Gerald Kuba

The two parameter ellipse problem

Mathematica Slovaca, Vol. 44 (1994), No. 5, 585--593

Persistent URL: http://dml.cz/dmlcz/136631

Math. Slovaca, **44** (1994), No. 5, 585–593

Mathematica Slovaca

THE TWO PARAMETER ELLIPSE PROBLEM

GERALD KUBA

(Communicated by Stanislav Jakubec)

ABSTRACT. The error term of the number of lattice points of an ellipse is studied.

1. Introduction

Consider an ellipse $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ in the Euclidian x, y-plane with $a \geq b \geq 1$, and let $R(a,b)$ be the number of lattice points (of the standard lattice $\mathbb{Z}^2$) inside or on the ellipse.

It is well known (see Fricker [2] or Krätzel [5]) that

$$R(a,b) = ab\pi + \Delta(a,b)$$

with

$$|\Delta(a,b)| \leq C(ab)^{\frac{1}{3}} \qquad \text{for} \quad a \geq b \geq 1\,,$$

(where C is an absolute constant), provided that the quotient $\frac{a}{b}$ is constant. Actually, the classical ellipse problem is a one parameter lattice point problem, where a fixed planar domain is "blown up".

In contrast, in this paper we consider the more general case where the major and minor axis of the ellipse are completely independent. In particular, we divide the "parameter domain" $a \geq b \geq 1$ in two subdomains in such a way that in the first subdomain (where $\frac{a}{b}$ is not too far away from a constant) the error term $\Delta(a,b)$ is qualitatively that of the circle problem (with the sharpest known exponent); in the second subdomain $\Delta(a,b)$ can be given in its true order of magnitude.

Thus, the objective of the present paper is a proof of the following results.

AMS Subject Classification (1991): Primary 11H56.
Key words: Ellipse, Lattice point, Error term.

THEOREM. *For* $a \geq b \geq 1$ *let*

$$R(a,b) = \#\left\{(x,y) \in \mathbb{Z}^2 \mid \frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1\right\}.$$

Furthermore, let $C > 0$ *be an absolute constant.*

Then we have

$$R(a,b) = ab\pi + \Delta(a,b)\,,$$

where the error term $\Delta(a,b)$ *can be estimated as follows:*

(i) *There exists a constant* c_1 *such that*

$$|\Delta(a,b)| \leq c_1 (ab)^{\frac{23}{73}} (\log ab)^{\frac{315}{146}}$$

for all a, b *in the region*

$$a \leq C b^{\frac{119}{100}} (\log ab)^{\frac{315}{100}}\,.$$

(ii) *There exists a constant* c_2 *such that*

$$|\Delta(a,b)| \leq c_2 \frac{a}{\sqrt{b}}$$

for all a, b *in the region*

$$a \geq C b^{\frac{119}{100}} (\log ab)^{\frac{315}{100}}\,.$$

(iii) *There exists a constant* $c_3 > 0$ *and a constant* C' *such that*

$$c_3 \frac{a}{\sqrt{b}} \leq |\Delta(a,b)|$$

for all sufficiently large $b \in \mathbb{N}$ *and all* $a \in \mathbb{R}$ *with*

$$a \geq C' b^{\frac{119}{100}} (\log ab)^{\frac{315}{100}}\,.$$

2. Preparation of the estimate

Let $a \geq b \geq 1$ throughout the paper, and let $H(a,b)$ be defined by

$$H(a,b) = \frac{2b}{\pi\sqrt{a}} \sum_{n=1}^{\infty} n^{-\frac{3}{2}} \sin\left(2\pi n a - \frac{\pi}{4}\right) + \frac{2a}{\pi\sqrt{b}} \sum_{n=1}^{\infty} n^{-\frac{3}{2}} \sin\left(2\pi n b - \frac{\pi}{4}\right).$$

In order to prove the Theorem it is sufficient to verify the following two propositions:

PROPOSITION 1. *For $a \geq b \geq 1$, we have*

$$R(a,b) = ab\pi + H(a,b) + O\Big(\frac{a}{b}\Big) + O\big((ab)^{\frac{1}{3}}\big)\,,$$

with absolute constants.

PROPOSITION 2. *If a, b satisfy the condition $1 \leq b \leq a \ll b^{\frac{13}{10}}$, we even have*

$$R(a,b) = ab\pi + H(a,b) + O\Big(\frac{a}{b}\Big) + O\big((ab)^{\frac{23}{73}}(\log ab)^{\frac{315}{146}}\big)\,,$$

where the constants only depend on the $\ll$-constant of the assumption.

Note that

$$H(a,b) = \Omega\Big(\frac{a}{\sqrt{b}}\Big)$$

for $(a,b) \in \mathbb{R}_+ \times \mathbb{N}$, where $a \gg b$ and b is sufficiently large.

Thus the expression $H(a,b) + O\Big(\frac{a}{b}\Big)$ in Propositions 1 and 2 can be replaced by a second main term $\begin{Bmatrix} O \\ \Omega \end{Bmatrix}\Big(\frac{a}{\sqrt{b}}\Big)$.

Furthermore, note that

$$\frac{a}{\sqrt{b}}\begin{Bmatrix} \ll \\ \gg \end{Bmatrix}(ab)^{\frac{1}{3}} \iff a\begin{Bmatrix} \ll \\ \gg \end{Bmatrix}b^{\frac{5}{4}}\,,$$

and

$$\frac{a}{\sqrt{b}}\begin{Bmatrix} \ll \\ \gg \end{Bmatrix}(ab)^{\frac{23}{73}}(\log ab)^{\frac{315}{146}} \iff a\begin{Bmatrix} \ll \\ \gg \end{Bmatrix}b^{\frac{119}{100}}(\log ab)^{\frac{315}{100}}\,.$$

Therefore we may use Proposition 2 to prove clause (i) of the Theorem. We then combine Proposition 1 and Proposition 2 to prove clause (ii) and (iii). Proposition 1 provides the results in the parameter domain $a \gg b^{\frac{51}{40}}$ and Proposition 2 in the domain $b^{\frac{119}{100}}(\log ab)^{\frac{315}{100}} \ll a \ll b^{\frac{51}{40}}$.

In order to establish Proposition 2 we make use of an essential tool from Huxley's "Discrete Hardy-Littlewood Method" in the shape presented in Huxley [3] and [4]. The following lemma is a combination of Huxley [4; Theorem 3 and Theorem 4]:

Let $\psi(\cdot)$ be defined by

$$\psi(z) = z - [z] - \frac{1}{2} \qquad (z \in \mathbb{R})$$

throughout the paper.

LEMMA 1. *Let M, M', T be positive real parameters satisfying $M \le M' < 2M$ and $M \le C_1 T^{\frac{83}{146}} (\log T)^{-\frac{63}{292}}$ with a constant C_1. Furthermore, let $F(t)$ be a four times continuously differentiable function on $1 \le t \le 2$ satisfying*

$$F'(t)\,,\ F''(t)\,,\ F^{(3)}(t)\,,\ F'(t)F^{(3)}(t) - 3\big(F''(t)\big)^2,\ F''(t)F^{(4)}(t) - 3\big(F^{(3)}(t)\big)^2 \neq 0$$

for all $1 \le t \le 2$. Then it follows that

$$\sum_{M \le k \le M'} \psi\Big(\frac{T}{M} F\Big(\frac{k}{M}\Big)\Big) \ll T^{\frac{23}{73}} (\log T)^{\frac{315}{146}} .$$

The O-constant depends on C_1 and on the range of values taken by the derivatives of the function F.

In order to verify Proposition 1 we use van der Corput's classical estimate of ψ-sums:

LEMMA 2. (see van der Corput [1]) *Let f be a real valued function, twice continuously differentiable on $[a,b] \subset \mathbb{R}$. Furthermore, let f'' be monotonous and nonzero on $[a,b]$. Then it follows that*

$$\sum_{a \le k \le b} \psi(f(k)) \ll \int_a^b |f''(t)|^{\frac{1}{3}}\ \mathrm{d}t + |f''(a)|^{-\frac{1}{2}} + |f''(b)|^{-\frac{1}{2}} .$$

where the O-constant is absolute.

The calculation of the main term in Propositions 1 and 2 is along the lines of the preparation of the classical ellipse problem. Clearly,

$$R(a,b) = 1 + 2[a] + 2[b] + 4S(a,b)\,,$$

with

$$S(a,b) = \left[\frac{a}{\sqrt{2}}\right]\left[\frac{b}{\sqrt{2}}\right] + \sum_{\frac{a}{\sqrt{2}} < x \le a} \left[b\sqrt{1 - \frac{x^2}{a^2}}\,\right] + \sum_{\frac{b}{\sqrt{2}} < y \le b} \left[a\sqrt{1 - \frac{y^2}{b^2}}\,\right].$$

Eliminating the Gauss brackets via the rounding error function ψ, using the Euler summation formula, and adding areas in an obvious way, we obtain

$$R(a,b) = ab\pi + I(a,b) + I(b,a) + \Psi(a,b) + \Psi(b,a) + O(1)\,,$$

with

$$I(\alpha,\beta) = -4\frac{\beta}{\alpha}\int\limits_{\frac{\alpha}{\sqrt{2}}}^{\alpha} \frac{x}{\sqrt{\alpha^2 - x^2}}\,\psi(x)\,\mathrm{d}x$$

and

$$\Psi(\alpha,\beta) = -4 \sum_{\frac{\alpha}{\sqrt{2}} < n \le \alpha} \psi\left(\beta\sqrt{1 - \frac{n^2}{\alpha^2}}\right)$$

for

$$\{\alpha,\beta\} = \{a,b\}\,.$$

3. Evaluation of the integrals $I(a,b)$ and $I(b,a)$

We substitute the lower limit of $I(\alpha,\beta)$ by 0 and obtain an error term $O\left(\frac{\beta}{\alpha}\right)$.

After an obvious substitution and insertion of the Fourier expansion of the function ψ we get

$$I(\alpha,\beta) = 2\beta\sum_{n=1}^{\infty}\frac{1}{n}J_1(2\pi n\alpha) + O\left(\frac{\beta}{\alpha}\right),$$

where J_1 is the Bessel function of first order, since

$$J_1(z) = \frac{2}{\pi}\int\limits_0^1 \frac{u}{\sqrt{1-u^2}}\sin zu\,\mathrm{d}u\,.$$

From the well known formula

$$J_1(z) = \sqrt{\frac{2}{\pi z}}\sin\left(z - \frac{\pi}{4}\right) + O\left(z^{-\frac{3}{2}}\right) \qquad (\,z \to \infty\,)$$

we obtain

$$I(\alpha,\beta) = \frac{2\beta}{\pi\sqrt{\alpha}}\sum_{n=1}^{\infty} n^{-\frac{3}{2}}\sin\left(2\pi n\alpha - \frac{\pi}{4}\right) + O\left(\frac{\beta}{\alpha}\right) \qquad \text{for} \quad \{\alpha,\beta\} = \{a,b\}\,.$$

Therefore we derive

$$R(a,b) = ab\pi + \Psi(a,b) + \Psi(b,a) + H(a,b) + O\left(\frac{a}{b}\right).$$

In order to complete the proof of Proposition 1 and Proposition 2, respectively, we have to estimate the ψ-sums.

4. Estimation of the sums $\Psi(a,b)$ and $\Psi(b,a)$

The estimation of $\Psi(\alpha,\beta)$ with the help of van der Corput's method (Lemma 2) is straightforward. This proves Proposition 1.

To execute Proposition 2 it is sufficient to verify the following proposition:

PROPOSITION 3. *For $1 \le \alpha \ll \beta^{\frac{13}{10}}$, we have*

$$\Psi(\alpha,\beta) \ll (\alpha\beta)^{\frac{23}{73}} (\log \alpha\beta)^{\frac{315}{146}} .$$

Proof. We can write

$$-\frac{1}{4}\Psi(\alpha,\beta) = S(\alpha,\beta;M_1) + O(M_1)$$

with

$$S(\alpha,\beta;M_1) = S = \sum_{\frac{\alpha}{\sqrt{2}} < n \le \alpha - M_1} \psi\big(g(n)\big) ,$$

where $M_1 \ge 2$ is to be exactly defined later, and

$$g(t) = \frac{\beta}{\alpha}\sqrt{\alpha^2 - t^2} \qquad \text{for} \quad t \in \left]\frac{\alpha}{\sqrt{2}}\,, \alpha - M_1\right] .$$

We note that

$$g'(t) = -\frac{\beta}{\alpha}\big(\alpha^2 - t^2\big)^{-\frac{1}{2}} t ,$$

$$g''(t) = -\frac{\beta}{\alpha}\big(\alpha^2 - t^2\big)^{-\frac{3}{2}} \alpha^2 ,$$

$$g^{(3)}(t) = -3\frac{\beta}{\alpha}\big(\alpha^2 - t^2\big)^{-\frac{5}{2}} \alpha^2 t ,$$

$$g^{(4)}(t) = -3\frac{\beta}{\alpha}\big(\alpha^2 - t^2\big)^{-\frac{7}{2}} \alpha^2 \big(4t^2 + \alpha^2\big) ,$$

and observe that for $t \in \left]\frac{\alpha}{\sqrt{2}}\,, \alpha - M_1\right]$ and $r \in \{1,2,3,4\}$,

$$|g^{(r)}| \asymp \frac{\beta}{\alpha}\big(\alpha^2 - t^2\big)^{\frac{1}{2}-r} \alpha^r \asymp \frac{\beta}{\sqrt{\alpha}}(\alpha - t)^{\frac{1}{2}-r} .$$

Furthermore, we see that for $t \in \left]\frac{\alpha}{\sqrt{2}}\,, \alpha - M_1\right]$,

$$g'(t), g''(t), g^{(3)}(t) \neq 0 ,$$

and

$$g'(t)g^{(3)}(t) - 3\big(g''(t)\big)^2 = -3\Big(\frac{\beta}{\alpha}\Big)^2\big(\alpha^2 - t^2\big)^{-2}\alpha^2 \neq 0\,,$$

and

$$g''(t)g^{(4)}(t) - 3\big(g^{(3)}(t)\big)^2 = 3\Big(\frac{\beta}{\alpha}\Big)^2\big(\alpha^2 - t^2\big)^{-5}\alpha^4(3t^2 + \alpha^2) \neq 0\,.$$

Now let $\tau = 1 - \dfrac{1}{\sqrt{2}}$, and define

$$f(u) := g\big([\alpha] - u\big)\,.$$

Then the new function f is certainly defined for $u \in [M_1, \tau\alpha[$, and for all such u,

$$|f^{(r)}(u)| \asymp \frac{\beta}{\sqrt{\alpha}} u^{\frac{1}{2}-r} \qquad (r = 1, 2, 3, 4)\,.$$

Of course, the "$\neq 0$-conditions" which hold for g also hold for f. The sum S may now be written

$$S = \tilde{S} + O(1)\,,$$

with

$$\tilde{S} = \sum_{M_1 \le m < \tau\alpha} \psi\big(f(m)\big)\,.$$

We split up the interval of summation by a geometric sequence $M_j = 2^j M_1$, for $j = 1, \dots, J$, with J such that $M_J = \tau\alpha$ and $M_1 \asymp (\alpha\beta)^{\frac{23}{73}}$.

Note that for β too large (i.e. $\beta \gg \alpha^{\frac{50}{23}+\varepsilon}$), $\Big]\dfrac{\alpha}{\sqrt{2}}\,,\, \alpha - M_1\Big] = \emptyset$ and $S = 0$ for sufficiently large α, and in this case the statement of Proposition 3 is trivially true.

We have

$$\tilde{S} = \sum_{j=1}^{J-1} S_j$$

with

$$S_j = \sum_{M_j \le m < M_{j+1}} \psi\big(f(m)\big) \qquad (1 \le j < J)\,.$$

Now we apply Lemma 1 to each of these sums S_j:

Put $M = M_j$, $M' = -[-2M] - 1$, $T = \dfrac{\beta}{\sqrt{\alpha}} M^{\frac{3}{2}}$, and $F(u) = \dfrac{M}{T} f(Mu)$.

Then $F(u)$ satisfies all conditions of Lemma 1. In addition we have $|F^{(r)}(u)| \asymp 1$ for $1 \le u \le 2$ and $r = 1, 2, 3, 4$.

To verify

$$M \le C_1 T^{\frac{83}{146}} \left(\log(T)\right)^{-\frac{63}{292}},$$

it is sufficient to show that

$$M \le C_2 T^{\frac{13}{23}},$$

since $\dfrac{13}{23} < \dfrac{83}{146}$.

According to the assumption of Proposition 3 there exists a constant C_3 with $\alpha \le C_3 \beta^{\frac{13}{10}}$ for all α, β to be considered.

This implies

$$\alpha \le C_4 \left(\frac{\beta}{\sqrt{\alpha}}\right)^{\frac{26}{7}}.$$

Since $M \le M_J \le \alpha$, we get

$$M \le C_2 \left(\frac{\beta}{\sqrt{\alpha}} M^{\frac{3}{2}}\right)^{\frac{13}{23}}.$$

The constant C_1 only depends on C_3.

Thus, all conditions of Lemma 1 are satisfied and we obtain

$$S_j \ll \left(\frac{\beta}{\sqrt{\alpha}} M_j^{\frac{3}{2}}\right)^{\frac{23}{73}} \left(\log\left(\frac{\beta}{\sqrt{\alpha}} M_j^{\frac{3}{2}}\right)\right)^{\frac{315}{146}} \le \left(\frac{\beta}{\sqrt{\alpha}}\right)^{\frac{23}{73}} \left(\log\left(\frac{\beta}{\sqrt{\alpha}} M_J^{\frac{3}{2}}\right)\right)^{\frac{315}{146}} M_j^{\frac{3}{2}\frac{23}{73}},$$

for all $1 \le j < J$. Summing all up, we get

$$\tilde{S} \ll \left(\frac{\beta}{\sqrt{\alpha}}\right)^{\frac{23}{73}} \left(\log\left(\frac{\beta}{\sqrt{\alpha}} M_J^{\frac{3}{2}}\right)\right)^{\frac{315}{146}} M_J^{\frac{3}{2}\frac{23}{73}},$$

since a finite geometric series is $\ll$ its largest term.

This completes the proof of Proposition 3, since $M_J \le \alpha$.

REFERENCES

[1] van der CORPUT, J. G.: *Zahlentheoretische Abschätzungen mit Anwendungen auf Gitterpunktsprobleme*, Math. Z. **17** (1913), 250–259.

[2] FRICKER, F.: *Einführung in die Gitterpunktlehre*, Birkhäuser Verlag, Basel-Boston-Stuttgart, 1982.

[3] HUXLEY, M. N.: *Exponential sums and lattice points*, Proc. London Math. Soc. (3) **60** (1990), 471–502.

[4] HUXLEY, M. N.: *Exponential sums and lattice points II*. Proc. London Math. Soc. (3) **66** (1993), 279–301.

[5] KRÄTZEL, E.: *Lattice Points*, Berlin, 1988.

Received January 17, 1994

Institut für Mathematik u.a.St.
Universität für Bodenkultur
A-1180 Wien
Austria
E-mail: Kuba@edv1.boku.ac.at