ACTA ARITHMETICA
180.2 (2017)

Localized quantitative criteria for equidistribution

by

Stefan Steinerberger (New Haven, CT)

1. Introduction

1.1. Introduction. Let $(x_n)_{n=1}^{\infty}$ be a sequence on $[0,1]$. A naturally associated object of interest is the behavior of gaps on a local scale. If the sequence consists of independently and uniformly distributed random variables, then

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ 1 \le m \ne n \le N : |x_m - x_n| \le \frac{s}{N} \right\} = 2s \quad \text{almost surely.}$$

Whenever a deterministic sequence $(x_n)_{n=1}^{\infty}$ has the same property, we say it has *Poissonian pair correlation*; this notion of pseudorandomness has been intensively investigated (see e.g. [5, 8, 9, 10]). Recently, Aistleitner, Lachmann & Pausinger [1] and Grepstad & Larcher [4] independently established that sequences with Poissonian pair correlation are uniformly distributed.

Theorem 1 (Aistleitner–Lachmann–Pausinger [1], Grepstad–Larcher [4]). *Let $(x_n)_{n=1}^{\infty}$ be a sequence on $[0,1]$ and assume that for all $s>0$,*

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ 1 \le m \ne n \le N : |x_m - x_n| \le \frac{s}{N} \right\} = 2s \quad \textit{a.s.}$$

Then $(x_n)_{n=1}^{\infty}$ is uniformly distributed.

An intuitive explanation is that any type of clustering produces many pairs (x_m, x_n) which are close to each other—the two available proofs are very different; the proof in [4] also implies a quantitative estimate on discrepancy. The purpose of this paper is to embed this result into an entire family of criteria that imply uniform distribution—this family of criteria

---

2010 *Mathematics Subject Classification*: Primary 11K06, 42A16; Secondary 42A82, 94A11.
Key words and phrases: uniform distribution, pair correlation, heat kernel, crystallization.
Received 10 April 2017.
Published online 9 August 2017.

DOI: 10.4064/aa170410-22-5

strengthen the result cited above and are applicable on general compact manifolds.

1.2. Qualitative results on the torus. We start by formulating our result in the special case of the one-dimensional torus $\mathbb{T}$ (normalized to have length 1). Let

$$\theta_t(x) = 1 + 2\sum_{n=1}^{\infty} e^{-4\pi^2 n^2 t}\cos(2\pi n x)$$

denote the Jacobi θ-function. We observe that $\theta_t(x) \geq 0$, $\theta_t(x) = \theta_t(1-x)$, and

$$\int_0^1 \theta_t(x)\,dx = 1, \qquad \theta_t(x) \sim \begin{cases} 1/\sqrt{t} & \text{for } |x| \lesssim \sqrt{t}, \\ 0 & \text{otherwise.} \end{cases}$$

This looks roughly like a Gaussian centered at 0 with standard deviation $\sim t^{1/2}$ (the profile indeed converges to that of a Gaussian as $t \to 0$). Note that the property $\theta_t(x) = \theta_t(1-x)$ implies that we never have to distinguish between points on the unit interval $[0,1]$ and points on the torus $\mathbb{T}$ of length 1. Our main result, restricted to the one-dimensional torus $\mathbb{T}$, is as follows.

Corollary 1. *Let $(x_n)_{n=1}^{\infty}$ be a sequence on $\mathbb{T}$. If there exists a sequence $(t_n)_{n=1}^{\infty}$ of positive and bounded real numbers such that*

$$\lim_{N\to\infty} \frac{1}{N^2}\sum_{m,n=1}^{N} \theta_{t_N}(x_n - x_m) = 1,$$

then $(x_n)_{n=1}^{\infty}$ is uniformly distributed on $\mathbb{T}$.

If $(t_n)_{n=1}^{\infty}$ additionally converges to 0, then we can simplify the criterion by replacing the Jacobi θ-function with a suitable scaled Gaussian: indeed, if $t_N \to 0$, then

$$\lim_{N\to\infty} \frac{1}{N^2}\sum_{m,n=1}^{N} \frac{1}{\sqrt{t_N}}\exp\left(-\frac{1}{t_N}(x_m - x_n)^2\right) = \sqrt{\pi}$$

immediately implies

$$\lim_{N\to\infty} \frac{1}{N^2}\sum_{m,n=1}^{N} \theta_{t_N}(x_n - x_m) = 1.$$

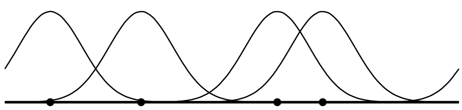

Fig. 1. Highly localized Gaussians evaluated at neighboring points

It is not difficult to see (and will be established as part of our argument) that for any $0 < s < t$ and any $x_1, \ldots, x_N \in \mathbb{T}$,

$$\frac{1}{N^2} \sum_{m,n=1}^{N} \theta_s(x_n - x_m) \geq \frac{1}{N^2} \sum_{m,n=1}^{N} \theta_t(x_n - x_m) \geq 1.$$

This shows that the criterion becomes more restrictive if the sequence of scales $(t_n)_{n=1}^{\infty}$ is made smaller. Conversely, if that sequence is taken to be the constant sequence, $t_n = 1$ for all $n \in \mathbb{N}$, the criterion becomes sharp and characterizes uniform distribution.

Corollary 2. *A sequence $(x_n)_{n=1}^{\infty}$ is uniformly distributed on $[0,1]$ or $\mathbb{T}$ if and only if*

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{m,n=1}^{N} \theta_1(x_n - x_m) = 1.$$

There is nothing special about $t = 1$ and the result holds if θ_1 is replaced by θ_c for any fixed $c > 0$. Our proof gives a little more information and shows that we could replace θ_1 by any function $\phi \in C^{\infty}(\mathbb{T})$ satisfying

$$\int_{\mathbb{T}} \phi(x)\, dx = 1, \qquad \phi(x) = \phi(1-x)$$

as well as

$$\int_{\mathbb{T}} \phi(x) e^{2\pi i k x}\, dx > 0 \quad \text{for all } k \in \mathbb{N}.$$

This result is related in spirit to the classical Bochner–Herglotz theorem, and variants exist on other topological groups (see e.g. [6]). We emphasize that Corollary 1, using the Jacobi θ-function, is a consequence of a result on general compact manifolds that does not assume any type of group structure.

1.3. Application to pair correlation. Theorem 1 states that if for all $s > 0$,

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ 1 \leq m \neq n \leq N : |x_m - x_n| \leq \frac{s}{N} \right\} = 2s,$$

then the sequence is uniformly distributed. A natural question is whether it is truly necessary to require the limit relation to hold for *all* $s > 0$. We can use Corollary 1 to show that it suffices to know it for all $s \in \mathbb{N}$ (sharper results could be obtained but this is not the focus of this paper). This should still be far from optimal and sharper criteria could be of interest.

Corollary 3 (Pair correlation). *Let $(x_n)_{n=1}^{\infty}$ be a sequence on $[0,1]$ and assume that for all $s \in \mathbb{N}$,*

$$\lim_{N\to\infty} \frac{1}{N} \# \left\{ 1 \leq m \neq n \leq N : |x_m - x_n| \leq \frac{s}{N} \right\} = 2s.$$

Then $(x_n)_{n=1}^{\infty}$ is uniformly distributed.

Corollary 2 cannot be directly applied to use pair correlation: for that we would be required to work at scale $t_N \sim N^{-2}$ and it is not too difficult to see that this is where the criterion has to stop working because the diagonal terms are already too large:

$$\frac{1}{N^2}\sum_{m,n=1}^{N}\theta_{t_N}(x_n - x_m) \geq \frac{1}{N^2}\sum_{n=1}^{N}\theta_{t_N}(x_n - x_n)$$
$$= \frac{\theta_{t_N}(0)}{N} \sim \frac{1}{N t_N^{1/2}} \gtrsim 1.$$

However, it is fairly easy to see that it is possible to make a slight adaption to the scale, essentially $t_N \sim f(N)N^{-2}$ for some suitably chosen and very slowly growing unbounded sequence $f(N)$ (growing at a speed depending on the speed of convergence of the pair correlation function). Many variants are conceivable; we prove the following natural generalization.

Corollary 4 (Weak pair correlation). *Let $(x_n)_{n=1}^{\infty}$ be a sequence on $[0,1]$, let $0 < \alpha < 1$ and assume that for all $s > 0$,*

$$\lim_{N\to\infty} \frac{1}{N^{2-\alpha}} \# \left\{ 1 \leq m \neq n \leq N : |x_m - x_n| \leq \frac{s}{N^{\alpha}} \right\} = 2s \qquad a.s.$$

Then $(x_n)_{n=1}^{\infty}$ is uniformly distributed.

We emphasize that this result is more widely applicable since the requirement of weak pair correlation is less stringent. Consider, for example, a sequence $(x_n)_{n=1}^{\infty}$ obtained by taking x_{2n-1} to be i.i.d. uniformly chosen random variables on $[0,1]$ and $x_{2n} = x_{2n-1}$. This sequence does not have Poissonian pair correlation but does have weak pair correlation for all $0 < \alpha < 1$ a.s. The argument naturally generalizes to other geometric spaces and can be roughly summarized as saying that whenever the quantities

$$\#\{1 \leq m, n \leq N : \|x_m - x_n\| \leq s t_N\} \quad \text{for some } t_N \to 0 \text{ and all } s > 0$$

behave as Poissonian random variables, then $(x_n)_{n=1}^{\infty}$ is uniformly distributed since criteria of this type can be used to determine the validity of the limit relation in our criterion.

1.4. A quantitative result. Our method is flexible enough to allow for the derivation of quantitative results. We only discuss the simplest case; the method applies to fairly general discrepancy systems on compact manifolds. Let now $x_1, \ldots, x_N \in \mathbb{T}$. *Discrepancy* is defined as the maximum deviation of uniform and empirical distribution on the set of all intervals $J \subset \mathbb{T}$:

$$D_N(\{x_1, \ldots, x_N\}) = \sup_{J \subset \mathbb{T}} \left| \frac{\#\{1 \leq i \leq N : x_i \in J\}}{N} - |J| \right|.$$

It is easy to see that D_N tends to 0 as $N \to \infty$ if and only if $(x_n)_{n=1}^{\infty}$ is uniformly distributed. We recall that for all $x_1, \ldots, x_N \in \mathbb{T}$,

$$\frac{1}{N^2} \sum_{m,n=1}^{N} \theta_t(x_n - x_m) \geq 1$$

and decreases in t.

COROLLARY 5 (Discrepancy bound). *There exists a universal constant $c > 0$ such that for any $x_1, \ldots, x_N \in \mathbb{T}$,*

$$D_N^2 \leq c\bigg(\frac{1}{N^2} \sum_{m,n=1}^{N} \theta_f(x_n - x_m) - 1\bigg) \quad \textit{where} \quad f = -\frac{D_N^2}{c \log D_N}.$$

We believe the result to be somewhat amusing but do not know whether it can be useful in a more general context. We note that the θ-function operates on spatial scale $\sim D_N(\log D_N)^{-1/2}$. Similar results can be obtained on general manifolds using the same argument. We point out a connection with crystallization problems: given N points on a manifold interacting via a nonlocal energy, minimizing configurations often arrange themselves into periodic structures (we refer to [3] for an introduction and to [2, 7] for results involving the Jacobi θ-function).

1.5. The general result. Let (M, g) be a smooth compact manifold. We use $e^{t\Delta}$ to denote the *heat kernel*, i.e. the semigroup that allows one to solve the heat equation

$$(\partial_t - \Delta_g)e^{t\Delta}u_0 = 0 \quad \text{ on } M \times [0, \infty],$$

and will apply it mostly to Dirac δ functions on the manifold. Note that classical short-time asymptotics show

$$[e^{t\Delta}\delta_x](y) \sim t^{-d/2} \exp\bigg(-\frac{d(x,y)^2}{4t}\bigg),$$

where $d(x, y)$ is the geodesic distance; we are therefore, for t sufficiently small, essentially dealing with Gaussians centered at x. We will prove the general inequality

$$\frac{1}{N^2} \sum_{m,n=1}^{N} (e^{t\Delta}\delta_{x_m})(x_n) \geq \frac{1}{\operatorname{vol}(M)},$$

and that asymptotic sharpness characterizes uniform distribution of $(x_n)_{n=1}^{\infty}$.

THEOREM 2. *Let (M, g) be a smooth compact manifold and let $(x_n)_{n=1}^{\infty}$ be a sequence on M. If there exists a bounded sequence of times $0 < t_N \leq C$*

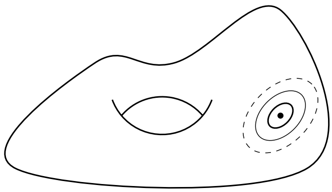

Fig. 2. $[e^{t\Delta}\delta_x](y)$ behaves like a Gaussian centered at y and for scale $\sim \sqrt{t}$

such that

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{m,n=1}^{N} (e^{t_N \Delta}\delta_{x_m})(x_n) = \frac{1}{\mathrm{vol}(M)},$$

then $(x_n)_{n=1}^{\infty}$ *is uniformly distributed on* (M,g). *Moreover, if* $t_N = c > 0$ *is constant, then this limit relation holds if and only if* $(x_n)_{n=1}^{\infty}$ *is uniformly distributed.*

It is immediately clear that on special manifolds such as $\mathbb{T}^d$ or $\mathbb{S}^{d-1}$ (where explicit formulas for the heat kernel are available), the result could be simplified and put into a similar form to the corollaries above on $[0,1]$ or $\mathbb{T}$. It is also not difficult to see that the condition in Theorem 2 can never be satisfied if t_N decays faster than $N^{-2/d}$ since

$$\frac{1}{N^2} \sum_{m,n=1}^{N} (e^{t_N \Delta}\delta_{x_m})(x_n) \ge \frac{1}{N^2} \sum_{m=1}^{N} (e^{t_N \Delta}\delta_{x_m})(x_m) \gtrsim \frac{1}{N} t_N^{-d/2},$$

which is ~ 1 for $t_N \sim N^{-2/d}$. We quickly sketch what happens if we apply the result on the torus: the heat kernel has the explicit form

$$[e^{t\Delta}(\delta_x)](y) = \theta_t(x-y).$$

The comparison between the Jacobi θ-function $\theta_t(x)$ for t small comes from the asymptotic estimate

$$\theta_t(x) \sim \frac{1}{\sqrt{4\pi t}} \exp\left(-\frac{x^2}{4t}\right).$$

It is easy to see that the error incurred is small (indeed, many orders smaller than would be required).

2. Proof of the main theorem

2.1. Warming up. Before going into details, we give a very simple argument for the monotonicity formula in the simplest case and prove that for any $0 < s < t$ and any $x_1, \ldots, x_N \in \mathbb{T}$,

$$\frac{1}{N^2} \sum_{m,n=1}^{N} \theta_s(x_n - x_m) \ge \frac{1}{N^2} \sum_{m,n=1}^{N} \theta_t(x_n - x_m) \ge 1.$$

The proof is straightforward: first, we rewrite the expression as

$$\frac{1}{N^2}\sum_{m,n=1}^{N}\theta_s(x_n-x_m)=\int_{\mathbb{T}}\theta_s*\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)\overline{\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)}\,dx.$$

The Plancherel identity yields

$$\int_{\mathbb{T}}\theta_s*\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)\overline{\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)}\,dx$$
$$=\sum_{\ell\in\mathbb{Z}}\theta_s*\widehat{\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)}(\ell)\overline{\widehat{\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)}(\ell)}.$$

We note that

$$\theta_t*\widehat{\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)}(\ell)=\widehat{\theta}_t(\ell)\widehat{\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)}(\ell)\quad\text{and}\quad\widehat{\theta}_t=\sum_{\ell\in\mathbb{Z}}e^{-4\pi^2\ell^2t}e^{2\pi i\ell x}.$$

Finally, an expansion into Fourier series

$$\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}=\sum_{\ell\in\mathbb{Z}}a_\ell e^{2\pi i\ell x}$$

with $a_\ell=a_{-\ell}$ and $a_0=1$ allows us to write

$$\int_{\mathbb{T}}\theta_t*\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)\overline{\left(\frac{1}{N}\sum_{n=1}^{N}\delta_{x_n}\right)}\,dx=\sum_{\ell\in\mathbb{Z}}e^{-4\pi^2\ell^2t}|a_\ell|^2=1+\sum_{\substack{\ell\in\mathbb{Z}\\ \ell\neq 0}}e^{-4\pi^2\ell^2t}|a_\ell|^2,$$

which is decreasing in t. On general manifolds, we will repeat the argument with the Fourier basis of L^2 being replaced by eigenfunctions of $-\Delta_g$. The semigroup properties of the heat kernel serve as a substitute for the behavior of convolution under Fourier transform.

2.2. Structure of the argument. We quickly outline the overall structure of the argument and will then divide the proof accordingly. The proof has five steps.

(1) We will start by showing that

$$\frac{1}{N^2}\sum_{m,n=1}^{N}(e^{t\Delta}\delta_{x_m})(x_n)\text{ is decreasing in }t.$$

Since we are dealing with a bounded sequence of times $0<t_N\leq C$, monotonicity implies that it suffices to prove the main result only for $t_n=C$

(2) If the sequence is not uniformly distributed, there exist a ball B and $\varepsilon>0$ such that for infinitely many N there are $(|B|+\varepsilon)N/\mathrm{vol}(M)$ out of the

first N elements contained in the ball B. We then consider, for a sufficiently small but fixed time $\delta = \delta_{B,\varepsilon} > 0$, the function

$$e^{\delta\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m}$$

and prove that its the average value in a small neighborhood of B is greater than $\mathrm{vol}(M)^{-1} + c_{\varepsilon,\delta,B}$ for some fixed $c_{\varepsilon,\delta,B} > 0$ and infinitely many $N \in \mathbb{N}$.

(3) The Cauchy–Schwarz inequality then implies

$$\left\| e^{\delta\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m} - \frac{1}{\mathrm{vol}(M)}\right\|_{L^2(M)} \ge c^*_{\varepsilon,\delta,B} \quad \text{for infinitely many } N \in \mathbb{N}.$$

(4) We use the spectral theorem, the eigenfunctions $(\phi_k)_{k=1}^{\infty}$ of the Laplace–Beltrami operator $-\Delta_g$ and inequalities related to the compactness of $e^{\delta\Delta} : L^1(M) \to H^s(M)$ to conclude that there exists a constant $N_0 \in \mathbb{N}$, depending only on B, ε, δ, such that for all N and all $x_1, \ldots, x_N \in M$,

$$\sum_{k\le N_0}\left|\left\langle e^{\delta\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m} - \frac{1}{\mathrm{vol}(M)}, \phi_k\right\rangle\right|^2 \ge \frac{1}{2}\left\| e^{\delta\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m} - \frac{1}{\mathrm{vol}(M)}\right\|^2_{L^2(M)}.$$

Combining these last two steps implies the existence of infinitely many $N \in \mathbb{N}$ such that

$$\sum_{k\le N_0}\left|\left\langle e^{\delta\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m} - \frac{1}{\mathrm{vol}(M)}, \phi_k\right\rangle\right|^2 \ge \frac{1}{2}(c^*_{\varepsilon,\delta,B})^2 > 0,$$

and finally we can use this to show that

$$\sum_{k\le N_0}\left|\left\langle e^{C\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m} - \frac{1}{\mathrm{vol}(M)}, \phi_k\right\rangle\right|^2 \ge c^{**}_{\varepsilon,\delta,B,N_0,C} > 0$$

for infinitely many $N \in \mathbb{N}$.

(5) We conclude by arguing that

$$\left\langle e^{2C\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m}, \frac{1}{N}\sum_{m=1}^{N}\delta_{x_m}\right\rangle \ge \frac{1}{\mathrm{vol}(M)} + \sum_{k\le N_0}\left|\left\langle e^{C\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m}, \phi_k\right\rangle\right|^2,$$

from which the result follows upon using

$$\left\langle e^{C\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m}-\frac{1}{\mathrm{vol}(M)},\phi_k\right\rangle=\left\langle e^{C\Delta}\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m},\phi_k\right\rangle\quad\text{for all } k\geq 1,$$

because these eigenfunctions are orthogonal to $\phi_0(x)=\mathrm{vol}(M)^{-1/2}$.

2.3. Proof of Theorem 2. Let (M,g) be given and consider the L^2-normalized Laplacian eigenfunctions

$$-\Delta_g\phi_n=\lambda_n\phi_n$$

as a basis of $L^2(M)$. Observe that $\lambda_0=0$, and $\phi_0=\mathrm{vol}(M)^{-1/2}$ is a constant function.

Step (1). We can rewrite the expression as

$$\frac{1}{N^2}\sum_{m,n=1}^{N}(e^{t_N\Delta}\delta_{x_m})(x_n)=\left\langle e^{t_N\Delta}\left(\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m}\right),\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m}\right\rangle.$$

This particular algebraic structure behaves well under the heat flow: for any $f\in C^\infty(M)$, we can write

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

and thus

$$\langle e^{t\Delta}f,f\rangle=\sum_{k=0}^{\infty}e^{-\lambda_k t}\langle f,\phi_k\rangle^2.$$

This quantity is obviously decreasing in t. Note that

$$\lim_{t\to\infty}\langle e^{t\Delta}f,f\rangle=\langle f,\phi_0\rangle^2=\left\langle f,\frac{1}{\mathrm{vol}(M)^{1/2}}\right\rangle^2=\frac{1}{\mathrm{vol}(M)}\Big(\int_M f\,dg\Big)^2,$$

which immediately implies, using a density argument, that for all $t>0$,

$$\left\langle e^{t\Delta}\left(\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m}\right),\frac{1}{N}\sum_{m=1}^{N}\delta_{x_m}\right\rangle\geq\frac{1}{\mathrm{vol}(M)}.$$

Step (2). Now assume that $(x_n)_{n=1}^{\infty}$ is not uniformly distributed but nonetheless

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{m,n=1}^{N}(e^{t_N\Delta}\delta_{x_m})(x_n)=\frac{1}{\mathrm{vol}(M)}.$$

The monotonicity of the expression under the heat flow implies that also

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{m,n=1}^{N}(e^{C\Delta}\delta_{x_m})(x_n)=\frac{1}{\mathrm{vol}(M)},$$

where C is the uniform upper bound on the sequence of times t_N. Not being uniformly distributed means there exists a geodesic ball $B \subset M$ and $\varepsilon_0 > 0$ such that

$$\left| \frac{\#\{1 \leq m \leq N : x_m \in B\}}{N} - \frac{|B|}{\mathrm{vol}(M)} \right| \geq \varepsilon_0 \quad \text{for infinitely many } N.$$

We shall assume that

$$\frac{\#\{1 \leq m \leq N : x_m \in B\}}{N} - \frac{|B|}{\mathrm{vol}(M)} \geq \varepsilon_0 \quad \text{for infinitely many } N,$$

because the other case implies the same estimate for another ball $B' \subseteq M \setminus B$ (possibly with a different value of ε_0). We may assume without loss of generality (by possibly making ε_0 smaller) that $|B| \leq 1/2$. Let B_δ denote the δ-neighborhood of B with $\delta > 0$ chosen so small that

$$\frac{|B_\delta|}{|B|} \leq 1 + \frac{\varepsilon_0}{100},$$

and let $t_0 > 0$ be chosen so small that

$$\inf_{z \in B} \int_{B_\delta} [e^{t_0 \Delta} \delta_z](x)\, dx \geq 1 - \frac{\varepsilon_0}{100}.$$

These two facts imply that for infinitely many N,

$$\int_{B_\delta} \left[e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right](x)\, dx \geq \left(1 - \frac{\varepsilon_0}{100}\right)(|B| + \varepsilon_0) \frac{1}{\mathrm{vol}(M)}.$$

This means that the average value satisfies

$$\begin{aligned} \frac{1}{|B_\delta|} \int_{B_\delta} \left[e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right](x)\, dx &\geq \left(1 - \frac{\varepsilon_0}{100}\right) \frac{|B| + \varepsilon_0}{|B_\delta|} \frac{1}{\mathrm{vol}(M)} \\ &\geq \left(1 - \frac{\varepsilon_0}{100}\right)\left(\frac{|B|}{|B_\delta|} + \frac{\varepsilon_0}{|B_\delta|}\right) \frac{1}{\mathrm{vol}(M)} \\ &\geq \left(1 - \frac{\varepsilon_0}{100}\right)\left(1 - \frac{\varepsilon_0}{100} + \varepsilon_0\right) \frac{1}{\mathrm{vol}(M)} \\ &\geq \left(1 + \frac{98\varepsilon_0}{100} - \frac{99\varepsilon_0^2}{10000}\right) \frac{1}{\mathrm{vol}(M)} > \frac{1}{\mathrm{vol}(M)}. \end{aligned}$$

Step (3). For general functions f, the Cauchy–Schwarz inequality implies

$$\begin{aligned} |B_\delta| \left| \frac{1}{|B_\delta|} \int_{B_\delta} f\, dx - \frac{1}{\mathrm{vol}(M)} \right| &= \left| \int_{B_\delta} \left(f - \frac{1}{\mathrm{vol}(M)} \right) dx \right| \\ &\leq \left(\int_{B_\delta} \left(f - \frac{1}{\mathrm{vol}(M)} \right)^2 dx \right)^{1/2} |B_\delta|^{1/2}, \end{aligned}$$

and therefore there exists a constant $\varepsilon_1 > 0$ (depending only on ε_0 and $|B_\delta|$) such that for infinitely many $N \in \mathbb{N}$,

$$\left\| \left[e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right](x) - \frac{1}{\mathrm{vol}(M)} \right\|_{L^2(M)} \geq \varepsilon_1.$$

Step (4). We will now prove that for any $t_0 > 0$, any $N \in \mathbb{N}$ and any set of points $x_0, x_1, \ldots, x_N$,

$$\left\| \nabla e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\|_{L^2(M)} \lesssim_{t_0,(M,g)} 1,$$

where $\lesssim_{t_0,(M,g)}$ denotes the existence of an implicit constant depending only on t_0 and (M, g) for some implicit constant that is both independent of N and of the actual set of points. It is easy to see that

$$\begin{aligned} \left\| \nabla e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\|_{L^\infty(M)} &= \left\| \frac{1}{N} \sum_{m=1}^{N} \nabla e^{t_0 \Delta} \delta_{x_m} \right\|_{L^\infty(M)} \\ &\leq \frac{1}{N} \sum_{m=1}^{N} \| \nabla e^{t_0 \Delta} \delta_{x_m} \|_{L^\infty(M)} \leq \sup_{x \in M} \| \nabla e^{t_0 \Delta} \delta_x \|_{L^\infty(M)} \lesssim_{(M,g),t_0} 1, \end{aligned}$$

which follows from the regularity of Green's function. By the same token

$$\begin{aligned} 1 = \left\| e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\|_{L^1(M)} &\leq \mathrm{vol}(M)^{1/2} \left\| e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\|_{L^2(M)} \\ &\lesssim_{(M,g)} \left\| \frac{1}{N} \sum_{m=1}^{N} e^{t_0 \Delta} \delta_{x_m} \right\|_{L^2(M)} \leq \sup_{x \in M} \| e^{t_0 \Delta} \delta_x \|_{L^2(M)} \lesssim_{(M,g),t_0} 1. \end{aligned}$$

We can now use the spectral theorem to write

$$1 \gtrsim_{(M,g),t_0} \left\| \nabla e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\|_{L^2(M)}^2 = \sum_{k=0}^{\infty} \lambda_k \left| \left\langle e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2.$$

However, at the same time, the eigenvalues of the Laplace–Beltrami operator are increasing and unbounded:

$$0 < \lambda_1 < \lambda_2 \leq \cdots \to \infty.$$

Weyl's law would give the asymptotic growth, but that is not necessary. Recall that for infinitely many $N \in \mathbb{N}$ we have

$$\left\| \left[e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right](x) - \frac{1}{\mathrm{vol}(M)} \right\|_{L^2(M)} \geq \varepsilon_1.$$

We can now argue that for $N_1 \geq 1$,

$$\begin{aligned}
\varepsilon_1^2 &\leq \left\| \left[e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right](x) - \frac{1}{\mathrm{vol}(M)} \right\|_{L^2(M)}^2 \\
&= \sum_{0 < \lambda_k \leq N_1} \left| \left\langle e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 + \sum_{\lambda_k > N_1} \left| \left\langle e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 \\
&\leq \sum_{0 < \lambda_k \leq N_1} \left| \left\langle e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 + \frac{1}{N_1} \sum_{\lambda_k > N_1} \lambda_k \left| \left\langle e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 \\
&\leq \sum_{0 < \lambda_k \leq N_1} \left| \left\langle e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 + \frac{1}{N_1} \left\| \nabla e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\|_{L^2(M)}^2.
\end{aligned}$$

Step (4) implies that this final gradient term is uniformly bounded for all N and all $x_1, \ldots, x_N \in \mathbb{T}$. This means that there exist $N_1 \in \mathbb{N}$ and $\varepsilon_2 > 0$ depending only on $(M, g), t_0, \varepsilon_1$ such that, for infinitely many N,

$$\sum_{0 < \lambda_k \leq N_1} \left| \left\langle e^{t_0 \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 \geq \varepsilon_2.$$

Step (5). Using representation in Fourier series, we easily deduce that for all $s > t_0$,

$$\sum_{0 < \lambda_k \leq N_1} \left| \left\langle e^{s \Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 \geq e^{-N_1^2 (s - t_0)} \varepsilon_2.$$

We conclude by arguing that, for infinitely many $N \in \mathbb{N}$,

$$\begin{aligned}
&\left\langle e^{2C\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\rangle = \left\langle e^{C\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, e^{C\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\rangle \\
&\quad = \sum_{k=0}^{\infty} \left| \left\langle e^{C\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 = \frac{1}{\mathrm{vol}(M)} + \sum_{k=1}^{\infty} \left| \left\langle e^{C\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 \\
&\quad \geq \frac{1}{\mathrm{vol}(M)} + \sum_{k \leq N_0} \left| \left\langle e^{C\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \phi_k \right\rangle \right|^2 \geq \frac{1}{\mathrm{vol}(M)} + e^{-N_1^2 (C - t_0)} \varepsilon_2. \quad \blacksquare
\end{aligned}$$

2.4. Proof of Corollary 4. This is a fairly straightforward diagonal argument. It is worthwhile to first study the behavior of various objects at the critical scale $t_N = N^{-2}$ corresponding to physical scale N^{-1}. We will only work with sequences $(t_n)_{n=1}^{\infty}$ of time scales that tend to 0 and will simplify exposition by using the form of the criterion employing the exponential function, which has a nice dilation symmetry. For any $\varepsilon > 0$, there are constants K and $(a_k)_{k=1}^K, (b_k)_{k=1}^K$ such that for all $N \in \mathbb{N}$,

$$\Big\| \exp(-N^2 y^2) - \sum_{k=1}^{K} a_k \chi_{|y| \le b_k/N} \Big\|_{L^\infty(\mathbb{R})} \le \varepsilon,$$

where χ denotes a characteristic function. This can be accomplished by the usual step function approximation. This naturally implies that

$$\left| \frac{1}{N^2} \sum_{m,n=1}^{N} N \exp(-N^2(x_m - x_n)^2) - \frac{1}{N} \sum_{m,n=1}^{N} \sum_{k=1}^{K} a_k \chi_{|x_m - x_n| \le b_k/N} \right| \le \frac{\varepsilon}{N}.$$

Assuming Poissonian pair correlation, we get

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \sum_{m,n=1}^{N} \sum_{k=1}^{K} a_k \chi_{|x_m - x_n| \le b_k/N} \\ &\quad = \sum_{k=1}^{K} a_k \frac{1}{N} \# \left\{ 1 \le m, n \le N : |x_m - x_n| \le \frac{b_k}{N} \right\} \to \sum_{k=1}^{K} a_k (1 + 2b_k). \end{aligned}$$

At the same time

$$\sum_{k=1}^{K} 2a_k b_k = \int_{\mathbb{R}} \sum_{k=1}^{K} a_k \chi_{|y| \le b_k} \, dy \sim \int_{\mathbb{R}} \exp(-y^2) \, dy = \sqrt{\pi}.$$

We see that the problem comes from the fact that our criterion sums over all pairs (x_m, x_n), including those for which $m = n$, while pair correlation only counts pairs (x_m, x_n) with $m \ne n$. At the same time, these diagonal terms have a nontrivial contribution in our theorem since

$$\frac{1}{N^2} \sum_{m=1}^{N} N \exp(-N^2 (x_m - x_m)^2) = 1.$$

However, it is easily seen that for the diagonal terms to contribute substantially,

$$\frac{1}{N^2} \sum_{m=1}^{N} \frac{1}{\sqrt{t_N}} \exp\left(-\frac{1}{t_N} (x_m - x_m)^2 \right) = \frac{1}{N^2 \sqrt{t_N}},$$

we do indeed require that $t_N \sim N^{-2}$. If it were to decay slightly slower, say $t_N = f(N) N^{-2}$ for an increasing unbounded sequence $f(N)$, then the diagonal terms disappear in the limit. We now simply define a sequence $(s_n)_{n=1}^{\infty}$ of integers by requiring them to have the property that for all $N \ge s_k$,

$$\left| \frac{1}{N} \# \left\{ 1 \le m \ne n \le N : |x_m - x_n| \le \frac{s}{N} \right\} - 2s \right| \le \frac{1}{2^k}$$

for all $s \in \{1, \dots, 2^k\}$. Clearly, the sequence $(s_n)_{n=1}^{\infty}$ is increasing and unbounded. We can then pick the times to be

$$t_N = s_\ell / N^2 \quad \text{where} \quad \ell = \max\{k \in \mathbb{N} : s_k \le N\}$$

and apply the criterion.

Finally, we show that weak pair correlation implies uniform distribution. We observe that because of $0 < \alpha < 1$ (and thus $N \ll N^{2-\alpha}$) we may include identical pairs (x_m, x_m) since there are only N of those and obtain

$$\lim_{N\to\infty} \frac{1}{N^{2-\alpha}} \# \left\{ 1 \le m, n \le N : |x_m - x_n| \le \frac{s}{N^\alpha} \right\} = 2s.$$

The argument is now completely straightforward: If $\alpha > 0$, we set the sequence of time scales to be

$$t_N = \frac{1}{N^{2\alpha}}$$

and wish to estimate

$$\frac{1}{N^2} \sum_{m,n=1}^{N} N^\alpha \exp(-N^{2\alpha}(x_m - x_n)^2).$$

The condition of weak correlation then implies

$$\frac{1}{N^2} \sum_{m,n=1}^{N} N^\alpha \exp(-N^{2\alpha}(x_m - x_n)^2) \to \int_0^\infty 2\exp(-s^2)\, ds = \sqrt{\pi},$$

and we obtain the result. The missing steps (approximation of the Gaussian by a step function, replacing the count of variables by the weak pair correlation condition) are standard and left to the reader.

3. Proof of Corollary 5

Lemma. *Let $t > 0$ and consider $\theta_t : \mathbb{T} \to \mathbb{R}_+$ given by*

$$\theta_t(x) = 1 + 2\sum_{n=1}^{\infty} e^{-4\pi^2 n^2 t} \cos(2\pi n x).$$

If $\varepsilon > 0$ and $x \ge 2\sqrt{\log(2/\varepsilon)}\sqrt{t}$, then

$$\int_{-x}^{x} \theta_t(y)\, dy \ge 1 - \varepsilon.$$

Proof. A simple topological argument allows us to compare the heat kernel on the torus with the heat kernel on the real line: on the torus we have the possibility of looping around, which we do not have on the real line. Therefore, for all $x \in \mathbb{R}$ and all $t > 0$,

$$1 + 2\sum_{n=1}^{\infty} e^{-4\pi^2 n^2 t}\cos(2\pi n x) \geq \frac{1}{\sqrt{4\pi t}} e^{-x^2/(4t)}.$$

The result then follows from the Chernoff bound

$$\int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\,dy \leq e^{-x^2/2}. \ \blacksquare$$

Proof of Corollary 5. Let $x_1, \ldots, x_N \in \mathbb{T}$ and assume that

$$D_N(\{x_1, \ldots, x_N\}) = \varepsilon.$$

Then there exists an interval $J \subset \mathbb{T}$ such that

$$\left| \frac{\#\{1 \leq m \leq N : x_m \in J\}}{N} - |J| \right| = \varepsilon.$$

We distinguish two cases:

$$\frac{\#\{1 \leq m \leq N : x_m \in J\}}{N} = |J| + \varepsilon \text{ or } |J| - \varepsilon.$$

We treat the first case; the second is essentially identical. We set

$$t = \frac{1}{100} \frac{\varepsilon^2}{\log(20/\varepsilon)}.$$

This choice guarantees, using the Lemma above, that

$$\int_{-\varepsilon/4}^{\varepsilon/4} \theta_t(y)\,dy \geq 1 - \frac{\varepsilon}{10}.$$

We will now consider the slightly larger interval J^* given as the $\varepsilon/4$-neighborhood of J. We see that, for infinitely many $N \in \mathbb{N}$,

$$\begin{aligned}
\frac{1}{|J^*|}\left\| \left[e^{(t/2)\Delta} \frac{1}{N}\sum_{m=1}^{N} \delta_{x_m} \right](x) \right\|_{L^1(J^*)} &\geq \left(1 - \frac{\varepsilon}{10}\right) \frac{1}{|J^*|} \left\| \frac{1}{N}\sum_{m=1}^{N} \delta_{x_m} \right\|_{L^1(J)} \\
&\geq \left(1 - \frac{\varepsilon}{10}\right) \frac{1}{|J| + \varepsilon/2} \left\| \frac{1}{N}\sum_{m=1}^{N} \delta_{x_m} \right\|_{L^1(J)} \geq \left(1 - \frac{\varepsilon}{10}\right) \frac{|J| + \varepsilon}{|J| + \varepsilon/2} \\
&\geq \left(1 - \frac{\varepsilon}{10}\right) \frac{1 + \varepsilon}{1 + \varepsilon/2} \geq 1 + \frac{\varepsilon}{10}.
\end{aligned}$$

We use the Cauchy–Schwarz inequality in the form

$$\left(\int_{J^*} (f(x) - 1)\,dx \right)^2 \leq \left(\int_{J^*} (f(x) - 1)^2\,dx \right) |J^*| \leq \left(\int_{\mathbb{T}} (f(x) - 1)^2\,dx \right)$$

to conclude that

$$\left\| \left[e^{(t/2)\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right](x) - 1 \right\|_{L^2(\mathbb{T})}^2 \geq \frac{\varepsilon^2}{100}.$$

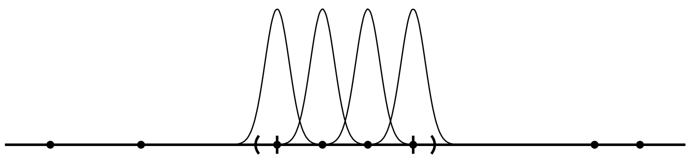

Fig. 3. Slightly too many points in an interval (bounded by straight lines) implies slightly too much L^1-mass of the heat kernel in a slightly larger interval (bounded by curved lines).

An explicit computation shows that

$$\begin{aligned}\left\| \left[e^{(t/2)\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right](x) - 1 \right\|_{L^2(\mathbb{T})}^2 &= \int_{\mathbb{T}} \left[e^{(t/2)\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right](x)^2 \, dx - 1 \\ &= \left\langle e^{(t/2)\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, e^{(t/2)\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\rangle - 1 \\ &= \left\langle e^{t\Delta} \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m}, \frac{1}{N} \sum_{m=1}^{N} \delta_{x_m} \right\rangle - 1 = -1 + \frac{1}{N^2} \sum_{m,n=1}^{N} \theta_t(x_n - x_m),\end{aligned}$$

and we can conclude the result.

The other case, of not enough points, follows analogously except that J^* is obtained from shrinking J. ■

Acknowledgments. The author is grateful to Christoph Aistleitner for several discussions.

References

[1] C. Aistleitner, T. Lachmann, and F. Pausinger, *Pair correlations and equidistribution*, arXiv:1612.05495 (2016).
[2] L. Bétermin, *Local optimality of cubic lattices for interaction energies*, arXiv:1611.07798 (2016).
[3] X. Blanc and M. Lewin, *The crystallization conjecture: a review*, EMS Surv. Math. Sci. 2 (2015), 255–306.
[4] S. Grepstad and G. Larcher, *On pair correlation and discrepancy*, arXiv:1612.08008 (2016).
[5] D. R. Heath-Brown, *Pair correlation for fractional parts of αn^2*, Math. Proc. Cambridge Philos. Soc. 148 (2010), 385–407.
[6] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Pure Appl. Math., Wiley-Interscience, New York, 1974.
[7] H. L. Montgomery, *Minimal theta functions*, Glasgow Math. J. 30 (1988), 75–85.

[8] Z. Rudnick and P. Sarnak, *The pair correlation function of fractional parts of polynomials*, Comm. Math. Phys. 194 (1998), 61–70.
[9] Z. Rudnick and A. Zaharescu, *A metric result on the pair correlation of fractional parts of sequences*, Acta Arith. 89 (1999), 283–293.
[10] Z. Rudnick and A. Zaharescu, *The distribution of spacings between fractional parts of lacunary sequences*, Forum Math. 14 (2002), 691–712.
[11] W. M. Schmidt, *Irregularities of distribution, VII*, Acta Arith. 21 (1972), 45–50.
[12] W. M. Schmidt, *Irregularities of distribution, IX*, Acta Arith. 27 (1975), 385–396.
[13] S. K. Zaremba, *La discrépance isotrope et l'intégration numérique*, Ann. Mat. Pura Appl. 87 (1970), 125–135.

Stefan Steinerberger
Department of Mathematics
Yale University
New Haven, CT 06511, U.S.A.
E-mail: stefan.steinerberger@yale.edu