1. Review of Modular Square Roots

The main issues of modular square roots are apparent in this multiplication table:

Table 1.1. The Multiplication Table for p = 17

·	**0**	1	2	**3**	4	**5**	**6**	**7**	8	9	**10**	**11**	**12**	13	**14**	15	16
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	0	2	4	6	8	10	12	14	16	1	3	5	7	9	11	13	15
3	0	3	6	9	12	15	1	4	7	10	13	16	2	5	8	11	14
4	0	4	8	12	16	3	7	11	15	2	6	10	14	1	5	9	13
5	0	5	10	15	3	8	13	1	6	11	16	4	9	14	2	7	12
6	0	6	12	1	7	13	2	8	14	3	9	15	4	10	16	5	11
7	0	7	14	4	11	1	8	15	5	12	2	9	16	6	13	3	10
8	0	8	16	7	15	6	14	5	13	4	12	3	11	2	10	1	9
9	0	9	1	10	2	11	3	12	4	13	5	14	6	15	7	16	8
10	0	10	3	13	6	16	9	2	12	5	15	8	1	11	4	14	7
11	0	11	5	16	10	4	15	9	3	14	8	2	13	7	1	12	6
12	0	12	7	2	14	9	4	16	11	6	1	13	8	3	15	10	5
13	0	13	9	5	1	14	10	6	2	15	11	7	3	16	12	8	4
14	0	14	11	8	5	2	16	13	10	7	4	1	15	12	9	6	3
15	0	15	13	11	9	7	5	3	1	16	14	12	10	8	6	4	2
16	0	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1

The most important thing to note is the main diagonal, which shows the square of each residue: the product of each residue by itself. I marked those in orange to help them stand out. The distinct residues on that diagonal are 0, 1, 2, 4, 8, 9, 13, 15, and 16 (which can also be written as 0, ±1, ±2, ±4, and ±8). They include the residue 0 and only half of the 16 nonzero residues: The so-called “QR (quadratic residues)”, the ones that possess square roots.

In general, for p-values that are odd primes, the number of quadratic residues is (p - 1)/2. The count makes good sense, since each nonzero residue that has a square root—called QRs (Quadratic Residues)—has two distinct square roots so the QRs show up twice on that main diagonal. Euler penned a result that lets you predict which residues are QRs, without having to draw up the whole table:

Euler's Criterion: Let p be an odd prime. The congruence

$$x^2 \equiv a \pmod{p}$$

is trivially solvable if a is a multiple of p. If p does *not* divide a then the congruence is solvable or not, i.e. a is QR or not, according as

$$a^{(p-1)/2} \equiv 1 \text{ or } -1 \pmod{p}$$

This ties in with a concept called *the Legendre Symbol*, denoted by $(a \mid p)$ and defined to equal +1 or -1 according as the congruence $x^2 \equiv a \pmod{p}$ is non-trivially solvable or not. $(a \mid p)$ is 0 in the trivial case. After Euler discovered his result, it became an efficient computational way to evaluate the Legendre Symbol, and is used that way to the present day.

These results and many others owe their discovery to the most important of all results in residue theory:

Theorem (Euler): If a and m are positive integers with GCD(a, m) = 1, *i.e.* a and m have no factors in common, then

Equation 1. $$a^{\varphi(m)} \equiv 1 \pmod{m}$$

where $\varphi(m)$ is called "the Euler Phi Function", defined for all positive integers m. A bit involved to calculate when m is not a prime number, it is easy when m is prime in which case $\varphi(m)$ is simply m - 1. Letting p be any prime integer we have:

Corollary: $a^p \equiv a \pmod{p}$

Corollary: $a^{p-1} \equiv 1 \pmod{p}$

Corollary: $a^{p-2} \equiv 1/a \pmod{p}$ [one of the best ways to calculate reciprocals mod p]

If the modulus m is large, then a^m can be astronomical in size even if a is modest. Luckily it was discovered early on that, by doing the reduction mod m as you go, you never need to work with numbers larger than m^2. Another immense efficiency was discovered: a^m for any positive integer can be calculated as a product of a small collection of squarings, *viz*. a, a^2, $(a^2)^2$, ... *etc*. For example, a^m with m being the largest number natural to a 32-bit computer such as an iPhone, can be calculated with at most 61 multiplications. And, by doing the reduction mod m after each multiplication, you never need more than double-precision (64-bit for the iPhone) arithmetic. You can imagine the savings in effort that

Euler, Legendre, Cauchy, and all gained. I think Newton would have plundered the mint for such facility!

That's pretty much the way things stood until the 20th century: There was still no good way to determine the actual *values* of square roots, even after the Euler-Legendre result let us know which residues *have* square roots. To appreciate this pathetic state of affairs, the first method I incorporated into the *Frrraction* iPhone app, before I found the work of Cipolla and Pocklington, had to be a brute-force search: Given an odd prime p and a residue a with (a | p) = +1, it simply generated all integers from 1 to the integer square root of p—a value quite easily computed—and stopped when it found one whose square (mod p) was equal to the given a. It was perfectly satisfactory for square-root values up to the tens of millions, but became annoyingly slow for larger results—and you never know ahead of time how big the result might be, only that it won't exceed m. Example: The square root of 18 (mod 2,147,483,647) is 2,147,287,039. I hadn't expected such a huge result and it took a long time using my early Frrraction algorithm, but no time at all using *Podlington Case 1* —when implemented using the powers-of-powers-of-2 trick— on the same iThing!

Effective Square-Root Algorithms

Since 1900 a number of decent square-root algorithms have been discovered. The simplest ones are restricted to special cases. Three of my favorite are these:

Pocklington Case 1: If p is an odd prime with $p \equiv 3 \pmod 4$ and r is QR, then the square root of r is $\pm r^{(p+1)/4}$.

Pocklington Case 2a: If p is an odd prime with $p \equiv 5 \pmod 8$, r is QR, and $r^{(p-1)/4} \equiv 1$, then the square root of r is $\pm r^{(p+3)/8}$.

Pocklington Case 2b: If p is an odd prime with $p \equiv 5 \pmod 8$, r is QR, and $r^{(p-1)/4} \equiv -1$, then let $s \equiv (4r)^{(p+3)/8}$. The square root of r is $\pm 1/2 \cdot s$ or, if s is odd, then the square root of r is $\pm 1/2 \cdot (s + p)$.

That's as far as I go with Pocklington. His Case 3 seems far more baroque than the perfectly general one I prefer:

The Cipolla Algorithm: If p is an odd prime and r is QR, then the square roots of r are given by $\pm\left(a+\sqrt{a^2-r}\right)^{(p+1)/2}$ where a is chosen to be any residue (mod p) for which $a^2 - r$ is QNR.

The parameter a is an artifact—a mathematical catalyst, really. It is essential to the method but due to Cipolla's insightful definition it does not appear in the final result. In practice it is not unique. Some choices of a produce one the square roots of r, others produce the other square root.

2. Introduction to Imaginary Modular Square Roots

As we saw in Section 1, computing square roots of quadratic residues was not exactly straight-forward, so we might expect matters to be even more interesting when we turn to quadratic *non*residues–since they are the residues traditionally considered to not possess square roots. This puts residues $\mathbf{Z}_m$ where the integers $\mathbf{Z}$ were, back in the days before the complex field $\mathbf{C}$ and $i = \sqrt{-1}$, when negative numbers had no square roots.

For a quick start, Table 2.1 is a complete square root table for residues mod 17. On the righthand side of the table are the familiar square roots for the quadratic residues 1, 2, 4, 8, -8≡9, -4≡13, -2≡15, and -1≡16. The lefthand side shows all the square roots for the quadratic nonresidues.

The QR side of the table is easier to use. To find the square root of a quadratic residue, find it in the header row, and look down to the ±m row to see its two square roots. For example, the square roots of 8 are 5 and -5.

Table 2.1.
QNR Square Root Table for mod p=17

p= 17	√3	√5	√6	√7	√-7	√-6	√-5	√-3
√3	±1	±8	±6	±5	±3	±7	±2	±4
√5	±2	±1	±5	±7	±6	±3	±4	±8
√6	±3	±7	±1	±2	±8	±4	±6	±5
√7	±7	±5	±8	±1	±4	±2	±3	±6
√-7	±6	±3	±2	±4	±1	±8	±5	±7
√-6	±5	±6	±4	±8	±2	±1	±7	±3
√-5	±8	±4	±3	±6	±7	±5	±1	±2
√-3	±4	±2	±7	±3	±5	±6	±8	±1

QR Square Root Table for mod p=17

p= 17	√1	√2	√4	√8	√-8	√-4	√-2	√-1
1	±1	±6	±2	±5	±3	±8	±7	±4

The QNR side of the table works the same as the QR side, with the added complication that the header column has $(p - 1)/2$ different entries, not just the simple single entry 1. To find the square root of a quadratic nonresidue, locate it in the header row of the lefthand table, then look down that column to your choice of a row–any choice is as good as any other. Note the two residues $\pm m$ shown in that entry. Then look over to the left at the header column for that row; it gives the imaginary multiplier v. The desired square root is the product $\pm m$ times v. For example, one of the forms for the two square roots of 7 is $\pm 5 \cdot \sqrt{3}$. This is easily confirmed by squaring the proposed square root:

$$(\pm 5\sqrt{3})^2 \equiv 25 \bullet 3 \equiv 8 \bullet 3 \equiv 24 \equiv 7 \pmod{17}$$

See? just as we set out to find: √7 is ±5 · √3. Somewhat disconcertingly, √7 is also ±2 · √6 and ±6 · √-5 to mention just a few of the alternative expressions provided by Table 2.1.

The "disconcerting ambiguity"

The resolution of that ambiguity is quite easy but first, let's see it in the familiar case of ordinary complex numbers, $\mathbf{C}$. Did you ever wonder why the imaginary unit i turned out to be √-1 instead of, say, $\sqrt{-\pi}$ or √-49 or the square root of any other negative number?

Neither did I, and nobody ever mentioned it to me even in graduate school. (I did used to tease my graduate students by saying, "Sure, you know that i is a square root of -1, but all numbers have *two* square roots. Do you know *which* of the two i is?" Logically, Euler or Cauchy or whoever could just have well have chosen i to be $\sqrt{-\pi}$ or whatever–all the alternatives are interchangeable as simple real multiples of each other. √-49 for instance is just 7 · √-1 or -7 · √-1, no big deal in either case–and we now know that it doesn't even matter whether we use the +7 or the -7. The only cost would have been the practical inconvenience of having accept a scaled version of the imaginary axis of the complex plane, carrying along a non-unity scale factor that in most cases wouldn't even be expressible by a finite number of digits. Some choices wouldn't have been so bad: √-4 for example, since √-4 is 2 · √-1 or -2 · √-1. Of course, circularly polarized electromagnetic fields would then be elliptically polarized. And wherever i shows up in a complex expression, we would have to replace it by $0.5 \cdot i_{new}$ by $-0.5 \cdot i_{new}$. Awkward, but not illogical.

Returning now to the residue field $\mathbf{Z}_p$. As with all those alternatives for $\mathbf{C}$, there are alternative imaginary units for $\mathbf{Z}_p$: Eight of them in the case of $\mathbf{Z}_{17}$: √3, √5, √7, *etc.*, and (p-1)/2 of them for $\mathbf{Z}_p$ when p is an odd prime. As in the discussion of multiples of i they are not independent of each other. They are all just the offspring of a single *imaginary principle*–carriers of that one principle, if you will. We could choose any one of those carriers, √3 for example, and eliminate all the rest as multiples of the chosen one: Thus,

$$2 \cdot \sqrt{5} \equiv 3 \cdot \sqrt{6} \equiv 7 \cdot \sqrt{7} \equiv 6 \cdot \sqrt{-7} \equiv 5 \cdot \sqrt{-6} \equiv 8 \cdot \sqrt{-5} \equiv 4 \cdot \sqrt{-3} \equiv \sqrt{3} \pmod{17}$$

The same could be done just as well with any of them. For a second example, multiply all those congruences through by the reciprocal of 2 to express all of them as multiples of √5:

$$2^{-1} \cdot 2 \cdot \sqrt{5} \equiv 2^{-1} \cdot 3 \cdot \sqrt{6} \equiv 2^{-1} \cdot 7 \cdot \sqrt{7} \equiv 2^{-1} \cdot 6 \cdot \sqrt{-7} \equiv \ldots \equiv 2^{-1} \cdot \sqrt{3} \equiv \sqrt{5} \pmod{17}$$

All residues of odd primes are units (have reciprocals), and 17 is an odd prime, so such expressions can be derived from Table 2.1 for any of the entries of the first column, and they all reduce with a little arithmetic to the form $\pm m \cdot v$. SInce $2^{-1} \equiv 9 \pmod{17}$ the above alternative forms of √5 are:

$$1 \cdot \sqrt{5} \equiv -7 \cdot \sqrt{6} \equiv -5 \cdot \sqrt{7} \equiv 3 \cdot \sqrt{-7} \equiv \ldots \equiv 9 \cdot \sqrt{3} \equiv \sqrt{5} \pmod{17}$$

Of course, these congruences are more easily confirmed than derived. For example, is $-7 \cdot \sqrt{6}$ truly congruent to $\sqrt{5}$? Square both sides and compare: $(-7 \cdot \sqrt{6})^2 \equiv 49 \cdot 6 \equiv 5 \equiv (\sqrt{5})^2 \pmod{17}$. Truly.

Why isn't $\sqrt{-1}$ among the choices for the imaginary carrier v?

Sometimes it is, sometimes it isn't. It all depends upon the modulus p. Remember: The discriminator between residues that have square roots and those that don't is the property of being a QR (quadratic residue) or a QNR (quadratic nonresidue), not the algebraic ± sign. The Legendre Symbol $(a|p)$, computable as

$$(a|p) = a^{\frac{p-1}{2}} \pmod{p} \text{ for odd prime } p$$

tells which kind a is. If $a \neq 0$ then the only values $(a|p)$ takes on are +1 and -1. If +1 then a is QR, otherwise it is QNR.

The square w of the imaginary carrier v must be QNR (or else $v \equiv \sqrt{w}$ would be a normal residue, not an imaginary residue). So what about -1? Easy to answer: -1 raised to an even power is +1. And -1 raised to an odd power is -1. The question becomes: Is half of p - 1 even or odd? If even, then:

$$\frac{p-1}{2} = 2k \text{ for some integer k, so}$$
$$p = 4k+1 \text{ for some integer k, so}$$
$$p \equiv 1 \pmod{4}$$

Neat, eh? How does p = 17 measure up? Well, 17/4 is 4+1/4, so 17 ≡ 1 (mod 4). What's the conclusion? As a residue class for p = 17, -1 is QR, so $\sqrt{-1}$ does not carry the imaginary principle for p = 17.

The other values for any integer mod 4 are 0, 2 and 3. p must be odd, so 0 and 2 are ruled out, leaving values of p congruent to 3 (mod 4) as the the moduli that offer $\sqrt{-1}$ as a choice for the imaginary carrier v.

p ≡ 3 (mod 4) if and only if p ≡ 4k + 3 for some integer k.

The first few of those values are 3, 7, 11, 15, 19, 23, *etc*. 15 is out, because it's not prime. 11 is a nice little number, let's look at Table 2.2, the p=11 version of Table 2.1.

Sure enough, -1 is QNR in $\mathbf{Z}_{11}$, so $v = \sqrt{-1}$ is a valid choice. Another ready observation is that, like the integers $\mathbf{Z}$ and the reals $\mathbf{R}$, $\mathbf{Z}_{11}$ comes close to using algebraic sign to distinguish between QR and QNR: If residue class r is QR the $-r$ is QNR, and *vice versa*. That's quite the opposite of $\mathbf{Z}_{17}$ where if r was QR then so was $-r$, and *vice versa*.

The way I determine which r-values to use for the headers of tables like Table 2.1 and Table 2.2 is to evaluate the Legendre Symbol for each residue class, one by one, and choose the r-values whose $(r|p)$ is -1.

Filling in the diagonal entries is always as easy as noting that $\sqrt{r} = \sqrt{r}$. The off-diagonals call for a bit more effort.

Table 2.2. QNR Square Roots

p= 11	√2	√6	√-4	√-3	√-1
√2	1		±3		
√6		1			
√-4			1		
√-3				1	
√-1			±9		1

Calculating $\sqrt{r}$ for QNR values of r

I got this whole imaginary residues idea from that Wikipedia presentation of Cipolla's Algorithm. As its amazing first step towards finding square roots of a QR residue r, it hunts until it stumbles upon a residue a such that $a^2 - r$ is QNR. I adapted this to the goal of this section by seeking a QNR a-value such that $A = a + r$ is QR, with the idea that $v = \sqrt{a}$ would then happily carry the imaginary principle. Following Cipolla, we plan to compute x_0 :

$$x_0 = \left(\sqrt{r+a} + \sqrt{a}\right)^{(p+1)/2} = \left(A + v\right)^{(p+1)/2}$$

letting $A = \sqrt{(r + a)}$. Continuing along the trail that was formed in 1907:

$$x_0^{\,2} = \left(A+v\right)^{p+1} = (A+v)(A+v)^p$$

but in any Z_p it is true that $(x+y)^p = x^p + y^p$.

Moreover, by Legendre, $x^p = x$ if x is QR,

and $x^p = -x$ if y is QNR so

$(A+v)^p = A - v$ and

$(A+v)(A+v)^p = (A+v)(A-v) = A^2 - v^2.$

Bingo. $x_0^{\,2} = (r+a) - a = r$

So the x_0 we compute from $x_0 = (A + v)^{(p+1)/2}$ is the desired imaginary square root of r.

For example, let's compute the entry of Table 2.2 for r = -4 (same as 7 (mod 11)). To start, we need a QNR value for a such that $r + a$ is QR. The following little table, obtained by liberal use of the Legendre Symbol to know which $\mathbf{Z}_{11}$ residues are QR and which are not, shows the possibilities:

Table 2.3. Possibilities for a

$a = v^2$		$r + a = A^2$	
2	QNR	9	QR
6	QNR	2	QNR
7	QNR	3	QR
8	QNR	4	QR
10	QNR	6	QNR

a = 6 and a = 10 are ruled out. a = 7 is the trivial case $\sqrt{-4} = \pm 1 \cdot \sqrt{-4}$, so let's go with a = 2 for this first example. After calculating the square root $A = \sqrt{(r + a)} = \sqrt{9} \equiv 3$, the

calculation of x_0 goes like this:

Table 2.4. Calculation of √-4 (mod 11)

$(A + v)^n$ with v = √2		
$(3 + v)^1$	≡	$3 + 1 \cdot v$
$(3 + v)^2$	≡	$0 + 6 \cdot v$
$(3 + v)^4$	≡	$6 + 0 \cdot v$
$x_0 \equiv (3 + v)^6$	≡	$0 + 3 \cdot v$ = √-4 = √7

[Hint: Calculate $(a + bv) \cdot (c + dv)$ as you would if v were i, except v^2 is 2 instead of -1.]

▣

We noted above that the QNRs 6 and 10 were ruled out of the running for v^2 because the corresponding values of $r + a$ were also QNR. This is a shortcoming of this method for calculating imaginary square roots, not a shortcoming of v = √6 or √10 as "carriers of the imaginary principle". In fact, the square roots of all QNRs are simple real multiples of each other and can therefore be freely substituted for one another. For instance, consider: $\sqrt{2} = a\sqrt{-1} \Leftrightarrow a^2 = \frac{2}{-1} \Leftrightarrow a = 3$. So, from the Table 2.4 result √7 ≡ 3√2 we see that an alternate expression is √7 ≡ 3 · 3√-1 ≡ 9√-1 (mod 11), a result easily verified by squaring both sides. The congruence $\sqrt{v} \equiv m \cdot \sqrt{w} \pmod{p}$ always yields a QR value for m when v and w are both QNR, so this observation provides the next section.

A simpler method for calculating √r for QNR values of r

Every entry of an imaginary square root table like Table 2.1 is just another congruence equation of the form

Equation 2.1. $\sqrt{r} \equiv m \cdot \sqrt{v} \Rightarrow m = \sqrt{(v/r)}$

where √r is a column header and √v is a row header, to be solved for m, given r and v. If m is a solution, then so too is -m., which accounts for both solutions expected of a quadratic residue equation.

As examples, we complete the bottom row of Table 2.2 by taking v to be -1 (same as 10) and r to be 2, 6, and -3 (same as 8), yielding m = 3, 4, and 5, respectively:

Table 2.5 last row.

p = 11	√2	√6	√r	√-3	√-1
√-1	±3	±4	±m	±5	±1

In these cases of $p \equiv 3 \pmod 4$, where the imaginary carrier can be the square root of -1, it's hard to resist the temptation to use the symbol i instead of v for the imaginary unit.

▣