ON A CONSTRUCTION FOR THE REPRESENTATION OF A POSITIVE INTEGER AS THE SUM OF FOUR SQUARES

L'Enseignement Mathématique, t. 33 (1987), p. 301-306

ON A CONSTRUCTION FOR THE REPRESENTATION OF A POSITIVE INTEGER AS THE SUM OF FOUR SQUARES

by G. ROUSSEAU

Hermite [4] gave a very simple construction for the representation of a prime of the form $4k+1$ as the sum of two squares, using continued fractions. This note is concerned with the extension of this construction to the domain of Gaussian integers.

In **1** we show that, as might be expected (although it does not seem to have been published previously), there is a similar construction in the domain of Gaussian integers for the representation of an arbitrary positive integer as the sum of four squares.

One might also expect some small additional complications in the four squares case, and indeed what is shown in **1** is that for a given positive integer m there is a representation of m, $2m$ or $3m$, from which one then passes to a representation of m. This is of course not difficult, but it is somewhat inelegant. However, in computations carried out by L. Rousseau and the author, it emerged that this subsidiary transformation is never in fact needed. The construction always yields a representation of m directly; in other words, the Hermite construction applies just as well for the construction of four squares representations as it does for two squares representations. This is proved in **2** by a somewhat more elaborate argument, using an old result of Auric [1] on continued fractions with Gaussian integer terms.

Some of the other known four squares constructions are discussed briefly in **3**.

1. We consider the extension of Hermite's construction to the domain of Gaussian integers. Any positive integer N may be expressed in the form $N = k^2m$, where m is square-free. In order to construct a representation for N as the sum of four squares, it is evidently sufficient to construct one for m. The congruence $a^2 + b^2 + 1 \equiv 0 \pmod{m}$ is solvable in rational integers, so, setting $\alpha = a + ib$, we have, in the domain $\mathbf{Z} + \mathbf{Z}i$ of Gaussian integers,

(1) $$\alpha\bar{\alpha} \equiv -1 \pmod{m}.$$

Performing the Euclidean algorithm in $\mathbf{Z} + \mathbf{Z}i$, beginning with $\gamma_0 = \alpha$, $\gamma_1 = m$, we obtain Gaussian integers κ_r, γ_{r+2} such that

$$\gamma_r = \kappa_r \gamma_{r+1} + \gamma_{r+2}, \qquad N(\gamma_{r+2}) \leqslant \frac{1}{2} N(\gamma_{r+1}) \qquad (r=0, 1, ..., n),$$

terminating with $\gamma_{n+2} = 0$. Thus α/m has the continued fraction development

$$\alpha/m = (\kappa_0; \kappa_1, ..., \kappa_{n-1}, \kappa_n) = \kappa_0 + 1/(\kappa_1 + 1/(\kappa_2 + ... + 1/\kappa_n) ...).$$

If $[\kappa_s, ..., \kappa_t]$ denotes the usual Gaussian bracket, then $\alpha_r = [\kappa_0, \kappa_1, ..., \kappa_r]$ and $\beta_r = [\kappa_1, ..., \kappa_r]$ are, respectively, the numerator and denominator of the rth convergent of the continued fraction for α/m. We have $\gamma_r = \gamma_{n+1}[\kappa_r, ..., \kappa_n]$, where $\gamma_{n+1} = (\alpha, m)$ is a unit. Also the following standard formulas may be established:

(2) $$\beta_r \gamma_{r+1} + \beta_{r-1} \gamma_{r+2} = \gamma_1;$$

(3) $$\beta_r \gamma_0 - \alpha_r \gamma_1 = (-1)^r \gamma_{r+2}.$$

From (3) we have $\alpha\beta_r \equiv (-1)^r \gamma_{r+2} \pmod{m}$, so that, in view of (1),

$$N(\beta_r) + N(\gamma_{r+2}) \equiv 0 \pmod{m}.$$

Thus, for any r, $N(\beta_r) + N(\gamma_{r+2}) = a_r^2 + b_r^2 + c_r^2 + d_r^2$, say, is a multiple of m.

We shall show that there exists r such that

$$0 < N(\beta_r) + N(\gamma_{r+2}) < 4m.$$

To this end we first show that

$$N(\beta_r) N(\gamma_{r+1}) < 4 N(\gamma_1) \qquad (r=0, 1, ..., n).$$

This is done by induction on r. Thus let $w_r = \beta_r \gamma_{r+1}/\gamma_1$; it is to be shown that $|w_r| < 2$. Clearly $w_0 = \beta_0 = 1$. In view of (2) we have $w_r = 1 - \beta_{r-1}\gamma_{r+2}/\gamma_1 = 1 - (\gamma_{r+2}/\gamma_r) w_{r-1}$; since, for $r > 0$, $|\gamma_{r+2}/\gamma_r| \leqslant \frac{1}{2}$, we have $|w_r| \leqslant 1 + \frac{1}{2}|w_{r-1}|$, so if $|w_{r-1}| < 2$ then $|w_r| < 2$, as required.

Now, since the $N(\gamma_{r+1})$ decrease monotonically from m^2 to zero, we may (if $m>1$) choose r so that $N(\gamma_{r+2}) < 2m \leqslant N(\gamma_{r+1})$. Then, on the one hand, $N(\gamma_{r+2}) < 2m$, while, on the other hand, $N(\beta_r) < 4m^2/N(\gamma_{r+1}) \leqslant 2m$. Thus $0 < N(\beta_r) + N(\gamma_{r+2}) < 4m$, as was to be shown.

Thus m, $2m$ or $3m$ is the sum of four squares, from which it easily follows that m is representable as the sum of four squares (cf. [2]). Indeed, if $2m = a^2 + b^2 + c^2 + d^2$ then by rearranging the terms if necessary we may suppose $a \equiv b \pmod 2$, $c \equiv d \pmod 2$, so that

$$m = \left(\frac{a+b}{2}\right)^2 + \left(\frac{a-b}{2}\right)^2 + \left(\frac{c+d}{2}\right)^2 + \left(\frac{c-d}{2}\right)^2,$$

where the terms on the right are integers, while if $3m = a^2 + b^2 + c^2 + d^2$ then by changing signs if necessary we may suppose $a, b, c, d \equiv 0$ or $1 \pmod 3$ and further, by rearranging the terms if necessary, that $a \equiv 0 \pmod 3$, $b \equiv c \equiv d \pmod 3$, so that

$$m = \left(\frac{b+c+d}{3}\right)^2 + \left(\frac{a-b+c}{3}\right)^2 + \left(\frac{a-c+d}{3}\right)^2 + \left(\frac{a-d+b}{3}\right)^2,$$

where again the terms on the right are integers.

2. We shall establish the following result:

THEOREM. *Let m be a square-free positive integer and let $\alpha\bar{\alpha} \equiv -1 \pmod m$. If α/m has the continued fraction development*

$$\alpha/m = (\kappa_0; \kappa_1, \kappa_2, \ldots, \kappa_n),$$

then for some $r \leqslant n$ we have

$$m = N(\beta_r) + N(\gamma_{r+2}) = a_r^2 + b_r^2 + c_r^2 + d_r^2,$$

where

$$a_r + ib_r = [\kappa_1, \ldots, \kappa_r], \quad c_r + id_r = [\kappa_{r+2}, \ldots, \kappa_n] \qquad (r = 0, 1, \ldots, n).$$

Thus a representation of m as the sum of four squares is obtained by performing the Euclidean algorithm for α/m, calculating β_r at each step, until an r is reached for which $N(\beta_r) + N(\gamma_{r+2}) = m$.

If all prime divisors of m are of the form $4k+1$ then α and the κ_r, γ_r may be taken as rational integers and so the construction reduces essentially to that of Hermite [4].

Proof. For any r, $N(\beta_r) + N(\gamma_{r+2}) = a_r^2 + b_r^2 + c_r^2 + d_r^2$ is a multiple of m. We must show that for some r the multiple of m in question is m itself. It is clear that $N(\beta_r) + N(\gamma_{r+2}) > 0$, since if $N(\gamma_{r+2}) = 0$ then $r = n$ and so $N(\beta_r) = N(\beta_n) = m^2 > 0$.

Auric [1] showed that

$$|\gamma_{r+2}/\gamma_r| \leqslant 1/3 \qquad (r=1, ..., n), \tag{4}$$

and that

$$|\beta_r \gamma_{r+1}| \leqslant 3m/2 \qquad (r=0, 1, ..., n). \tag{5}$$

If $|\gamma_0 \gamma_1| < 3m/2$ then it is easy to see that $m \leqslant 3$ and $N(\gamma_0) + N(\gamma_2) = m$. In the contrary case, the $|\gamma_r \gamma_{r+1}|$ decrease monotonically from $|\gamma_0 \gamma_1| \geqslant 3m/2$ to zero. Hence we may choose r so that

$$|\gamma_{r+1} \gamma_{r+2}| < 3m/2 \leqslant |\gamma_r \gamma_{r+1}|. \tag{6}$$

For each r let $N_r = N(\beta_r) + N(\gamma_{r+2})$; then $N_r = k_r m$ for some positive integer k_r. By Euler's identity and (2) we have

$$\begin{aligned} N_{r-1}N_r &= N(\beta_{r-1}\bar{\beta}_r - \gamma_{r+1}\bar{\gamma}_{r+2}) + N(\beta_r\gamma_{r+1} + \beta_{r-1}\gamma_{r+2}) \\ &= |\beta_{r-1}\bar{\beta}_r - \gamma_{r+1}\bar{\gamma}_{r+2}|^2 + m^2. \end{aligned}$$

Now

$$|\beta_{r-1}\bar{\beta}_r| = \frac{|\beta_{r-1}\gamma_r|\,|\beta_r\gamma_{r+1}|}{|\gamma_r\gamma_{r+1}|} \leqslant 3m/2,$$

by (5) and (6), and

$$|\gamma_{r+1}\bar{\gamma}_{r+2}| < 3m/2,$$

by (6). If $3m/2 \leqslant |\gamma_{r+1}|^2$ then

$$|\beta_{r-1}\bar{\beta}_r| = \frac{|\beta_{r-1}\gamma_r|\,|\beta_r\gamma_{r+1}|}{|\gamma_{r+1}|^2}\left|\frac{\gamma_{r+1}}{\gamma_r}\right| \leqslant 3m/(2\sqrt{2}),$$

by (5). If $|\gamma_{r+1}|^2 < 3m/2$ then

$$|\gamma_{r+1}\bar{\gamma}_{r+2}| = |\gamma_{r+1}|^2\left|\frac{\gamma_{r+2}}{\gamma_{r+1}}\right| \leqslant 3m/(2\sqrt{2}).$$

Also

$$|\beta_{r-1}\bar{\beta}_r\gamma_{r+1}\bar{\gamma}_{r+2}| = |\beta_{r-1}\gamma_r|\,|\beta_r\gamma_{r+1}|\left|\frac{\gamma_{r+2}}{\gamma_r}\right| \leqslant 3m^2/4,$$

by (4) and (5).

Hence we have

$$\begin{aligned} N_{r-1}N_r &\leqslant |\beta_{r-1}\bar{\beta}_r|^2 + |\gamma_{r+1}\bar{\gamma}_{r+2}|^2 + 2|\beta_{r-1}\bar{\beta}_r\gamma_{r+1}\bar{\gamma}_{r+2}| + m^2 \\ &\leqslant 9m^2/8 + 9m^2/4 + 3m^2/2 + m^2 < 6m^2. \end{aligned}$$

It follows that $k_{r-1}k_r = 1, 2, 3, 4$, or 5. But if $k_{r-1}k_r = 4$ then $3m^2 = |\beta_{r-1}\bar{\beta}_r - \gamma_{r+1}\bar{\gamma}_{r+2}|^2$, so that $3m^2$ is the sum of two rational integral squares, which is impossible. Hence $k_{r-1}k_r$ is a prime or unity and so one of the factors is unity, i.e., $N(\beta_{r-1}) + N(\gamma_{r+1}) = m$ or $N(\beta_r) + N(\gamma_{r+2}) = m$. This completes the proof.

The above argument shows that for any α (not necessarily satisfying (1)) there exists r such that

$$N(\beta_r) + N(\gamma_{r+2}) < \sqrt{\frac{47}{8}}\, m = 2 \cdot 4238\ldots\, m\,.$$

Computation shows that for $m \leqslant 200$ and arbitrary α there exists r such that

$$N(\beta_r) + N(\gamma_{r+2}) < 2m\,.$$

It would be of interest to know if the same is true for all m.

3. Among the other known constructions for the representation of a positive integer as the sum of four squares is that provided by the reduction theory of positive definite Hermitian forms ([5]). Equivalent to this is Smith's construction [6], based on representing the fraction m/α as a conjugate-symmetric continued fraction

$$m/\alpha = (\kappa_0; \kappa_1, \ldots, \kappa_n, \bar{\kappa}_n, \ldots, \bar{\kappa}_1, \bar{\kappa}_0)\,, \tag{7}$$

which gives, in view of (2),

$$m = N([\kappa_0, \ldots, \kappa_n]) + N([\kappa_0, \ldots, \kappa_{n-1}])\,.$$

Another construction is known from the theory of integral quaternions. Namely, if (1) holds then the (right) g.c.d. $q = (m, \alpha + j)$ exists and $m = N(q)$ (here we may interpret "integral" either in the sense of Lipschitz or of Hurwitz; in either case q may be taken as having rational integral coefficients). This may be seen from Smith's construction; thus, with the notation as in (7), set

$$q_r = [\kappa_r, \ldots, \kappa_n, \bar{\kappa}_n, \ldots, \bar{\kappa}_0] + (-1)^{r-1}\,[\kappa_0, \ldots, \kappa_{r-2}]j\,;$$

then we have

$$q_0 = m, \quad q_1 = \alpha + j\,,$$

$$q_r = \kappa_r q_{r+1} + q_{r+2} \quad (r = 0, 1, \ldots, n-1)\,,$$

$$q_n = (\kappa_n + (-1)^{n+1} j)\, q_{n+1}\,,$$

from which

$$(m, \alpha + j) = q_{n+1} = [\bar{\kappa}_n, \ldots, \bar{\kappa}_0] + (-1)^n\,[\kappa_0, \ldots, \kappa_{n-1}]\,j\,;$$

i.e.,

$$N(q_{n+1}) = N([\kappa_0, \ldots, \kappa_n]) + N([\kappa_0, \ldots, \kappa_{n-1}]) = m\,.$$

If only the existence of a representation is required, rather than an explicit construction, then, as is the case in other contexts, the use of continued fractions may be replaced by the use of Dirichlet's pigeonhole principle or results in the geometry of numbers which follow from it (cf. Brauer and Reynolds [2], Davenport [3]). Indeed the existence of a solution of $\alpha\xi \equiv \eta \pmod{m}$ with $0 < N(\xi) + N(\eta) < 4m$ is a consequence of Minkowski's theorem on linear forms, while the existence of a solution with $0 < N(\xi) + N(\eta) < 2m$ follows from Minkowski's theorem on convex bodies.

The author is indebted to L. Rousseau for assistance with computational work in relation to this paper.

REFERENCES

[1] AURIC, A. Essai sur la théorie des fractions continues. *J. Math. pures et appl. (5), 8* (1902), 387-431.
[2] BRAUER, A. and T. L. REYNOLDS. On a Theorem of Aubry-Thue. *Canadian J. Math. 3* (1951), 367-374.
[3] DAVENPORT, H. The Geometry of Numbers. *Math. Gazette 31* (1947), 206-210.
[4] HERMITE, C. Note sur un théorème relatif aux nombres entiers. *Œuvres I*, Gauthier-Villars, Paris (1905), 264.
[5] —— Sur la théorie des formes quadratiques, 2. *Œuvres I*, Gauthier-Villars, Paris (1905), 234-263.
[6] SMITH, H. J. S. De fractionibus quibusdam continuis. *Collected Math. Papers, II*, O.U.P., Oxford (1894), 287-309.

(Reçu le 10 avril 1987)

G. Rousseau

Mathematics Department
The University.
Leicester LE1 7RH
(England)