MATHEMATICS OF COMPUTATION
Volume 88, Number 316, March 2019, Pages 1007–1019
http://dx.doi.org/10.1090/mcom/3349
Article electronically published on May 29, 2018

DIRICHLET'S PROOF OF THE THREE-SQUARE THEOREM: AN ALGORITHMIC PERSPECTIVE

PAUL POLLACK AND PETER SCHORN

Abstract. The *Gauss–Legendre three-square theorem* asserts that the positive integers n expressible as a sum of three integer squares are precisely those not of the form $4^k(8m+7)$ for any nonnegative integers k, m. In 1850, Dirichlet gave a beautifully simple proof of this result using only basic facts about ternary quadratic forms. We explain how to turn Dirichlet's proof into an algorithm; if one assumes the Extended Riemann Hypothesis (ERH), there is a random algorithm for expressing $n = x^2+y^2+z^2$, where the expected number of bit operations is $O((\lg n)^2(\lg\lg n)^{-1} \cdot M(\lg n))$. Here, $M(r)$ stands in for the bit complexity of multiplying two r-bit integers. A random algorithm for this problem of similar complexity was proposed by Rabin and Shallit in 1986; however, their analysis depended on both the ERH and on certain conjectures of Hardy–Littlewood type.

1. Introduction

Given a positive integer n, how can we recognize if n is a sum of three integer squares, i.e., of the form $x^2+y^2+z^2$ with $x, y, z \in \mathbb{Z}$? And if we decide that it is representable, how can we efficiently find the integers x, y, and z?

The first question has a classical answer. According to the *Gauss–Legendre three-square theorem*, n is a sum of three squares precisely when n is *not* of the form $4^k(8m+7)$ for any nonnegative integers k, m. Attention to the second question seems to be more recent. In 1986, Rabin and Shallit [RS86] proposed a random algorithm for writing n as a sum of three squares. They had noted already that when p is a prime congruent to 1 mod 4, it is easy to find representations of p and $2p$ in the form x^2+y^2 (see §1 of [RS86]). Thus, if one can locate a z for which $n-z^2$ has the form p or $2p$, one can quickly derive a representation $n = x^2+y^2+z^2$. Carefully elaborated on, this idea leads to an algorithm where the expected number of bit operations is $O((\lg n)^2(\lg\lg n)\cdot M(\lg n))$. Here $\lg n$ denotes the number of bits required to represent n, while $M(r)$ represents the bit complexity of multiplying two r-bit numbers.

There is one catch, however. The algorithm of [RS86] depends on unproved number-theoretic conjectures; these are needed both to guarantee that the algorithm eventually succeeds and to carry out the runtime analysis. One assumption Rabin and Shallit employ is that for all large enough $n \equiv 2 \pmod 4$, the proportion of $x \in [1, \sqrt{n}]$ with $n - x^2$ prime is $\gg (\log n \cdot \log\log n)^{-1}$. This estimate would follow by combining two well-believed hypotheses in number theory: Hardy

Received by the editor July 16, 2017, and, in revised form, July 18, 2017, and December 4, 2017.

2010 *Mathematics Subject Classification.* Primary 11E25; Secondary 11Y50.

The research of the first-named author was supported by NSF award DMS-1402268.

and Littlewood's Conjecture H from [HL23] and the Extended Riemann Hypothesis (the Riemann Hypothesis for Dirichlet L-functions).* Sadly, a proof of either hypothesis appears far out of reach, and unconditionally we cannot even prove that all large $n \equiv 2 \pmod 4$ admit at least one expression in the form $x^2 + p$.

In this note, we lay out an alternative method for representing n as a sum of three squares rooted in Dirichlet's 1850 proof [Dir50] of the three-square theorem. Assuming ERH (but *not* any conjectures of Hardy–Littlewood type), we arrive at a random algorithm with similar theoretical complexity to the Rabin–Shallit algorithm.

Our new algorithm is primarily of theoretical interest. In practice, the fastest algorithm we are aware of for representing integers as sums of three squares remains the method of Rabin and Shallit (actually a particular derandomized variant of their method). See §5.2 below for further discussion.

2. Dirichlet's proof

In this section, we sketch Dirichlet's proof of the sufficiency half of the three-square theorem, in a guise convenient for later reference in §3.

2.1. **Preliminaries on ternary quadratic forms.** By a *ternary quadratic form* F, we mean a polynomial $F(X, Y, Z)$ having the shape

$$F(X,Y,Z) = a_{11}X^2 + a_{22}Y^2 + a_{33}Z^2 + 2a_{12}XY + 2a_{13}XZ + 2a_{23}YZ, \tag{1}$$

where $a_{11}, a_{22}, a_{33}, a_{12}, a_{13}, a_{23}$ are integers. The *matrix M_F corresponding to F* is the unique 3×3 symmetric matrix for which

$$F\left(\left[\begin{smallmatrix} X \\ Y \\ Z \end{smallmatrix}\right]\right) = [X,Y,Z] M_F \left[\begin{smallmatrix} X \\ Y \\ Z \end{smallmatrix}\right].$$

(We identify $F(X, Y, Z)$ with $F(\left[\begin{smallmatrix} X \\ Y \\ Z \end{smallmatrix}\right])$.) Explicitly,

$$M_F = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}.$$

If F and G are two ternary quadratic forms, we say that F and G are *equivalent* if there is an $A \in \mathrm{SL}_3(\mathbb{Z})$ with

$$M_G = A^t M_F A. \tag{2}$$

Said differently, F and G are equivalent if

$$G\left(\left[\begin{smallmatrix} X \\ Y \\ Z \end{smallmatrix}\right]\right) = F\left(A \left[\begin{smallmatrix} X \\ Y \\ Z \end{smallmatrix}\right]\right) \tag{3}$$

for some $A \in \mathrm{SL}_3(\mathbb{Z})$. We say an integer n *is represented by F* if $F(x, y, z) = n$ for some $x, y, z \in \mathbb{Z}$. It follows from (3) that if F and G are equivalent, then the set of integers represented by F coincides with the set of integers represented by G. We define the *determinant* Δ_F of F as the determinant of the associated matrix M_F; it is clear from (2) that equivalent forms have the same determinant.

*The statement of Conjecture H involves an Euler product factor of size $\asymp L(1, (\frac{4n}{\cdot}))^{-1}$. Under ERH, $L(1, (\frac{4n}{\cdot})) \ll \log\log n$; this explains the "$\log\log n$" appearing above in the conjectured lower bound on the proportion of x values. In place of this ERH-conditional bound on $L(1, (\frac{4n}{\cdot}))$, one could substitute the elementary estimate $L(1, (\frac{4n}{\cdot})) \ll \log n$. This would imply that the expected number of bit operations in the Rabin–Shallit algorithm is $O((\lg n)^3 M(\lg n))$, conditional on Conjecture H alone.

A ternary quadratic form F is said to be *positive-definite* if $F(x, y, z) > 0$ for all integers x, y, z not all zero. It is easy to see from (3) that every form equivalent to a positive-definite form is positive-definite. It can be shown [Lan58, Theorem 182, pp. 156–157] that if F is written as in (1), then F is positive-definite if and only if

$$a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} > 0, \quad \text{and} \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} > 0.$$

Note that the final determinant here is precisely the definition of Δ_F.

Finally, we need the notion of a *quasi-reduced form*. For a ternary quadratic form F, the adjoint F^* is the ternary quadratic form associated to $-\operatorname{adj}(M_F)$, where adj is the adjugate matrix. We write

$$M_{F^*} = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{12} & A_{22} & A_{23} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}.$$

A positive-definite ternary quadratic form F is called *quasi-reduced* if

(i) $|a_{11}| < 2\Delta_F^{1/3}$,
(ii) $|A_{33}| < 2\Delta_F^{2/3}$,
(iii) $|a_{12}| \le \frac{1}{2}a_{11}$,
(iv) $|A_{13}| \le \frac{1}{2}|A_{33}|$,
(v) $|A_{23}| \le \frac{1}{2}|A_{33}|$.

(Lagarias defines what it means to be quasi-reduced for all ternary quadratic forms, not only the positive-definite ones, but we will not need the more general notion.) This is a weakening of the notion of "reduced form" discussed by Gauss in Section V of the *Disquisitiones* [Gau86]. Gauss proves in Art. 272–275 that every form is equivalent to a (not-necessarily-unique!) reduced form. *A fortiori*, each positive-definite ternary quadratic form is equivalent to a quasi-reduced form.

While Gauss gives an explicit algorithm for reduction in [Gau86], it is not obvious that his algorithm has polynomial running time. This was Lagarias's motivation for introducing quasi-reduced forms; he shows in [Lag80] that quasi-reduction *can* be done in polynomial time, by a variant of Gauss's procedure.

We conclude this section with a lemma that ties all of these notions back to sums of three squares.

Lemma 1. *The unique positive-definite quasi-reduced ternary form F of determinant 1 is $X^2 + Y^2 + Z^2$.*

Proof. Since $a_{11} > 0$ (as F is positive-definite), condition (i) in the definition of a quasi-reduced form shows that $a_{11} = 1$. Condition (iii) then forces $a_{12} = 0$. Hence,

$$-A_{33} = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} = a_{22}.$$

Since F is positive-definite, $-A_{33} > 0$. So (ii) forces $a_{22} = 1$, and $A_{33} = -1$. Now (iv) and (v) imply that $A_{13} = 0$ and $A_{23} = 0$. But $A_{13} = a_{22}a_{13} - a_{12}a_{23} = a_{13}$, while $A_{23} = a_{11}a_{23} - a_{12}a_{13} = a_{23}$. The only entry of M_F not yet determined is a_{33}; but $\det(M_F) = 1$ forces $a_{33} = 1$. Hence, $a_{11} = a_{22} = a_{33} = 1$, while $a_{12} = a_{13} = a_{23} = 0$. Thus, $F = X^2 + Y^2 + Z^2$. □

2.2. **Dirichlet's proof of sufficiency in the three-square theorem.** Let n be a positive integer not of the form $4^k(8m+7)$. Write $n = 4^k n_0$, where $4 \nmid n_0$. Thus, $n_0 \equiv 1, 2, 3, 5, \text{ or } 6 \pmod 8$. It suffices to find a representation of n_0 as $x^2+y^2+z^2$, for then $n = (2^k x)^2 + (2^k y)^2 + (2^k z)^2$. In other words, we can (and will) assume that $n \equiv 1, 2, 3, 5, \text{ or } 6 \pmod 8$.

Our goal is to construct a positive-definite ternary quadratic form F of determinant 1 which represents n. After quasi-reducing F, we see (Lemma 1) that F is equivalent to $X^2 + Y^2 + Z^2$. Since equivalent forms represent the same integers, n is a sum of three squares, as desired.

We search for $F(X, Y, Z)$ among ternary forms having the shape

$$F(X, Y, Z) = a_{11}X^2 + a_{22}Y^2 + nZ^2 + 2a_{12}XY + 2XZ,$$

so that

$$M_F = \begin{bmatrix} a_{11} & a_{12} & 1 \\ a_{12} & a_{22} & 0 \\ 1 & 0 & n \end{bmatrix}.$$

For each of these forms, $F(0, 0, 1) = n$, so that F represents n. Write

$$d := a_{11}a_{22} - a_{12}^2.$$

Expanding $\det(M_F)$ by minors along the third column, we see that

$$\Delta_F = \det(M_F) = dn - a_{22}.$$

Keeping in mind the conditions on a form to be positive-definite, accomplishing our goal reduces to finding integers a_{11}, a_{12}, a_{22} satisfying

$$a_{11} > 0, \quad d = a_{11}a_{22} - a_{12}^2 > 0, \quad \text{and} \quad a_{22} = dn - 1. \tag{4}$$

The conditions (4) imply that $d > 0$ and that $-d$ is a square modulo $dn - 1$. Conversely, if D is any positive integer for which $-D$ is a square modulo $Dn - 1$, we can put $a_{22} = Dn - 1$, choose a_{12} with $a_{12}^2 \equiv -D \pmod{Dn-1}$, and define a_{11} so that

$$a_{11}a_{22} = D + a_{12}^2.$$

Then (4) holds with $d = D$. This is clear except possibly for the first condition in (4). To see that inequality, note that since the right-hand side of the last display is positive, a_{11} and a_{22} are nonzero and of the same sign. But $a_{22} = Dn - 1 \geq 0$, and hence $a_{11}, a_{22} > 0$.

Thus, the proof will be complete if we can show that there is some $D > 0$ with $-D$ congruent to a square modulo $Dn - 1$. Dirichlet deduces this from his theorem on primes in an arithmetic progression. There are several cases to consider according to the residue class of n modulo 8.

Case I (n is even; so $n \equiv 2 \text{ or } 6 \pmod 8$). We choose an auxiliary prime

$$p \equiv 2n - 1 \pmod{4n}.$$

Such a prime p has the form $Dn - 1$ for some $D \equiv 2 \pmod 4$; moreover, $p \equiv 3 \pmod 8$. By quadratic reciprocity and the usual supplementary laws, we get

$$\begin{aligned} \left(\frac{-D}{p}\right) &= \left(\frac{-2}{p}\right)\left(\frac{D/2}{p}\right) = \left(\frac{D/2}{p}\right) = \left(\frac{p}{D/2}\right)(-1)^{(\frac{D}{2}-1)/2} \\ &= \left(\frac{-1}{D/2}\right)(-1)^{(\frac{D}{2}-1)/2} = 1. \end{aligned}$$

Thus, $-D$ is a square modulo $p = Dn - 1$.

Case II ($n \equiv 1 \pmod 8$). Choose

$$p \equiv 6n - 1 \pmod{8n}.$$

Then $p = Dn - 1$ for some $D \equiv 6 \pmod 8$. Moreover, $p \equiv 5 \pmod 8$. Thus,

$$\left(\frac{-D}{p}\right) = \left(\frac{-2}{p}\right)\left(\frac{D/2}{p}\right) = -\left(\frac{D/2}{p}\right) = -\left(\frac{p}{D/2}\right) = -\left(\frac{-1}{D/2}\right) = 1,$$

using in the last step that $D/2 \equiv 3 \pmod 4$. Again, $-D$ is a square modulo $p = Dn - 1$.

Case III ($n \equiv 3 \pmod 8$). Choose

$$p \equiv \frac{5n-1}{2} \pmod{4n}.$$

Then $2p = Dn - 1$ for some $D \equiv 5 \pmod 8$. Moreover, $p \equiv 3 \pmod 4$. Thus,

$$\left(\frac{-D}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{D}{p}\right) = -\left(\frac{D}{p}\right) = -\left(\frac{p}{D}\right) = -\left(\frac{2p}{D}\right)\left(\frac{2}{D}\right) = -\left(\frac{-1}{D}\right)\left(\frac{2}{D}\right) = 1.$$

Hence, $-D$ is a square modulo $p = \frac{Dn-1}{2}$. Obviously, $-D \equiv (-D)^2 \pmod 2$, and thus $-D$ is also a square modulo $2p = Dn - 1$.

Case IV ($n \equiv 5 \pmod 8$). Choose

$$p \equiv \frac{3n-1}{2} \pmod{4n}.$$

Then $2p = Dn - 1$ for some $D \equiv 3 \pmod 8$. As in the last case, $p \equiv 3 \pmod 4$. Therefore,

$$\left(\frac{-D}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{D}{p}\right) = -\left(\frac{D}{p}\right) = \left(\frac{p}{D}\right) = \left(\frac{2p}{D}\right)\left(\frac{2}{D}\right) = \left(\frac{-1}{D}\right)\left(\frac{2}{D}\right) = 1.$$

As in the previous case, we conclude that $-D$ is a square modulo $2p = Dn - 1$.

The above method of selecting D is slightly different from the original choices of Dirichlet. Our modifications ensure that the auxiliary prime p always lands in either the residue class 3 mod 4 or 5 mod 8; this will be convenient later.

3. An efficient random algorithm under ERH

We now extract from Dirichlet's argument a procedure for writing a given n as a sum of three squares. It is enough to treat $n \equiv 1, 2, 3, 5,$ or $6 \pmod 8$. Then the basic steps are as follows.

(1) Find an auxiliary prime p from the appropriate arithmetic progression.
(2) With D as above, compute a square root a_{12} of $-D$ modulo $Dn - 1$.
(3) With $a_{22} = Dn - 1$ and $a_{11} = (D + a_{12}^2)/a_{22}$, we know that the form $F = a_{11}X^2 + a_{22}Y^2 + nZ^2 + 2a_{12}XY + 2XZ$ is equivalent to $F_0 = X^2 + Y^2 + Z^2$. Using Lagarias's algorithm for quasi-reduction, explicity compute a matrix $A \in \mathrm{SL}_3(\mathbb{Z})$ with

$$M_{F_0} = A^t M_F A.$$

(4) Output the third column $\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]$ of A^{-1}. Since

$$x^2 + y^2 + z^2 = F_0\left(\left[\begin{smallmatrix} x \\ y \\ z \end{smallmatrix}\right]\right) = F_0\left(A^{-1}\left[\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right]\right) = F\left(\left[\begin{smallmatrix} 0 \\ 0 \\ 1 \end{smallmatrix}\right]\right) = n,$$

we have found a representation of n as a sum of three squares.

We now discuss how to carry out steps (1)–(4) efficiently, working under the assumption of the ERH.

In what follows, the reader can take $M(r)$ as any one of the three functions r^2, $r \cdot \lg r \cdot \lg\lg r$, or $r \cdot \lg r \cdot 2^{3\lg^* r}$. These represent the bit-complexity of r-bit multiplication from the naive perspective, from Schönhage–Strassen, and from Fürer–Harvey, van der Hoeven–Lecerf, respectively. See [BZ11, §1.3 and §2.3.3] for a discussion of multiplication methods; the more recent Fürer–Harvey–van der Hoeven–Lecerf work is described in [Für07, Für09], [HHL16].

3.1. **Steps (1) and (2).** Let a, q be coprime integers with $q \geq 2$ and $0 < a < q$. Under ERH, one can show that for all real $x \geq 2q^3$, there are $\gg x/(\varphi(q)\log x)$ primes $p \equiv a \pmod q$ with $p \leq x$ (see, e.g., §3 of [PTn]). Taking $x = 2q^3$, we see that if we choose a nonnegative integer $k < 2q^2$ at random, $p = kq + a$ will be prime with probability

$$\gg \frac{q}{\varphi(q)} \frac{1}{\log(2q^3)} \gg \frac{q}{\varphi(q)} \frac{1}{\lg q},$$

and so we expect to stumble across a prime within $O(\frac{\varphi(q)}{q} \lg q) = O(\lg q)$ iterations of this process.

Returning to our context and taking $q = 4n$ or $8n$ (according to whether we are in Cases I, III, IV or Case II), we expect our auxiliary prime p to be located within $O(\lg n)$ random trials.

Suppose that we have found p. Recall that $p \equiv 3 \pmod 4$ or $p \equiv 5 \pmod 8$. If $p \equiv 3 \pmod 4$, then a square root of $-D$ modulo p is given by $(-D)^{(p+1)/4}$ modulo p. On the other hand, if $p \equiv 5 \pmod 8$, then $(-D)^{(p+3)/4} \equiv \pm D \pmod p$, and a square root of $-D$ modulo p is given by either

$$(-D)^{(p+3)/8} \mod p, \quad \text{or} \quad 2^{(p-1)/4}(-D)^{(p+3)/8} \mod p.$$

(These facts follow from Euler's criterion for quadratic residuacity; cf. [BS96, Exercise 7.9.1, p. 188].) In all cases, a square root of $-D$ modulo p can be computed in time $O((\lg p)M(\lg p)) = O((\lg n)M(\lg n))$. If $p = Dn - 1$, we have our desired square root of $-D$ modulo $Dn - 1$. If $p = \frac{Dn-1}{2}$, we keep the root if its parity agrees with that of $-D$, otherwise we add p to it; the result is a square root of $-D$ modulo $Dn - 1$.

We have left out an important detail. When searching for primes, we expect to find one in $O(\lg n)$ steps, but we have not said anything about how to recognize when we have found one. To avoid a (relatively) expensive deterministic test for primality, we combine steps (1) and (2) as follows. Let $q = 4n, 8n, 4n$, or $4n$, and $a = 2n - 1, 6n - 1, \frac{5n-1}{2}$, or $\frac{3n-1}{2}$, according to whether we are in Case I, II, III, or IV, respectively.

(1′) For a random nonnegative integer $k < 2q^2$, let $p = qk + a$.

(2′) Attempt to find a square root of $-D$ modulo p as if p were prime. That is:

If $p \equiv 3 \pmod 4$, compute the least remainder R of $(-D)^{(p+1)/4}$ modulo p and test if $R^2 \equiv -D \pmod p$.

If $p \equiv 5 \pmod 8$, compute the least remainder R of $(-D)^{(p+3)/8}$ modulo p and test if $R^2 \equiv \pm D \bmod p$. If not, p is composite, and we go back to $(1')$. If $R^2 \equiv +D \bmod p$, replace R by $R \cdot 2^{(p-1)/4}$, reduced mod p, and check if now $R^2 \equiv -D \pmod p$.

If we fail to find a square root of $-D$ modulo p in this way, then p is not prime, and we return to step $(1')$. Otherwise, we let $a_{12} = R$. If we are in cases III or IV, we replace a_{12} with $a_{12} + p$ if $a_{12} \not\equiv -D \pmod 2$.

Under ERH, $(1')$ results in a prime p with probability $\gg 1/\lg n$. Moreover, when p is prime, $(2')$ is guaranteed to find a square root of $-D$ modulo $Dn-1$. Thus, we expect to run through steps $(1')$ and $(2')$ only $O(\lg n)$ times. The expected number of bit operations for this part of the algorithm is therefore $O((\lg n)^2 M(\lg n))$.

Note that p may not be prime at the termination of $(1')$ and $(2')$. However, the primality of p is not needed for steps (3) and (4).

3.2. **Steps (3) and (4).** Here our analysis is based on the following result of Lagarias and Eisenbrand–Rote. For a matrix M, we let $\|M\|$ denote the largest of the absolute values of the entries of M. For a ternary quadratic form F, we put $\|F\| = \|M_F\|$.

Proposition 2. *Let F be a positive-definite ternary quadratic form. The modified Gauss reduction algorithm described on* [Lag80, p. 165] *produces a quasi-reduced form F_{qred} equivalent to F and a matrix $A \in \mathrm{SL}_3(\mathbb{Z})$ with*

$$M_{F_{\mathrm{qred}}} = A^t M_F A.$$

The algorithm terminates in $O((\lg\|F\|)(\lg\lg\|F\|)\cdot M(\lg\|F\|))$ bit operations. Moreover,

$$\|A\| = O((\lg\|F\|)(\lg\lg\|F\|)).$$

Proof sketch. For positive-definite forms, this is a strengthened version of [Lag80, Theorem 4.9]. This modified result follows from Lagarias's analysis in [Lag80] together with the observation on [ER01, p. 6] that the number of times the algorithm cycles through Lagarias's steps 1, 2, and 3 is actually $O(\lg\lg\|F\|)$ and not simply $O(\lg\|F\|)$ as estimated in [Lag80]. □

It is easy to see that our method of carrying out steps (1) and (2) results in a form F with $\lg\|F\| = O(\lg n)$. So by Proposition 2, we can find F_0 and A — completing step (3) — in $O((\lg n)(\lg\lg n)\cdot M(\lg n))$ bit operations.

Once we have A, it is easy to compute the third-column entries x, y, and z of A^{-1}. Indeed, since A is known to have determinant 1, each of x, y, z is given by a 2×2 minor of A (up to sign). Using the bound on $\|A\|$ in Proposition 2, we see that computing X, Y, Z requires $O(M(\lg n \lg\lg n))$ operations. This is asymptotically negligible compared to the work required in step (3). Thus, the number of bit operations required for steps (3) and (4) is $O((\lg n)(\lg\lg n)\cdot M(\lg n))$.

3.3. **Wrapping up.** Assembling our results, we have shown the following.

Theorem 3 (assuming ERH)**.** *There is a random algorithm for writing any $n \neq 4^k(8m+7)$ as a sum of three squares for which the expected number of bit operations is $O((\lg n)^2 M(\lg n))$.*

Notice that our invocation of ERH functions similarly to that of [RS86]; ERH is used both to show that the algorithm eventually terminates and to estimate its expected running time.

A slightly better runtime can be gained at the cost of mildly complicating the algorithm. The idea is to randomly generate p in a way that ensures that it is free of small prime factors, thus giving it a "leg up" in terms of its odds of being prime.

To this end, we modify the algorithm to begin with a precomputaton. We compute the list of primes $\ell \le \log n$ and record, for each of these ℓ, whether or not ℓ divides q. (We are continuing to use a and q with the same meanings as in §3.1.) We then compute

$$L := \prod_{\substack{\ell \le \log n \\ \ell \nmid q}} \ell.$$

This can all be carried out in $O((\lg n)^{3/2} M(\lg n))$ bit operations. Note that estimates from prime number theory imply that $L < q^{1.1}$ (see, for instance, §8.8 of [BS96]).

We replace Step (1′) with the following procedure:

(1″) For each prime $\ell \le \log n$ not dividing q, choose $a_\ell \in \{1, 2, \dots, \ell - 1\}$ uniformly at random. Let a_0 be the unique integer in $(0, qL]$ satisfying the simultaneous congruences

$$a_0 \equiv a \pmod{q} \quad \text{and} \quad a_0 \equiv a_\ell \pmod{\ell} \quad \text{for all } \ell \mid L.$$

With $K = \lceil 2q^2/L \rceil$, put

$$p = a_0 + qLk,$$

where the integer k is selected uniformly at random from $[0, K)$.

Selecting a_ℓ, computing the resulting integer a_0, and selecting p can all be done in $O((\lg n)^2)$ bit operations (one reference for the complexity of the Chinese Remainder Theorem calculation is [BS96, Corollary 5.5.3, p. 105]).

The upshot of (1″) is that p is chosen uniformly at random from the set

$$\mathcal{S} := \{u \le qLK : u \equiv a \pmod{q},\ u \text{ has no prime factors} \le \log n\}.$$

We have

$$\#\mathcal{S} = LK \prod_{p \mid L} (1 - 1/p) = K\varphi(L). \tag{5}$$

Moreover, each prime in $[2, qLK]$ that is congruent to $a \pmod{q}$ is an element of $\mathcal{S}$. (Each prime congruent to a modulo q exceeds $\log n$, since $a \ge \frac{3n-1}{2} \ge \log n$.) As $qLK \ge 2q^3$, ERH implies that there are

$$\gg \frac{qLK}{\varphi(q) \log(qLK)} \gg \frac{qLK}{\varphi(q) \lg n} \tag{6}$$

primes up to qLK that are congruent to $a \bmod q$. Comparing (5) and (6), the probability that a randomly chosen $p \in \mathcal{S}$ is prime is

$$\gg \frac{1}{\lg n} \frac{qL}{\varphi(q)\varphi(L)} = \frac{1}{\lg n} \prod_{\ell \mid qL} (1 - 1/\ell)^{-1} \ge \frac{1}{\lg n} \prod_{\ell \le \log n} (1 - 1/\ell)^{-1} \gg \frac{\lg \lg n}{\lg n}.$$

Here the product over all primes $\ell \le \log n$ was estimated by Mertens' theorem; cf. [BS96, Theorem 8.8.6, p. 234]. So we expect to find a prime within $O(\lg n / \lg \lg n)$ random trials.

Thus, we expect to run through (1″) and (2′) only $O(\lg n/\lg\lg n)$ times. As noted above, each run through (1″) requires $O((\lg n)^2)$ operations. Since each iteration of (2′) requires $O(\lg n \cdot M(\lg n))$ operations, the total number of bit operations expected in the precomputation, (1′), and (2″) is

$$\ll (\lg n)^{3/2} M(\lg n) + \left(\frac{\lg n}{\lg\lg n}\right) \cdot (\lg n)^2 + \left(\frac{\lg n}{\lg\lg n}\right) \cdot (\lg n \cdot M(\lg n))$$
$$\ll (\lg n)^2 (\lg\lg n)^{-1} \cdot M(\lg n).$$

This dominates the running time estimate for Steps (3) and (4). We conclude that we may replace $(\lg n)^2 M(\lg n)$ in Theorem 3 with $(\lg n)^2(\lg\lg n)^{-1}M(\lg n)$.

Theorem 3′ (Assuming ERH)**.** *There is a random algorithm for writing any $n \neq 4^k(8m+7)$ as a sum of three squares for which the expected number of bit operations is $O((\lg n)^2(\lg\lg n)^{-1}M(\lg n))$.*

4. Life without ERH

If we are not believers in ERH, what can be shown? In that case, we are unable to prove the existence of a random algorithm for finding three-square representations that runs in expected polynomial time. We can, however, show that the problem is in a certain sense no more difficult than factoring, for which random algorithms that run in subexponential time are known (for instance, [LP92]).

Recall that to represent an $n \equiv 1, 2, 3, 5,$ or $6 \pmod 8$ as a sum of three squares, we constructed a positive-definite ternary form F of determinant 1 representing n. The construction depended on finding an auxiliary prime in an appropriate arithmetic progression, and ERH was used to show that one could expect to find such a prime reasonably quickly by random sampling. Alternatively, we can construct F by the following method of Gauss, as described in Flath [Fla89, Chapter 5]. For simplicity, we restrict attention to the cases when $n \equiv 1$ or $2 \pmod 4$. Gauss proves that there is a primitive, positive-definite binary quadratic form $f(x, y) = ax^2 + 2bxy + cy^2$ of discriminant $-4n$ with $\gcd(a, 4n) = 1$ and $\left(\frac{a}{p}\right) = \left(\frac{-1}{p}\right)$ for every odd prime p dividing n. It is then elementary to produce integers u, v, w for which

$$\begin{bmatrix} a & b & u \\ b & c & v \\ u & v & w \end{bmatrix}$$

has determinant 1; see [Fla89, Lemma 8.1]. In fact, the integers u, v, w are easily computed from a square root of $-a$ modulo n. (The Legendre symbol conditions ensure that a root exists.) The ternary quadratic form corresponding to the adjugate of the displayed matrix can be shown to be a positive-definite form F of determinant 1 with $F(0, 0, 1) = n$.

If we suppose that n is factored completely, then this can be made into an efficient algorithm for writing n as a sum of three squares. As explained in [Fla89], the Legendre symbol conditions correspond to placing f in a certain genus of the class group of primitive, positive-definite forms of discriminant $-4n$. Given the factorization of n, an algorithm of Lagarias/Bosma–Stevenhagen (see [Lag80b, §4], [BS96c]) can be used to produce such a form f (in fact, a reduced form f) in expected polynomial time. Since we know the factorization of n, we can also take

square roots modulo n in expected polynomial time (using, e.g., the algorithm of Tonelli–Shanks along with Hensel lifting and the Chinese Remainder Theorem). This allows us to construct F in expected polynomial time. Once we have F, we can find a three-square representation of n by reduction. We assumed here that $n \equiv 1, 2 \pmod 4$, but similar arguments work when $n \equiv 3 \pmod 8$ (cf. the paragraph starting at the bottom of p. 178 of [Fla89]). To summarize, we have sketched a proof of the following theorem.

Theorem 4. *There is an algorithm which, given the prime factorization of a positive integer n not of the form $4^k(8m+7)$, returns a representation of n as a sum of three squares in expected polynomial time.*

The theoretical situation here might be contrasted with the problem of four-square representations, studied in [RS86] and [PTn]. In that problem, one can prove unconditionally [PTn] that there is a random algorithm with expected running time $O((\lg n)^2(\lg\lg n)^{-1}M(\lg n))$, matching our ERH-conditional complexity bound for three-square representations. Moreover, in the four-square problem, given the factorization of n one can unconditionally produce a representation in deterministic polynomial time.

5. Concluding remarks

5.1. **Is randomness necessary?** We suspect that there is a *deterministic* polynomial time algorithm for representing integers as sums of three squares. Indeed, we believe this is true for the derandomized version of our algorithm where the auxiliary "prime", rather than being chosen at random from the integers in the appropriate progression $a \bmod q$, is chosen as the smallest integer $p \equiv a \bmod q$ with no small factors (say, no prime factors up to $\log n$) for which the procedure of $(2')$ for computing a square root of $-D$ modulo p succeeds. That this runs in polynomial time would follow from a conjecture of Heath-Brown in [HB78] that the least prime in each coprime residue class modulo q is $O(q(\lg q)^2)$. Heath-Brown's conjecture is supported by heuristic reasoning (see [BH93] and [PLS17]) but seems to lie very deep; even on ERH, the best we can show is that these primes are $O((\varphi(q)\lg q)^2)$. (See [BS96b] and [LLS15] for numerically explicit versions of the ERH-conditional estimate.)

5.2. **From theory to practice.** While the emphasis in this paper has been theoretical, the algorithms are also computer-practical, especially the deterministic variant just described. An implementation of ours in `PARI/GP` [PARI] takes about 300 milliseconds on average to represent numbers with ≈ 500 decimal digits.

While this may seem quite speedy, in our experiments a derandomized version of the Rabin–Shallit algorithm exhibits better performance. Recall that the Rabin–Shallit algorithm looks for primes 1 mod 4, or the double of such, having the form $n - x^2$ with x chosen *randomly* from $[0, \sqrt{n}]$. (We assume here that we have already reduced to the cases $n \equiv 1, 2, 3, 5, 6 \pmod 8$.) In the derandomization we have in mind, we instead start by taking x as large as possible, meaning the largest integer not exceeding $\lfloor\sqrt{n}\rfloor$ having feasible parity, and successively decrease x until a prime is found. We have implemented a version of this in `PARI/GP`, with the task of primality testing handled by the built-in function `ispseudoprime` and the task

of representing primes 1 mod 4 as sums of two squares handled by `qfbsolve`. It appears to average a mere 30 ms (roughly) per 500 digit input.†

5.3. **Beyond $X^2+Y^2+Z^2$.** There are a handful of other positive-definite ternary quadratic forms for which Dirichlet's method has been used to classify all representable integers. For instance, in 1927 Dickson [Dic27b] treated the six forms

$$X^2+Y^2+2Z^2, \quad X^2+Y^2+3Z^2, \quad X^2+2Y^2+2Z^2,$$
$$X^2+2Y^2+3Z^2, \quad X^2+2Y^2+4Z^2, \quad X^2+2Y^2+5Z^2.$$

These forms appeared in earlier work of Ramanujan, who in each case correctly conjectured the set of represented integers, but did not present proofs. For further examples, see [Dic27c, Dic27a] and cf. [BDTT16]. Because of the similarity of the proofs to Dirichlet's, the methods of this paper could be adapted to show that the analogue of Theorem 3 (and 3′) holds for any of these forms replacing $X^2+Y^2+Z^2$, although each individual form requires an ad hoc argument. (One extra difficulty is that not all of these forms are the unique quasi-reduced forms in their equivalence class, but this could be overcome by employing the theory of Eisenstein-reduced forms.)

A theorem of Jones [Jon31] asserts that every positive integer *locally representable* by a positive-definite ternary quadratic form F, meaning representable over the p-adic integers $\mathbb{Z}_p$ for every prime p, is represented over $\mathbb{Z}$ by some form in the same genus of F. From a high level point of view, the success of Dirichlet's method for the forms listed above depends on the fact that all of these forms are alone in their genus (i.e., the forms have class number 1). This suggests the following attractive problem, whose solution would allow one to unify the arguments alluded to above.

Problem. Fix a positive-definite ternary quadratic form F. Give an efficient algorithm which, for any positive integer n locally represented by F, outputs a form $\tilde{F}$ in the same genus as F along with a triple of integers x, y, z satisfying $\tilde{F}(x,y,z) = n$.

Acknowledgments

We thank Pete L. Clark and Enrique Treviño for helpful conversations, and we thank the referee for several suggestions that improved the exposition.

References

[BDTT16] S. Blackwell, G. Durham, K. Thompson, and T. Treece, *A generalization of a method of Mordell to ternary quadratic forms*, Int. J. Number Theory **12** (2016), no. 8, 2081–2105. MR3562015

[BH93] E. Bach and L. Huelsbergen, *Statistical evidence for small generating sets*, Math. Comp. **61** (1993), no. 203, 69–82. MR1195432

[BS96] E. Bach and J. Shallit, *Algorithmic Number Theory. Vol. 1*, Efficient Algorithms, Foundations of Computing Series, MIT Press, Cambridge, MA, 1996. MR1406794

[BS96b] E. Bach and J. Sorenson, *Explicit bounds for primes in residue classes*, Math. Comp. **65** (1996), no. 216, 1717–1735. MR1355006

†The `ispseudoprime` command relies on the Baillie–Pomerance–Selfridge–Wagstaff compositeness test, which is expected to falsely declare some composites prime (but no such examples are known). Thus, our implementation leaves open the possibility that a nonprime is fed to `qfbsolve`, potentially resulting in an error or an incorrect representation of n. But any incorrect representation is easily discovered, and one can simply continue the prime search in those cases.

[BS96c] W. Bosma and P. Stevenhagen, *On the computation of quadratic* 2*-class groups*, J. Théor. Nombres Bordeaux **8** (1996), no. 2, 283–313. MR1438471

[BZ11] R. P. Brent and P. Zimmermann, *Modern Computer Arithmetic*, Cambridge Monographs on Applied and Computational Mathematics, vol. 18, Cambridge University Press, Cambridge, 2011. MR2760886

[Dic27a] L. E. Dickson, *Ternary quadratic forms and congruences*, Ann. of Math. (2) **28** (1926/27), no. 1-4, 333–341. MR1502786

[Dic27b] L. E. Dickson, *Integers represented by positive ternary quadratic forms*, Bull. Amer. Math. Soc. **33** (1927), no. 1, 63–70. MR1561323

[Dic27c] L. E. Dickson, *Quaternary quadratic forms representing all integers*, Amer. J. Math. **49** (1927), no. 1, 39–56. MR1506600

[Dir50] G. Lejeune Dirichlet, *Über die Zerlegbarkeit der Zahlen in drei Quadrate* (German), J. Reine Angew. Math. **40** (1850), 228–232. MR1578694

[ER01] F. Eisenbrand and G. Rote, *Fast reduction of ternary quadratic forms*, Cryptography and lattices (Providence, RI, 2001), Lecture Notes in Comput. Sci., vol. 2146, Springer, Berlin, 2001, pp. 32–44. MR1903885

[Fla89] D. E. Flath, *Introduction to Number Theory*, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1989. MR972739

[Für07] M. Fürer, *Faster integer multiplication*, STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing, ACM, New York, 2007, pp. 57–66. MR2402428

[Für09] M. Fürer, *Faster integer multiplication*, SIAM J. Comput. **39** (2009), no. 3, 979–1005. MR2538847

[Gau86] C. F. Gauss, *Disquisitiones Arithmeticae*, Springer-Verlag, New York, 1986. Translated and with a preface by Arthur A. Clarke; Revised by William C. Waterhouse, Cornelius Greither and A. W. Grootendorst and with a preface by Waterhouse. MR837656

[HB78] D. R. Heath-Brown, *Almost-primes in arithmetic progressions and short intervals*, Math. Proc. Cambridge Philos. Soc. **83** (1978), no. 3, 357–375. MR0491558

[HHL16] D. Harvey, J. van der Hoeven, and G. Lecerf, *Even faster integer multiplication*, J. Complexity **36** (2016), 1–30. MR3530637

[HL23] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio numerorum'; III: On the expression of a number as a sum of primes*, Acta Math. **44** (1923), no. 1, 1–70. MR1555183

[Jon31] B. W. Jones, *The regularity of a genus of positive ternary quadratic forms*, Trans. Amer. Math. Soc. **33** (1931), no. 1, 111–124. MR1501578

[Lag80] J. C. Lagarias, *Worst-case complexity bounds for algorithms in the theory of integral quadratic forms*, J. Algorithms **1** (1980), no. 2, 142–186. MR604862

[Lag80b] J. C. Lagarias, *On the computational complexity of determining the solvability or unsolvability of the equation* $X^2 - DY^2 = -1$, Trans. Amer. Math. Soc. **260** (1980), no. 2, 485–508. MR574794

[Lan58] E. Landau, *Elementary Number Theory*, Chelsea Publishing Co., New York, N.Y., 1958. Translated by J. E. Goodman. MR0092794

[LLS15] Y. Lamzouri, X. Li, and K. Soundararajan, *Conditional bounds for the least quadratic non-residue and related problems*, Math. Comp. **84** (2015), no. 295, 2391–2412. MR3356031

[LP92] H. W. Lenstra Jr. and C. Pomerance, *A rigorous time bound for factoring integers*, J. Amer. Math. Soc. **5** (1992), no. 3, 483–516. MR1137100

[PARI] PARI/GP, version `2.9.2`. `http://pari.math.u-bordeaux.fr/`, Bordeaux, 2017.

[PLS17] J. Li, K. Pratt, and G. Shakan, *A lower bound for the least prime in an arithmetic progression*, Q. J. Math. **68** (2017), no. 3, 729–758. MR3698292

[PTn] P. Pollack and E. Treviño, *Finding the four squares in Lagrange's theorem*, Integers 18A (2018), 16 pages (electronic).

[RS86] M. O. Rabin and J. O. Shallit, *Randomized algorithms in number theory*, Comm. Pure Appl. Math. **39** (1986), no. S, suppl., S239–S256. Frontiers of the mathematical sciences: 1985 (New York, 1985). MR861490

Department of Mathematics, University of Georgia, Athens, Georgia 30602
Email address: pollack@uga.edu

Culmannstrasse 77, CH-8006 Zurich, Switzerland
Email address: peter.schorn@acm.org