Report on Presentation: Equidistribution and the Gauss Circle Problem

Krits Costa
Jon Ruiz Martínez
ETH Zurich

1 Introduction

This paper aims to give an introduction and few findings to the Gauss circle problem, named after the German mathematician Johann Carl Friedrich Gauss for being the first one to make significant progress on a solution (1834). The problem is the following:
Given a circle of radius $R \geq 0$ in $\mathbb{R}^2$ and centered at the origin, the goal is to find the number of lattice points defined by $N(R)$ inside of it, in other words, try to find $\overline{B_0(R)} \cap \mathbb{Z}^2$. Here are the first values of $N(R)$ for R an integer between 0 and 12.

R	$N(R)$
0	1
1	5
2	13
3	29
4	49
5	81
6	113
7	149
8	197
9	253
10	317
11	377
12	441

First results showed that $N(R)$ was approximately equal to the area of the circle πR^2, which can indeed be observed with the table above. There is however an error term depending on the radius of the circle and defined by $E(R)$, which is actually the very core of this problem and for which Gauss found a first upper bound: $|E(R)| \leq 2\sqrt{2}\pi R$. There has been since then a lot of new and more accurate approximations for this bound, the error being initially in the form of $O(R^\theta)$, mathematicians have been trying to minimise θ (Gauss showed $E(R) = O(R)$), but the English Godfrey Harold Hardy found in 1915 a lower bound for it being 1/2. As for the upper bound, many found smaller and smaller ones throughout the XXth century, the period ending with british Martin Neil Huxley finding the best known bound until then, $131/208 \approx 0.6298$ (2000). Although it is conjectured that the correct error is $|E(R)| = O(R^{1/2+\epsilon})$ for any $\epsilon > 0$, the best improvement on the upper bound we have today is still "far off" this result and attributed to belgian Jean Bourgain and English Nigel Watt (whose Ph.D. advisor was Huxley!), they found in 2017 that $\theta = 517/824 + \epsilon$ for any $\epsilon > 0$. It is interesting to note that this improvement is very recent and the topic is still well studied today, all we know is $1/2 < \theta \leq 517/824 \approx 0.6274$, so there will probably still be quite some work to do before being able to prove the conjecture.

Our goal throughout this paper will be to first prove that $E(R) = O(R)$ and then $E(R) = o(R)$ (implying $N(R) = \pi R^2 + o(R)$). In order to show the latter, we will need a proposition that we will prove as well.

2 Background

To understand the following, it is advised to have basic knowledge of analysis (be familiar with the Landau notation, limits, definition of the Riemann integral, parametrized curves, continuity, density, taylor expansions), linear algebra (basis, orthogonality) and measure theory (measure space, null set).
We will also need a few objects from differential geometry but their exact meaning is not relevant for this paper, we define them here:

$$\mathrm{T}^1\mathbb{R}^2 = \{(x, v) : x \in \mathbb{R}^2, v \in \mathbb{R}^2, ||v|| = 1\} = \mathbb{R}^2 \times \mathbb{S}^1$$

is the unit tangent bundle of $\mathbb{R}^2$ and

$$\mathrm{T}^1\mathbb{T}^2 = \{(x, v) : x \in \mathbb{T}^2, v \in \mathbb{R}^2, ||v|| = 1\} = \mathbb{T}^2 \times \mathbb{S}^1$$

is the unit tangent bundle of $\mathbb{T}^2$.
Here $\mathbb{S}^1$ is the unit circle in the plane and $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ is the torus in the space.

3 Main Results

3.1 Proposition 5.1

For any $R > 0$ let

$$N(R) = |\{n \in \mathbb{Z}^2 : ||n|| \leq R\}|$$

then

$$N(R) = \pi R^2 + O(R).$$

In this section we will show that the error term as defined in the introduction is $O(R)$, i.e. the error term grows at most as fast as R when R goes to infinity.

PROOF: To prove this first result, we need to define the unit square S beforehand. Let us consider $S = [-\frac{1}{2}, \frac{1}{2}) \times [-\frac{1}{2}, \frac{1}{2})$, then $S + \{n \in \mathbb{Z}^2 : ||n|| \leq R\} := G$ represents all the unit squares whose center is a lattice point inside a circle of radius R (see Figure 1: the blue square is S shifted by (-2,-1) $\in \{n \in \mathbb{Z}^2 : ||n|| \leq R\}$, assuming R big enough for the purpose of the representation in the figure) , hence there are as many squares as there are lattice points inside the circle and the sum of areas of all these squares (each has area 1) is exactly equal to the sum of lattice points that we are interested in ($N(R)$).
The trick here is to consider two other circles: one smaller of radius $R - 1/\sqrt{2}$ and one bigger of radius $R + 1/\sqrt{2}$, it is clear that any point inside the smaller circle is contained in G as by

diminishing the radius by half the diagonal of a square, we make sure that any point in G that is not in $B_R(0)$ now is in $B_{R-1/\sqrt{2}}(0)$. For analogous reasons, any point in G is contained inside the bigger circle (see Figure 1). Therefore

$$B_{R-1/\sqrt{2}}(0) \subseteq G \subseteq B_{R+1/\sqrt{2}}(0)$$

and by taking areas (as explained above the area of the grid G is exactly $N(R)$)

$$(R - \frac{1}{\sqrt{2}})^2\pi \leq N(R) \leq (R + \frac{1}{\sqrt{2}})^2\pi$$

which allows us to conclude $N(R) = \pi R^2 + O(R)$ as we wanted. $\square$

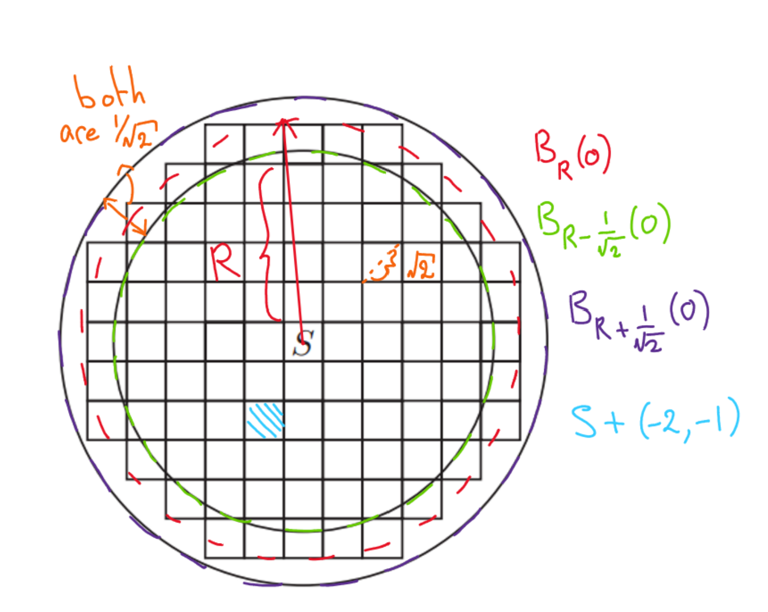

Figure 1: The grid G enclosed by $B_{R-1/\sqrt{2}}(0)$ and $B_{R+1/\sqrt{2}}(0)$

3.2 Proposition 5.2

Let

$$\begin{aligned} \gamma_R : [0,1] &\longrightarrow \mathrm{T}^1\mathbb{R}^2 \\ t &\longmapsto \left(Re^{2\pi it}, e^{2\pi it}\right) \end{aligned}$$

be the constant speed parametrization of the outward tangent vectors on the circle of radius R. Then

$$\int_0^1 f\left(\gamma_R(t)\right)\mathrm{d}t \to \int_{\mathrm{T}^1\mathbb{T}^2} f(x,v)\mathrm{d}(x,v)$$

as $R \to \infty$*, for every* $f \in C(\mathrm{T}^1\mathbb{T}^2)$.

In this section we will prove an integral transformation result that we will need to show the most important result we will discuss in this paper, namely *Theorem 5.3* in the next section.

PROOF: The first thing to notice here is that γ_R is constituted of two circles, one of radius R in the first component and a unit circle in the second component. As t goes from 0 to 1, $\gamma_R(t) = (Re^{2\pi it}, e^{2\pi it}) := (x_R(t), v(t))$ passes through all directions with constant speed so that the only real thing we have to show is that the positional part $x_R(t)$ restricted to any subinterval $[\alpha, \beta] \subseteq [0, 1]$ equidistributes on $\mathbb{T}^2$ as R goes to infinity.
Recall that equidistribution here means that the distribution of $x_R(t)$ is homogeneous in the sense that the quantity of points in any subinterval (or in general measurable subset with respect to Jordan measure[1]) of the considered interval (measurable set) is proportional to its length (measure). Typically, in a discrete setting (for a sequence), a good way to check for equidistribution is to verify the given sequence can be applied on a sum as a sample to calculate the integral of any Riemann-integrable function. Concretely, if $(x_n)_{n\in\mathbb{N}}$ is the aforementioned sequence with $x_n \in X$, where X is a measurable set w.r.t. the Jordan measure, we want to verify:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_X f(x)\mathrm{d}(x)$$

for every Riemann-integrable function $f : X \to \mathbb{C}$. However, in a continuous setting such as here with $x_R(t)$, the analogue is to check the above equality with an integral on the left hand side instead of a sum and for any continuous f. Concretely, what we want to show is

$$\frac{1}{\beta - \alpha} \int_\alpha^\beta f(x_R(t))\mathrm{d}t \to \int_{\mathbb{T}^2} f(x)\mathrm{d}(x) \qquad (*)$$

as $R \to \infty$ $\forall f \in C(\mathbb{T}^2)$.

Indeed, if we manage to show the above, we could split the interval [0,1] into subintervals

$$[0, \frac{1}{n}] \cup [\frac{1}{n}, \frac{2}{n}] \cup \ldots \cup [\frac{n-1}{n}, 1]$$

and use the continuity of $f \in C(\mathrm{T}^1\mathbb{T}^2)$ together with the equidistribution of $x_R(t)$ to conclude that

$$\begin{aligned}
\lim_{R\to\infty} \textstyle\int_0^1 f\left(\gamma_R(t)\right)\mathrm{d}t &= \lim_{R\to\infty} \sum_{j=0}^{n-1} \int_{j/n}^{(j+1)/n} f(x_R(t), e^{2\pi it})\mathrm{d}t \\
&\overset{(1)}{=} \lim_{n,R\to\infty} \sum_{j=0}^{n-1} \int_{j/n}^{(j+1)/n} f(x_R(t), e^{2\pi ij/n})\mathrm{d}t \\
&= \lim_{n,R\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \frac{1}{(1/n)} \int_{j/n}^{(j+1)/n} f(x_R(t), e^{2\pi ij/n})\mathrm{d}t \\
&\overset{(2)}{=} \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \int_{\mathbb{T}^2} f(x, e^{2\pi ij/n})\mathrm{d}(x) \\
&= \lim_{n\to\infty} \int_{\mathbb{T}^2} \frac{1}{n} \sum_{j=0}^{n-1} f(x, e^{2\pi ij/n})\mathrm{d}(x) \\
&\overset{(3)}{=} \int_{\mathrm{T}^1\mathbb{T}^2} f(x, v)\mathrm{d}(x, v)
\end{aligned}$$

[1]According to wikipedia (Peano-Jordan measure): "In mathematics, the Peano-Jordan measure, also known as Jordan content, is an extension of the notion of size (lenght, area, volume) to shapes more complicated than, for example, a triangle, disk, or parallelepiped."

where (1) comes from the uniform continuity (continuous on compact interval) of the function $g : [0,1] \to \mathrm{T}^1\mathbb{R}^2$ defined as $g(s) := f(x_R(t), e^{2\pi is})$ $\forall t \in [0,1]$. Indeed, the uniform continuity implies $\forall \epsilon < 0$ $\exists \delta < 0$ such that $\forall s \in [j/n, (j+1)/n]$

$$|s - j/n| < \delta \implies |g(s) - g(j/n)| < \epsilon.$$

As $n \to \infty$, $|s-j/n| \leq |(j+1)/n - j/n| = 1/n \to 0$ and we are just left with $|g(s)-g(j/n)| < \epsilon$ for any $\epsilon < 0$, which means

$$g(s) = g(j/n) \iff f(x_R(t), e^{2\pi is}) = f(x_R(t), e^{2\pi ij/n}).$$

(2) simply comes from the equidistribution argument we claimed earlier with $[\alpha, \beta] = [j/n, (j+1)/n]$ and (3) comes from the definition of the Riemann integral as $e^{2\pi ij/n}$ clearly is equidistributed over $\mathbb{S}^1$, so

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(x, e^{2\pi ij/n}) = \int_{\mathbb{S}^1} f(x,v)\mathrm{d}(v)$$

and together with the integral over $\mathbb{T}^2$, we obtain an integral over $\mathbb{T}^2 \times \mathbb{S}^1$, which is exactly $\mathrm{T}^1\mathbb{T}^2$.

Thus, it just remains to prove the claim (*) of the equidistribution of $x_R(t)$. To prove this, notice that the set of functions

$$e_n(x) = e^{2\pi i(n_1x_1+n_2x_2)}$$

is dense in $C(\mathbb{T}^2)$ (recall Fourier series in $\mathbb{R}/\mathbb{Z}$), meaning any $f \in C(\mathbb{T}^2)$ can be expressed as a linear combination of e_n's. Hence, thanks to linearity of the integral, we just have to show equidistribution for functions e_n.
This trivially holds for $n = (0,0)$ since $e_0(x) = 1$ so the left hand side is 1 and the right hand side as well (for example write $\mathbb{T}^2$ as $\mathbb{S}^1 \times \mathbb{S}^1$ and easily find 1 for the decomposed integral). Let us now focus on the case where $n \neq (0,0)$ and fix n as such.

First of all, without loss of generality, we may assume that n is never orthogonal to

$$\begin{bmatrix} -sin(2\pi t) \\ cos(2\pi t) \end{bmatrix}$$

for $t \in [\alpha, \beta]$. This assertion will reveal its purpose a bit later in the proof. To see it, let $(t_m)_{m=0}^M \subseteq [\alpha, \beta]$ be the set of points where the assumption fails and assume, again w.l.o.g., that t_m is increasing. This set is finite since there cannot be infinitely many points in $[\alpha, \beta]$ for which the above vector is orthogonal to n. We can then we split the interval into finitely many subintervals

$$[\alpha, \beta] = [\alpha, t_0) \cup (t_0, t_1) \cup \ldots \cup (t_M, \beta] \cup \bigcup_{m=0}^{M} \{t_n\}$$

and notice that the last set of points is a null set and has therefore no impact on the process so that we may proceed with applying the non-orthogonality argument on each subinterval written above (if the equidistribution in (*) holds for any such subinterval, it also does for the entire interval). For that very reason, let's assume

$$|n \cdot \begin{bmatrix} -sin(2\pi t) \\ cos(2\pi t) \end{bmatrix}| = |-n_1 sin(2\pi t) + n_2 cos(2\pi t)| \geq \kappa$$

for some $\kappa > 0$ and all $t \in [\alpha, \beta]$.

Secondly, we split the interval $[\alpha, \beta]$ into $m = \lfloor R^{3/4} \rfloor$ subintervals

$$[\alpha_1, \alpha_2] \cup [\alpha_2, \alpha_3] \cup \ldots \cup [\alpha_m, \alpha_{m+1}]$$

with $\alpha_j = \alpha + (j-1)\frac{\beta-\alpha}{m}$. By the Taylor expansion (around α_j) we have

$$\begin{aligned} x_R(t) = R \begin{bmatrix} cos(2\pi t) \\ sin(2\pi t) \end{bmatrix} &= R \left(\begin{bmatrix} cos(2\pi\alpha_j) \\ sin(2\pi\alpha_j) \end{bmatrix} + (t-\alpha_j)2\pi \begin{bmatrix} -sin(2\pi\alpha_j) \\ cos(2\pi\alpha_j) \end{bmatrix} + O\left(\frac{1}{m^2}\right) \right) \\ &= R \begin{bmatrix} cos(2\pi\alpha_j) \\ sin(2\pi\alpha_j) \end{bmatrix} + R(t-\alpha_j)2\pi \begin{bmatrix} -sin(2\pi\alpha_j) \\ cos(2\pi\alpha_j) \end{bmatrix} + O\left(\frac{1}{R^{1/2}}\right) \end{aligned}$$

for all $t \in [\alpha_j, \alpha_{j+1}]$ and $j = 1, ..., m$. As $n \neq 0$ is fixed and $R \to \infty$, this gives

$$\begin{aligned} \left| \frac{1}{\beta-\alpha} \int_\alpha^\beta e_n(x_R(t))\mathrm{d}t \right| &= \left| \frac{1}{\beta-\alpha} \sum_{j=1}^m \int_{\alpha_j}^{\alpha_{j+1}} e_n \left(R \begin{bmatrix} cos(2\pi\alpha_j) \\ sin(2\pi\alpha_j) \end{bmatrix} + R(t-\alpha_j)2\pi \begin{bmatrix} -sin(2\pi\alpha_j) \\ cos(2\pi\alpha_j) \end{bmatrix} \right) \right| + o(1) \\ &\leq \frac{1}{\beta-\alpha} \sum_{j=1}^m \left| e^{2\pi i R[n_1(cos(2\pi\alpha_j)+2\pi\alpha_j sin(2\pi\alpha_j))+n_2(sin(2\pi\alpha_j)-2\pi\alpha_j cos(2\pi\alpha_j))]} \right| \\ &\left| \int_{\alpha_j}^{\alpha_{j+1}} e^{4\pi^2 iRt(-n_1 sin(2\pi\alpha_j)+n_2 cos(2\pi\alpha_j)} \mathrm{d}t \right| + o(1) \\ &= \frac{1}{\beta-\alpha} \sum_{j=1}^m \left| \int_{\alpha_j}^{\alpha_{j+1}} e^{4\pi^2 iRt(-n_1 sin(2\pi\alpha_j)+n_2 cos(2\pi\alpha_j))} \mathrm{d}t \right| + o(1) \\ &\leq \frac{1}{\beta-\alpha} \sum_{j=1}^m \frac{2}{4\pi^2 R|-n_1 sin(2\pi\alpha_j)+n_2 cos(2\pi\alpha_j)|} + o(1) \end{aligned}$$

Here, we finally see the purpose of our assumption on the non-orthogonality because we want to be able to bound the denominator in the last inequality, and we can as we assumed that $|-n_1 sin(2\pi\alpha_j) + n_2 cos(2\pi\alpha_j)| \geq \kappa$. Thus we get

$$\lim_{R\to\infty} \left| \frac{1}{\beta-\alpha} \int_\alpha^\beta e_n(x_R(t))\mathrm{d}t \right| \leq \lim_{R\to\infty} \frac{1}{\beta-\alpha} \sum_{j=1}^m \frac{2}{4\pi^2 R\kappa} = \lim_{R\to\infty} \frac{1}{\beta-\alpha} \frac{2m}{4\pi^2 R\kappa} = \lim_{R\to\infty} \frac{1}{\beta-\alpha} \frac{2}{4\pi^2 R^{1/4}\kappa} = 0$$

Finally, as we wanted, we obtain $\frac{1}{\beta-\alpha} \int_\alpha^\beta e_n(x_R(t))\mathrm{d}t \to 0$ as $R \to \infty$, which allows us to conclude that (*) is true for any $f \in C(\mathbb{T}^2)$ by linear combinations of functions e_n and ends the proof of the proposition. □

3.3 Theorem 5.3

$$N(R) = \pi R^2 + \mathrm{o}(R)$$

In this section we will be improving the error term. As we have seen before, Gauss proved that we can write the error term as $\mathcal{O}(R)$. We will now prove that this error term can be written as $o(R)$, i.e. the error term grows strictly slower than R as R goes to infinity.

PROOF: In order to get this result, we will have to find a function defined on a tangent

bundle with the property that the integral along a line segment relates with the difference between the area calculation and the lattice point count. We will then be able to use the equidistribution result we proved in *Proposition 5.2* .
We will now define functions in the tangent bundle of S

$$\mathrm{T}^1 S = \left[-\frac{1}{2}, \frac{1}{2}\right) \times \left[-\frac{1}{2}, \frac{1}{2}\right) \times \mathbb{S}^1$$

recall that $S = \left[-\frac{1}{2}, \frac{1}{2}\right) \times \left[-\frac{1}{2}, \frac{1}{2}\right)$ Let's define a function *h(x,v)* as the area of the polygon determined by S and the half-space with x in its boundary and v as an outward normal. If 0 is also in the polygon we subtract 1. We also characterise the line denote that goes through x and is normal to v as *L(x,v)*

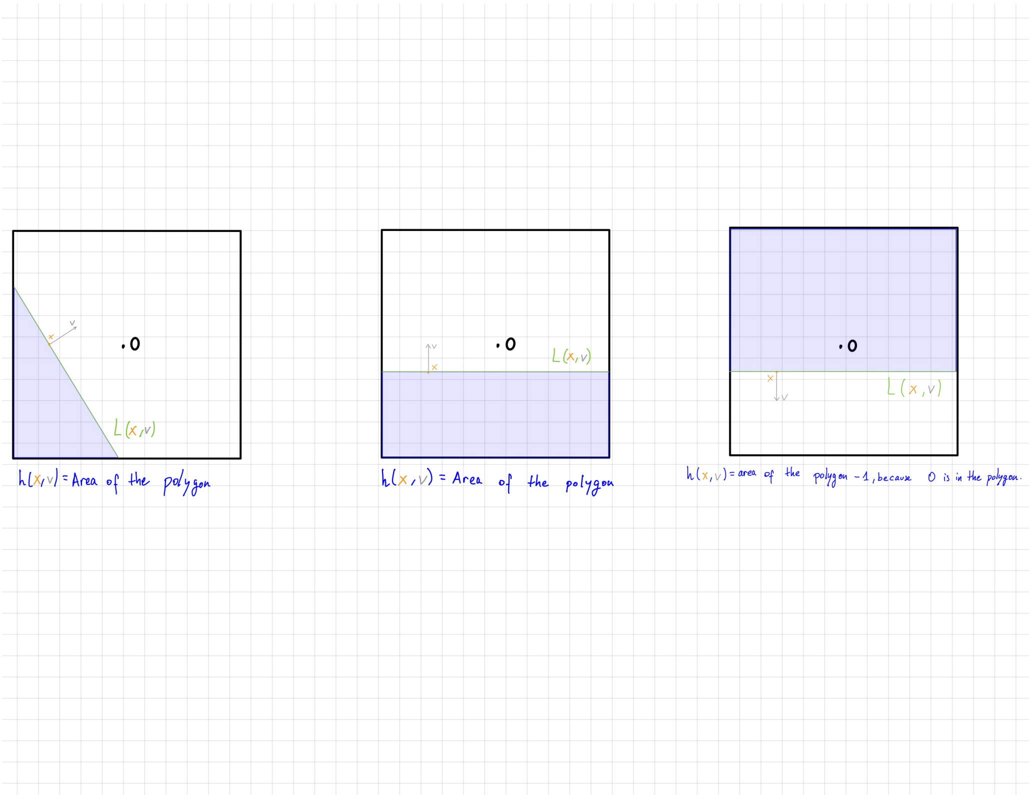

Figure 2: Description of the function $h(x, v)$ and $L(x, v)$

Now we'll define a function $f : S \times \mathbb{S}^1 \to \mathbb{R}$ as $f(x, v) = \frac{h(x,v)}{\text{length of } L(x,v)}$.
Claim: f is Riemann integrable and $\int_{\mathrm{T}^1\mathbb{T}^2} f \, \mathrm{d}(x, v) = 0$
Proof of the claim:

1. f is bounded:
   h is bounded by $\frac{1}{2}$. If the area of the polygon is bigger than $\frac{1}{2}$, then 0 must also be contained in the polygon and thus one is subtracted from the area. So if 0 is contained in the polygon h is smaller than zero and if 0 isn't contained in the polygon then is h bounded by $\frac{1}{2}$.
   The length of L is small if and only if it is close to a corner of S. In that case the polygon is a triangle and as we can see in Figure 2 we can denote two of its sides as a and b. The length of L can then be written as length of $L(x, v) = \sqrt{a^2 + b^2}$ and $h(x, v) = \frac{a \cdot b}{2}$. Then we can

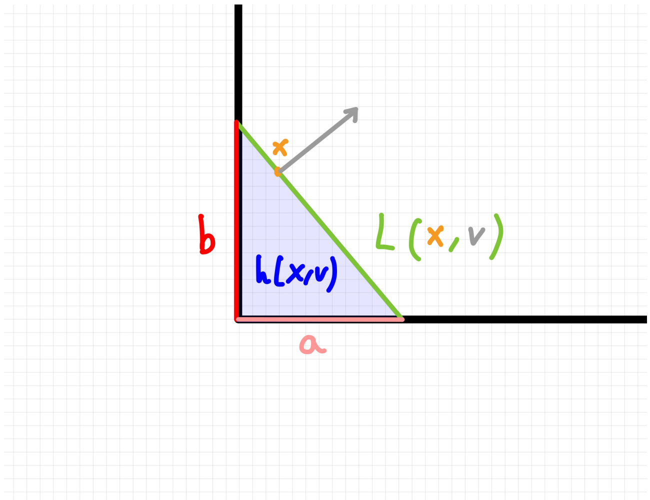

Figure 3: Description of the length of $L(x,v)$ when its length is small

write f as $f(x,v) = \frac{h(x,v)}{\text{length of } L(x,v)} = \frac{\frac{a\cdot b}{2}}{\sqrt{a^2+b^2}} \leq \frac{a^2+b^2}{\sqrt{a^2+b^2}} = a^2+b^2 = \text{length of } L(x,v) < \infty$. The inequality follows from $(a-b)^2 \geq 0 \Leftrightarrow a^2+b^2-2a\cdot b \geq 0 \Leftrightarrow a^2+b^2 \geq 2a\cdot b \geq \frac{a\cdot b}{2}$

2. The set of discontinuities is a null set:
The function f is discontinuous in the set $\{(x,v) \mid 0 \in L(x,v)\}$ as one can see from Figure 3. There may be other discontinuities in the border of S. Thus the set of discontinuities is contained in
$$(\partial S) \times \mathbb{S}^1 \cup \{(x,v) \mid 0 \in L(x,v)\}$$
which is a null set.
This two points prove that f is Riemann integrable.

3. $\int_{\mathrm{T}^1\mathbb{T}^2} f \; \mathrm{d}(x,v) = 0$
First notice that *f(x,v)=-f(x,-v)*. Then we can split $\mathrm{T}^1\mathbb{T}^2 = \mathbb{T}^2 \times \mathbb{S}^1$ into two disjoint sets $\mathbb{T}^2 \times \mathbb{S}^{1^+}$ and $\mathbb{T}^2 \times \mathbb{S}^{1^-}$ where S^{1^-} represents the lower half and S^{1^+} represents the upper half of the circle. Then $\int_{\mathrm{T}^1\mathbb{T}^2} f \; \mathrm{d}(x,v) = \int_{\mathbb{T}^2\times\mathbb{S}^{1+}} f \; \mathrm{d}(x,v) + \int_{\mathbb{T}^2\times\mathbb{S}^{1-}} f \; \mathrm{d}(x,v) = \int_{\mathbb{T}^2\times\mathbb{S}^{1+}} f \; \mathrm{d}(x,v) - \int_{\mathbb{T}^2\times\mathbb{S}^{1+}} f \; \mathrm{d}(x,v) = 0$ $\square$

Because f is Riemann integrable we know that for any $\epsilon > 0$, there are continuous functions $f_-, f_+ \in C\left(\mathrm{T}^1\mathbb{T}^2\right)$ with the properties that $f_- \leq f \leq f_+$ and

$$\int_{\mathrm{T}^1\mathbb{T}^2} (f_+ - f_-) \; \mathrm{d}(x,v) \leq \epsilon$$

Since $\int_{\mathrm{T}^1\mathbb{T}^2} f \; \mathrm{d}(x,v) = 0$

$$\int_{\mathrm{T}^1\mathbb{T}^2} f_- \; \mathrm{d}(x,v) \leq 0 \leq \int_{\mathrm{T}^1\mathbb{T}^2} f_+ \; \mathrm{d}(x,v)$$

and

$$\int_{\mathrm{T}^1\mathbb{T}^2} f_+ \; \mathrm{d}(x,v) \leq \int_{\mathrm{T}^1\mathbb{T}^2} f_- \; \mathrm{d}(x,v) + \epsilon \leq \epsilon$$

and

$$\int_{\mathrm{T}^1\mathbb{T}^2} f_- \; \mathrm{d}(x,v) \geq \int_{\mathrm{T}^1\mathbb{T}^2} f_+ \; \mathrm{d}(x,v) - \epsilon \geq -\epsilon$$

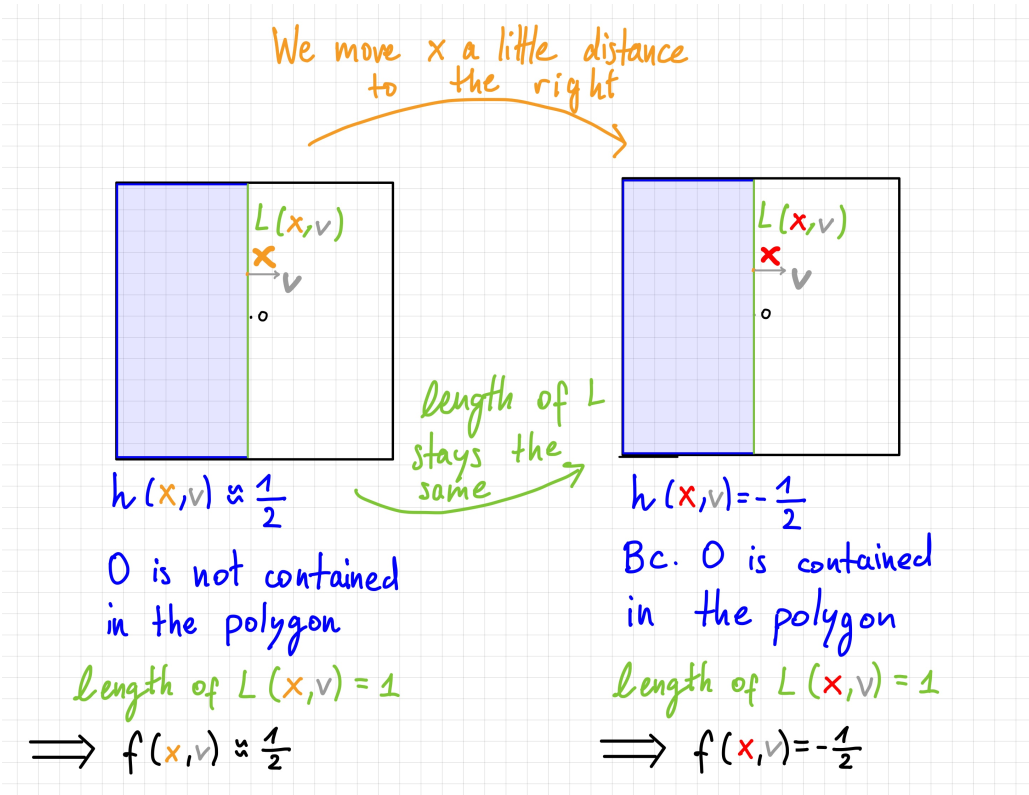

Figure 4: Discontinuities in $\{(x,v) \mid 0 \in L(x,v)\}$

Thus it holds

$$-\varepsilon \leqslant \int_{\mathrm{T}^1\mathbb{T}^2} f_- \mathrm{d}(x,v) \leqslant 0 \leqslant \int_{\mathrm{T}^1\mathbb{T}^2} f_+ \mathrm{d}(x,v) \leqslant \varepsilon$$

We can apply Proposition 5.2 to f_- and f_+ so if R is big enough then

$$\begin{aligned} -2\varepsilon \leqslant \textstyle\int_{\mathrm{T}^1\mathbb{T}^2} f_- \mathrm{d}(x,v) - \varepsilon \leqslant \int_0^1 f_-\left(\gamma_R(t)\right) \mathrm{d}t \leqslant \int_0^1 f\left(\gamma_R(t)\right) \mathrm{d}t \\ \leqslant \textstyle\int_0^1 f_+\left(\gamma_R(t)\right) \mathrm{d}t \leqslant \int_{\mathrm{T}^1\mathbb{T}^2} f_+ \mathrm{d}(x,v) + \varepsilon \leqslant 2\varepsilon, \end{aligned}$$

Recall that $\gamma_R(t) = (Re^{2\pi i t}, e^{2\pi i t})$.
We define a new path $\bar{\gamma}_R$ as the path that joins all points of $\gamma_R(t)$ that intersect with the grid with a line as we can see in Figure 4. $\gamma_R(t)$ and $\bar{\gamma}_R$ are uniformly $O(R^{-1})$-close as we can deduce from Figure 5. Thus it for R large enough that
$-3\varepsilon \leqslant \int_0^1 f_-\left(\bar{\gamma}_R(t)\right) \mathrm{d}t \leqslant \int_0^1 f\left(\bar{\gamma}_R(t)\right) \mathrm{d}t \leqslant \int_0^1 f_+\left(\bar{\gamma}_R(t)\right) \mathrm{d}t \leqslant 3\varepsilon$
This holds since $\int_0^1 f_+\left(\bar{\gamma}_R(t)\right) \mathrm{d}t = \int_0^1 (f_+\left(\gamma_R(t)\right) + (f_+\left(\bar{\gamma}_R(t)\right) - f_+\left(\gamma_R(t)\right))) \mathrm{d}t \leq 2\epsilon + \epsilon = 3\epsilon$.
With the same argument we can justify that $\int_0^1 f_-\left(\bar{\gamma}_R(t)\right) \mathrm{d}t \geq -3\epsilon$ It follows that

$$\int_0^1 f\left(\bar{\gamma}_R(t)\right) \mathrm{d}t = \mathrm{o}(1)$$

as $R \to \infty$.
Let P_k be the polygon as in Figure 6 and let I_k be a subinterval of [0,1] such that $\gamma_R(I_k)$ is between two points of the grid. Then we can write

$$\int_{I_k} f\left(\bar{\gamma}_R(t)\right) \mathrm{d}t = \left(\text{area of } P_k - \nVdash_{P_k}(n_k)\right) \frac{|I_k|}{\text{length of } L_k}$$

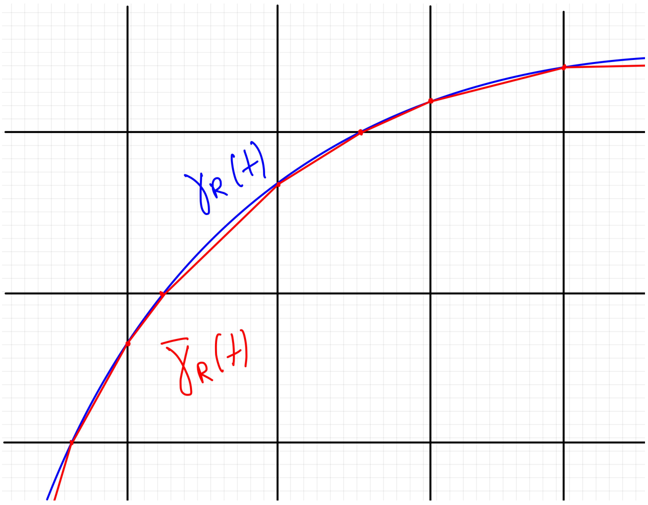

Figure 5: $\gamma_R(t)$ and $\bar{\gamma}_R$

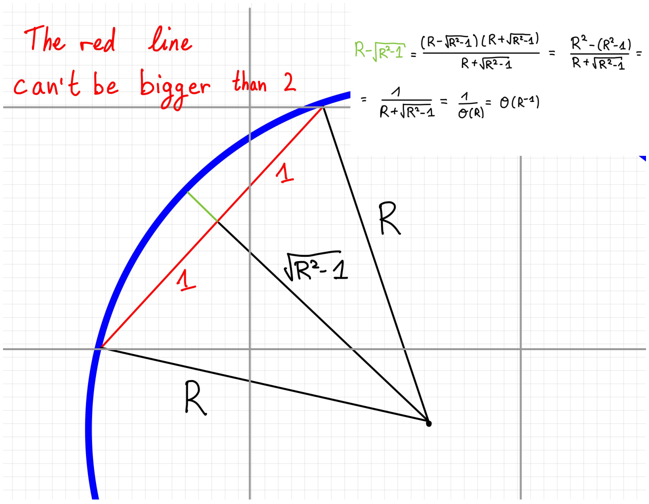

Figure 6: Justification why $\gamma_R(t)$ and $\bar{\gamma}_R$ are uniformly $O(R^{-1})$-close

.

By construction, the length $|I_k|$ of I_k is $\frac{\phi_k}{2\pi}$ where ϕ_k is the angle of the arc on the circle of radius R corresponding to I_k, so $\phi_k = \mathrm{O}\left(R^{-1}\right)$ because $\phi_k \cdot R =$ length of $\gamma_R(I_k)$. On the other hand,

$$\text{length of } L_k = 2R\sin\frac{\phi_k}{2} = 2R\left(\frac{\phi_k}{2} + \mathrm{O}\left(\phi_k^3\right)\right).$$

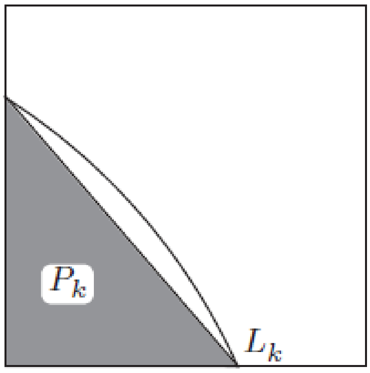

Figure 7: The polygon P_k

Therefore

$$\begin{aligned}
\int_0^1 f\left(\bar{\gamma}_R(t)\right) \mathrm{d}t =& \sum_{k=1}^{K} (\text{ area of } P_k - \#_{P_k}(n_k)) \frac{|I_k|}{\text{length of } L_k} \\
=& \frac{1}{2\pi R}\left(\sum_{k=1}^{K} (\text{ area of } P_k - \#_{P_k}(n_k)) \frac{\phi_k}{\phi_k + \mathrm{O}\left(\phi_k^3\right)}\right) \\
=& \frac{1}{2\pi R}\left(\sum_{k=1}^{K} (\text{ area of } P_k - \#_{P_k}(n_k)) + \sum_{k=1}^{K} \mathrm{O}\left(\phi_k^3\right)\right) \\
=& \frac{1}{2\pi R} (\text{ area of polygon enclosed by } \bar{\gamma}_R \\
& \text{-no. of lattice points inside }) + \mathrm{O}\left(R^{-3}\right).
\end{aligned}$$

Let's recall that $\int_0^1 f\left(\bar{\gamma}_R(t)\right) \mathrm{d}t = o(1)$. So multiplying by R in both sides we get

$$\text{area of polygon enclosed by } \bar{\gamma}_R \text{ - no. of lattice points inside } \bar{\gamma}_R = o(R)$$

We have proven our theorem but for a polygon determined by $\bar{\gamma}_R$ that approximates the circle of radius R. To extrapolate this result to the circle we claim that the area of the polygon enclosed by $\bar{\gamma}_R$ differs from the area of the circle by O(1). We also claim that the number of points inside the circle but outside the polygon is o(R). These claims prove the theorem for the circle since

$$\begin{aligned}
& \text{area of polygon enclosed by } \gamma_R \text{ - no. of lattice points inside } \gamma_R \\
=\ & \text{area of polygon enclosed by } \bar{\gamma}_R + \mathrm{O}(1) \text{ - no. of lattice points inside } \bar{\gamma}_R + o(R) \\
=\ & o(R)
\end{aligned}$$

From Figure 7 we can deduce why the first claim is true. In order to prove the second claim, notice that any lattice point n inside the circle but outside the polygon satisfies

$$0 \leqslant R - \|n\| = \mathrm{O}\left(R^{-1}\right)$$

.

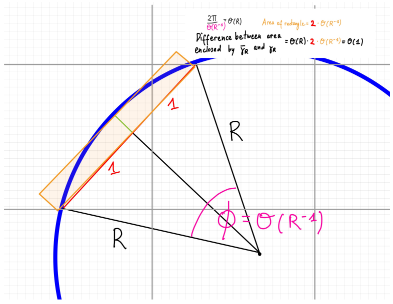

Figure 8: Justification of the first claim

This follows from Figure 5.
Fix $\delta > 0$ and let $g(x,v) = \mathbb{1}_{B_{2\delta}(0)}(x)$ be the characteristic function of the 2δ-ball around the identity $0 \in \mathbb{T}^2$ (but considered as a function on $\mathrm{T}^1\mathbb{T}^2$). If now $\|n\|$ lies in $[R-\delta, R]$, then there is a corresponding subinterval of length $\gg \frac{\delta}{R}$ such that $g\left(\gamma_R(t)\right) = 1$ for all t in that interval. Therefore

$$N_{R,\delta} = \left|\left\{n \in \mathbb{Z}^2 \mid R - \delta \leqslant \|n\| \leqslant R\right\}\right| \ll \frac{R}{\delta} \int_0^1 g\left(\gamma_R(t)\right) \mathrm{d}t.$$

The inequality holds since $\int_0^1 g\left(\gamma_R(t)\right) \mathrm{d}t$ is one in an interval of length much bigger than $\frac{\delta}{R}$ for every element in $N_{R,\delta}$. Thus $\int_0^1 g\left(\gamma_R(t)\right) \mathrm{d}t \gg \frac{\delta}{R} \cdot N_{R,\delta}$
By construction, g is Riemann integrable and so, for sufficiently large R,

$$N_{R,\delta} \ll \frac{R}{\delta} \int_{\mathrm{T}^1\mathbb{T}^2} g(x,v) \mathrm{d}(x,v) \ll \frac{R}{\delta}\delta^2 = R\delta. \quad \square$$

This proves the claim, and hence the theorem.

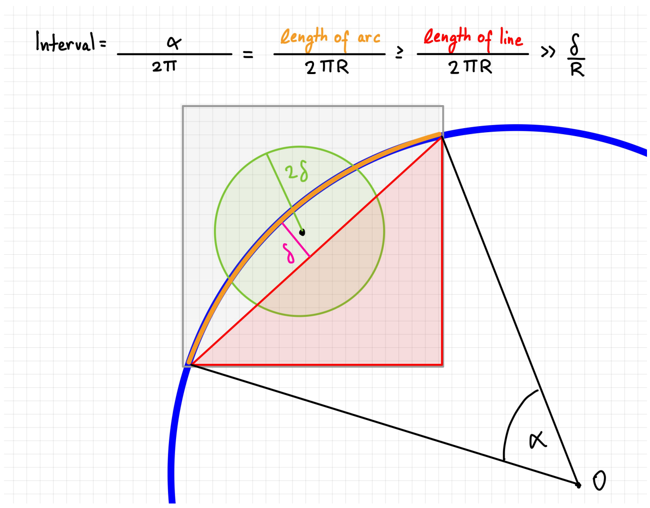

Figure 9: Intuition interval