CHAPTER III: The 20th Century

In the previous two chapters, we explored some of the history of Fermat's Last Theorem, discussed some of the mathematics involved, and walked through a few special cases. We will now sketch out the eventual proof of the theorem. The logic of the proof is quite simple, but the mathematical terms/concepts/objects/theory involved are quite advanced. So in the first two sections of Chapter 3, we will cover the basics. This will admittedly be done with not as much depth as the topics deserve - each one could easily comprise a graduate class in their own right.

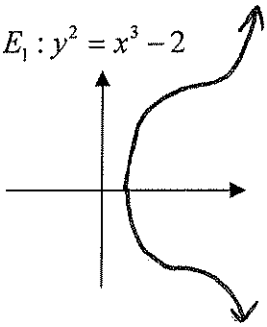
It is sort of a misnomer to call this chapter "The 20th Century". It's true that the developments we will outline took place in that century, but much of the mathematics involved came much earlier. As we saw in Chapter 2, the attempts at proving Fermat's Last Theorem ran into a problem in unique factorization. As the value of n increased, the difficulty grew very rapidly and before long it was clear that a new approach was necessary. That new approach was taken in 1986 by Gerhard Frey, but it involved mathematical objects that had been studied for centuries.

Section 1: Elliptic Curves

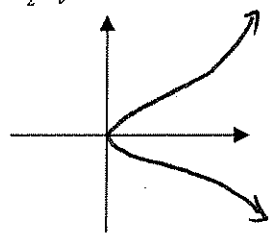
Speaking of misnomers, elliptic curves are sort of misnamed as well. An elliptic curve is a cubic equation of the form $y^2 = x^3 + ax^2 + bx + c$.

Example 3.1.1 A few examples of elliptic curves are (with their graphs):

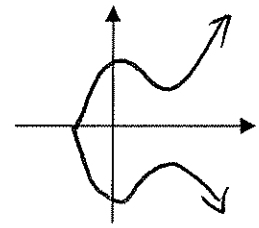
$E_1 : y^2 = x^3 - 2$



$E_2 : y^2 = x^3 + x$



$E_3 : y^2 = x^3 - 4x^2 + 16$



Note that the graph of an equation like this is NOT an ellipse. So what gives? To explain this name, we need to delve into a good bit of history.

Solving equations like $y^2 = x^3 - 2$ actually goes back as far as Diophantus. However, around 250 AD variables and equations were not used like they are today. In fact, Diophantus studied his equations using words such as "Find a number for which its cube reduced by two is a square." Once algebraic notation was developed in the 16th century (by Viète among others), René Descartes and Fermat (in the 17th century) began the field of mathematics now known as analytic geometry. It was essentially a marriage between the classical field of geometry and the new algebra.

René Descartes (1596-1650) was one of the most influential thinkers in the history of mankind yet he only published one work on mathematics. He was also a philosopher and scientist. Around 1633, Descartes was working on a book about the nature of the universe when he heard of the fate of Galileo. (Recall that The Church had condemned Galileo for advancing a theory that the Church viewed as contradictory to the Bible.) Descartes abandoned his work, and started working on *A Discourse on the Method of Rightly Conducting the Reason and Seeking Truth in the Sciences*. This philosophical treatise explained Descartes' views on science in general. It contained three appendices, the third of which was entitled *La geometrie*, and it was in this appendix that Descartes outlined his version of analytic geometry.

In *La geometrie*, Descartes described his coordinate system, which we now call the Cartesian coordinate system in his honor. His idea that you could identify every point in the plane with two coordinates was revolutionary. It allowed problems of geometry to be translated into problems involving variables that algebra could easily handle. There is an interesting anecdote about how Descartes came up with this idea. Legend has it that he was watching a fly crawling on his ceiling. He realized that he could describe the position of the fly by noting how far it was from each of the two nearby walls. Whether this is true or not, this new ability to relate a group of points to an algebraic equation was a powerful tool in advancing both algebra and geometry.

Fermat developed analytic geometry around the same time as Descartes, but from the opposite point of view. Whereas Descartes started with points and attempted to construct curves geometrically and then derive their equations, Fermat would start with the equation and derive the graph. It is this approach that more closely resembles our current method. Indeed the differing approaches soon caused a controversy. After receiving a copy of *La geometrie* sent by fellow Frenchman Jean Beugrand, Fermat paid it little attention. Finally, a friend and colleague Marin Mersenne asked Fermat to render an opinion of Descartes' work. Fermat said the Descartes was "groping about in the shadows" and a feud began between the two brilliant men.

As mathematicians (following Descartes' approach) set out to determine the circumference of the graph of an ellipse, they stumbled across functions of the form $y^2 = x^3 + ax^2 + bx + c$ and therefore named them "elliptic curves". But it was Fermat who then took equations of that form and tried to study their solutions. What we now call algebraic geometry is simply the study of the solutions to algebraic curves (like elliptic curves). To a modern algebraic geometer, elliptic curves are *abelian varieties of genus one,* but we'll stick to calling them elliptic curves.

Back to our examples. Each of them has solutions in the integers, for example

E_1 has the solutions (3,5) and (3,-5)

(Do you recognize this equation? See Exercise 1.2.7)

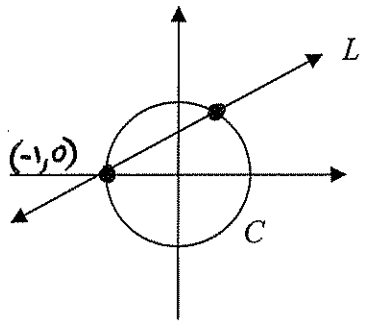
E_2 has the solution (0,0)

E_3 has the solutions (0,4) and (4,4).

These solutions were just found by trial and error, but what if we wanted to find more solutions? What if we wanted rational solutions? That is algebraic geometry. Let's illustrate how this works with a simple example.

Example 3.1.2 Consider the equation $x^2 + y^2 = 1$. This is obviously just the circle C of radius 1 centered at the origin (0,0). We want to find all rational solutions to this equation. In other words, speaking geometrically, we want to find all points on the circle with rational coordinates. There are four obvious ones: $(1,0)$, $(0,1)$, $(-1,0)$, and $(0,-1)$. But what about all the others?

Let m be any rational number. Consider the line L going through the point $(-1,0)$ with slope m. This line is given by the equation $L: y = m(x+1)$.



It is clear from the graph that the intersection $C \cap L$ consists of exactly two points, one of which is $(-1,0)$. We want the other point. To find the intersection of C and L, we just need to solve the system of equations

$$x^2 + y^2 = 1 \text{ and } y = m(x+1).$$

If we substitute the second equation into the first and simplify, we get

$$\begin{aligned} x^2 + (m(x+1))^2 &= 1 \\ x^2 + m^2(x^2 + 2x + 1) &= 1 \\ (m^2+1)x^2 + 2m^2x + (m^2 - 1) &= 0 \end{aligned}$$

This is just a quadratic equation, so we could use the quadratic formula to solve it. But since we already know one solution ($x = -1$), we can divide the polynomial on the left by $x+1$ to find the other root. Using long division,

$$\frac{(m^2+1)x^2 + 2m^2x + (m^2-1)}{x+1} = (m^2+1)x + (m^2-1),$$

which means that the other solution is $x=\frac{1-m^2}{1+m^2}$. If we substitute that value into $y=m(x+1)$, we find the other coordinate $y=m\left(\frac{1-m^2}{1+m^2}+1\right)=\frac{2m}{1+m^2}$. Thus, for every rational number m, $\left(\frac{1-m^2}{1+m^2},\frac{2m}{1+m^2}\right)$ is a rational solution to $x^2+y^2=1$. This in fact yields all rational solutions (except $(-1,0)$), since if (x_1,y_1) is a rational solution to $x^2+y^2=1$, the line through (x_1,y_1) and $(-1,0)$ will have rational slope.

Theorem 3.1.3 Every rational point on the circle $x^2+y^2=1$ (except $(-1,0)$) is of the form

$$\left(\frac{1-m^2}{1+m^2},\frac{2m}{1+m^2}\right) \text{ for some rational number } m.$$

This is what algebraic geometers do for many different curves. We can even relate this to our work with Pythagorean triples. Since m is a rational number, let $m=\frac{v}{u}$. Substituting this into Theorem 3.1.3 and simplifying, we get $\left(\frac{u^2-v^2}{u^2+v^2},\frac{2uv}{u^2+v^2}\right)$. This means that $\left(\frac{u^2-v^2}{u^2+v^2}\right)^2+\left(\frac{2uv}{u^2+v^2}\right)^2=1$. Clearing denominators gives us our characterization of primitive Pythagorean triples from Section 2.1.

Exercise 3.1.4 Find all rational solutions to the equation $y^2=x-1$.

Let's try this same procedure on $y^2=x^3+17$. This equation has several integer solutions: $(2,5)$, $(-2,3)$, and $(-1,4)$. We draw lines through the point $(-2,3)$ and see what other points we find. Points on this line satisfy the equation $y=m(x+2)+3$ and if we substitute that into $y^2=x^3+17$, we get

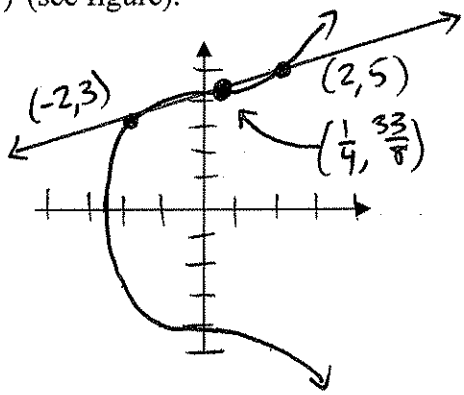
$$\begin{aligned}(m(x+2)+3)^2&=x^3+17\\(mx+2m+3)^2&=x^3+17\\m^2x^2+2m(2m+3)x+(2m+3)^2&=x^3+17\\0&=x^3-m^2x^2-(2m(2m+3))x+17-(2m+3)^2\\0&=(x+2)(x^2-(m^2+2)x-(2m^2+6m-4))\end{aligned}$$

In this case, the second factor is not quite as easy to deal with. As a quadratic factor, it MAY have rational solutions, but there is no guarantee of that. But if we knew one of the roots of the quadratic factor was rational, we would know the other was also. So perhaps we chose the

wrong line to intersect with our curve. We only knew one point the line $y = m(x+2)+3$. If instead we use the line connecting $(2,5)$ and $(-2,3)$ we get the line $y = \frac{1}{2}x+4$. Intersecting THIS line with the curve $y^2 = x^3+17$ yields the cubic equation $0 = x^3 - \frac{1}{4}x^2 - 4x+1$. Since we know two of the rational solutions to this equation, we can be certain the third solution is rational also. This equation has $x = 2$ and $x = -2$ as two roots, so we can factor it as

$$x^3 - \tfrac{1}{4}x^2 - 4x+1 = (x-2)(x+2)(x-\tfrac{1}{4}).$$

Of course, by eliminating variable slope, we are not finding ALL rational points on $y^2 = x^3+17$, but we did find a new one $(\frac{1}{4}, \frac{33}{8})$ (see figure).



Using this new point (and the symmetry of the graph to find more rational points), we can repeat this procedure and find many rational points on $y^2 = x^3+17$.

Exercise 3.1.5 We now know of many rational points on $y^2 = x^3+17$. Given were $(2,5)$, $(-2,3)$, and $(-1,4)$. We found $(\frac{1}{4}, \frac{33}{8})$. Since the curve is symmetric with respect to the x-axis, $(2,-5)$, $(-2,-3)$, $(-1,-4)$, and $(\frac{1}{4}, -\frac{33}{8})$ are also rational points. Repeat the above to find two more.

So we now know we can find infinitely many rational points on $y^2 = x^3+17$. But can we find all of them? Are there rational points on $y^2 = x^3+17$ that cannot be found in this way…given enough time? Nope.

Theorem 3.1.6 (Mordell's Theorem) Let $E: y^2 = x^3 + ax^2 + bx + c$ be an elliptic curve where a, b, c are integers such that the *discriminant*

$$\Delta(E) = -4a^3c + a^2b^2 - 4b^3 - 27c^2 + 18abc$$

is not zero.* Then every rational point on E can be found using a finite number of rational point in the method described above.

(* If $\Delta(E) = 0$, then E contains a double or triple root and the curve will either cross itself or have a cusp. These curves are more difficult to handle.)

Exercise 3.1.7 Who was Mordell? Write 2-3 paragraphs about him/her/them.

Exercise 3.1.8 Above, we stated "Since we know two of the rational solutions to this equation, we can be certain the third solution is rational also." Show this. In other words, show that for any cubic equation, if you know two solutions are rational, then all three are.