Collect. Math. **49**, 2–3 (1998), 283–291

Expressing a number as the sum of two coprime squares

Warren Dicks*

Departament de Matemàtiques, Universitat Autònoma de Barcelona,
08193 Bellaterra (Barcelona), Spain
E-mail: dicks@mat.uab.es

Joan Porti*

Laboratoire Emile Picard, CNRS UMR 5580, Université Paul Sabatier,
31062 Toulouse Cedex 4, France
E-mail: porti@picard.ups-tlse.fr

Dedicat a la memòria d'en Ferran Serrano

Abstract

We use hyperbolic geometry to study the limiting behavior of the average number of ways of expressing a number as the sum of two coprime squares. An alternative viewpoint using analytic number theory is also given.

1. Introduction

Let $\mathbb{N}$ denote the set of all finite cardinals, and $\mathbb{N}_+$ the set of all nonzero finite cardinals.

Poincaré series of Fuchsian groups have critical exponent one, which is connected to the fact that they have the whole circle at infinity as their limit set, and this circle has Hausdorff dimension one. For the Fuchsian group $PSL_2(\mathbb{Z})$, a Poincaré series is given by $P(s) = 2 \sum\limits_{n \geq 1} \frac{b_n}{n^s}$, where, for each $n \in \mathbb{N}_+$,

$$b_n = \left| \{(c, d) \in \mathbb{N}^2 \mid c, d \text{ coprime}, c^2 + d^2 = n\} \right|,$$

* Partially supported by the DGICYT (Spain) through grant number PB93-0900.

the number of *primitive* representations of n as the sum of two squares. Thus $P(s)$ converges if $\mathrm{Re}(s) > 1$ and diverges if $\mathrm{Re}(s) < 1$, which suggests that if we put

$$c_n = \sum_{k=1}^{n} b_k = \left| \{(c,d) \in \mathbb{N}^2 \mid c,d \text{ coprime}, c^2+d^2 \leq n\} \right|,$$

for $n \in \mathbb{N}_+$, then there should be restraints on the behavior of $\frac{c_n}{n}$, that is, the average number of ways of expressing a natural number as the sum of two coprime squares. We found no explicit reference to this behavior in the literature. On omitting the coprimeness condition, one has a sequence which was studied by Gauss, and converges to $\frac{\pi}{4}$. Thus, he was interested in the number of integer lattice points in a circle of radius $\sqrt{n}$, while we are interested in the subset of those points which are "visible from the origin", and lie in the first quadrant. Resorting to simple computer calculations, we were intrigued to see that $\frac{c_n}{n}$ seemed to be quite docile; for example, the thousandth term is $\frac{478}{1000}$ and the ten thousandth term is $\frac{4772}{10000}$. Eventually, using hyperbolic geometry, we were able to prove that the $\frac{c_n}{n}$ converge to $\frac{3}{2\pi} = 0.477465....$ These geometric techniques gave even more information, which we now describe.

For $m, n \in \mathbb{N}_+$, let

$$c_{m,n} = \left| \{(c,d) \in \mathbb{N}^2 \mid c,d \text{ coprime}, c^2+d^2 \leq n, c \in m\mathbb{N}\} \right|.$$

For example, $c_{1,n} = c_n$. Define $\psi\colon \mathbb{N}_+ \to \mathbb{N}_+$ by $\psi(m) = m \prod_{p|m}(1+\frac{1}{p})$, where the product ranges over all prime numbers p which divide m.

Our main result is as follows.

Theorem 1.1

For each $m \in \mathbb{N}_+$, $c_{m,n} \sim \frac{3n}{2\pi\psi(m)}$ as n tends to infinity.

That is, $\lim_{n\to\infty} \frac{c_{m,n}}{n} = \frac{3}{2\pi\psi(m)}$; we do not discuss the error term $c_{m,n} - \frac{3n}{2\pi\psi(m)}$ at all. Our proof uses the action of the modular group $PSL_2(\mathbb{Z})$ on the upper half plane $\{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$, viewed as the hyperbolic plane $\mathbb{H}^2$. We show that the asymptotic behavior of $c_{m,n}$ is equivalent to the asymptotic behavior of subsets of certain congruence subgroups of $PSL_2(\mathbb{Z})$. The asymptotic behavior can then be computed in terms of the covolume of these subgroups. For example, for $m = 1$, the theorem says that $c_n \sim \frac{3n}{2\pi}$, which is related to the fact that the volume of $PSL_2(\mathbb{Z})\backslash\mathbb{H}^2$ is $\frac{\pi}{3}$.

The referee has kindly provided us with an alternative proof of Theorem 1.1 using analytic number theory, and Section 4 describes this proof and the relationship with Eisenstein series.

2. The Farey tessellation

Let $\mathbb{H}^2$ denote the upper half plane $\{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$, viewed as the hyperbolic plane with the metric $ds^2 = \frac{d|z|^2}{\mathrm{Im}(z)^2}$. There is a natural boundary at infinity $\partial\mathbb{H}^2 \cong S^1$ which we identify with $\mathbb{R} \cup \{\infty\}$.

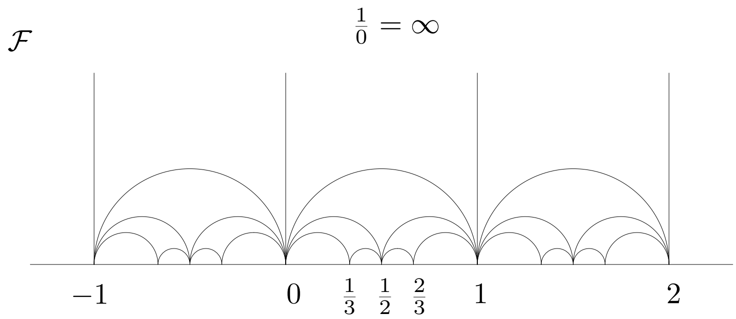

Figure 2.1. The Farey tessellation

The *Farey tessellation*, $\mathcal{F}$, of $\mathbb{H}^2$ is a tessellation by ideal triangles with vertices lying in $\mathbb{Q} \cup \{\infty\}$, as in Figure 2.1. Two elements $\frac{p}{q}$ and $\frac{r}{s}$ of $\mathbb{Q} \cup \{\infty\}$ in lowest terms (so $\infty = \frac{\pm 1}{0}$) are joined by a (geodesic) edge in the Farey tessellation if and only if $ps - qr = \pm 1$. Every edge having vertices $\frac{p}{q}$ and $\frac{r}{s}$ is adjacent to the triangles with third vertices $\frac{p+r}{q+s}$ and $\frac{p-r}{q-s}$. Using the Euclidean algorithm, one can prove that this defines a tesselation, $\mathcal{F}$, and it is clear that $\mathcal{F}$ is preserved by the action of $PSL_2(\mathbb{Z})$.

Let R be the region $\{z \in \mathbb{H}^2 \mid 0 \leq \mathrm{Re}(z) \leq 1\}$. Let $E(\mathcal{F})$ denote the set of oriented edges of the Farey tessellation, and $E(\mathcal{F}|_R)$ the set of those oriented edges which lie in R. Each $e \in E(\mathcal{F}|_R)$ has an initial vertex $\frac{a}{c}$ and a terminal vertex $\frac{b}{d}$, both in $[0,1] \cup \{\infty\}$, and we may assume these are in lowest terms, and that a, b, c, d lie in $\mathbb{N}$; here we set $f(e) = (c,d)$ to define a map

$$f\colon E(\mathcal{F}|_R) \to \left\{(c,d) \in \mathbb{N}^2 \mid c,d \text{ coprime}\right\}.$$

Lemma 2.1

i) The action of $PSL_2(\mathbb{Z})$ on $E(\mathcal{F})$ is transitive and faithful.
ii) The map $f\colon E(\mathcal{F}|_R) \to \{(c,d) \in \mathbb{N}^2 \mid c,d$ coprime$\}$ is two-to-one.

Proof. Assertion i) is well-known and easy to prove. Assertion ii) is equivalent to the following fact, which is also easy to prove: given two coprime non-negative integers c and d, there are exactly two pairs $(a,b) \in \mathbb{N}^2$ such that $ad - bc = \pm 1$ and $\frac{a}{c}, \frac{b}{d} \in [0,1] \cup \{\infty\}$. $\square$

Corollary 2.2

$$\begin{aligned} &\left|\,\{(c,d) \in \mathbb{N}^2 \mid c,d \text{ coprime},\, c^2 + d^2 = n\}\,\right| \\ &= \frac{1}{2}\left|\,\left\{\gamma \in PSL_2(\mathbb{Z}) \mid \gamma i \in R \text{ and } Im(\gamma i) = \frac{1}{n}\right\}\,\right|. \end{aligned}$$

Proof. Let e_0 be the edge joining $0 = \frac{0}{1}$ to $\infty = \frac{1}{0}$, and consider any $\gamma \in PSL_2(\mathbb{Z})$ such that γe_0 is contained in the interior of R. This is equivalent to $\gamma i \in R$. If $f(\gamma e_0) = (c,d)$, then the imaginary part of γi is $\frac{1}{c^2+d^2}$. The result now follows from the preceding lemma. $\square$

3. The geometric proof

In this section we prove Theorem 1.1. We will need the following.

Theorem (Nicholls [4])

Let Γ be a Fuchsian group of finite covolume, $r > 0$ a real number, $x \in \mathbb{H}^2$, and θ an interval in $\partial\mathbb{H}^2$. Let $B(x,r)$ be the hyperbolic ball in $\mathbb{H}^2$ with center x and radius r, $B(x,r,\theta)$ the set of points of $B(x,r)$ whose projection to $\partial\mathbb{H}^2$ from x lies in θ, and $w_x(\theta)$ the angle of this cone, $B(x,r,\theta)$, at the point x. Then, for any $y \in \mathbb{H}^2$,

$$\left|\,\{\gamma \in \Gamma \mid \gamma y \in B(x,r,\theta)\}\,\right| \sim \frac{w_x(\theta)}{2\pi}\frac{\mathrm{vol}B(x,r)}{\mathrm{vol}(\Gamma\backslash\mathbb{H}^2)} \qquad \text{as } r \text{ tends to } \infty\,. \ \square$$

We shall use Nicholls' Theorem about balls to obtain the following result about horoballs.

Proposition 3.1

Let Γ be a Fuchsian group of finite covolume such that ∞ is a parabolic point for Γ, let $a < b$ and $t > 0$ be real numbers, and let

$$X(a,b,t) = \left\{z \in \mathbb{H}^2 \mid \mathrm{Im}(z) \geq \frac{1}{t}, a \leq \mathrm{Re}(z) \leq b\right\}.$$

Then, for any $w \in \mathbb{H}^2$,

$$\big|\{\gamma \in \Gamma \mid \gamma w \in X(a,b,t)\}\big| \sim \frac{\operatorname{vol}(X(a,b,t))}{\operatorname{vol}(\Gamma\backslash\mathbb{H}^2)} \qquad \textit{as } t \textit{ tends to } \infty\,.$$

Proof. There exists a parabolic point (for Γ) c between a and b, and the desired formula is additive, so we may assume that one of a or b is parabolic, and by symmetry we may assume that b is parabolic.

By conjugating by an isometry of $\mathbb{H}^2$ fixing ∞, we may assume that $a = 0$ and $b = 1$, so 1 is a parabolic point.

Let $m \in \mathbb{N}_+$, and let $t > 0$.

Define P_m to be the cone consisting of the points in $\mathbb{H}^2$ whose radial projection from mi to $\partial\mathbb{H}^2$ lies in the unit interval $[0, 1]$. Set

$$\begin{aligned} P_{m,t} &= X(0,1,t) \cap P_m, \\ V_{m,t} &= \Big\{z \in P_m \mid d(z, mi) \le d\big(mi, \frac{i}{t}\big)\Big\}, \\ W_{m,t} &= \big\{z \in P_m \mid d(z, mi) \le d(mi, z_{m,t})\big\}, \end{aligned}$$

where d denotes the hyperbolic distance, and $z_{m,t}$ is the point of the geodesic joining mi to 1 whose imaginary part is $\frac{1}{t}$; see Figure 3.1. Notice that $V_{m,t} \subseteq P_{m,t} \subseteq W_{m,t}$.

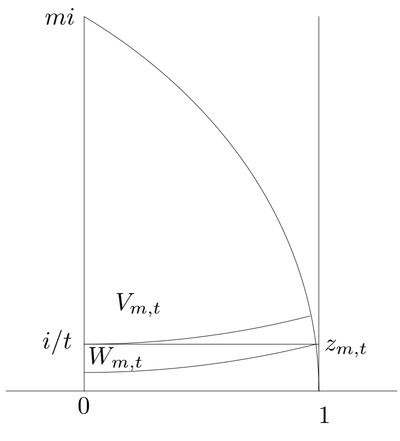

Figure 3.1

Let $p_{m,t}$ (resp. $v_{m,t}$, $w_{m,t}$) denote the number of those elements γ of Γ such that γw belongs to $P_{m,t}$ (resp. $V_{m,t}$, $W_{m,t}$). Notice that $v_{m,t} \le p_{m,t} \le w_{m,t}$.

Since 1 and ∞ are parabolic points, for sufficiently large m, the set

$$\left\{\gamma \in \Gamma \mid \gamma w \in X(0,1,t) - P_{m,t}\right\}$$

has uniformly bounded cardinality. Hence, for sufficiently large m,

$$\left|\{\gamma \in \Gamma \mid \gamma w \in X(0,1,t)\}\right| \sim p_{m,t} \qquad \text{as } t \text{ tends to } \infty\,.$$

We claim that the following hold:
(i) For fixed m, $\frac{v_{m,t}}{t}$ (resp. $\frac{w_{m,t}}{t}$) tends to a limit $\bar{v}_m$ (resp. $\bar{w}_m$) as t tends to ∞.
(ii) Both $\bar{v}_m$ and $\bar{w}_m$ tend to $1/\mathrm{vol}(\Gamma\backslash\mathbb{H}^2)$ as m tends to infinity.

Since $\mathrm{vol}(X(0,1,t)) = t$, these two claims imply the result.

Since $d(mi, \frac{i}{t}) = \log(mt)$, and the volume of the ball of radius $\log(mt)$ is $\pi(mt + \frac{1}{mt} - 2)$, it follows from Nicholl's Theorem that, for fixed m, $\frac{v_{m,t}}{t}$ converges to

$$\bar{v}_m = \frac{\alpha_m m}{2\mathrm{vol}(\Gamma\backslash\mathbb{H}^2)}\,,$$

where α_m is the angle of the cone P_m at mi. Similarly, since

$$d(mi, z_{m,t}) = \log\left(\left(m+\frac{1}{m}\right)t + \frac{1}{2}\sqrt{\left(m+\frac{1}{m}\right)^2 t^2 - 4}\,\right),$$

for fixed m, $\frac{w_{m,t}}{t}$ converges to

$$\bar{w}_m = \frac{\alpha_m(m+\frac{1}{m})}{2\mathrm{vol}(\Gamma\backslash\mathbb{H}^2)}\,.$$

This proves Claim (i).

Finally, it follows from the equality

$$\sin(\alpha_m) = \frac{2}{m+\frac{1}{m}}$$

that $\bar{v}_m$ and $\bar{w}_m$ both converge to $1/\mathrm{vol}(\Gamma\backslash\mathbb{H}^2)$, which proves Claim (ii). $\square$

Proof of Theorem 1.1. Let $m \in \mathbb{N}_+$. The *congruence group of level* m is defined as

$$\Gamma_0(m) = \left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) \mid c \in m\mathbb{Z}\right\}.$$

Let $P\Gamma_0(m)$ denote its projectivization, $\Gamma_0(m)/\{\pm 1\}$. Consider the oriented edge e_0 of the Farey tessellation joining ∞ to 0. Let $\mathcal{F}(m)$ denote the restriction of the Farey tessellation whose set of oriented edges is the $P\Gamma_0(m)$-orbit of e_0, so

$$E\big(\mathcal{F}(m)\big) = P\Gamma_0(m)e_0 \,.$$

Let

$$f_m\colon E\big(\mathcal{F}(m)|_R\big) \to \big\{(c,d) \in \mathbb{N}^2 \mid c,d \text{ coprime}, c \in m\mathbb{N}\big\}$$

be the map obtained by restricting the two-to-one map

$$f\colon E(\mathcal{F}|_R) \to \big\{(c,d) \in \mathbb{N}^2 \mid c,d \text{ coprime}\big\}$$

of Lemma 2.1. It follows that f_m is also two-to-one. The same proof as for Corollary 2.2 shows that

$$\begin{aligned} c_{m,n} &= \big|\, \{(c,d) \in \mathbb{N}^2 \mid c,d \text{ coprime}, c^2+d^2 \leq n, c \in m\mathbb{N}\}\big| \\ &= \frac{1}{2}\left|\left\{\gamma \in P\Gamma_0(m) \mid \gamma i \in R \text{ and } \mathrm{Im}(\gamma i) \geq \frac{1}{n}\right\}\right| \\ &= \frac{1}{2}\,\big|\,\{\gamma \in P\Gamma_0(m) \mid \gamma i \in X(0,1,n)\big| \,. \end{aligned}$$

It now follows from Proposition 3.1 that

$$c_{m,n} \sim \frac{\mathrm{vol}(X(0,1,n))}{2\mathrm{vol}(P\Gamma_0(m)\backslash \mathbb{H}^2)} \qquad \text{as } n \text{ tends to } \infty \,.$$

It remains to calculate these volumes. It is well-known and easy to prove that

$$\mathrm{vol}(X(0,1,n)) = n \,,$$

and that $\mathrm{vol}(PSL_2(\mathbb{Z})\backslash\mathbb{H}^2) = \frac{\pi}{3}$. The index of $P\Gamma_0(m)$ in $PSL_2(\mathbb{Z})$ is the value $\psi(m)$ described in Section 1; see, for example, [3, Proposition 9.3]. Hence

$$\mathrm{vol}(P\Gamma_0(m)\backslash\mathbb{H}^2) = \frac{\pi\psi(m)}{3} \,.$$

This proves $c_{m,n} \sim \frac{3n}{2\pi\psi(m)}$ as n tends to ∞. □

4. The referee's proof using analytic number theory

Let $m \in \mathbb{N}_+$, let $z \in \mathbb{H}^2$, and let Γ denote $\Gamma_0(m)$, the congruence group of level m.

The *Eisenstein series* of weight 0, and level m, for the cusp ∞, evaluated at z, is defined as

$$E_m(z,s) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \mathrm{Im}(\gamma z)^s, \text{ for } \mathrm{Re}(s) > 1\,,$$

where Γ_∞ denotes the Γ-stabilizer of ∞.

This series has an analytic continuation to the whole complex s-plane, with no singularities on $\mathrm{Re}(s) = 1$ except for a simple pole at $s = 1$ with constant residue equal to $\frac{3}{\pi\psi(m)}$; see, for example, [1, p. 31] and [2, p. 239].

We now take $z = i$. A straightforward calculation shows that

$$E_m(i,s) = 2\sum_{n \geq 1} \frac{b_{m,n}}{n^s}\,, \text{ for } \mathrm{Re}(s) > 1\,,$$

where, for all n in $\mathbb{N}_+$,

$$b_{m,n} = \left| \{(c,d) \in \mathbb{N}^2 \mid c, d \text{ coprime}, c^2 + d^2 = n, c \in m\mathbb{N}\} \right|,$$

so $c_{m,n} = \sum\limits_{k=1}^{n} b_{m,k}$.

Since the $b_{m,n}$ are non-negative, we can apply the step-function version of Ikehara's Theorem [5, Theorem 17, p. 130], and deduce that $\lim\limits_{n\to\infty} \frac{c_{m,n}}{n} = \frac{3}{2\pi\psi(m)}$, as desired.

Acknowledgments. All the figures and numerical results have been produced using the Wolfram Research, Inc. computer software system *Mathematica*®, Version 2.2.

This work took place during a visit of the second-named author to the Centre de Recerca Matemàtica of the Institut d'Estudis Catalans, and he is grateful for the hospitality he received.

We thank the referee for many useful suggestions and references to the literature, and for the content of Section 4. We also thank Joaquim Bruna for much useful information about Tauberian theorems.

References

1. Jürgen Fischer, *An approach to the Selberg Trace Formula via the Selberg Zeta-Function,* Lecture Notes in Math., Vol. 1253, Springer, Berlin, 1987.
2. B.H. Gross and D.B. Zagier, Heegner points and derivatives of L-series, *Invent. Math.* **84** (1980), 225–326.
3. Anthony W. Knapp, *Elliptic Curves,* Princeton University Press, Princeton N.J., 1992.
4. Peter Nicholls, A lattice point problem in hyperbolic space, *Mich. Math. J.* **30** (1983), 273-287.
5. Norbert Wiener, *The Fourier integral and certain of its applications*, Dover, New York, 1958.