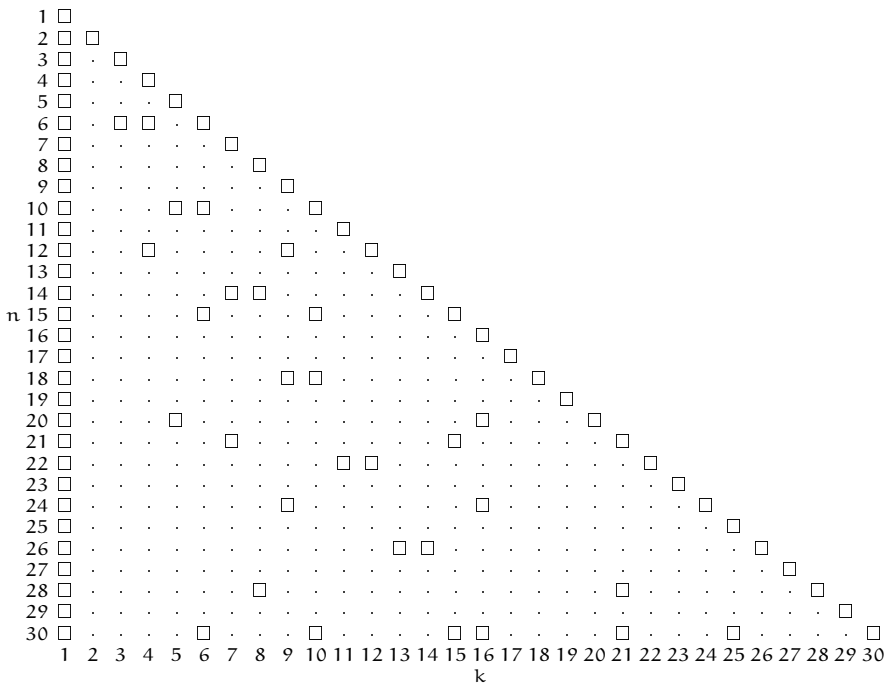
Counting self-square numbers

Definition 1 For positive integers n and k, we say that k is *self-square modulo* n if $k^2 \equiv k \pmod{n}$.

The following table shows self-square numbers modulo a few small n. The cell in the nth row and kth column contains "□" if k is self-square modulo n, and "·" otherwise. Since whether k is self-square modulo n depends only on k's residue modulo n, only the cells for $1 \le k \le n$ are shown.[1]



The objective of this note is to prove the following result on the number of squares in each row of this table.

Definition 2 For any positive integer n, let $\omega(n)$ denote the number of distinct primes that divide n.

Proposition 3 For any positive integer n, there are $2^{\omega(n)}$ distinct (i.e., non-congruent) self-square numbers modulo n.

Many of the arguments proceed by exploiting the algebraic properties of the greatest common divisor and least common multiple functions. Befitting this algebraic approach, these functions are treated as binary operations on the

[1] The rows' left-right symmetry is more apparent with this range of k than with the more usual range $0 \le k < n$.

nonnegative integers[2] and written with an infix syntax: the greatest common divisor of a and b is denoted "$a \gcd b$"; their least common multiple, "$a \operatorname{lcm} b$". These operations have lower precedence than the usual arithmetic operations; for example, "$a \gcd b + 2c$" means "$a \gcd (b + 2c)$", not "$(a \gcd b) + 2c$". Neither gcd nor lcm has precedence over the other, so when both appear it is necessary to parenthesize explicitly; for example, the mutual distributivity of these operations is rendered by:

$$a \gcd (b \operatorname{lcm} c) = (a \gcd b) \operatorname{lcm} (a \gcd c)$$
$$a \operatorname{lcm} (b \gcd c) = (a \operatorname{lcm} b) \gcd (a \operatorname{lcm} c)$$

When $a \gcd b = 1$ we say that a and b are *coprime*, and write $a \perp b$.

We will need the following properties of gcd and lcm, which we make no effort to justify, assuming them to be familiar from number theory. When they are used in proofs and calculations, they will be identified by the names given here. For any nonnegative integers a, b, and c:

$a \gcd (b \gcd c) = (a \gcd b) \gcd c$	(gcd is associative)
$a \gcd b = b \gcd a$	(gcd is commutative)
$a \gcd a = a$	(gcd is idempotent)
$a(b \gcd c) = ab \gcd ac$	(multiplication distributes over gcd)
$a \gcd 1 = 1$	(1 dominates gcd)
$a \operatorname{lcm} 1 = a$	(1 is the identity of lcm)
a divides $b \iff a \gcd b = a$	(gcd divisibility criterion)
$a \perp b \iff a \operatorname{lcm} b = ab$	(coprime factors)
$a \gcd b = a \gcd a + b$	(Euclid's step)

Euclid's step is so named because it is the observation underlying Euclid's algorithm.

From the extended version of that algorithm we also know that $a \gcd b$ can be realized as a linear combination of a and b, that is, for any nonnegative integers a and b, there exist integers s and t such that $as - bt = a \gcd b$. We will need the following refinement of this result for positive numbers.

Proposition 4 For any positive integers a and b, there exist positive integers s and t such that $as - bt = a \gcd b$.

Proof We describe a method for constructing such integers s and t.

Since a and b are positive, they are nonnegative; use the extended Euclid's algorithm to find integers s and t such that $as - bt = a \gcd b$. Then perform the following procedure with s and t: if s and t are both positive, stop; otherwise replace s with $s + b$ and t with $t + a$, and repeat.

[2] Not, note, on the positive integers. Thus we must define gcd and lcm so that 0 is an acceptable argument; the only definition which accords with the properties described is $a \gcd 0 = a$ and $a \operatorname{lcm} 0 = 0$ for any nonnegative integer a. (In particular, $0 \gcd 0 = 0 \operatorname{lcm} 0 = 0$.)

Since a and b are positive, each iteration of this procedure strictly increases s and t. By the well-ordering principle, a strictly increasing sequence of integers is eventually positive; so eventually s and t will be positive, and the procedure will terminate.

Moreover, since

$$a(s+b) - b(t+a) = as - bt\,,$$

each iteration of this procedure preserves the value of $as - bt$; so when it terminates, s and t will not only be positive but will still satisfy $as - bt = a \gcd b$, as desired. □

The following simple propositions illustrate the techniques of proof that will be used most in what follows.

Proposition 5 $a \perp a + 1$ for any nonnegative integer a.

Proof We calculate that, for any nonnegative integer a,

$$\begin{aligned} & a \gcd a + 1 \\ &= a \gcd 1 && \text{(Euclid's step, with } a, b := a, 1) \\ &= 1 && \text{(1 dominates gcd)} \end{aligned}$$

and so $a \perp a + 1$ by definition. □

In the second line of this proof, the phrase "with $a, b := a, 1$" should be read as "with a and b replaced by a and 1 respectively".[3] This notation serves to describe, when clarity requires, how to instantiate a universal statement to justify the equality at hand. A similar example appears in the next proof.

Proposition 6 For any nonnegative integers a, b, and n: if $a \equiv b \pmod{n}$, then $n \gcd a = n \gcd b$.

Proof Without loss of generality, $a \leq b$ (so that $b - a$ is a nonnegative integer, hence a suitable argument to gcd). If $a \equiv b \pmod{n}$, that is, n divides $b - a$, then by the gcd divisibility criterion, $n = n \gcd b - a$, and so

$$\begin{aligned} & n \gcd a \\ &= n \gcd b - a \gcd a \\ &= n \gcd b - a \gcd b && \text{(Euclid's step, with } a, b := b - a, a) \\ &= n \gcd b \end{aligned}$$

as claimed. □

Note that, per the precedence rules discussed above, in the second line of this proof the expression "$n \gcd b - a \gcd a$" means "$n \gcd (b - a) \gcd a$"; note

[3] It is proper to specify that a is to be replaced by a because the a in the statement of Euclid's step is a dummy variable of a universal quantification, not the same a as in the proof of proposition 5.

also that we use the associativity of gcd implicitly by not choosing between "$(n \gcd b - a) \gcd a$" and "$n \gcd (b - a \gcd a)$". Likewise in the third line.

Definition 7 For positive integers x, y, and n: the ordered pair (x, y) is a *coprime factorization of* n if $xy = n$ and $x \perp y$.

Proposition 8 A positive integer n has $2^{\omega(n)}$ coprime factorizations.

Proof Consider the following procedure: Write n as a product of powers of distinct primes. Assign each of the $\omega(n)$ prime powers that appear in that product either to x or to y. Let the value of x be the product of the prime powers assigned to x, and let the value of y be the product of the prime powers assigned to y. (Note that, if no prime powers are assigned to x, then this procedure sets $x = 1$, and likewise for y.)

The procedure involves making $\omega(n)$ choices, each between 2 options, so there are $2^{\omega(n)}$ ways to perform this procedure. The properties of prime factorizations make it easy to see that each way of performing this procedure constructs a coprime factorization of n, and that each coprime factorization of n arises from exactly one way of performing this procedure. □

Now we can begin proving proposition 3. To each k which is self-square modulo n and satisfies $1 \leq k \leq n$, assign the pair $(n \gcd k, n \gcd k - 1)$. For example, with $n = 60$:

k	$60 \gcd k$	$60 \gcd k - 1$
1	1	60
16	4	15
21	3	20
25	5	12
36	12	5
40	20	3
45	15	4
60	60	1

Proposition 9 below will show that, since each k here is self-square modulo n, the pairs assigned to them are coprime factorizations of n. Proposition 10 below will show that, moreover, every coprime factorization of n is assigned to exactly one such k. Thus this assignment establishes a one-to-one correspondence between the self-square numbers from 1 to n and the coprime factorizations of n. Since there are $2^{\omega(n)}$ coprime factorizations of n (proposition 8 above), there are $2^{\omega(n)}$ self-square numbers modulo n, from 1 to n, which is proposition 3.

Proposition 9 For nonnegative integers k and n: k is self-square modulo n if and only if $(n \gcd k, n \gcd k - 1)$ is a coprime factorization of n.

Proof First note that, for any k (self-square or not),

$$\begin{aligned}
&(n \gcd k) \gcd (n \gcd k - 1) \\
&= n \gcd n \gcd k \gcd k - 1 && \text{(gcd is associative and commutative)} \\
&= n \gcd n \gcd 1 && (k \perp k - 1 \text{ by proposition 5}) \\
&= 1 && \text{(1 dominates gcd, twice)}
\end{aligned}$$

and so $n \gcd k$ and $n \gcd k - 1$ are coprime. Thus it suffices to show that k is self-square modulo n if and only if $(n \gcd k)(n \gcd k - 1) = n$.

Next, note that, for any k (again, self-square or not),

$$\begin{aligned}
&(n \gcd k)(n \gcd k - 1) \\
&= (n \gcd k) \operatorname{lcm} (n \gcd k - 1) && \text{(coprime factors)} \\
&= n \gcd (k \operatorname{lcm} k - 1) && \text{(gcd distributes over lcm)} \\
&= n \gcd k(k - 1) && \text{(coprime factors; } k \perp k - 1 \text{ by proposition 5)} \\
&= n \gcd k^2 - k\,.
\end{aligned}$$

Thus

$$\begin{aligned}
&n = (n \gcd k)(n \gcd k - 1) \\
&\iff n = n \gcd k^2 - k && \text{(by the calculation above)} \\
&\iff n \text{ divides } k^2 - k && \text{(gcd divisibility criterion)} \\
&\iff k^2 \equiv k \pmod{n} && \text{(definition of “}\equiv\text{”)} \\
&\iff k \text{ is self-square modulo } n && \text{(definition of “self-square”)}
\end{aligned}$$

which completes the proof. □

Proposition 10 If (x, y) is a coprime factorization of n, then there exists a unique integer k with the following properties:

1. $1 \leq k \leq n$;
2. $x = n \gcd k$;
3. $y = n \gcd k - 1$; and
4. k is self-square modulo n.

Proof Let (x, y) be a coprime factorization of n.

(Existence of k.) Since $x \perp y$ and both are positive, by proposition 4 there exist positive integers s and t such that $xs - yt = 1$. Let k be the (unique) integer satisfying $k \equiv xs \pmod{n}$ and $1 \leq k \leq n$.

k satisfies property 1 by construction. For property 2, we compute that

$$
\begin{aligned}
& n \gcd k \\
&= n \gcd xs && \text{(proposition 6)} \\
&= xy \gcd xs && (xy = n \text{ by hypothesis}) \\
&= x(y \gcd s) && \text{(multiplication distributes over gcd)} \\
&= x(y \gcd yt \gcd s) && \text{(gcd divisibility criterion; } y \text{ divides } yt) \\
&= x(y \gcd yt \gcd xs \gcd s) && \text{(gcd divisibility criterion; } s \text{ divides } xs) \\
&= x(y \gcd yt \gcd yt + 1 \gcd s) && (xs = yt + 1 \text{ by construction}) \\
&= x(y \gcd 1 \gcd s) && \text{(proposition 5)} \\
&= x1 && \text{(1 dominates gcd, twice)} \\
&= x
\end{aligned}
$$

The computation demonstrating property 3 is similar. Property 4 then follows from properties 2 and 3, by proposition 9.

(Uniqueness of k.) Suppose k and j have all four properties specified. Without loss of generality, $k \geq j$ (so that $k - j$ is a nonnegative integer and can be used as an argument to gcd). Note first that then

$$
\begin{aligned}
& x \gcd k - j \\
&= n \gcd j \gcd k - j && (j \text{ has property 2}) \\
&= n \gcd j \gcd k && (\text{Euclid's step, with } a, b := j, k - j) \\
&= x \gcd k && (j \text{ has property 2}) \\
&= n \gcd k \gcd k && (k \text{ has property 2}) \\
&= n \gcd k && (\text{gcd is idempotent}) \\
&= x && (k \text{ has property 2})
\end{aligned}
$$

Similarly, by using Euclid's step with $a, b := j - 1, k - j$, we obtain

$$y \gcd k - j = y \, .$$

Thus

$$
\begin{aligned}
& n \gcd k - j \\
&= xy \gcd k - j && (xy = n \text{ by hypothesis}) \\
&= (x \operatorname{lcm} y) \gcd k - j && \text{(coprime factors)} \\
&= (x \gcd k - j) \operatorname{lcm} (y \gcd k - j) && \text{(gcd distributes over lcm)} \\
&= x \operatorname{lcm} y && \text{(as shown above)} \\
&= xy && \text{(coprime factors)} \\
&= n && (xy = n \text{ by hypothesis})
\end{aligned}
$$

and so, by the gcd divisibility criterion, n divides $k - j$, that is, $k \equiv j \pmod{n}$. Since $1 \leq k \leq n$ and $1 \leq j \leq n$ (property 1), it follows that $k = j$. □