FINDING THE FOUR SQUARES IN LAGRANGE'S THEOREM

For Jeff Shallit on his 60th birthday

Paul Pollack[1]
Department of Mathematics, University of Georgia, Athens, GA 30602, USA
pollack@uga.edu

Enrique Treviño
Department of Mathematics and Computer Science, Lake Forest College, Lake Forest, IL 60045, USA
trevino@lakeforest.edu

Received: , Revised: , Accepted: , Published:

Abstract

In 1986, Rabin and Shallit presented three randomized algorithms to compute, given a positive integer n, integers X, Y, Z, W with $X^2 + Y^2 + Z^2 + W^2 = n$. The fastest of the three has expected runtime $O((\log n)^2)$, but this runtime analysis assumes the truth of the Extended Riemann Hypothesis. (Here we measure runtime not by bit operations, but by the number of "basic operations" one must carry out on numbers of size $\lessapprox n$.) The other two algorithms admit slightly worse runtime estimates but are unconditional, in the sense that no unproved hypotheses are used in the proof of correctness or the running-time analysis. In this paper we explain how to modify their algorithms to do slightly better. We give two algorithms for this problem with expected runtime $O((\log n)^2(\log\log n)^{-1})$; the first is easily described but depends on ERH, while the latter is unconditional but slightly involved.

1. Introduction

A fundamental result in number theory, claimed by Bachet in 1621 and proved by Lagrange in 1770 [10], is that every positive integer n can be written as a sum of four squares, i.e., expressed in the form $X^2 + Y^2 + Z^2 + W^2$ for integers X, Y, Z, W. In this article, we consider the computational problem of finding X, Y, Z, W given n.

[1]Research of the first-named author is supported by NSF award DMS-1402268.

As far as we are aware, Rabin and Shallit were the first authors to describe a provably efficient algorithm for this problem, in 1986 [18]. Their article presents three randomized algorithms for expressing n as a sum of four squares, with running time complexity $O((\lg n)^2)$, $O((\lg n)^2 \lg\lg n)$, and $O((\lg n)^2(\lg\lg n)^2)$.[2] Here our convention for complexity estimates follows [18], so that complexity counts not bit operations but arithmetic operations on numbers of size $n^{O(1)}$, where an "arithmetic operation" means computing $n \pm m, n \cdot m, \lfloor n/m \rfloor$, or the least nonnegative remainder when n is divided by m, denoted here $n \bmod m$.[3]

The fastest of the three algorithms alluded to above comes with an important caveat: it depends on the Extended Riemann Hypothesis (meaning the Riemann Hypothesis for Dirichlet L-functions). Without ERH, it is not guaranteed to terminate, although when it does terminate, its output is correct. The two slightly slower algorithms are unconditional.

In this paper, we present two randomized algorithms for the four-squares problem, each of which is slightly faster than the fastest of the algorithms in [18]. Our two randomized algorithms have complexity $O((\lg n)^2(\lg\lg n)^{-1})$. The first, simpler algorithm depends on ERH, while the second is more complicated, but is independent of any unproved hypotheses. While the core ideas of the algorithms are borrowed from [18], we are able to shave off double-logarithmic factors by paying more attention to the effect of the small prime divisors of n on the magnitude of functions like $\varphi(n)/n$.

2. Preliminaries on the Gaussian integers and integral quaternions

Our argument depends crucially on convenient arithmetic properties possessed by two particular "rings of integers". The first of these rings is the Gaussian integers $\mathbb{Z}[i] := \{a + bi : a, b \in \mathbb{Z}\} \subset \mathbb{C}$. The second is possibly less familiar. Recall that the ring of real quaternions is the (non-commutative!) $\mathbb{R}$-algebra with $\mathbb{R}$-basis 1, i, j, k and multiplication determined by $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $jk = -kj = i$, and $ki = -ik = j$. We work in the subring $\mathbb{H}$ of (Hurwitz) integral quaternions, defined by

$$\mathbb{H} := \Big\{\frac{1}{2}(a + bi + cj + dk) : a, b, c, d \in \mathbb{Z}, a \equiv b \equiv c \equiv d \pmod 2\Big\}.$$

In this section we quickly review the facts we need about $\mathbb{Z}[i]$ and $\mathbb{H}$. Proofs of the results we will assume about $\mathbb{Z}[i]$ can be found in introductory texts on abstract

[2] We write $\lg n$ for the number of binary digits in n; e.g., $\lg 0 = \lg 1 = 1$, while $\lg 1957 = 11$.

[3] For the reader who prefers to count bit operations, it is safe to multiply all of our final complexity estimates by $M(\lg n)$, where $M(n)$ is the complexity of multiplying two numbers with n bits.

algebra (see, e.g., [4, Chapter 8]); for $\mathbb{H}$, section 3 of [18] may be consulted.

In both $\mathbb{Z}[i]$ and $\mathbb{H}$, an important role is played by the *norm* of an element. If $\alpha = a + bi \in \mathbb{Z}[i]$, its *conjugate* is defined as $a - bi$. Similarly, the conjugate of $a + bi + cj + dk \in \mathbb{H}$ is defined as $a - bi - cj - dk$. For α in either system, we define the norm by $N\alpha = \alpha\bar{\alpha}$. Concretely, if $\alpha = a + bi \in \mathbb{Z}[i]$, then $N\alpha = a^2 + b^2$, and if $\alpha = a + bi + cj + dk \in \mathbb{H}$, then $N\alpha = a^2 + b^2 + c^2 + d^2$. In both systems, $N\alpha$ takes values in the nonnegative integers, and $N\alpha = 0$ if and only if $\alpha = 0$. A fundamental fact is that the norm is multiplicative: for all α, β, we have $N(\alpha\beta) = N\alpha \cdot N\beta$.

It is immediate that n is a sum of two squares if and only if $n = N\alpha$ for some $\alpha \in \mathbb{Z}[i]$. The corresponding fact holds for sums of four squares relative to $\mathbb{H}$: n is a sum of four squares if and only if $n = N\alpha$ for some $\alpha \in \mathbb{H}$. In this latter statement, the "only if" direction is clear, but the "if" direction perhaps less so, since $\mathbb{H}$ contains elements of the form $\alpha = \frac{1}{2}(a + bi + cj + dk)$ with a, b, c, d odd. However, every such α can be multiplied by an $\epsilon \in \mathbb{H}$ of norm 1 in such a way that the product $\epsilon\alpha$ has integer components; indeed, if we choose $\epsilon_a, \epsilon_b, \epsilon_c, \epsilon_d \in \{\pm 1\}$ so that

$$\epsilon_a \equiv a, \quad \epsilon_b \equiv -b, \quad \epsilon_c \equiv -c, \quad \epsilon_d \equiv -d \pmod 4,$$

then $\epsilon = \frac{1}{2}(\epsilon_a + \epsilon_b i + \epsilon_c j + \epsilon_d k)$ has the required property. Since the norm of α is the same as the norm of $\epsilon\alpha$, every norm of an element of $\mathbb{H}$ is the norm of a quaternion with integral components, and so is a sum of four squares.

Both $\mathbb{Z}[i]$ and $\mathbb{H}$ possess division algorithms. Specifically, if $\alpha, \beta \in \mathbb{Z}[i]$ and $\beta \neq 0$, then there are $\gamma, \delta \in \mathbb{Z}[i]$ with

$$\alpha = \beta\gamma + \delta, \quad N\delta \leq \frac{1}{2} N\beta.$$

The same statement holds with $\mathbb{H}$ replacing $\mathbb{Z}[i]$. In the setting of $\mathbb{H}$, one could also consider dividing α by β "on the other side", and the analogous result holds; there are $\gamma, \delta \in \mathbb{H}$ with

$$\alpha = \gamma\beta + \delta, \quad N\delta \leq \frac{1}{2} N\beta.$$

The theory of the Euclidean algorithm now implies that every ideal of $\mathbb{Z}[i]$ is principal, as is every right-ideal and left-ideal of $\mathbb{H}$.[4]

The principality of ideals implies the existence of gcds. Specifically, if $\alpha, \beta \in \mathbb{Z}[i]$, every generator of the ideal (α, β) is a gcd. Similarly, in $\mathbb{H}$, every generator of the left-ideal generated by α, β is a greatest common right-divisor (gcrd)[5] of α, β, and every generator of the corresponding right-ideal is a greatest common left-divisor

[4] For us, a right-ideal is an additive subgroup that absorbs multiplication on the right. Note that in $\mathbb{Z}[i]$, the collections of left-ideals and right-ideals coincide, since $\mathbb{Z}[i]$ is commutative.

[5] We say that γ is a right-divisor of θ if there is a δ such that $\theta = \delta\gamma$. Left-divisors are defined analogously.

(gcld) of α, β. The gcd is unique only up to multiplication by a unit (multiplication on the left in the case of a gcrd, and on the right in the case of a gcld). In both systems, the units are precisely the elements of norm 1; these are $\pm 1, \pm i$ in $\mathbb{Z}[i]$ and

$$\pm 1, \ \pm i, \ \pm j, \ \pm k, \ \frac{1}{2}(\pm 1 \pm i \pm j \pm k)$$

in $\mathbb{H}$.

Division is easy from a computational standpoint. Specifically, if α, β are given with $N\alpha, N\beta$ having $O(\lg n)$ bits, one can determine γ, δ in $O(1)$ arithmetic operations on integers having $O(\lg n)$ bits.[6] Now following the Euclidean algorithm, one determines a gcd (or gcrd/gcld in the case of $\mathbb{H}$) in $O(\lg n)$ arithmetic operations.

We will make essential use of the following two propositions, which tell us that certain gcd (resp., gcrd) computations yield two-square (resp., four-square) representations.

Lemma 1. *Let n be an odd positive integer. If $n \mid N(a+bi)$, where $\gcd(a,b) = 1$, then every gcd of n and $a+bi$ has norm n.*

Lemma 2. *Let n be an odd positive integer. If $n \mid N(a+bi+cj+dk)$, where $\gcd(a,b,c,d) = 1$, then every gcrd of n and $a+bi+cj+dk$ has norm n.*

We prove only Lemma 2. The proof of Lemma 1 is similar but simpler (since $\mathbb{Z}[i]$ is commutative).

Proof sketch of Lemma 2. Let α be a gcrd of n and $a+bi+cj+dk$, so that α generates the left ideal $A := \mathbb{H} \cdot n + \mathbb{H} \cdot (a+bi+cj+dk)$. Let $\bar{A} = \{\bar{\kappa} : \kappa \in A\}$. Recalling that conjugation is an antiautomorphism of $\mathbb{H}$ (meaning that $\overline{\alpha\beta} = \overline{\beta} \cdot \overline{\alpha}$), we see that $\bar{A} = n \cdot \mathbb{H} + (a-bi-cj-dk) \cdot \mathbb{H}$ and $\bar{A}$ is generated as a right ideal by $\bar{\alpha}$. Consider the product $A\bar{A}$, by which we mean $\{\mu\nu : \mu \in A, \nu \in \bar{A}\}$. Since A consists of the left-multiples of α and $\bar{A}$ consists of the right-multiples of $\bar{\alpha}$, one finds that $A\bar{A}$ is the two-sided ideal consisting of all multiples of $N\alpha$. (The term "multiples" is unambiguous here, since $\mathbb{Z}$ lies in the center of $\mathbb{H}$.) On the other hand,

$$A\bar{A} = \{(\beta n + \gamma(a+bi+cj+dk))(n\delta + (a-bi-cj-dk)\epsilon) : \beta, \gamma, \delta, \epsilon \in \mathbb{H}\}.$$

By a direct calculation,

$$\begin{aligned}&(\beta n + \gamma(a+bi+cj+dk))(n\delta + (a-bi-cj-dk)\epsilon) \\ &= n(\beta\delta n + \beta(a-bi-cj-dk)\epsilon + \gamma(a+bi+cj+dk)\delta + \frac{N(a+bi+cj+dk)}{n}\gamma\epsilon).\end{aligned}$$

We see from this that n divides every element of $A\bar{A}$. Hence, $n^{-1}A\bar{A}$ is also a two-sided ideal of $\mathbb{H}$. Taking $\beta = \delta = 1$ and $\epsilon = \gamma = 0$, we see that $n^{-1}A\bar{A}$ contains

[6] See [18, §3] for the details of how to efficiently implement division in $\mathbb{Z}[i]$ and $\mathbb{H}$.

n. Taking $\beta = 1, \delta = 0$, and $\gamma = 0$ shows that $n^{-1}A\bar{A}$ contains

$$(a - bi - cj - dk)\epsilon$$

for all ϵ. Similarly, taking $\delta = 1, \beta = 0$ and $\epsilon = 0$, we find that $n^{-1}A\bar{A}$ contains

$$\gamma(a + bi + cj + dk)$$

for all γ. Since $n^{-1}A\bar{A}$ is an ideal, these last two results together imply that

$$(a - bi - cj - dk)\epsilon + \overline{(a - bi - cj - dk)\epsilon} \in n^{-1}A\bar{A}$$

for all choices of ϵ. Now letting $\epsilon = 1, i, j, k$, we find that $2a, 2b, 2c, 2d \in n^{-1}A\bar{A}$. By assumption, a, b, c, d generate the unit ideal of $\mathbb{Z}$, and so $n^{-1}A\bar{A}$ contains 2. Since $n^{-1}A\bar{A}$ also contains n, and n is odd, $n^{-1}A\bar{A}$ is all of $\mathbb{H}$. Hence, $A\bar{A}$ consists precisely of the multiples of n. Since $A\bar{A}$ also consists precisely of the multiples of $N\alpha$, and both $N\alpha$ and n are positive integers, we must have $n = N\alpha$. □

We suspect Lemmas 1 and 2 are classical. However, they are perhaps not as well-known as they could be. As some evidence, Lemma 2 implies that the repeated splitting procedure described on pp. S246, S251 of [18] is never necessary.

3. An ERH-conditional algorithm

Our ERH-conditional algorithm is a riff on the ERH-conditional algorithm described in [18], which was discovered by Rabin in 1977 (see the interview [20]). Here is a sketch of that algorithm. First, we can assume that the number n we are trying to represent is odd. To see this, write $n = 2^e n'$ with n' odd. If $X'^2 + Y'^2 + Z'^2 + W'^2 = n'$, then $X^2 + Y^2 + Z^2 + W^2 = n$ for X, Y, Z, W defined by

$$(1 + i)^e(X' + Y'i + Z'j + W'k) = X + Yi + Zj + Wk.$$

The computation of e here, as well as the computation of X, Y, Z, W from e and X', Y', Z', W', requires only $O(\lg n)$ steps.

To represent an odd n as a sum of four squares, we look for a prime $p \equiv -1 \pmod{n}$, $p \equiv 1 \pmod{4}$ smaller than $(2n)^5$. Since $p \equiv 1 \pmod{4}$, we can write $p = A^2 + B^2$. Suppose we have computed A, B. Then

$$n \mid p + 1 = A^2 + B^2 + 1 = N(A + Bi + j).$$

By Lemma 2, we recover a four squares representation of n by computing $\gcd(n, A + Bi + j)$, where we assume the gcrd has been multiplied by a unit in $\mathbb{H}$ to have integer

components. Without the restriction that $p < (2n)^5$, the existence of such a prime p follows by Dirichlet's theorem on primes in arithmetic progressions. So we see that these ideas lead immediately to a quick proof of Lagrange's theorem (but one that assumes the theorem of Dirichlet, proved nearly 70 years later).

Rabin realized that (under ERH) one can quickly find the needed prime p by making randomized choices. Indeed, under ERH, among all integers up to $(2n)^5$ that are $\equiv -1 \pmod{n}$ and $\equiv 1 \pmod{4}$, the proportion of primes is $\gg \frac{n}{\varphi(n)} \cdot \frac{1}{\lg n} \geq \frac{1}{\lg n}$. So we expect to hit a prime p in $O(\lg n)$ trials.

We now present our modified algorithm for representing an odd n as a sum of four squares. The new idea is, rather than simply using the trivial lower bound of 1 on $\frac{n}{\varphi(n)}$, to exploit that $\frac{n}{\varphi(n)}$ is large when n is divisible by many small primes; this allows us to reduce the expected number of trials to $O(\lg n / \lg\lg n)$. We assume that $n > 20$; for $n \leq 20$ it is trivial to find a four-squares representation.

(1) [Precomputation] Determine the primes not exceeding $\log n$ and compute their product M.

(2) [Random trials] Choose an odd number $k < n^5$ at random, and let

$$p = Mnk - 1.$$

(Notice that $p \equiv 1 \pmod{4}$, since $2 \parallel M$ and n, k are odd.) For a randomly chosen $u \in [1, p-1]$, compute $s = u^{(p-1)/4} \bmod p$ and test if $s^2 \equiv -1 \pmod{p}$. If so, continue to the next step. Otherwise, restart this step.

(3) [Denouement] Compute $A + Bi := \gcd(s + i, p)$. Then compute $\operatorname{gcrd}(A + Bi + j, n)$, normalized to have integer components. Write this gcrd as $X + Yi + Zj + Wk$, and output the representation $n = X^2 + Y^2 + Z^2 + W^2$.

We now explain why the algorithm is correct and why the expected number of required arithmetic operations is $O((\lg n)^2 (\lg\lg n)^{-1})$.

First, we address the correctness of the output. At the conclusion of (2), $p \mid N(s+i)$, and so Lemma 1 gives that $A^2 + B^2 = p$. Since $p + 1 = Mnk$, we have

$$n \mid p + 1 = A^2 + B^2 + 1 = N(A + Bi + j).$$

Now Lemma 2 shows that every gcrd of n and $A + Bi + j$ has norm n. This completes the correctness proof.

We remark that although we used the letter p, it is not necessary to assume here that p is prime. (But it is unlikely we would have found $s^2 \equiv -1 \pmod{p}$ unless p were prime.)

Now we address the complexity. Each integer in $[2, \log n]$ can be tested for primality in $O((\lg n)^{1/2})$ operations (by trial division), and so the complete list of primes in

$[2, \log n]$ can be found with $O((\lg n)^{3/2})$ operations.[7] We can then compute M using $O(\lg n)$ multiplications. (By the prime number theorem, all of the partial products arising in the computation of M are of size $n^{O(1)}$.) Hence, the precomputation step can be carried out with $O((\lg n)^{3/2})$ arithmetic operations.

In order to continue, we must recall two consequences of the ERH.

(a) For each real x and each pair of integers a, q with $q > 0$, let $\pi(x; q, a)$ denote the count of primes $p \le x$ with $p \equiv a \pmod q$. If a, q are relatively prime and $x \ge 2$, then

$$\left|\pi(x; q, a) - \frac{1}{\varphi(q)} \int_2^x \frac{\mathrm{d}t}{\log t}\right| \le \sqrt{x}(\log x + 2 \log q).$$

(See Oesterlé [12] for a more general result.)

(b) (Bach and Sorenson [2]) For $q \ge 2$ and a coprime to q, the least prime $p \equiv a \pmod q$ satisfies $p \le 2(q \log q)^2$.

We will use the following crude consequence of (a) and (b): For $q \ge 2$ and $x \ge 2q^3$,

$$\pi(x; q, a) \gg \frac{x}{\varphi(q) \log x}.$$

(We leave the task of deducing this statement from (a) and (b) to the reader.)

We now apply this lower bound on $\pi(x; q, a)$ to estimate the number of primes $p \equiv -1 \pmod{Mn}$, $p \equiv 1 \pmod 4$ not exceeding Mn^6. The two congruence conditions place p in a coprime residue class modulo $2Mn$. It is known (see [17]) that

$$\prod_{\substack{\ell \text{ prime} \\ \ell \le T}} \ell \le e^{1.02T} \tag{1}$$

for all $T > 0$. Taking $T = \log n$ yields $M \le n^{1.02}$. Using this, it is straightforward to check that $Mn^6 \ge 2(2Mn)^3$ for $n \ge 18$. So under ERH, the number of primes p as above is

$$\gg \frac{Mn^6}{\varphi(2Mn) \log(Mn^6)} \gg \frac{n^5}{\log n} \cdot \frac{Mn}{\varphi(Mn)} = \frac{n^5}{\log n} \prod_{\ell \mid Mn} (1 - 1/\ell)^{-1} \gg n^5 \frac{\log\log n}{\log n},$$

using in the last step that Mn is divisible by all primes up to $\log n$ and that the sum of the reciprocals of those primes is $\log\log\log n + O(1)$. Thus, if an odd k is selected at random from the positive integers $\le n^5$, then $p = Mnk - 1$ is prime with probability $\gg \log\log n / \log n$. Moreover, when p is prime, half of the values of

[7] It would be more efficient to use the sieve of Eratosthenes here, but this part of the algorithm is not the bottleneck.

$u \in [1, p-1]$ (namely, the quadratic nonresidues) are such that $u^{(p-1)/4}$ is a square root of -1 mod p.

It follows that each trial step has probability $\gg \log\log n/\log n$ of success, so that the expected number of trials is $O(\log n/\log\log n)$. The most expensive step in an individual trial is the computation of $u^{(p-1)/4} \bmod p$; by repeated squaring, this can be done with $O(\lg p) = O(\lg n)$ operations. Thus, the expected number of operations required in step (2) is $O((\lg n)^2(\lg\lg n)^{-1})$, which is acceptable. (We have used here that log and lg have the same order of magnitude in the range of n we are considering, as do log log and lg lg.)

Finally, (3) requires $O(\lg n)$ arithmetic operations, using the Euclidean algorithm to compute the gcds. Putting everything together, we see that the expected total number of arithmetic operations required by the algorithm is $O((\lg n)^2(\lg\lg n)^{-1})$.

4. An unconditional algorithm

In the ERH-conditional algorithm, we find a prime p from the progression 1 modulo 4 that is also congruent to -1 modulo n. Writing $p = x^2 + y^2$ gives a solution to $n \mid x^2+y^2+1$, allowing us to extract a four-square representation from Lemma 2. In our unconditional algorithm, we instead search for x, y for which $(-(x^2+y^2)) \bmod n$ can be quickly expressed as a sum of two squares, say z^2+w^2. Then $n \mid x^2+y^2+z^2+w^2$ and a four-square representation can again be obtained from Lemma 2. (Of course, we need n odd and $\gcd(x, y, z, w) = 1$ to apply Lemma 2; we will arrange below for these conditions to hold.)

We began the previous section by mentioning that we can easily reduce to considering only odd n. In fact, for an acceptable computational cost, we can reduce to the more special case of odd n having no prime factors $\ell \le \log n$ with $\ell \equiv 1 \pmod 4$. As in the previous section, we assume that $n > 20$, so that $\log n > 3$.

To see how this reduction goes, first note that we can flag each number in $[1, \log n]$ as prime or composite using $O((\lg n)^{3/2})$ operations. Since our goal is to produce an algorithm requiring $O((\lg n)^2(\lg\lg n)^{-1})$ operations, such a computation is acceptable. In another $O(\lg n)$ operations, we can compute $X^2 + Y^2$ for all pairs X, Y with $0 \le X, Y \le (\log n)^{1/2}$. We record, for $\ell = 2$ and for the primes $\ell \le \log n$ with $\ell \equiv 1 \pmod 4$, integers X_ℓ, Y_ℓ with

$$\ell = X_\ell^2 + Y_\ell^2.$$

Writing "$\ell = \square + \square$" for the condition that $\ell = 2$ or $\ell \equiv 1 \pmod 4$, we factor

$$n = \left(\prod_{\substack{\ell \le \log n \\ \ell = \square + \square}} \ell^{e_\ell} \right) n',$$

where n' is odd and coprime to all primes $\ell \equiv 1 \pmod 4$, $\ell \le \log n$. This can be done with $O(\lg n)$ operations: For each ℓ, we repeatedly divide n by ℓ until the division leaves a remainder. For a given ℓ, this requires $e_\ell + 1$ divisions, and so the total number of divisions required is

$$\sum_{\substack{\ell \le \log n \\ \ell = \square + \square}} (e_\ell + 1) = \sum_{\substack{\ell \le \log n \\ \ell = \square + \square}} e_\ell + \sum_{\substack{\ell \le \log n \\ \ell = \square + \square}} 1 \le \frac{\log n}{\log 2} + \log n,$$

which is $O(\lg n)$. If $X'^2 + Y'^2 + Z'^2 + W'^2 = n'$, then $X^2 + Y^2 + Z^2 + W^2 = n$, where

$$(X' + Y'i + Z'j + W'k) \prod_{\substack{\ell \le \log n \\ \ell = \square + \square}} (X_\ell + Y_\ell i)^{e_\ell} = X + Yi + Zj + Wk.$$

Computing X, Y, Z, W, given the e_ℓ and the numbers X', Y', Z', W', requires another $O(\lg n)$ operations. Collecting the estimates, we may perform this reduction at the cost of $O((\lg n)^{3/2})$ steps.

Before describing our unconditional algorithm, which is a variant of the method described in §3 of [18], we make some observations. From now on, $n > 20$, n is odd, and n has no prime factors $\ell \le \log n$ from the arithmetic progression 1 mod 4. Let

$$P = \prod_{\substack{\ell \le \log n \\ \ell \equiv 3 \pmod 4}} \ell,$$

and let

$$N = n \cdot P / \gcd(P, n).$$

We have already computed the primes up to $\log n$, and so computing P and N requires only $O(\lg n)$ operations. Note that $n \le N \le n^{O(1)}$, where the second inequality comes from the prime number theorem (or (1)).

The following lemma is a special case of Lemma 3.2 on p. S247 of [18], and was seemingly first proved by Hermite in 1854 [8].

Lemma 3. *Let N be an odd positive integer. For every a coprime to N, the number of solutions x, y mod N to $x^2 + y^2 \equiv a \pmod N$ is precisely*

$$N \prod_{\ell \mid N} \left(1 - \left(\frac{-1}{\ell} \right) \frac{1}{\ell} \right),$$

where $\left(\frac{-1}{\ell}\right)$ *is the Legendre symbol.*

From the primes $\ell \le \log n$, our N is divisible by none that are congruent to 1 (mod 4) and by all that are congruent to 3 (mod 4). Writing n' for the largest divisor of n supported on primes exceeding $\log n$, we deduce that

$$\prod_{\ell | N} \left(1 - \left(\frac{-1}{\ell}\right)\frac{1}{\ell}\right) \ge \prod_{\substack{\ell \le \log n \\ \ell \equiv 3 \pmod 4}} \left(1 + \frac{1}{\ell}\right) \prod_{\substack{\ell | N \\ \ell > \log n}} \left(1 - \frac{1}{\ell}\right)$$

$$\gg (\log\log n)^{1/2} \prod_{\substack{\ell | N \\ \ell > \log n}} \left(1 - \frac{1}{\ell}\right) \ge (\log\log n)^{1/2} \cdot \left(1 - \frac{1}{\log n}\right)^{\omega(n')},$$

where, as usual, $\omega(\cdot)$ counts the number of distinct prime divisors. (In moving from the first to the second line, we used that the sum of the reciprocals of the primes congruent to 3 (mod 4) up to T, for $T \ge 2$, is $\frac{1}{2}\log\log T + O(1)$. See, e.g., [11, pp. 449–450]. We also used that N and n have the same prime divisors exceeding $\log n$.) Notice that $(\log n)^{\omega(n')} \le n'$, so that $\omega(n') \le \log n'/\log\log n < \log n$. Consequently,

$$\left(1 - \frac{1}{\log n}\right)^{\omega(n')} \ge \left(1 - \frac{1}{\log n}\right)^{\log n} \ge \left(1 - \frac{1}{3}\right)^3 > \frac{1}{4}.$$

(We used here that $\log n \ge 3$ and that $(1 - 1/T)^T$ is increasing for $T \ge 3$.) Thus, each congruence $x^2 + y^2 \equiv a \pmod N$, with $\gcd(a, N) = 1$, has $\gg N(\log\log n)^{1/2}$ solutions (x, y) mod N.

The algorithm begins by selecting x, y at random from $[1, N]$ and computing

$$r := (-(x^2 + y^2)) \bmod N.$$

We show below (Lemma 4) that there are $\gg N(\log\log n)^{1/2}/\log N$ integers in $[1, N]$ that have the form $r_1 p$, where r_1 is a product of primes $\ell \le \log n$ with $\ell \equiv 1 \pmod 4$, and $p > \log n$ is a prime congruent to 1 modulo 4 not dividing N. All of these numbers $r_1 p$ are coprime to N. Using the concluding result of the last paragraph, we see that the number of choices for x, y where r lands on one of the numbers $r_1 p$ is

$$\gg N\frac{(\log\log n)^{1/2}}{\log N} \cdot N(\log\log n)^{1/2} \gg N^2\frac{\log\log n}{\log n}.$$

Thus, with x, y chosen at random from $[1, N]$, we expect to have $r = r_1 p$ within $O(\log n/\log\log n)$ trials.

Having located $r = r_1 p$, we note that it is easy to compute a two-squares representation of r_1:

$$u^2 + v^2 = r_1, \quad \text{where} \quad u + vi := \prod_{\ell^{v_\ell} \| r_1} (X_\ell + Y_\ell i)^{v_\ell}. \tag{2}$$

Determining the exponents v_ℓ and computing u, v requires only $O(\lg n)$ operations, by arguments similar to those appearing at the start of this section.

Suppose we have written $p = U^2 + V^2$, and let $z + wi = (u + vi)(U + Vi)$, so that $z^2 + w^2 = r_1 p$. Then

$$-(x^2 + y^2) \equiv r = r_1 p = z^2 + w^2 \pmod{N},$$

so that

$$n \mid N \mid x^2 + y^2 + z^2 + w^2. \tag{3}$$

We show below (see Lemma 5) that $\gcd(x, y, z, w) = 1$ — in fact, that $\gcd(z, w) = 1$ — so that by Lemma 2 a four squares representation of n is obtained by computing $\operatorname{gcrd}(n, x + yi + zj + wk)$ (normalized to have integer components).

As it stands, the above description is incomplete. We have glossed over how we determine whether $r = r_1 p$ for some r_1, p as above, and how to write p as a sum of two squares. To circumvent these issues, we amend the initial (trial) steps of the algorithm to the following, denoted (T):

(T) For a random choice of $x, y \in [1, N]$, compute $r := (-(x^2 + y^2)) \bmod N$. Immediately choose another x, y unless $r \equiv 1 \pmod 4$ and $\gcd(r, N) = 1$. Once an r satisfying these conditions is found, determine r_1 by

$$r = r_1 p, \quad \text{where} \quad r_1 = \prod_{\substack{\ell \le \log n \\ \ell \equiv 1 \pmod 4 \\ \ell^{v_\ell} \| r}} \ell^{v_\ell}. \tag{4}$$

(Finding r_1 requires only $O(\lg n)$ operations.) If $p = 1$, we stop and declare victory.

Now assume that $p > 1$. In that case, we choose $u \in [1, p - 1]$ at random and test whether $s := u^{(p-1)/4} \bmod p$ is a square root of $-1 \bmod p$. If not, we restart (T) with another random choice of x, y.

When p is a prime, s is a square root of -1 half the time. Thus (from Lemma 4) the random trial (T) succeeds with probability $\gg \log\log n / \log n$. So we expect to do only $O(\log n / \log\log n)$ trials before (T) succeeds. Each trial requires $O(\lg n)$ operations, and so we expect (T) to require $O((\lg n)^2 / \lg\lg n)$ operations.

When (T) succeeds, we know a pair x, y with

$$-(x^2 + y^2) \equiv r_1 p,$$

we have the factorization of r_1 in the form (4), and either $p = 1$ or we have a square root s of -1 modulo p. Compute the u, v in (2), so $u^2 + v^2 = r_1$. If $p = 1$, put $U = 1$ and $V = 0$; otherwise, calculate $\gcd(p, s + i) = U + Vi$. In either case, $p = U^2 + V^2$.

With $z + wi = (u + vi)(U + Vi)$, we have, as in (3), $n \mid x^2 + y^2 + z^2 + w^2$. By Lemma 5 below, $\gcd(z, w) = 1$, and so, by Lemma 2, we can extract a four-squares representation of n by computing $\operatorname{gcrd}(n, x + yi + zj + wk)$. These concluding steps require $O(\lg n)$ operations in total, and so the expected number of steps in the entire (amended) algorithm is $O((\lg n)^2/\lg\lg n)$, as desired.

We have two outstanding debts: proving that there are indeed many values of $r_1 p$, and proving that $\gcd(z, w) = 1$.

Lemma 4. *Let n and N be odd integers with $20 < n \le N$. The number of positive integers $R \le N$ which admit a decomposition $R = r_1 p$ where r_1 is a product of primes $\ell \le \log n$, $\ell \equiv 1 \pmod 4$, and $p > \log n$ is a prime congruent to $1 \pmod 4$ with $p \nmid N$, is*

$$\gg N \frac{\sqrt{\log\log n}}{\log N}.$$

Proof. We first prove the lemma when n is sufficiently large. For each r_1 that can be written as a product of primes $\ell \le \log n$, $\ell \equiv 1 \pmod 4$, we count the number of corresponding choices for p. For a given r_1, the number of choices for p is at least

$$\pi(N/r_1; 4, 1) - \pi(\log n; 4, 1) - \omega(N).$$

We obtain a lower bound by summing this expression over $r_1 \le N^{1/2}$. By the prime number theorem for arithmetic progressions, for all large N and all $r_1 \le N^{1/2}$,

$$\pi(N/r_1; 4, 1) \gg \frac{N/r_1}{\log(N/r_1)} \gg \frac{N}{r_1 \log N}.$$

This last quantity is $\gg \frac{N^{1/2}}{\log N}$. On the other hand,

$$\pi(\log n; 4, 1) \le \log n \le \log N \quad \text{and} \quad \omega(N) \le \log N/\log 3 < \log N.$$

Since $\log N = o(N^{1/2}/\log N)$, as $N \to \infty$, we see that the number of p corresponding to a given r_1 is $\gg N/(r_1 \log N)$, uniformly for $r_1 \le N^{1/2}$ (once n, and hence N, is large). Thus, the total number of values of $r = r_1 p$ we produce this way is

$$\gg \frac{N}{\log N}\left(\sum_{r_1} \frac{1}{r_1} - \sum_{r_1 > N^{1/2}} \frac{1}{r_1}\right),$$

where the sums on r_1 are over numbers composed of primes $\ell \le \log n, \ell \equiv 1 \pmod 4$. Now

$$\sum_{r_1} \frac{1}{r_1} = \prod_{\substack{\ell \le \log n \\ \ell \equiv 1 \pmod 4}} \left(1 + \frac{1}{\ell} + \frac{1}{\ell^2} + \cdots\right) \gg \sqrt{\log\log n}.$$

(We use here that the sum of the reciprocals of the primes congruent to 1 (mod 4) up to T, for $T \geq 2$, is $\frac{1}{2}\log\log T + O(1)$. See [11, pp. 449–450].) To handle the sum on $r_1 > N^{1/2}$, note that every such r_1 is $(2\log r_1)$-smooth (since $\log n < 2\log r_1$). But the number of integers $t \in (1, T]$ that are $(2\log t)$-smooth is $T^{o(1)}$, as $T \to \infty$. (This can be deduced from [21, Theorem 5.2, p. 513].) It now follows by partial summation that the sum on $r_1 > N^{1/2}$ appearing above is $O(1)$. Assembling these estimates yields the lemma for all sufficiently large n, say $n \geq n_0$.

It is clear that the lemma also holds when $20 < n < n_0$ (adjusting the implied constant appropriately), provided that there are $\gg N/\log N$ values of R for all choices of n, N. To obtain this lower bound, we consider the contribution from values of R with $r_1 = 1$. This includes all numbers $R = p$ with $p \in (N/3, 2N/3]$ and $p \equiv 1 \pmod 4$. Indeed, since $N \geq n > 20$, we have $p > N/3 > \log N \geq \log n$. And it cannot be that $p \mid N$: since $N/p \in [3/2, 3)$, for p to divide N we would need $N = 2p$, contradicting that N is odd. The proof is completed by recalling that the number of $p \in (N/3, 2N/3]$ with $p \equiv 1 \pmod 4$ is $\gg N/\log N$. Indeed, it was proved by Erdős [6] that the number of primes congruent to 1 (mod 4) in $(T, 2T]$ is $\gg T/\log T$ for all $T \geq \frac{13}{2}$. □

Lemma 5. $\gcd(z, w) = 1$.

Proof. In our algorithm, $z + wi = (u + vi)(U + Vi)$, where $u + vi$ is a product of certain of the numbers $X_\ell + Y_\ell i$, and either $U + Vi = 1$ (in the case $p = 1$) or $U + Vi = \gcd(p, s + i)$ for some square root s of -1 mod p. Note that p is not necessarily prime, but we do know that p has no prime factors up to $\log n$. When $p > 1$, the fact that -1 has a square root mod p implies that every prime factor of p belongs to the residue class 1 modulo 4.

Suppose now that the prime q divides z, w. Then $q \mid z + wi$ in $\mathbb{Z}[i]$. The Gaussian primes $X_\ell + Y_\ell i$, as well as all the Gaussian primes dividing $U + Vi$, lie above primes 1 modulo 4. Hence, $q \equiv 1 \pmod 4$, and $q = \pi\bar{\pi}$ with π and $\bar{\pi}$ nonassociated Gaussian primes. Then π and $\bar{\pi}$ both divide $z + wi$. By construction, $u + vi$ is divisible by at most one of π and $\bar{\pi}$. (For each ℓ, at most one of the two primes above ℓ divides $u + vi$, namely $X_\ell + Y_\ell i$.) So either π and $\bar{\pi}$ both divide $U + Vi$ or one of $\pi, \bar{\pi}$ divides $u + vi$ while the other divides $U + Vi$. If $U + Vi = 1$, each of these scenarios is absurd, so we may assume $U + Vi = \gcd(p, s + i)$ with s a square root of -1 mod p.

If π and $\bar{\pi}$ both divide $U + Vi$, then $q = \pi\bar{\pi} \mid U + Vi = \gcd(p, s + i)$. But then $q \mid s + i$, which is absurd. If one of $\pi, \bar{\pi}$ divides $u + vi$ while the other divides $U + Vi$, then $N(u + vi)$ and $N(U + Vi)$ are both divisible by $q = N(\pi) = N(\bar{\pi})$. Since $N(u + vi)$ is a product of primes not exceeding $\log n$, it must be that $q \leq \log n$. But $N(U + Vi) = p$ has no prime factors in $[2, \log n]$, so again we have a contradiction. □

5. Concluding remarks

It would, of course, be desirable to possess a *deterministic* polynomial-time algorithm for computing a representation of n as a sum of four squares. Such an algorithm is available when n is prime (a result of Bumby [3]) and so, via quaternion multiplication, whenever we are given the prime factorization of n.

It seems difficult to prove that a representation of n can always be found in deterministic polynomial time (without prior information on the factorization of n). This would follow from the conjecture of Heath-Brown [7] that the least prime congruent to $a \pmod q$, when $\gcd(a, q) = 1$, is $\ll q(\lg q)^2$. For reasons already discussed, we can restrict attention to odd n. For $k = 1, 3, 5, 7, \ldots$, use the AKS test [1] to decide whether $p = 2nk - 1$ is prime. Assuming the truth of Heath-Brown's conjecture (with $q = 4n$), we are certain to hit upon a prime p within $O((\lg n)^2)$ steps. Use Schoof's algorithm [19] to compute the number of $\mathbb{F}_p$-points on the elliptic curve $y^2 = x^3 - x$. If this is N, let $a = \frac{1}{2}(p + 1 - N)$. Then $a \in \mathbb{Z}$ and $p = a^2 + b^2$ for some b (see, e.g., [9, Theorem 5, p. 307]), whose value is easily found from p, a. We have that $n \mid 2nk = p + 1 = a^2 + b^2 + 1$, and we find a four-squares representation of n by computing $\operatorname{gcrd}(n, a + bi + j)$.

Unconditionally, we can show that a positive proportion of numbers can be written as a sum of four squares in deterministic polynomial time. Certainly all integers of the form $4^k + p$, where p is a prime congruent to $1 \pmod 4$, have this property. (In fact, these numbers can be quickly expressed as a sum of *three* squares!) That these numbers comprise a set of positive lower density can be proved by an easy modification of a method of Romanov [16].

Under ERH, we can do better: *Almost all* natural numbers (by which we mean asymptotically 100%) can be expressed as a sum of four squares in deterministic polynomial time. Assuming ERH, Prachar showed that for a certain absolute constant C, almost all n admit a representation in the form

$$n = p + m^2, \quad \text{where} \quad 0 \le m < (\log n)^C.$$

(This is a special case of [14, Satz 1]. The constant C could be computed from the proof but is not given explicitly in [14].) When n is congruent to $2 \pmod 4$ and not of the form $m^2 + 2$, every such representation of n has $p \equiv 1 \pmod 4$. It follows that almost all $n \equiv 2 \pmod 4$ can be expressed as a sum of four squares in deterministic polynomial time. Elementary arguments now suffice to transition from almost all $n \equiv 2 \pmod 4$ to almost all positive integers n (cf. the discussion in the middle of p. S244 of [18]).

We conclude with a word about sums of higher powers. Confirming a 1770 conjecture of Waring, Hilbert showed in 1909 that for each fixed $k \ge 2$, every positive integer

can be written as a sum of $O_k(1)$ nonnegative kth powers. Hilbert's proof goes by reducing the case of arbitrary k to Lagrange's four-square theorem (corresponding to $k = 2$). It seems interesting to note that this reduction can be made computationally effective: for every fixed $k \geq 2$, whenever one has an algorithm that runs in expected (respectively, deterministic) polynomial time for representing an arbitrary positive integer as a sum of four squares, one gets an algorithm running in expected (respectively, deterministic) polynomial time for representing an arbitrary positive integer as a sum of $O_k(1)$ nonnegative kth powers. This is clear, for instance, from the simplified solution to Waring's problem described by Dress in [5]. To make the algorithm explicit, one needs an explicit form of the "fundamental Hilbert identities". The existence of these identities is proved nonconstructively in [5] (following Ellison), but a constructive proof, due to Hausdorff and Stridsberg, can be given by means of the Hermite polynomials. Details can be found in G. J. Rieger's dissertation [15]. See also [13] for an exposition of the ideas of Rieger and Dress.[8]

Acknowledgements

We thank the referees for detailed comments on the manuscript that led to improvements in the readability.

References

[1] M. Agrawal, N. Kayal, and N. Saxena, PRIMES is in P, *Ann. of Math.* (2) **160** (2004), 781–793.

[2] E. Bach and J. Sorenson, Explicit bounds for primes in residue classes, *Math. Comp.* **65** (1996), 1717–1735.

[3] R. T. Bumby, Sums of four squares, *Number Theory (New York, 1991–1995)*, Springer, New York, 1996, pp. 1–8.

[4] D. S. Dummit and R. M. Foote, *Abstract Algebra*, third ed., John Wiley & Sons, Inc., Hoboken, NJ, 2004.

[5] F. Dress, Méthodes élémentaires dans le problème de Waring pour les entiers, Université de Provence, Marseille, 1971, *Journées Arithmétiques Françaises, Mai 1971.*

[6] P. Erdős, Über die Primzahlen gewisser arithmetischer Reihen, *Math. Z.* **39** (1935), 473–491.

[8] Using [13], the last occurrence of $O_k(1)$ can be replaced with $(2k+1)^{2000k^5}$.

[7] D. R. Heath-Brown, Almost-primes in arithmetic progressions and short intervals, *Math. Proc. Cambridge Philos. Soc.* **83** (1978), 357–375.

[8] C. Hermite, Sur la théorie des formes quadratiques. Second mémoire, *J. Reine Angew. Math.* **47** (1854), 343–368.

[9] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, second ed., Graduate Texts in Mathematics, vol. 84, Springer-Verlag, New York, 1990.

[10] J.-L. Lagrange, Démonstration d'un théoréme d'arithmétique, *Nouv. Mém. Acad. Roy. Sc. de Berlin* (1770), 123–133. Also in *Oeuvres de Lagrange* **3** (1869), pp. 189–201.

[11] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen. 2 Bände*, second edition, Chelsea Publishing Co., New York, 1953.

[12] J. Oesterlé, Versions effectives du théorème de Chebotarev sous l'hypothèse de Riemann généralisée, *Astérisque* **61** (1979), 165–167.

[13] P. Pollack, On Hilbert's solution of Waring's problem, *Cent. Eur. J. Math.* **9** (2011), 294–301.

[14] K. Prachar, Über Zahlen, die sich als Summe einer Primzahl und einer "kleinen" Potenz darstellen lassen, *Monatsh. Math.* **68** (1964), 409–420.

[15] G. J. Rieger, Zur Hilbertschen Lösung des Waringschen Problems: Abschätzung von $g(n)$, *Mitt. Math. Sem. Giessen.* (1953), no. 44, 35 pages.

[16] N. P. Romanov, Über einige Sätze der additiven Zahlentheorie, *Math. Ann.* **109** (1934), 668–678.

[17] J. B. Rosser and L. Schoenfeld, Approximate formulas for some functions of prime numbers, *Illinois J. Math.* **6** (1962), 64–94.

[18] M. O. Rabin and J. O. Shallit, Randomized algorithms in number theory, *Comm. Pure Appl. Math.* **39** (1986), no. S, suppl., S239–S256.

[19] R. Schoof, Elliptic curves over finite fields and the computation of square roots mod p, *Math. Comp.* **44** (1985), 483–494.

[20] D. Shasha, An Interview with Michael Rabin, *Commun. ACM* **53** (2010), no. 2, 37–42.

[21] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, third ed., Graduate Studies in Mathematics, vol. 163, American Mathematical Society, Providence, RI, 2015.