DERIVING THE EQUATION OF AN ELLIPSE

The definition of ellipse given in the textbook is

An ellipse is the set of all points in a plane such that the sum of the distances to two fixed points is a constant.

More explicitly, suppose we are given a plane, two points \( F_1 \) and \( F_2 \) in the plane, and a constant \( L > F_1 F_2 \). (We are using the standard geometric notation for the length of a line segment: \( F_1 F_2 \) denotes the distance from \( F_1 \) to \( F_2 \).) Then the set of all points \( P \) such that

\[
PF_1 + PF_2 = L
\]

(1)

is an ellipse. The points \( F_1 \) and \( F_2 \) are called the foci of the ellipse (singular: focus), and the midpoint of the line segment \( F_1 F_2 \) is its centre. (It might be helpful to take a glance at Figure 1 here. Concentrate on the points \( F_1, F_2 \) and \( P \) for now.)

![Figure 1: The geometry of an ellipse](image)

This definition of an ellipse is, of course, a geometric one. Exercise 62 in Chapter 12 of the textbook asks you to make the connection between this geometric definition and algebra by deriving an algebraic equation for an ellipse. It is not too hard to write down such an equation; the difficulty
is to re-write it in the most useful form. If you are having difficulty with the algebra, here is an outline that you can try to follow. The class website suggests a numerical example from the handout; you should probably work through that example before doing this exercise.

1. Set up a co-ordinate system in which the centre of the ellipse is at the origin, and the $x$-axis passes through the two foci. Call the distance from the origin to either focus $c$, so that the foci have co-ordinates $(-c, 0)$ and $(c, 0)$. Let $P = (x, y)$ be an arbitrary point on the ellipse. Explain why Equation (1) can be expressed algebraically as

$$\sqrt{(x - (-c))^2 + (y - 0)^2} + \sqrt{(x - c)^2 + (y - 0)^2} = L. \quad (2)$$

2. Show that subtracting the second square root in Equation (2) from both sides of the equation and then squaring leads to

$$(x + c)^2 + y^2 = L^2 - 2L\sqrt{(x - c)^2 + y^2} + (x - c)^2 + y^2, \quad (3)$$

and that this equation can be rewritten as

$$2L\sqrt{(x - c)^2 + y^2} = L^2 - 4cx \quad (4)$$

3. Show that squaring Equation (4) and re-arranging leads to

$$4(L^2 - 4c^2)x^2 + 4L^2y^2 = L^2(L^2 - 4c^2). \quad (5)$$

4. Explain why we have to take $L > F_1F_2$ in the definition of an ellipse, and why this means that the quantity $L^2 - 4c^2$ is positive.

5. Explain why we are allowed to divide Equation (5) by $L^2(L^2 - 4c^2)$, and show that after this division the equation can be written in the standard form

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad (6)$$

where we have defined

$$a := \frac{L}{2} \quad \text{and} \quad b := \frac{\sqrt{L^2 - 4c^2}}{2}. \quad (7)$$

6. Use Equation (6) to find the co-ordinates of the points $P_1$ and $P_2$ where the ellipse intersects the co-ordinate axes. (See Figure 1.) Hence give a geometric interpretation for the values $a$ and $b$. 

2
The computation was a lot of work, but it was worth it in the end! We have written the equation of the ellipse in a form from which we can read off the geometric information we need to sketch the ellipse quickly and efficiently—at least for an ellipse centred at the origin of the co-ordinate system. If the centre of the ellipse is at some other point, say the point \((h,k)\), we can apply our ‘shifting rule’ to find the equation. It will be

\[
\frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1.
\]

(8)

For future reference, it may be useful to note that the definitions of \(a\) and \(b\) in Equation (7) imply

\[
c^2 = a^2 - b^2.
\]

(9)

Although we arrived at this last equation algebraically, it is really a relation between three geometric distances. Can you see a way of deriving this relationship geometrically? (Hint: Consider the fact that \(F_1P_2 + F_2P_2 = L\).)