A NONLOCAL FUNCTIONAL PROMOTING LOW-DISCREPANCY POINT SETS

STEFAN STEINERBERGER

ABSTRACT. Let \( X = \{ x_1, \ldots, x_N \} \subset \mathbb{T}^d \cong [0, 1]^d \) be a set of \( N \) points in the \( d \)-dimensional torus that we want to arrange as regularly possible. The purpose of this paper is to introduce a curious energy functional

\[
E(X) = \sum_{1 \leq m, n \leq N, m \neq n} d \prod_{k=1}^{d} (1 - \log (2 \sin(\pi |x_{m,k} - y_{m,k}|)))
\]

and to suggest that moving a set \( X \) into the direction \(-\nabla E(X)\) may have the effect of increasing regularity of the set in the sense of decreasing discrepancy. We numerically demonstrate the effect for Halton, Hammersley, Kronecker, Niederreiter and Sobol sets. Lattices in \( d = 2 \) are critical points of the energy functional, some (possibly all) are strict local minima.

1. INTRODUCTION

1.1. Introduction. This paper is partially motivated by earlier results about how to distribute points on a manifold in a regular way. One idea (from [9, 13]) is to not construct these points a priori but instead use (local) minimizers of an energy functional. For example, suppose we want to distribute \( N \) points on the two-dimensional torus \( \mathbb{T}^2 \) in a way that is good for numerical integration. One way of doing this is by trying to find local minimizers of the energy functional

\[
F(X) = \sum_{1 \leq m, n \leq N, m \neq n} e^{-cN^{-1} \|x_i - x_j\|^2},
\]

where \( c \sim 1 \) is a constant. These point configurations are empirically shown [9] to be better at integrating trigonometric polynomials than commonly used classical constructions, the reason for that being a connection between the Gaussian and the heat kernel (which, in itself, can be interpreted as a mollifier in Fourier space dampening high oscillation). This method is also geometry independent and works on general compact manifolds (with \( \|x_i - x_j\| \) replaced by the geodesic distance).

1.2. The problem. We were curious whether there is any way to proceed similarly in the problem of finding low-discrepancy sets of points. Suppose \( X \subset [0, 1]^2 \) is a set \( \{ x_1, \ldots, x_N \} \) of \( N \) distinct points. A classical question is how would to arrange them so as to minimize the star discrepancy \( D_N^*(X) \) defined by

\[
D_N^*(X) = \max_{0 \leq x, y \leq 1} \left| \frac{\# \{ 1 \leq i \leq N : x_i \leq x \wedge y_i \leq y \}}{n} - xy \right|.
\]

2010 Mathematics Subject Classification. 11L03, 42B05, 82C22.

Key words and phrases. Low discrepancy sequence, gradient flow, energy functional.

The author is supported by the NSF (DMS-1763179) and the Alfred P. Sloan Foundation.
A seminal result of Schmidt [11] implies that
\[ D^*_N \gtrsim \log \frac{N}{N}. \]
Many constructions of sets are known that attain this growth, we refer to the classical textbooks by Dick & Pillichshammer [5], Drmota & Tichy [6] and Kuiers & Niederreiter [7] for descriptions. Some of the classical configurations are also used as examples in this paper. The problem is famously unsolved in higher dimensions where the best known constructions satisfy \( D_N \lesssim (\log N)^{d-1} N^{-1} \) but no matching lower bound exists (see [2, 3, 4]). Indeed, there is not even consensus as to whether the best known constructions attain the optimal growth or whether there might be more effective constructions in \( d \geq 3 \).

![Figure 1. Left: 50 points of a Niederreiter sequence with \( D^*_N \sim 0.082 \). Right: gradient flow produces a (similar) set, \( D^*_N \sim 0.061 \).](image)

We were interested in whether it is possible to assign a notion of 'energy' to a set of points that vaguely corresponds to discrepancy in the sense that moving the points in such a way that perturbations of the points decreasing the energy also often decrease discrepancy. What would be of interest is

1. a notion of 'energy' is fast to compute
2. is often helpful in improving existing point sets
3. and may have the potential to lead to new constructions.

We believe this questions to be of some interest. The purpose of this paper is to derive one functional that seems to work very well in practice. Indeed, it works shockingly well: when applied to the classical low discrepancy constructions, it always seems to further decrease discrepancy (though sometimes, when the sets are already well distributed, only by very little). We provide a heuristic explanation in §3.3. There might be many other such functionals (possibly related to different kinds of math, e.g. [1, 8, 10]) and we believe that constructing and understanding them could be quite interesting indeed.

**Open Problem.** Construct other energy functionals whose gradient flow has a beneficial effect on discrepancy. What can be rigorously proven? Can they be used for numerical integration? How do they scale in the dimension?
1.3. **Outline of the paper.** §2 introduces the energy functional and the main result. §3 explains how the energy functional was derived, describes the one-dimensional setting and relates it to Fourier analysis. A proof of the main result is given in §4. Numerical examples of how the energy functional acts on well-known constructions are given throughout the paper – these examples are all two-dimensional (for simplicity of exposition). We emphasize that the examples of point sets are all essentially picked at random, the functional does seem to work at an overwhelming level of generality and we invite the reader to try it on their own favorite sets.

2. **An energy functional**

2.1. **The functional.** Given a set \( X = \{x_1, \ldots, x_N\} \subset \mathbb{T}^d \cong [0, 1]^d \) of \( N \) points in the \( d \)-dimensional torus where each point is given by \( x_n = (x_n,1, \ldots, x_n,d) \in \mathbb{T}^d \), we introduce the energy function \( E : ([0, 1]^d)^N \to \mathbb{R} \) via

\[
E(X) = \sum_{1 \leq m, n \leq N} \prod_{k=1}^d \left( 1 - \log \left( 2 \sin \left( \pi \frac{|x_m,k - y_m,k|}{|x_n,k - y_n,k|} \right) \right) \right).
\]

We note that, for \( 0 \leq x, y \leq 1 \) we have that

\[
1 - \log \left( 2 \sin \pi |x - y| \right) \geq 1 - \log 2
\]

and so every term in the product is always positive. We also note that if two different points \( x_i, x_j \) have the same \( k \)-th coordinate, then the functional is not defined and we set \( E(X) = \infty \) in that case. In practice, we can always perturb points ever so slightly to avoid that scenario. We note that the functional has an interesting hyperbolic structure: it very much likes to avoid having too many points that have very similar coordinates. This makes sense since such points can be easily captured by a thin (hyper-)rectangle. We now first discuss how to actually minimize it in practice and then discuss our main result.

**Figure 2.** Left: 128 points of the Halton sequence in base 2 and 3 having \( D_N \sim 0.032 \). Right: evolution of the gradient flow changes the set a tiny bit to one with discrepancy \( D_N^* \sim 0.025 \).
2.2. How to compute things. We are using the standard gradient descent: if $f : \mathbb{R}^d \to \mathbb{R}$ is a differentiable function, gradient descent is trying to find a (local) minimum by defining an iterative sequence of points via

$$x_{n+1} = x_n - \alpha \nabla f(x_n),$$

were $\alpha > 0$ is the step-size. This is exactly how we proceed as well. The gradient $\nabla E$ can be computed explicitly and

$$\frac{\partial E}{\partial x_{n,i}} = \sum_{m = 1}^{N} \left( \prod_{k = 1}^{d} (1 - \log (2 \sin (\pi |x_{m,k} - y_{m,k}|))) \right) h(x_{n,i} - x_{m,i}),$$

where

$$h(x) = -\pi \cot (x) \text{sign}(x).$$

This allows us to compute

$$\frac{\partial E}{\partial x_n} = \left( \frac{\partial E}{\partial x_{n,1}}, \frac{\partial E}{\partial x_{n,2}}, \ldots, \frac{\partial E}{\partial x_{n,d}} \right)$$

which is the infinitesimal direction in which we have to move $x_n$ to get the largest increase in the energy functional. Since we are interested in decreasing it, we replace

$$x_n \leftarrow x_n - \alpha \frac{\partial E}{\partial x_n}.$$

The algorithm is somewhat sensitive to the choice of $\alpha$ (this is not surprising and a recurring theme for gradient methods): it has to be chosen so small that the first order approximation is still somewhat valid, however, if it is chosen too small, then convergence becomes very slow and one needs more iterations to converge. In practice, for point sets containing $\sim 100$ points, we worked with $\alpha \sim 10^{-5}$ which usually leads to a local minimum within less than a hundred iterations. The cost of computing a gradient step is of order $O(n^2d)$ and thus not at all unreasonable.

![Figure 3](image-url)

**Figure 3.** Left: 50 points of a Sobol sequence with $D_N^* \sim 0.063$. Right: evolution of the flow leads to a set with $D_N^* \sim 0.057$. 
2.3. **Lattices.** We observe that if the initial point set is already very well distributed, then minimizing the energy tends to have very little effect on both the set and the discrepancy. There is one setting where this behavior is especially pronounced. We will consider lattice rules of the type

\[ X_N = \left\{ \left( \frac{n}{N}, \frac{an}{N} \right) : 0 \leq n \leq N - 1 \right\}, \]

where \( a, N \in \mathbb{Z} \) are coprime and \( \{x\} = x - \lfloor x \rfloor \) is the fractional part.

**Theorem.** Every lattice rule \( X_N \) is a critical point of the energy functional. Moreover, if \( a^2 \equiv 1 \pmod{N} \), then \( X_N \) is a strict local minimum.

We do not know whether the condition \( a^2 \equiv 1 \pmod{N} \) is necessary, it seems like it should not be; we comment on this at the end of the paper. Several of the classical point sets (i.e. Sobol sequences) barely move under the gradient flow – is it maybe true that many classical sequences have a local minimum nearby?

![Figure 4. Left: 64 Halton points (base 2, 5) with \( D_N^* \sim 0.65 \). Right: the gradient flow leads to a set with \( D_N^* \sim 0.045 \).](image)

### 3. Heuristic Derivation of the Energy Functional

We first give a one-dimensional argument to avoid notational overload and then derive the analogous quantity for higher dimensions in §3.2.

#### 3.1. **One dimension.**

Our derivation is motivated by the Erdős-Turan inequality bounding the discrepancy \( D_N \) of a set \( \{x_1, \ldots, x_N\} \subset [0,1] \) by

\[ D_N \lesssim \frac{1}{N} + \sum_{k=1}^{N} \left\lfloor \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i k x_n} \right\rfloor, \]

where \( k \) is arbitrary. We can bound this from above by

\[ \sum_{k=1}^{N} \left\lfloor \frac{1}{N} \sum_{n=1}^{N} e^{2\pi i k x_n} \right\rfloor \leq \frac{1}{2} \sum_{k=1}^{N} \left( \frac{1}{k} \sum_{n=1}^{N} e^{2\pi i k x_n} \right)^2 + \frac{1}{2} k N \sum_{n=1}^{N} e^{2\pi i k x_n}. \]
Using merely this upper bound, we want to make sure that the second term is small. This second term simplifies to

\[ \frac{1}{N} \sum_{k=1}^{N} \left( \frac{1}{k} \left| \sum_{n=1}^{N} e^{2\pi i k x_n} \right|^2 \right) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{k} \sum_{n,m=1}^{N} e^{2\pi i k (x_n - x_m)} \]

Ignoring the scaling factor \( N^{-1} \), we decouple into diagonal and off-diagonal terms and obtain

\[ \sum_{k=1}^{N} \frac{1}{k} \sum_{n,m=1}^{N} e^{2\pi i k (x_n - x_m)} = \sum_{k=2}^{N} \frac{N}{k} + \sum_{k=1}^{N} \frac{1}{k} \sum_{m,n}^{N} \cos \left( 2\pi k (x_m - x_n) \right) \]

The first term is a fixed constant and thus independent of the actual points, the second sum can be written as

\[ \sum_{k=1}^{N} \frac{1}{k} \sum_{m,n=1}^{N} \cos \left( 2\pi k (x_m - x_n) \right) = \sum_{m,n=1}^{N} \frac{1}{k} \sum_{m \neq n}^{N} \cos \left( 2\pi k (x_m - x_n) \right) \]

The inner sum can now be simplified by letting the limit go to infinity since

\[ \sum_{k=1}^{\infty} \frac{\cos \left( 2\pi k x \right)}{k} = -\log \left( 2 \sin \left( \pi |x| \right) \right) \]

This suggests that we should really try to minimize the functional

\[ E(x) = \sum_{m,n=1}^{N} -\log \left( 2 \sin \left( \pi |x_m - x_n| \right) \right) \]

3.2. Higher dimensions. The general case follows from the Erdős-Turan-Koksma inequality and the heuristic outlined above for the one-dimensional case. We recall that the Erdős-Turan-Koksma inequality allows us to bound the discrepancy of a set \( \{x_1, \ldots, x_N\} \subset [0,1]^d \) by

\[ D_N \lesssim_d \frac{1}{M+1} + \sum_{\|k\|_\infty \leq M} \frac{1}{r(k)} \frac{1}{N} \left| \sum_{\ell=1}^{N} e^{2\pi i (k \cdot x_\ell)} \right| \]

where \( r: \mathbb{Z}^d \to \mathbb{N} \) is given by

\[ r(k) = \prod_{j=1}^{d} \max \{1, k_j\} \]

We obtain

\[ \sum_{\|k\|_\infty \leq M} \frac{1}{r(k)} \frac{1}{N} \left| \sum_{\ell=1}^{N} e^{2\pi i (k \cdot x_\ell)} \right| \leq \sum_{\|k\|_\infty \leq M} \frac{1}{r(k)} \frac{1}{N} \left| \sum_{\ell=1}^{N} e^{2\pi i (k \cdot x_\ell)} \right|^2 + \sum_{\|k\|_\infty \leq M} \frac{1}{r(k)} \frac{1}{N} \left| \sum_{\ell=1}^{N} e^{2\pi i (k \cdot x_\ell)} \right|^2 \]
The second sum we can expand into
\[ \sum_{\|k\|_\infty \leq M} \frac{1}{r(k)} \frac{1}{N} \left| \sum_{\ell=1}^N e^{2\pi i \langle k, x_\ell \rangle} \right|^2 = \frac{1}{N} \sum_{m,n=1}^N \prod_{j=1}^d \left( 1 + \sum_{k \in \mathbb{Z}, k \neq 0} \frac{1}{k} e^{2\pi i (x_{m,j} - x_{n,j})} \right). \]

Letting \( M \to \infty \), we can simplify every one of these terms to
\[ \sum_{k \in \mathbb{Z}, k \neq 0} \frac{e^{2\pi i k (x_{m,j} - x_{n,j})}}{k} = -\log (2 \sin (\pi |x_{m,j} - x_{n,j}|)) \]
and we obtain the general form of the energy functional.

3.3. A heuristic explanation. The Erdős-Turan-Koksma inequality shows
\[ D_N \lesssim d \frac{1}{M+1} + \sum_{\|k\|_\infty \leq M} \frac{1}{r(k)} \frac{1}{N} \left| \sum_{\ell=1}^N e^{2\pi i \langle k, x_\ell \rangle} \right| \]
We know that the best possible behavior is on the scale of \( D_N \lesssim (\log N)^{d-1} N^{-1} \) (or possibly even smaller). This suggests that the exponential sums cannot typically be that large, it should be roughly at scale \( \sim 1 \) most of the time. If it happens to be a lot larger many times, the sequence is presumably not all that regularly distributed. Since one would expect most Fourier coefficients to be at scale \( \sim 1 \) (especially for sets that are already well distributed), one might assume that the actual impact of squaring the expression is not that large. Of course this is very heuristic and will be false in many cases, a more rigorous explanation is needed. We conclude by establishing a rigorous bound.

Lemma. We have, for \( \{x_1, \ldots, x_N\} \subseteq \mathbb{T}^d \),
\[ \sum_{\|k\|_\infty \leq N} \frac{1}{r(k)} \left| \sum_{\ell=1}^N e^{2\pi i \langle k, x_\ell \rangle} \right|^2 \lesssim_d E(X) \]
Proof. The argument outlined above already establishes the result except for one missing ingredient: for all $0 < x < 1$, there is a uniform bound

$$\max_{n \in \mathbb{N}} \sum_{k=1}^{n} \frac{\cos(2\pi kx)}{k} \lesssim -\log |\sin(\pi x)|.$$ 

We can assume w.l.o.g. that $0 < x < 1/2$. We use Abel summation to write

$$\sum_{k=1}^{n} \frac{\cos(2\pi kx)}{k} = (n + 1) \frac{\cos(2\pi nx)}{n} + \int_{1}^{n} [k + 1] \left( \frac{\cos(2\pi kx)}{k^2} + \frac{2\pi x \sin(2\pi kx)}{k} \right) dk.$$ 

The first term is $O(1)$, it remains to treat the integral. The first term has the structure of an alternating Leibniz series with the first root being at $kx = 1/4$. Thus

$$\int_{1}^{n} [k + 1] \frac{\cos(2\pi kx)}{k^2} dk \lesssim \int_{1}^{1/(4x)} [k + 1] \frac{\cos(2\pi kx)}{k^2} dk \lesssim \int_{1}^{1/(4x)} \frac{\cos(2\pi kx)}{k} dk \lesssim \log(1/x).$$

The second integral simplifies to

$$\int_{1}^{n} [k + 1] \frac{2\pi x \sin(2\pi kx)}{k} dk = 2\pi x \int_{1}^{n} [k + 1] \frac{\sin(2\pi kx)}{k} dk \lesssim 1.$$ 

$\square$

Figure 6. Left: 50 points of the Hammersley sequence in base 3 with $D_N \sim 0.064$. Right: evolution of the flow leads to a set with $D_N \sim 0.042$. The Hammersley sets in base 2 and 5 barely move under the flow.
3.4. A one-dimensional result. Things are usually simpler in one dimension (though also less interesting because the optimal constructions are trivial and given by equispaced points). We have the following basic result.

**Proposition.** Let \( (x_n) \) be a sequence in \( T \cong [0, 1] \). If

\[
\limsup_{N \to \infty} \frac{1}{N^2} \sum_{1 \leq m \neq n \leq N} \left( 1 - \log \left( 2 \sin \left( \pi |x_m - x_n| \right) \right) \right) = 1,
\]

then the sequence is uniformly distributed.

**Proof.** The proof is similar in spirit to the main argument in [12], we refer to that paper for definition of the Jacobi \( \theta \)-function and the main idea. We define a one-parameter family of functions via

\[
f_N(t, x) = \sum_{k=1}^N \theta_t(x - x_k)
\]

where \( \theta_t \) is the Jacobi \( \theta \)-function. In particular

\[
\lim_{t \to 0^+} f_N(t, x) = \frac{1}{N} \sum_{k=1}^N \delta_{x_k} \quad \text{in the sense of weak convergence.}
\]

Defining

\[
g(x) = 1 - \log \left( 2 \sin \left( \pi |x| \right) \right),
\]

we can define the function

\[
h(t) = \langle g \ast f_N(t, x), f_N(t, x) \rangle
\]

is monotonically decaying in time. This is seen by applying the Plancherel theorem

\[
h(t) = \sum_{k \in \mathbb{Z}} \hat{g}(k) |\widehat{f_N(t, x)}(k)|^2
\]

\[
= \sum_{k \in \mathbb{Z}} \hat{g}(k) e^{-4\pi^2 k^2 t} \left| \sum_{\ell=1}^N e^{-2\pi i k x_\ell} \right|^2
\]

and using \( \hat{g}(k) > 0 \). We can now take the limit \( t \to \infty \) and obtain that

\[
h(t) \geq \hat{g}(0) N^2 = N^2.
\]

As for the second part of the argument, suppose that \( (x_n) \) is not uniformly distributed. Weyl's theorem implies that there exist \( \varepsilon > 0, k \in \mathbb{N} \) such that

\[
\left| \sum_{\ell=1}^N e^{-2\pi i k x_\ell} \right|^2 \geq \varepsilon \quad \text{for infinitely many } n.
\]

Then, however,

\[
h(1) \geq \hat{g}(0) N^2 + e^{-4\pi^2 k^2} |\hat{g}(k)|^2 \geq (1 + \delta) N^2
\]

for some \( \delta > 0 \) and infinitely many \( N \). \( \square \)
4. Proof of the Theorem

4.1. An Inequality. We first prove an elementary inequality.

**Lemma.** Let $0 < x, y < 1$. Then

$$2\left|\cot(\pi x) \cot(\pi y)\right| < (1 - \log (2 \sin (\pi x))) \csc^2 (\pi y) + \csc^2 (\pi x)(1 - \log (2 \sin (\pi y))).$$

**Proof.** The right-hand side is always positive, we can thus assume w.l.o.g. that $0 < x, y < 1/2$. Multiplying with $\sin^2 (\pi x) \sin^2 (\pi y)$ on both sides leads to the equivalent statement $A \leq B$, where

$$A = 2 \sin (\pi x) \cos (\pi x) \sin (\pi y) \cos (\pi y)$$

and

$$B = (1 - \log (2 \sin (\pi x))) \sin^2 (\pi x) + \sin^2 (\pi y)(1 - \log (2 \sin (\pi y))).$$

We use $2ab \leq a^2 + b^2$ to argue that

$$A \leq \sin^2 (\pi x) \cos^2 (\pi x) + \sin^2 (\pi y) \cos^2 (\pi y).$$

The result then follows from the inequality

$$\cos^2 (\pi x) < 1 - \log (2 \sin (\pi x)) \quad \text{for all } 0 < x \leq \frac{1}{2}$$

which can be easily seen by elementary methods. \hfill \Box

---

**Figure 7.** Left: 100 points of a lattice rule with discrepancy $D_N^* \sim 0.08$. Right: evolution of the gradient flow leads to a set with half the discrepancy $D_N^* \sim 0.04$.

4.2. Proof of the Theorem.

**Proof of the Theorem.** The symmetries of the sequence and the energy functional imply that it is sufficient to show that the energy is locally convex around the point in $(0, 0)$. This means we want to show that

$$f(\varepsilon, \delta) = \sum_{n=1}^{N-1} \left(1 - \log \left(2 \sin \left(\pi \left|\frac{n}{N} - \varepsilon\right| \right)\right) \left(1 - \log \left(2 \sin \left(\pi \left|\frac{an}{N} - \delta\right| \right)\right)\right)$$

\hfill \Box
is strictly positive for all $\varepsilon, \delta$ sufficiently small. We can assume $|\varepsilon|, |\delta| < N^{-1}$, expand the first term in $\varepsilon$ up to second order and note that

$$1 - \log(2 \sin(\pi(x - \varepsilon))) = (1 - \log(2 \sin(\pi x))) + \pi \cot(\pi \varepsilon) + \pi^2 \csc^2(\pi x) \frac{\varepsilon^2}{2} + O(\varepsilon^3).$$

This shows that

$$\frac{\partial}{\partial \varepsilon} g(\varepsilon, 0)|_{\varepsilon=0} = \sum_{n=1}^{N-1} \cot\left(\frac{n\pi}{N}\right)\left(1 - \log\left(2 \sin\left(\pi \left\{\frac{n}{N}\right\}\right)\right)\right).$$

We group the summand $n$ and $N-n$ and observe that $\cot$ is odd on $(0, \pi)$ while the second summand is even, therefore the sum evaluates to 0. The other derivative

$$\frac{\partial}{\partial \varepsilon} g(0, \delta)|_{\delta=0} = \sum_{n=1}^{N-1} \cot\left(\pi \left\{\frac{n}{N}\right\}\right)\left(1 - \log\left(2 \sin\left(\pi \left\{\frac{n}{N}\right\}\right)\right)\right)$$

vanishes for exactly the same reason and therefore the lattice is a critical point.

It remains to show that it is a local minimizer which requires an expansion up to second order. This expansion naturally decouples into three sums, where

$$(I) = \frac{\pi^2 \varepsilon^2}{2} \sum_{n=1}^{N-1} \csc^2\left(\pi \left\{\frac{n}{N}\right\}\right)\left(1 - \log\left(2 \sin\left(\pi \left\{\frac{n}{N}\right\}\right)\right)\right)$$

$$(II) = \pi^2 \varepsilon \delta \sum_{n=1}^{N-1} \cot\left(\pi \left\{\frac{n}{N}\right\}\right) \cot\left(\pi \left\{\frac{n}{N}\right\}\right)$$

$$(III) = \frac{\pi^2 \delta^2}{2} \sum_{n=1}^{N-1} \csc^2\left(\pi \left\{\frac{n}{N}\right\}\right)\left(1 - \log\left(2 \sin\left(\pi \left\{\frac{n}{N}\right\}\right)\right)\right)$$

Figure 8. Left: 101 points combined from a Halton sequence ($x \leq 0.5$) and a Sobol sequence ($x \geq 0.5$) with $D_N^* \sim 0.042$. Right: gradient flow leads to a set with discrepancy $D_N^* \sim 0.034$. Most of the (little) activity is around $\{x = 0.5\}$. 
We can now argue that \( (II) \) is bounded by
\[
\left| \sum_{n=1}^{N-1} \varepsilon \delta \cot \left( \frac{\pi n}{N} \right) \cot \left( \left\{ \frac{an}{N} \right\} \right) \right| \leq \left( \frac{\varepsilon^2}{2} + \frac{\delta^2}{2} \right) \left| \sum_{n=1}^{N-1} \cot \left( \frac{\pi n}{N} \right) \cot \left( \left\{ \frac{an}{N} \right\} \right) \right|
\leq \frac{\varepsilon^2}{2} \left| \sum_{n=1}^{N-1} \cot \left( \frac{\pi n}{N} \right) \cot \left( \left\{ \frac{an}{N} \right\} \right) \right| + \frac{\delta^2}{2} \left| \sum_{n=1}^{N-1} \cot \left( \frac{\pi n}{N} \right) \cot \left( \left\{ \frac{an}{N} \right\} \right) \right|.
\]

The Lemma implies that we can bound the first term by
\[
\sum_{n=1}^{N-1} \left| \cot \left( \frac{\pi n}{N} \right) \cot \left( \left\{ \frac{an}{N} \right\} \right) \right| \leq \frac{1}{2} \sum_{n=1}^{N-1} \csc^2 \left( \frac{\pi n}{N} \right) \left( 1 - \log \left( 2 \sin \left( \frac{\pi n}{N} \right) \right) \right) + \frac{1}{2} \sum_{n=1}^{N-1} \csc^2 \left( \frac{\pi an}{N} \right) \left( 1 - \log \left( 2 \sin \left( \frac{\pi an}{N} \right) \right) \right)
\]

We finally use the algebraic structure and argue that if \( a^2 \equiv 1 \mod N \), then
\[
n \rightarrow a \cdot n \quad \text{is an involution mod } N
\]
and that implies that both sums are actually the same sum written in a different order. The arising sum is actually the term we are given in \( (I) \). The argument for the third sum is identical and altogether we conclude that
\[
(II) \leq (I) + (III)
\]
which implies the desired result. \( \square \)

It remains an open question whether the same result \( (II) \leq (I) + (III) \) remains true in general. Basic numerical experiments suggest that this should be the case.

We can reformulate the problem by writing out the quadratic form and computing its determinant. The relevant question is then whether \( (1)(2) \geq (3)^2 \), where
\[
(1) = \sum_{n=1}^{N-1} \csc^2 \left( \frac{\pi n}{N} \right) \left( 1 - \log \left( 2 \sin \left( \frac{\pi n}{N} \right) \right) \right)
\]
\[
(2) = \sum_{n=1}^{N-1} \cot \left( \frac{\pi n}{N} \right) \cot \left( \left\{ \frac{an}{N} \right\} \right)
\]
\[
(3) = \sum_{n=1}^{N-1} \csc^2 \left( \frac{\pi an}{N} \right) \left( 1 - \log \left( 2 \sin \left( \frac{\pi an}{N} \right) \right) \right).
\]

**References**

[1] A. Baernstein II, A minimum problem for heat kernels of flat tori. Extremal Riemann surfaces (San Francisco, CA, 1995), 227–243, Contemp. Math., 201, Amer. Math. Soc., Providence, RI, 1997.

[2] D. Bilyk, Roths orthogonal function method in discrepancy theory, Uniform Distribution Theory 6 (2011), no. 1, 143–184.

[3] D. Bilyk and M. Lacey, On the small ball Inequality in three dimensions, Duke Math. J. 143 (2008), no. 1, 81–115.

[4] D. Bilyk, M. Lacey and A. Vagharshakyan, On the small ball inequality in all dimensions, J. Funct. Anal. 254 (2008), no. 9, 2470–2502.
[5] J. Dick and F. Pillichshammer, Digital nets and sequences. Discrepancy theory and quasi-Monte Carlo integration. Cambridge University Press, Cambridge, 2010.
[6] M. Drmota, R. Tichy, Sequences, discrepancies and applications. Lecture Notes in Mathematics, 1651. Springer-Verlag, Berlin, 1997.
[7] L. Kuipers and H. Niederreiter, Uniform distribution of sequences. Pure and Applied Mathematics. Wiley-Interscience, New York-London-Sydney, 1974.
[8] H. Montgomery, Minimal theta functions. Glasgow Math. J. 30 (1988), no. 1, 75–85.
[9] J. Lu, M. Sachs and S. Steinerberger, Quadrature Points via Heat Kernel Repulsion, arXiv:1804.02227
[10] B. Oesting and J. Marzuola, Spectrally optimized pointset configurations. Constr. Approx. 46 (2017), no. 1, 1–35.
[11] W. Schmidt, Irregularities of distribution. VII. Acta Arith. 21 (1972), 45–50.
[12] S. Steinerberger, Localized Quantitative Criteria for Equidistribution, arxiv, Acta Arithmetica, 180, 183–199.
[13] S. Steinerberger, Spectral Limitations of Quadrature Rules and Generalized Spherical Designs, arXiv:1708.08736
[14] W. H. Young, On a certain series of Fourier, Proc. London Math. Soc. 11 (1912), 357–366

Department of Mathematics, Yale University, New Haven, CT 06511, USA
E-mail address: stefan.steinerberger@yale.edu