 2, we naturally expect that $n^{1/3}$ can be replaced by $\mathcal{O}(\log n)$. The scaling of the L^1 -discrepancy suggests that $\mathcal{O}(\sqrt{\log n})$ may be a plausible guess. It is not clear to us whether it could be as small as $\mathcal{O}(1)$. Theorem 2 breaks the square root barrier, in particular, it shows in a concrete way that Kritzinger sequences are more regular than i.i.d. random variables (for which this quantity is $\sim n^{1/2}$ w.h.p.).

1.5. A curious inequality. A new technical ingredient of the proof is a curious new inequality for real-valued functions.

Lemma (Main Lemma). *Let $g : [0, 1] \rightarrow \mathbb{R}$ be continuous and $g \not\equiv 0$. Then*

$$\max_{0 \leq z \leq 1} \left| \int_0^z g(x)x \, dx - \int_z^1 g(x)(1-x) \, dx \right| \geq \frac{1}{16} \frac{1}{\|g\|_{L^2}^2} \left(\max_{0 \leq x \leq 1} \left| \int_0^x g(y)dy \right| \right)^3.$$

This inequality is sharp up to at most a factor of 8: take g to be the characteristic function on $[0, \varepsilon]$. Then

$$\max_{0 \leq z \leq 1} \left| \int_0^z g(x)x \, dx - \int_z^1 g(x)(1-x) \, dx \right| = \int_0^\varepsilon g(x)x \, dx = \frac{\varepsilon^2}{2}$$

while

$$\frac{1}{16} \frac{1}{\|g\|_{L^2}^2} \left(\max_{0 \leq x \leq 1} \left| \int_0^x g(y)dy \right| \right)^3 = \frac{1}{16\varepsilon} \cdot \varepsilon^3 = \frac{\varepsilon^2}{16}.$$

This inequality plays a role in a part of the proof that is somewhat modular and could relatively easily be exchanged by something else. Any lower bound

$$\max_{0 \leq z \leq 1} \left| \int_0^z g(x)x \, dx - \int_z^1 g(x)(1-x) \, dx \right| \geq \text{informative quantity}$$

could be used to derive a regularity statement for Kritzinger sequences.

1.6. Related Results. The problem of irregularities of distributions in dimensions $d \geq 2$ has proven to be challenging [2, 3, 5, 6]. Indeed, at this point there is not even a clear consensus on what the optimal scale of regularity should be. Most examples of highly regular sequences seem to follow the spirit of van der Corput sequences or generalized Kronecker sequences (we refer to the textbooks [9, 12, 13, 17]). This naturally suggests investigating the possibility of new constructions. Such constructions were given by the author in [23, 24] and by Brown and the author in [7, 8]. Pausinger [20] showed that a large class of greedy constructions can replicate the van der Corput sequence: this means that, at least for some suitable initial conditions, they can produce sequences at the theoretical limit of regularity. One of the most interesting example of a greedy sequence was recently given by Ralph Kritzinger [18]: the arising sequence is, empirically, as regular as any of the other greedy sequences but additionally consists of rational numbers which greatly simplifies computation. Most of these greedy constructions can easily be shown to have regularity $\lesssim n^{-1/2}$, a barrier which has resisted improvement. A notable exception is [25] where additional structure in the complex plane could be used to show that a classic greedy constructions for polynomials has optimal rate up to possibly logarithmic factors (see also work of Beck [4] resolving a question of Erdős [14]). The Wasserstein distance has been introduced to the study of irregularities of distributions in [26] and was then further developed in [7]. We especially emphasize work of C. Graham [15] who established an irregularities of distributions phenomenon for $d = 1$. An explicit inequality relating the W_2 distances to the Green Energy [27] has lead to the following rather general principle (established by Brown and the author in [7]): on compact manifolds in dimension $d \geq 3$ normalized to have volume 1, there always exist greedy sequences $(x_k)_{k=1}^\infty$ obtained by minimizing the Green energy for which $W_2\left(\frac{1}{n} \sum_{k=1}^n \delta_{x_k}, dx\right) \lesssim n^{-1/d}$ uniformly in n .

2. PROOFS

§3.1 gives two proofs of Theorem 1. §2.2 derives a useful alternative representation, §2.3 proves an Lemma which is enough to prove a regularity rate of $\lesssim n^{1/2}$. §2.4 proves the Main Lemma which is then used in §2.5 to prove Theorem 2.

2.1. Proof of Theorem 1.

Proof. Let $0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq 1$ be given. Observe that

$$\begin{aligned} W_2\left(\frac{1}{n} \sum_{k=1}^n \delta_{x_k}, dx\right)^2 &= \sum_{k=1}^n \int_{\frac{k-1}{n}}^{\frac{k}{n}} (x_k - z)^2 dz \\ &= \sum_{k=1}^n \int_{\frac{k-1}{n}}^{\frac{k}{n}} x_k^2 - 2x_k z + z^2 dz \\ &= \frac{1}{3} + \frac{1}{n} \sum_{k=1}^n x_k^2 - \sum_{k=1}^n x_k \frac{2k-1}{n^2}. \end{aligned}$$

Rescaling by a factor of n^2 , we have

$$n^2 \cdot W_2\left(\frac{1}{n} \sum_{k=1}^n \delta_{x_k}, dx\right)^2 = \frac{n^2}{3} + n \sum_{k=1}^n x_k^2 - \sum_{k=1}^n x_k (2k-1).$$

Let us now consider the original set and let us add an additional point x . We introduce the variables such that

$$\{x_1, \dots, x_n, x\} = \{y_1, \dots, y_{n+1}\}$$

and $y_1 \leq y_2 \leq \dots \leq y_{n+1}$. A computation shows that

$$\begin{aligned} (n+1)^2 \cdot W_2^2 \left(\frac{1}{n+1} \sum_{k=1}^{n+1} \delta_{y_k}, dx \right) &= \frac{(n+1)^2}{3} + (n+1) \sum_{k=1}^{n+1} y_k^2 - \sum_{k=1}^{n+1} y_k (2k-1) \\ &= n^2 \cdot W_2^2 \left(\frac{1}{n} \sum_{k=1}^n \delta_{x_k}, dx \right) + \frac{(n+1)^2 - n^2}{3} \\ &\quad + (n+1)x^2 + \sum_{k=1}^n x_k^2 \\ &\quad + \sum_{k=1}^n x_k \cdot (2k-1) - \sum_{k=1}^{n+1} y_k \cdot (2k-1). \end{aligned}$$

Among these terms only two, $(n+1)x^2$ and the last sum, actually depend on x . The next step consists in simplifying the difference between the last two sums to exploit some additional cancellation. There are now three cases: (1) $x < x_1$, (2) $x > x_n$ or (3) there exists j such that

$$x_j \leq x < x_{j+1}.$$

We will only deal with case (3), the other two cases can be dealt with analogously. Under this assumption, we necessarily have

$$y_i = \begin{cases} x_i & \text{if } i \leq j \\ x & \text{if } i = j+1 \\ x_{i-1} & \text{if } i \geq j+2 \end{cases}$$

Thus we have

$$\begin{aligned} \sum_{k=1}^{n+1} y_k (2k-1) - \sum_{k=1}^n x_k (2k-1) &= (2j+1)x + 2 \sum_{k=j+1}^n x_k \\ &= x + 2 \sum_{k=1}^n \max \{x, x_k\}. \end{aligned}$$

Collecting all the terms that actually depend on x , our goal is to minimize

$$F(x) = (n+1)x^2 - x - 2 \sum_{i=1}^n \max \{x, x_i\}.$$

which is exactly the Kritzinger functional. \square

Kritzinger's functional can be derived from trying to minimize the L^2 -discrepancy

$$\int_0^1 |\# \{1 \leq k \leq n : x_k \leq x\} - nx|^2 dx.$$

In hindsight, the fact that these functionals are related is perhaps not too surprising: the fact that Wasserstein distances can be equivalently defined over empirical cumulative distribution functions is a feature of us working in one dimension (see,

for example, Vallender [28]). Another quick proof of Theorem 1 could be given by appealing to the formula (see e.g. [18, Equation 10]) that for ordered points

$$\int_0^1 |\# \{1 \leq k \leq n : x_k \leq x\} - nx|^2 dx = n \sum_{k=1}^n \left(x_k - \frac{2k-1}{2n} \right)^2 + \frac{1}{12}$$

while

$$n^2 \cdot W_2 \left(\frac{1}{n} \sum_{k=1}^n \delta_{x_k}, dx \right)^2 = \frac{n^2}{3} + n \sum_{k=1}^n x_k^2 - \sum_{k=1}^n x_k (2k-1).$$

This quickly implies that

$$\int_0^1 |\# \{1 \leq k \leq n : x_k \leq x\} - nx|^2 dx - n^2 \cdot W_2^2 \left(\frac{1}{n} \sum_{k=1}^n \delta_{x_k}, dx \right) = c_n$$

for a constant c_n that is independent of the set of points. Therefore, minimizing one of the functionals is equivalent to minimizing the other. The explicit form of F allows to prove that minimizers are rational numbers.

Proposition (Kritzinger [18]). *Let $0 \leq x_1 \leq x_2 \leq \dots \leq x_n \leq 1$ and let*

$$F(x) = (n+1)x^2 - x - 2 \sum_{k=1}^n \max \{x, x_k\}.$$

Then the minimizer is different from all the existing points and of the form $x = j/(2n)$ for some odd integer $j \in 2\mathbb{Z} + 1$.

Proof. Note that the function F is continuous, not differentiable in the points x_1, \dots, x_n but differentiable everywhere else. In particular, if x is different from the n points, then

$$F'(x) = 2(n+1)x - 1 - 2 \cdot \# \{1 \leq k \leq n : x_k < x\}.$$

The only candidates for roots outside of the existing points are numbers of the form

$$x = \frac{1 + 2 \# \{1 \leq k \leq n : x_k < x\}}{2n+2}$$

which is exactly of the desired form. However, we also note that F' is piecewise linear between the points. If it were now the case that F assumes its global minimum in x_k , then we would have to have $F'(x_k - \varepsilon) \leq 0$ and $F'(x_k + \varepsilon) \geq 0$ for sufficiently small values of ε which is ruled out by the explicit form of F' . \square

2.2. Some Preparatory Work. Since we now know that minimizing the W_2 cost is equivalent to minimizing the L^2 -discrepancy, we will instead work with this new, renormalized functional. To this end we introduce, for any given set of n points in the unit interval, the counting function

$$f_n(x) = \# \{1 \leq k \leq n : x_k \leq x\}.$$

We will also use, throughout the rest of the paper, the rescaled version

$$g_n(x) = \# \{1 \leq k \leq n : x_k \leq x\} - nx.$$

Assume x_{n+1} is being added. A straightforward computation shows that

$$\begin{aligned} \int_0^1 (f_{n+1}(x) - (n+1)x)^2 dx &= \int_0^{x_{n+1}} ((f_n(x) - nx) - x)^2 dx \\ &\quad + \int_{x_{n+1}}^1 ((f_n(x) - nx) + (1-x))^2 dx \\ &= \int_0^1 (f_n(x) - nx)^2 dx - 2 \int_0^{x_{n+1}} (f_n(x) - nx)x dx \\ &\quad + 2 \int_{x_{n+1}}^1 (f_n(x) - nx)(1-x) dx + \frac{x_{n+1}^3}{3} + \frac{(1-x_{n+1})^3}{3}. \end{aligned}$$

Note that some of these terms do not depend on x_{n+1} while others do. Collecting only those terms that depend on x_{n+1} shows that one needs to solve

$$E(z) + \frac{z^3 + (1-z)^3}{3} \rightarrow \min,$$

where the function E is defined as

$$E(z) = -2 \int_0^z (f_n(x) - nx)x dx + 2 \int_z^1 (f_n(x) - nx)(1-x) dx.$$

For the remainder of the paper, we will work with this representation.

2.3. An Elementary Lemma. The purpose of this section is to introduce a basic Lemma and to use it to prove the missing part of Kritzinger's Theorem: that

$$\int_0^1 (f_n(x) - nx)^2 dx \leq \frac{n}{3} + c,$$

where $c > 0$ only depends on the initial configuration. Subsequent arguments will then both use and refine this step.

Lemma (Basic Form). *Let $g : [0, 1] \rightarrow \mathbb{R}$ be continuous. Then*

$$\max_{0 \leq z \leq 1} \int_0^z g(x)x dx - \int_z^1 g(x)(1-x) dx \geq 0.$$

Proof. Using integration by parts

$$\begin{aligned} \int_0^z g(x)x dx - \int_z^1 g(x)(1-x) dx &= \int_0^z g(x)x dx + \int_z^1 g(x)(x-1) dx \\ &= \int_0^1 g(x)xdx - \int_z^1 g(x)dx. \end{aligned}$$

Introducing the anti-derivative $G : [0, 1] \rightarrow \mathbb{R}$ via $G(s) = \int_0^s g(x)dx$ we have, integrating again by parts,

$$\int_0^1 g(x)xdx = G(x)x|_0^1 - \int_0^1 G(x)dx = G(1) - \int_0^1 G(x)dx$$

as well as $\int_z^1 g(x)dx = G(1) - G(z)$. Thus

$$\int_0^z g(x)x dx - \int_z^1 g(x)(1-x) dx = G(z) - \int_0^1 G(x)dx.$$

Every function has to be at least as large as its average somewhere and

$$\max_{0 \leq z \leq 1} G(z) - \int_0^1 G(x) dx \geq 0.$$

□

One sees that the proof is merely a sequence of identities: the part where one can hope to gain is the last step which is clearly only sharp whenever G is constant. The elementary Lemma now implies the missing part of Corollary 1.

Proposition. *We have*

$$\int_0^1 (f_{n+1}(x) - (n+1)x)^2 dx \leq \int_0^1 (f_n(x) - nx)^2 dx + \frac{1}{3}.$$

Proof. Note that we have

$$\begin{aligned} \int_0^1 (f_{n+1}(x) - (n+1)x)^2 dx &= \int_0^1 (f_n(x) - nx)^2 dx - 2 \cdot E(x_{n+1}) \\ &\quad + \frac{x_{n+1}^3 + (1 - x_{n+1})^3}{3}. \end{aligned}$$

There exists x_{n+1} such that $E(x_{n+1}) \geq 0$ from which the result follows since

$$\forall 0 \leq z \leq 1 \quad \frac{z^3 + (1 - z)^3}{3} \leq \frac{1}{3}$$

□

2.4. A refined Lemma. The purpose of this section is to establish what is perhaps the core of the improvement: instead of merely using that a function has to exceed its average, as in the proof of the basic Lemma, we quantify the gain from below.

Lemma (Main Lemma). *Let $g : [0, 1] \rightarrow \mathbb{R}$ be continuous and $g \not\equiv 0$. Then*

$$\max_{0 \leq z \leq 1} \int_0^z g(x)x dx - \int_z^1 g(x)(1-x) dx \geq \frac{1}{16} \frac{1}{\|g\|_{L^2}^2} \left(\max_{0 \leq x \leq 1} \left| \int_0^x g(y) dy \right| \right)^3.$$

Proof. Using again the antiderivative

$$G(x) = \int_0^x g(y) dy$$

the computations in the basic Lemma already established that

$$\max_{0 \leq z \leq 1} \int_0^z g(x)x dx - \int_z^1 g(x)(1-x) dx = \max_{0 \leq z \leq 1} G(z) - \int_0^1 G(x) dx \geq 0.$$

We will work with this expression instead. In order to emphasize that the integral in this expression is truly a constant, the average value, we will abbreviate

$$\bar{G} = \int_0^1 G(x) dx \in \mathbb{R}$$

We start by arguing that if G assumes a value much smaller than the average, then it also has to assume values that are larger than the average. The difficulty here is that ‘much smaller than the average’ is a pointwise statement which we have to turn into a statement about the integral. This is summarized in the following statement.

Fact. *We have*

$$\max_{0 \leq z \leq 1} G(z) - \bar{G} \geq \frac{1}{8\|g\|_{L^2}^2} \cdot \left(\bar{G} - \min_{0 \leq x \leq 1} G(x) \right)^3.$$

Proof of the Fact. For simplicity of exposition, we abbreviate

$$M = \bar{G} - \min_{0 \leq x \leq 1} G(x) \geq 0.$$

Let us assume the minimum is assumed in

$$G(z_1) = \min_{0 \leq x \leq 1} G(x).$$

Let J be the shortest interval containing both z_1 and a point $0 \leq z_2 \leq 1$ for which

$$G(z_2) = G(z_1) + \frac{M}{2}.$$

The existence of such a point z_2 follows from the fact that G is continuous and

$$G(z_1) \leq G(z_1) + \frac{M}{2} \leq G(z_1) + M = \bar{G}$$

and the function G assumes both the value $G(z_1)$ and the value \bar{G} and the intermediate value theorem applies. If $M = 0$ there is nothing to prove, thus we can assume $M \neq 0$ implying $z_1 \neq z_2$. The shortest interval J containing both z_1 has positive length and z_1, z_2 are its endpoints. Using Cauchy-Schwarz on J ,

$$\begin{aligned} \frac{M}{2} &= |G(z_2) - G(z_1)| = \left| \int_J g(x) dx \right| \\ &\leq |J|^{1/2} \left(\int_J g(x)^2 dx \right)^{1/2} \leq |J|^{1/2} \left(\int_0^1 g(x)^2 dx \right)^{1/2} = |J|^{1/2} \cdot \|g\|_{L^2}. \end{aligned}$$

This provides an upper bound on M . Our next goal will be to show that $|J|$ cannot be too long: if it is long, then $\max_{0 \leq z \leq 1} G(z) - \bar{G}$ has to be large which is exactly what we are trying to show. This follows at once from

$$\begin{aligned} \bar{G} &= \int_0^1 G(x) dx = \int_J G(x) dx + \int_{[0,1] \setminus J} G(x) dx \\ &\leq \left(\bar{G} - \frac{M}{2} \right) \cdot |J| + (1 - |J|) \cdot \left(\bar{G} + \left[\max_{0 \leq z \leq 1} G(z) - \bar{G} \right] \right) \\ &\leq \bar{G} - \frac{M}{2} \cdot |J| + \left[\max_{0 \leq z \leq 1} G(z) - \bar{G} \right] \end{aligned}$$

from which we infer that

$$\max_{0 \leq z \leq 1} G(z) - \bar{G} \geq |J| \cdot \frac{M}{2}.$$

We can now combine these bounds. We have

$$\frac{M^2}{4\|g\|_{L^2}^2} \leq |J| \leq \frac{2}{M} \cdot \left[\max_{0 \leq z \leq 1} G(z) - \bar{G} \right]$$

which implies that

$$\max_{0 \leq z \leq 1} G(z) - \bar{G} \geq \frac{M^3}{8\|g\|_{L^2}^2}.$$

□

The fact will now allow us to conclude the proof of the Main Lemma. Consider the values assumed by G . Since G is continuous, we have

$$\{G(x) : 0 \leq x \leq 1\} = [A, B],$$

where $A \leq \bar{G} \leq B$. We introduce $Y = B - A$. As a first observation, we have

$$\max_{0 \leq x \leq 1} \left| \int_0^x g(y) dy \right| = \max_{0 \leq x \leq 1} |G(x)|.$$

Note that $G(0) = 0$ and thus $A \leq 0 \leq B$ from which we deduce that

$$\frac{Y}{2} \leq \max_{0 \leq x \leq 1} \left| \int_0^x g(y) dy \right| = \max_{0 \leq x \leq 1} |G(x)| \leq Y.$$

We will now show that a large value of Y will imply that G has to exceed its maximum by at least a little bit (depending on Y). For this, we write $Y = Y_1 + Y_2$, where $Y_1 = B - \bar{G}$ and $Y_2 = \bar{G} - A$. Then, tautologically,

$$\max_{0 \leq z \leq 1} G(z) - \bar{G} = Y_1.$$

Simultaneously, appealing to the fact proved above, we have

$$\max_{0 \leq z \leq 1} G(z) - \bar{G} \geq \frac{Y_2^3}{8\|g\|_{L^2}^2}.$$

Altogether, we have

$$\max_{0 \leq z \leq 1} G(z) - \bar{G} \geq \max \left\{ Y_1, \frac{Y_2^3}{8\|g\|_{L^2}^2} \right\} = \max \left\{ Y_1, \frac{(Y - Y_1)^3}{8\|g\|_{L^2}^2} \right\}$$

which we turn into a smoother object via

$$\max_{0 \leq z \leq 1} G(z) - \bar{G} \geq \frac{1}{2} \left(Y_1 + \frac{(Y - Y_1)^3}{8\|g\|_{L^2}^2} \right).$$

Differentiation shows that

$$\frac{\partial}{\partial Y_1} \left[Y_1 + \frac{(Y - Y_1)^3}{8\|g\|_{L^2}^2} \right] = 1 + \frac{3}{8} \frac{(Y - Y_1)^2}{\|g\|_{L^2}^2} \geq 1$$

which shows that the minimum is assumed for $Y_1 = 0$. Thus

$$\max_{0 \leq z \leq 1} G(z) - \bar{G} \geq \frac{1}{16} \frac{Y^3}{\|g\|_{L^2}^2}.$$

Altogether, we obtain, as desired,

$$\max_{0 \leq z \leq 1} G(z) - \bar{G} \geq \frac{1}{16} \frac{1}{\|g\|_{L^2}^2} \left(\max_{0 \leq x \leq 1} \left| \int_0^x g(y) dy \right| \right)^3.$$

□

2.5. Proof of Theorem 2.

Proof. Suppose now that $(x_k)_{k=1}^\infty$ is a Kritzinger sequence. We note that

$$\int_0^1 (f_n(x) - nx)^2 dx = \int_0^1 g_n(x)^2 dx \leq \frac{n}{3} + c,$$

where c only depends on the initial configuration. Suppose now there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have

$$\max_{0 \leq x \leq 1} \left| \int_0^x g_n(y) dy \right| \geq 2 \cdot n^{1/3}.$$

We will use this to deduce a contradiction. We have

$$\begin{aligned} \int_0^1 g_{n+1}(x)^2 dx &\leq \int_0^1 g_n(x)^2 dx + \frac{1}{3} \\ &\quad - \max_{0 \leq z \leq 1} 2 \left(\int_0^z g_n(x)x dx - \int_z^1 g_n(x)(1-x) dx \right). \end{aligned}$$

The Main Lemma implies

$$\begin{aligned} \max_{0 \leq z \leq 1} 2 \left(\int_0^z g_n(x)x dx - \int_z^1 g_n(x)(1-x) dx \right) &\geq \frac{1}{8} \frac{1}{\|g_n\|_{L^2}^2} \left(\max_{0 \leq x \leq 1} \left| \int_0^x g_n(y) dy \right| \right)^3 \\ &\geq \frac{1}{8} \frac{1}{\frac{n}{3} + c} 8n = \frac{n}{n/3 + c}. \end{aligned}$$

However, for n sufficiently large (say $n \geq N_1$ where N_1 depends only on c), this implies, for all $n \geq \max(N, N_1)$

$$\int_0^1 g_{n+1}(x)^2 dx \leq \int_0^1 g_n(x)^2 dx - 2$$

which leads to a contradiction since these integrals are always positive. \square

Remark. We note that this allows for a small refinement: the argument shows that for all N sufficiently large, there always exist $N \leq n \leq 100N$ for which

$$\max_{0 \leq x \leq 1} \left| \int_0^x g_n(y) dy \right| \leq 2 \cdot n^{1/3}.$$

REFERENCES

- [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] J. Beck, A two-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution. Compositio Math. 72 3, 269–339 (1989).
- [3] J. Beck and W. Chen, Irregularities of Distribution, Cambridge Tracts in Mathematics (No. 89), Cambridge University Press, 1987.
- [4] J. Beck, The modulus of polynomials with zeros on the unit circle: A problem of Erdos, Annals of Mathematics, 134 (1991), p. 609–651
- [5] D. Bilyk and M. Lacey, On the small ball Inequality in three dimensions, Duke Math. J. 143 (2008), no. 1, 81–115.
- [6] D. Bilyk, M. Lacey and A. Vagharshakyan, On the small ball inequality in all dimensions, J. Funct. Anal. 254 (2008), no. 9, 2470–2502.
- [7] L. Brown and S. Steinerberger, Positive-definite Functions, Exponential Sums and the Greedy Algorithm: a curious Phenomenon, Journal of Complexity 61 (2020), 101485
- [8] L. Brown and S. Steinerberger, On the Wasserstein distance between classical sequences and the Lebesgue measure, Trans. Amer. Math. Soc. 373 (2020), p. 8943–8962.

- [9] B. Chazelle, The discrepancy method. Randomness and complexity. Cambridge University Press, Cambridge, 2000.
- [10] J. van der Corput, Verteilungsfunktionen I, Proc. Akad. Wetensch. , 38 (1935), 813–821.
- [11] J. van der Corput, Verteilungsfunktionen II, Akad. Wetensch., Proc. 38 (1935), 1058–1068
- [12] J. Dick and F. Pillichshammer, Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration. Cambridge University Press, Cambridge, 2010.
- [13] M. Drmota, R. Tichy, Sequences, discrepancies and applications. Lecture Notes in Mathematics, 1651. Springer-Verlag, Berlin, 1997.
- [14] P. Erdős, Some unsolved problems, Michigan Math. J. 4 (1957), p. 291–300
- [15] C. Graham, Irregularity of distribution in Wasserstein distance. Journal of Fourier Analysis and Applications, 26 (2020), p. 1-21.
- [16] G. Larcher and F. Puchhammer, An improved bound for the star discrepancy of sequences in the unit interval, Unif. Distrib. Theory 11 (2016), p. 1–14
- [17] L. Kuipers and H. Niederreiter, Uniform distribution of sequences. Pure and Applied Mathematics. Wiley-Interscience, New York-London-Sydney, 1974.
- [18] R. Kritzinger, Uniformly distributed sequences generated by a greedy minimization of the L_2 discrepancy, arXiv:2109.06298.
- [19] V. Ostromoukhov, Recent progress in improvement of extreme discrepancy and star discrepancy of one-dimensional sequences. MCQMC 2008, pages 561–572.
- [20] F. Pausinger, Greedy energy minimization can count in binary: point charges and the van der Corput sequence, Annali di Matematica Pura ed Applicata 200 (2021), p. 165–186.
- [21] K. F. Roth, On irregularities of distribution. Mathematika 1, 73–79 (1954).
- [22] W. Schmidt, Irregularities of distribution. VII. Acta Arith. 21 (1972), 45–50.
- [23] S. Steinerberger, A Nonlocal Functional promoting Low-Discrepancy Point Sets, Journal of Complexity 54 (2019), 101410
- [24] S. Steinerberger, Dynamically Defined Sequences with Small Discrepancy, Monatshefte fur Mathematik 191 (2020), p. 639–655
- [25] S. Steinerberger, Polynomials with zeros on the unit circle: regularity of Leja sequences, Mathematika 67 (2021), p. 553–568
- [26] S. Steinerberger, Wasserstein distance, Fourier series and applications, Monatshefte fur Mathematik 194 (2021), p. 305-338
- [27] S. Steinerberger, A Wasserstein inequality and minimal Green energy on compact manifolds, Journal of Functional Analysis 281(2021):109076
- [28] S. Vallender, Calculation of the Wasserstein Distance Between Probability Distributions on the Line Theory of Probability & Its Applications 18 (1974), p. 435.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195, USA
Email address: steinerb@uw.edu