THE UNIVERSITY
of LIVERPOOL

בס"ד

Dissertation
Affine Area Parallels and Symmetry Sets

Daniel S. Littlestone

January 7, 2005

In loving memory of Papa Tony

יונה ב"ר שמואל הלוי ז"ל

Contents

List of Figures

5

6

1 Introduction

This study is a continuation of my last work concerning area functions and singularity theory. The idea of the last piece of work was two fold. Firstly, to introduce some of the more important concepts involved with singularity theory, such as contact, height functions and distance-squared functions, using specific examples, and secondly to define and discuss the significance of area functions, specifically investigating affine area parallels and symmetry sets. The latter part of that study consisted of a fairly substantial report regarding equilateral triangles and some important results were also presented relating to closed plane curves.

This dissertation is an extension of the second part of the mini dissertation. We can split up the work on closed plane curves into two categories. That of convex plane curves where the curvature is continually positive, and that of non-convex plane curves, which have an even number of changes of sign of curvature, and therefore contain inflections. In the last study convex plane curves were discussed in some detail, but more results are presented in this report, concerning the classification of the area function. The discussion of non-convex plane curves is rather extensive, and so a whole chapter is devoted to this topic.

Looking at non-convex plane curves adds a lot more complications to the subject, which are discussed in detail. First of all, we find that a chord can cut the curve in more than two places, and so areas outside of the closed curve can be separated by the chord, as well as areas inside. This idea brings us on to the discussion of algebraic area, as opposed to numerical area that was required for convex plane curves, and opens up possibilities of a zero area, where the area enclosed by the chord outside of the curve is equal to that inside the curve. With all possibilities of non-convex curves, the local area parallel, near the inflections, is calculated using power series, by assigning local variables to the significant parts of the curve. The results are also presented geometrically, by way of examples, both local and global.

The discussion on equilateral triangles in the previous study was relatively straight forward for reasons similar to that of convex curves. Clearly, there are no non-convex possibilities involved with triangles. For this reason, I have included a relatively significant chapter discussing quadrilaterals. Similar to the closed plane curves we can split up the subject into two: convex and non-convex. We can simplify our investigation by comparing the quadrilaterals with closed plane curves, by rounding off the vertices. Then the comparison is more acceptable.

Furthermore, we find there are other complications involved, specifically concerning the possibility of affine transformations. We know that any three points in the plane can be affinely transformed onto any other three points in the plane. For this reason, any triangle can be affinely transformed onto any other triangle, which means that a study of equilateral triangles alone is sufficient. With quadrilaterals, however, we only have a limited number of degrees of freedom concerning affine transformations, and so there may be more possibilities to consider. In all cases, the area parallel and symmetry set are discussed.

This field of mathematics is a rather modern topic of study. Indeed, the preponderance of documents concerning affine area parallels, symmetry sets and medial axes have been published within the last five years. A fairly substantial investigation has taken place recently by Niethammer of the Georgia Institute of Technology, Betelu of the University of North Texas, Sapiro of the University of Minnesota, Tannenbaum of the Georgia Institute of Technology and Giblin of the University of Liverpool, in their paper entitled Area-Based Medial Axis of Planar Curves, where a new definition of affine invariant medial axis of planar curves is introduced. This paper was accepted in February last year. Credit must also go to Leandro Estrozi for his work on this subject, specifically concerning his program that plots area parallels and symmetry sets, which has become invaluable during this study.

Thanks must go to my family and various members of the department for their continual support during the last year, and special thanks must go to my tutor Peter Giblin for all of his help and encouragement, particularly during the research for this dissertation and the mini-dissertation. My enjoyment of mathematics has developed more than I thought possible over the last year, largely due to his guidance and support, and I feel privileged to have had this opportunity to work with him so closely over the last year.

2 Closed Plane Curves

2.1 A Review of Convex Curves

Let us begin by recalling some of the information from the Mini-Dissertation concerning convex curves. Using the information that we have already noted, we will be able to extend the study into a more general form, specifically looking at non-convex curves, that is, closed curves which contain inflections. So, initially let us consider a convex closed plane curve which is defined by a polar equation $r = r(\theta)$, with $r > 0$. The pole X is inside the curve, with a chord at an angle α from this pole, as is shown in the diagram.
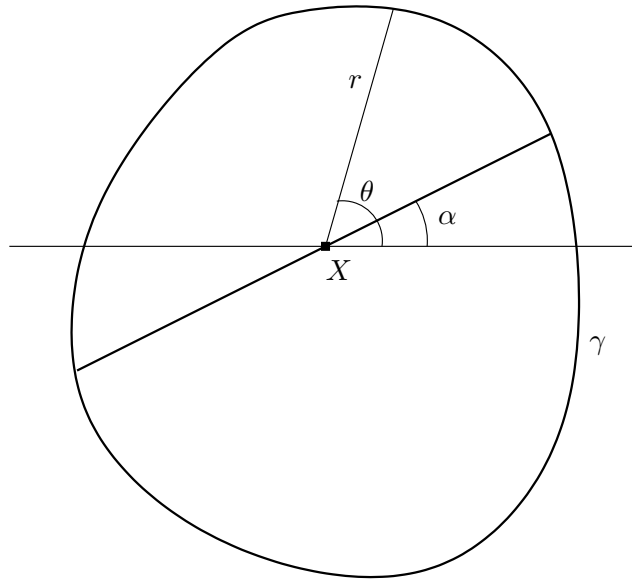
Figure 1: Curve γ with polar equation $r = r(\theta)$ relative to the pole X.

2.2 The Area Function

In the Mini-Dissertation, several important results were stated and verified. Particularly, the area function A in terms of α was confirmed to be of the form

$$A(\alpha) = \frac{1}{2} \int_{\alpha}^{\alpha+\pi} r^2 d\theta$$

using some general theory concerning the area of a sector. Moreover, using the rules of differentiating integrals, with the respective variable contained in the limits rather than the integrand, the first derivative of the area function was found to be of the

form

$$A'(\alpha) = \frac{1}{2}r(\alpha+\pi)^2 - \frac{1}{2}r(\alpha)^2$$

When this was equated to zero, we saw that $r(\alpha+\pi) = r(\alpha)$, giving the result that the pole X is the midpoint of the chord if and only if $A'(\alpha) = 0$. Furthermore, we found that differentiating again gives

$$A''(\alpha) = r(\alpha+\pi)r'(\alpha+\pi) - r(\alpha)r'(\alpha)$$

and when this was equated to zero, at the same time as the first derivative being equal to zero, and assuming that $r(\alpha)$ is non-zero, that is it has some positive length, we saw that $r'(\alpha+\pi) = r'(\alpha)$.

The general form of the gradients of the tangents that meet the curve at the endpoints of the chord were also found. In polar coordinates, $x = r(\theta)\cos\theta$ and $y = r(\theta)\sin\theta$. We found that the gradient m_1, where $\theta = \alpha$ was of the form

$$m_1 = \frac{r'(\alpha)\sin\alpha + r(\alpha)\cos\alpha}{r'(\alpha)\cos\alpha - r(\alpha)\sin\alpha}$$

and that m_2, where $\theta = \alpha+\pi$, referring to the tangent line at the opposite end of the chord, was of the form

$$m_2 = \frac{-r'(\alpha+\pi)\sin\alpha - r(\alpha+\pi)\cos\alpha}{-r'(\alpha+\pi)\cos\alpha + r(\alpha+\pi)\sin\alpha}$$

2.3   Classifying The Area Function

Now that the scene is set, we can go on to describe some further consequences concerning the second derivative of the area function, specifically, classifying the area cut off by the chord. In order to simplify the calculations, the value of α is set at zero, giving

$$m_1 = \frac{r(0)}{r'(0)} \qquad \text{and} \qquad m_2 = \frac{r(\pi)}{r'(\pi)}$$

Let's assume that we have m_1 on the right of the pole and m_2 on the left, as shown below in Figure 2.
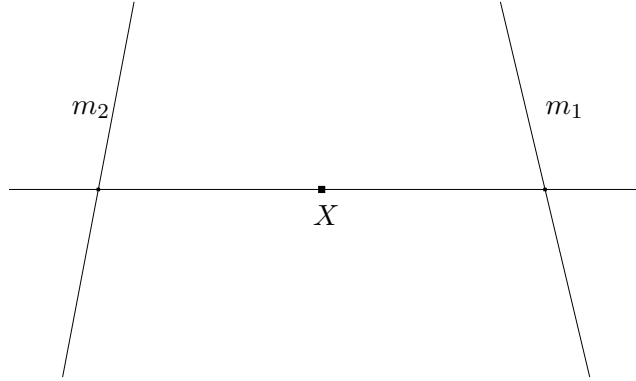
Figure 2: The two tangent lines relative to the pole X.

We are currently assuming that the gradients are *not* the same. In fact, there are several cases here, which we will look at in turn in order to try to classify the behaviour. It is important to point out at this stage, that we are assuming that $A'(\alpha) = 0$, which means that the pole X is the midpoint of the chord, as is stated above. This implies that, when classifying the second derivative, $r(0) = r(\pi)$, since this is simply the distance from the pole to the curve. We will look at the cases where the tangent lines meet above the horizontal.

2.3.1   Case 1: $m_1 > m_2 > 0$

If gradient m_1 is greater than m_2, with both m_1 and m_2 positive, then the two tangent lines will meet above the horizontal. If this is the case, then we have

$$\frac{r(0)}{r'(0)} > \frac{r(\pi)}{r'(\pi)} > 0$$

We can then divide both sides by $r(0) = r(\pi)$ to get

$$\frac{1}{r'(0)} > \frac{1}{r'(\pi)} > 0$$

and multiply up, keeping in mind that, since m_1 and m_2 are positive, $r'(0)$ and $r'(\pi)$ are positive. We then have

$$r'(\pi) > r'(0) > 0$$

In summary, if $r'(\pi) > r'(0)$, with both $r'(\pi)$ and $r'(0)$ positive, then the tangent lines will meet above the horizontal. This is shown geometrically in Figure 3.

11

2.3.2   Case 2: $m_2 < m_1 < 0$

If gradient m_2 is less than m_1, with both m_1 and m_2 negative, then the two tangent lines will meet above the horizontal. In this case

$$\frac{1}{r'(\pi)} < \frac{1}{r'(0)} < 0$$

which implies that

$$r'(0) < r'(\pi) < 0$$

In summary, if $r'(\pi) > r'(0)$, with both $r'(\pi)$ and $r'(0)$ negative, then the tangent lines will meet above the horizontal. This is shown geometrically in Figure 4.

2.3.3   Case 3: $m_1 < 0 < m_2$

The final case that ensures that the tangent lines will meet above the horizontal is where m_1 is negative and m_2 is positive. This implies that

$$\frac{1}{r'(0)} < 0 < \frac{1}{r'(\pi)}$$

and so

$$r'(\pi) > 0 > r'(0)$$

In summary, if $r'(\pi) > r'(0)$, with $r'(\pi)$ positive and $r'(0)$ negative, then the tangent lines will meet above the horizontal. This is shown geometrically in Figure 5.

These are the only three cases that ensure that the tangent lines meet above the horizontal. We can now use this information to classify the second derivative. There is a minimum of area if and only if $A''(0) > 0$. Then

$$r(\alpha + \pi)r'(\alpha + \pi) - r(\alpha)r'(\alpha) > 0$$

But we set α to be zero, so

$$r(\pi)t'(\pi) - r(0)r'(0) > 0$$

Rearranging, and cancelling $r(0)$ with $r(\pi)$, we have

$$r'(\pi) > r'(0)$$

We see that this situation occurs in all 3 cases and so we find that there is a minimum of area if and only if the tangent lines meet above the horizontal. Otherwise there is a maximum of area.
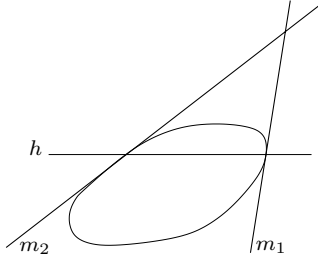
Figure 3: The first case where the tangent lines meet above the horizontal. We have m_1 and m_2 both positive, with $m_1 > m_2$.
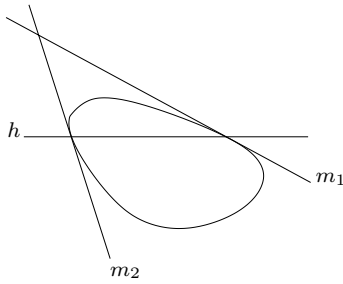
Figure 4: The second case where the tangent lines meet above the horizontal. We have m_1 and m_2 both negative, with $m_1 > m_2$.
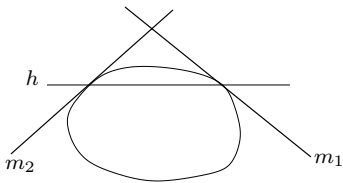
Figure 5: The third case where the tangent lines meet above the horizontal. We have m_1 negative and m_2 positive.

2.4   Parallel Tangent Lines

Having found the required conditions that ensure a maximum or minimum of area, let us now look at the special degenerate case where the second derivative is equated to zero. As we showed earlier

$$
\begin{aligned}
A''(\alpha) &= r(\alpha)r'(\alpha + \pi) - r(\alpha)r'(\alpha) \\
&= r(\alpha)\left(r'(\alpha + \pi) - r'(\alpha)\right)
\end{aligned}
$$

Assuming the radius $r(\alpha)$ is non-zero, that is, it has some positive length, then we can see quite clearly that $r'(\alpha + \pi) = r'(\alpha)$ if and only if $A'(\alpha) = A''(\alpha) = 0$. Then substituting this into the gradient formula stated in Section 2.2, we have

$$
m_2 = \frac{r'(\alpha)\sin\alpha + r(\alpha)\cos\alpha}{r'(\alpha)\cos\alpha - r(\alpha)\sin\alpha} = m_1
$$

This shows that the gradients of the tangent lines at the two ends of the chord are the same. So we have proved that the tangents at the two ends of the chord are parallel, *and* the pole X is the midpoint of this chord, if and only if $A'(\alpha) = A''(\alpha) = 0$. This is illustrated in Figure 6.
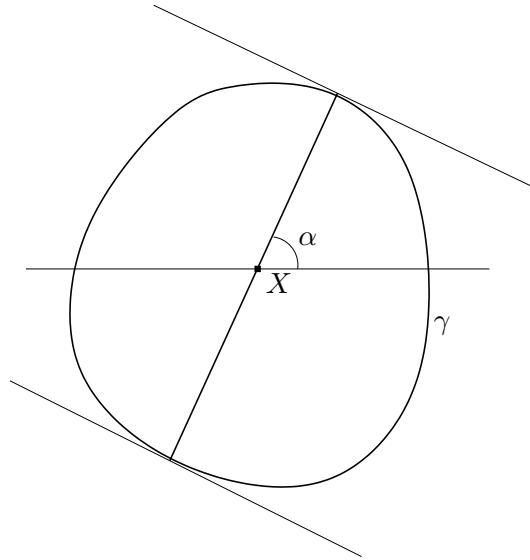
Figure 6: Curve γ together with the tangent lines at the ends of the chord.

14

The diagram shows the specific case in which $A'(\alpha) = A''(\alpha) = 0$ and the tangent lines at each end of the chord are parallel. In addition to this, the pole X is the midpoint. This is in fact a very important case, as it is where we have an A_2 singularity. It is these parts of the area parallel that correspond to the cusps. This will be discussed in Section 4, where we consider the area parallel of various quadrilaterals in a fair amount of detail.

We also find that this is completely analogous to the formation of cusps on the Euclidean parallel, as the distance-squared function has its first two derivatives equal to zero, but the third derivative non-zero, and therefore has an A_2 singularity.

3   Non-Convex Curves

3.1   Algebraic Area

In the Mini-Dissertation, we restricted our study to that of convex curves, due to the simple fact that any further material is so much more substantial. So we will now begin to study some non-convex closed plane curves. These are curves which contain at least two inflections. Indeed, as we have seen before on numerous occasions, inflections will always come in pairs.

The difficulty we have with non-convex curves is that it can be far more problematic when cutting off an area with a chord. If we cut off an area in a continually convex part of the curve, then we will not encounter problems in finding that part of the area parallel. However when the chord moves round to any non-convex sections of the curve, we will encounter complications. Figure 7 illustrates.
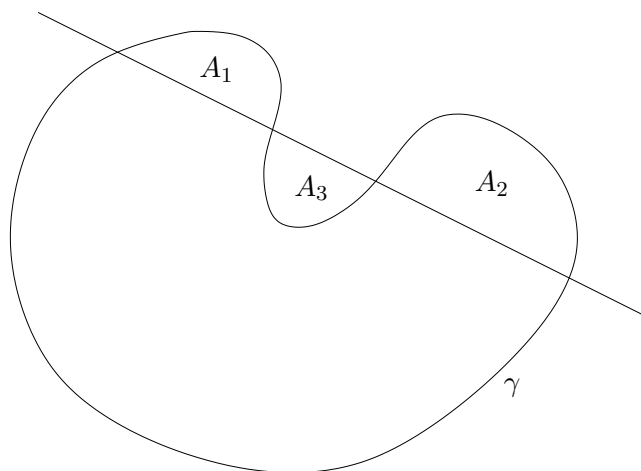
Figure 7: A non-convex curve, with a chord cutting off a fixed area.

As can be seen in the diagram, with the chord positioned as it is, we are initially unsure as to what area is being cut off. Are we to take each of A_1, A_2 and A_3 as positive areas? Or perhaps we are to take A_3 as a"negative" area.

Imagine that we were to take two curves $y = f(x)$ and $y = g(x)$, with three intersections, isolating two areas, one above the line $g(x)$ and one below, as shown in Figure 8.

The area that we calculate by integration is called the algebraic area. We take any area that is enclosed above the line $y = g(x)$ as a positive area and any area below as a negative area. We therefore take A_2 as a negative area. It must be noted here that the algebraic area is purely theoretical, since you can not have, in practice,
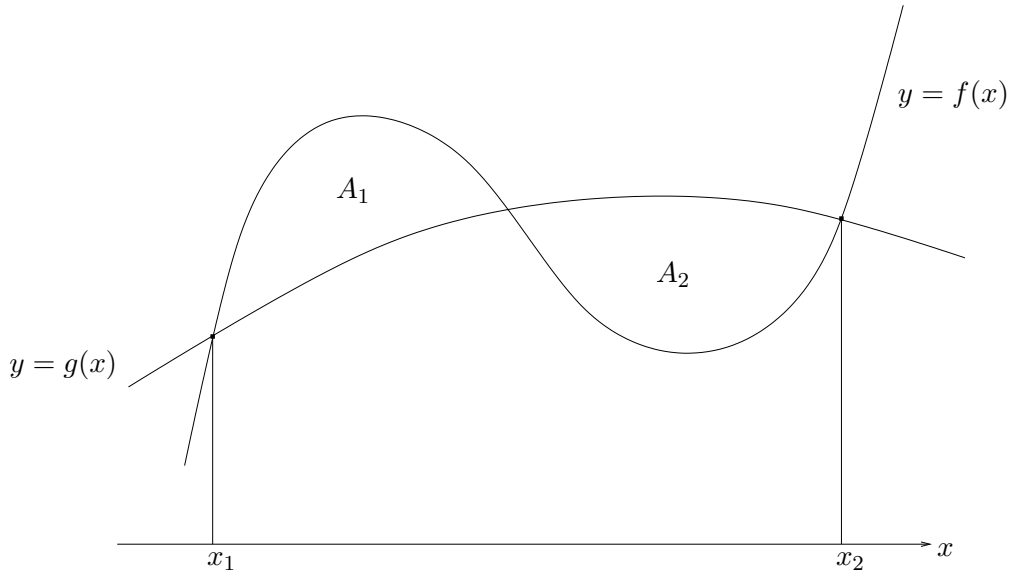
Figure 8: Intersection of two curves.

a negative area. The integral is given by

$$A_1 - A_2 = \int_{x_1}^{x_2} (f(x) - g(x))\mathrm{d}x$$

The same principle is used when calculating area parallels. Any area cut off by the chord which is *inside* the curve is counted as a positive area, and any area that is cut off by the chord but is *outside* the closed curve is treated as a negative area. The algebraic area illustrated in Figure 7 is therefore given by $A_1 + A_2 - A_3$.

3.2   Defining Local Coordinates

In order to find the area parallel of the general case of non-convex curves we will start by looking at the local area that contains the inflections. Let us take a curve $y = f(x)$ on Cartesian axes given by x and y. We will have a zero at $x = a$, with a negative, and a double zero at the origin. Let us also initially enclose an area between the curve and the x-axis, so that the line $y = 0$ is our chord. We will call this $y = L(x)$. As usual, we may move this chord around the curve as long as we always enclose the same area between $y = L(x)$ and $y = f(x)$, but we must remember that if the chord goes past the inflection point, then it will cut off both a positive area and a negative area, and, as discussed in the previous section, we must use the algebraic area.

We will only be moving the chord in this local region, so it is a good idea to introduce some local variables. In the local region of the origin, where the x-axis

17

is tangent to the curve $y = f(x)$ we will use the variable s. The one endpoint of the chord will be restricted to this section of the curve, and will have coordinates $(s, f(s))$. In the local region of the curves intersection with the x-axis at the point $x = a$, we will use the variable t. We can also redefine the curve in this area as $y = g(x - a)$. The opposite endpoint of the chord will be restricted to this section of the curve, and will have coordinates $(t+a, g(t))$. Figure 9 illustrates this information.
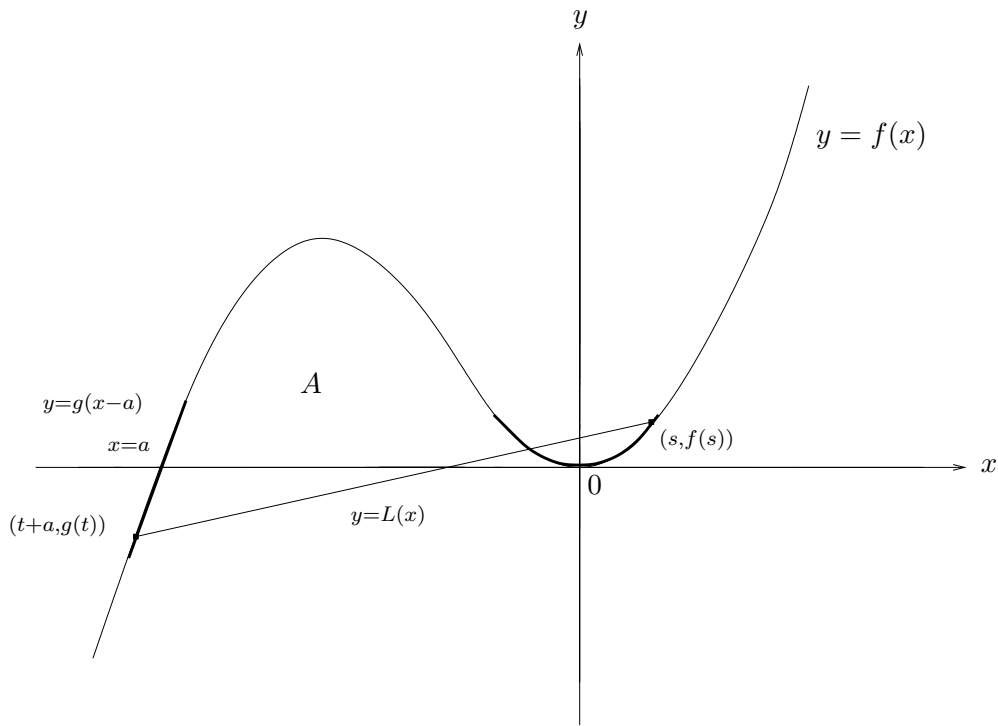
Figure 9: A section of the non-convex curve.

Here we can see the local regions are clearly marked by bold lines. Our task is now to observe how the endpoint of the chord near $x = a$ moves as we move the endpoint near the origin along its corresponding local region of the curve.

3.3   The Relationship Between the Two Endpoints of the Chord

Figure 10 shows us the algebraic area that we must find once the chord $L(x)$ has moved beyond the origin. It is important to remember that the algebraic area cut off will always be the same as the initial area that was cut off before the chord started to move.

Using this diagram, and comparing it with Figure 8 we can see that we must position our chord $L(x)$ so that the area between the x-axis and the curve $y = f(x)$ between a and 0 is the same as the difference of the two areas A_1 and A_2, since this

18
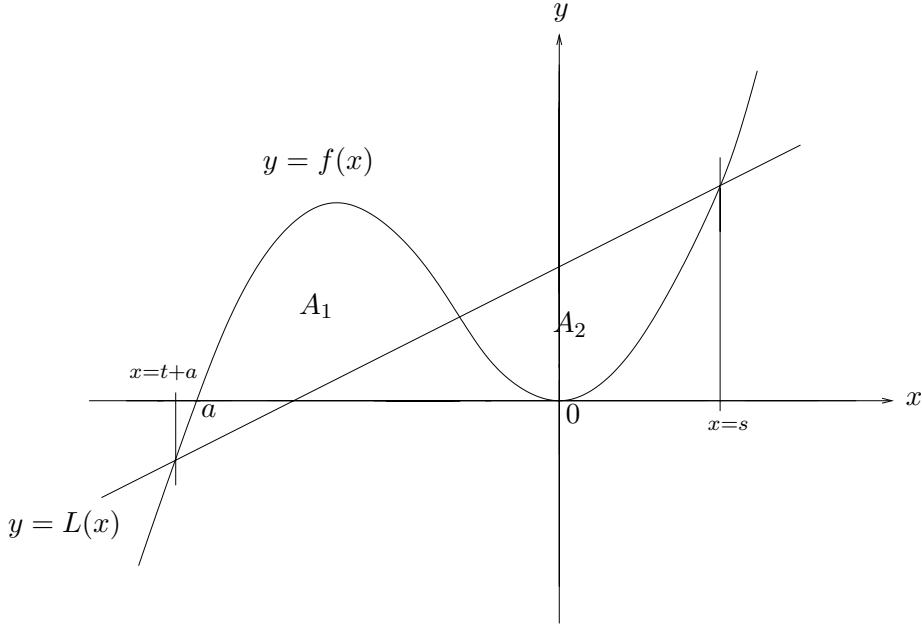
Figure 10: The area under the curve $y = f(x)$ between a and 0 is the same as the difference of areas A_1 and A_2.

is the algebraic area. That is to say

$$A_1 - A_2 = \int_{x=t+a}^{x=s} (f(x) - L(x))\mathrm{d}x = \int_{x=a}^{x=0} f(x)\mathrm{d}x$$

Hence, rearranging gives

$$\int_{t+a}^{s} L(x)\mathrm{d}x = \int_{t+a}^{s} f(x)\mathrm{d}x - \int_{a}^{0} f(x)\mathrm{d}x$$

Now, let's consider the right hand side. In words, the first integral gives us the area between the curve $y = f(x)$ and the x-axis, from $x = t + a$ up to $x = s$. As can be seen from Figure 10, most of this area is above the x-axis, but the small area between $x = t + a$ and $x = a$ is below the axis. We can split the integral up into the sum of three integrals to make up the given area. We must use the fact that the area between $x = t + a$ and $x = a$ is locally modelled by $y = g(x - a)$, as is shown in Figure 9, giving us

$$\int_{t+a}^{s} f(x)\mathrm{d}x = \int_{t+a}^{a} g(x - a)\mathrm{d}x + \int_{a}^{0} f(x)\mathrm{d}x + \int_{0}^{s} f(x)\mathrm{d}x$$

19

We therefore have

$$\int_{t+a}^{s} L(x)\mathrm{d}x = \int_{t+a}^{a} g(x-a)\mathrm{d}x + \int_{a}^{0} f(x)\mathrm{d}x + \int_{0}^{s} f(x)\mathrm{d}x - \int_{a}^{0} f(x)\mathrm{d}x$$

and we can not only simplify this by cancelling, but also use the substitution $u = x - a$, giving us

$$\int_{t+a}^{s} L(x)\mathrm{d}x = \int_{t}^{0} g(u)\mathrm{d}u + \int_{0}^{s} f(x)\mathrm{d}x$$

This has given us a rather important relationship. We shall now use this crucial equation to find the connection between s and t.

In order to evaluate the left hand side, we must find an expression for $L(x)$. We know that $L(x)$ is a straight line that connects the coordinates $(t + a, g(t))$ and $(s, f(s))$. (Verification of this is trivial from Figure 9). We can therefore right down the equation of the line.

$$\frac{y - f(s)}{x - s} = \frac{f(s) - g(t)}{s - (t + a)}$$

This can be rearranged to give

$$y = L(x) = \left(\frac{x - s}{s - t - a} \right) \left(f(s) - g(t) \right) + f(s)$$

$$\begin{aligned}
\int_{t+a}^{s} L(x)\mathrm{d}x &= \int_{t+a}^{s} \left(\frac{x - s}{s - t - a} \right) \left(f(s) - g(t) \right) \mathrm{d}x + \int_{t+a}^{s} f(s)\mathrm{d}x \\
&= \int_{t+a}^{s} \frac{x(f(s) - g(t))}{s - t - a}\mathrm{d}x - \int_{t+a}^{s} \frac{s(f(s) - g(t))}{s - t - a}\mathrm{d}x + \int_{t+a}^{s} f(s)\mathrm{d}x \\
&= \frac{x^2(f(s) - g(t))}{2(s - t - a)} \bigg|_{t+a}^{s} - \frac{xs(f(s) - g(t))}{s - t - a} \bigg|_{t+a}^{s} + xf(s) \bigg|_{t+a}^{s} \\
&= \frac{s^2(f(s) - g(t))}{2(s - t - a)} - \frac{(t + a)^2(f(s) - g(t))}{2(s - t - a)} - \frac{s^2(f(s) - g(t))}{s - t - a} \\
&\quad + \frac{s(t + a)(f(s) - g(t))}{s - t - a} + sf(s) - (t + a)f(s) +
\end{aligned}$$

This can be greatly simplified, giving

$$\int_{t+a}^{s} L(x)\mathrm{d}x = \frac{(s - t - a)(f(s) + g(t))}{2}$$

20

Since the part of the curve $y = f(x)$ that corresponds to the variable s is locally quadratic, we shall define $f(s)$ so that it models a quadratic curve. That is, we shall define $f(s)$ as a Taylor series beginning with a quadratic term as following.

$$f(s) = f_2 s^2 + f_3 s^3 + f_4 s^4 + \dots$$

Here, all of the f_i are constants. Similarly, the part of the curve that corresponds to the variable t is locally linear, so we will define $g(t)$ so that it models a straight line. $g(t)$ will therefore be a Taylor series beginning with a linear term. Hence, we have

$$g(t) = g_1 t + g_2 t^2 + g_3 t^3 + \dots$$

where all of the g_j are constants.

We must now look at the right hand side of the relationship. Clearly

$$
\begin{aligned}
\int_0^s f(x)\mathrm{d}x &= \int_0^s f_2 x^2 + f_3 x^3 + f_4 x^4 + \dots \mathrm{d}x \\
&= \left[f_2 \frac{x^3}{3} + f_3 \frac{x^4}{4} + f_4 \frac{x^5}{5} + \dots \right]_0^s \\
&= f_2 \frac{s^3}{3} + f_3 \frac{s^4}{4} + f_4 \frac{s^5}{5} + \dots - 0 \\
&= \sum_{n=2}^{\infty} f_n \frac{s^{n+1}}{n+1}
\end{aligned}
$$

A similar argument shows that

$$\int_t^0 g(u)\mathrm{d}u = -\sum_{n=1}^{\infty} g_n \frac{t^{n+1}}{n+1}$$

Now, in order to find the relationship between s and t, we must expand out both sides of the equation. When expanding the left we have

$$
\begin{aligned}
\frac{(s - t - a)(f(s) + g(t))}{2} &= \frac{sf(s)}{2} - \frac{tf(s)}{2} - \frac{af(s)}{2} + \frac{sg(t)}{2} - \frac{tgt}{2} - \frac{ag(t)}{2} \\
&= f_2 \frac{s^3}{2} + f_3 \frac{s^4}{2} + f_4 \frac{s^5}{2} - f_2 \frac{ts^2}{2} - f_3 \frac{ts^3}{2} - f_4 \frac{ts^4}{2} - f_2 \frac{as^2}{2} \\
&\quad - f_3 \frac{as^3}{2} - f_4 \frac{as^4}{2} + g_1 \frac{st}{2} + g_2 \frac{st^2}{2} + g_3 \frac{st^3}{2} - g_1 \frac{t^2}{2} \\
&\quad - g_2 \frac{t^3}{2} - g_3 \frac{t^4}{2} - g_1 \frac{at}{2} - g_2 \frac{at^2}{2} - g_3 \frac{at^3}{2} + \dots
\end{aligned}
$$

Looking at the right, we have

$$f_2\frac{s^3}{3} + f_3\frac{s^4}{4} + f_4\frac{s^5}{5} - g_1\frac{t^2}{2} - g_2\frac{t^3}{3} - g_3\frac{t^4}{4} + \cdots$$

Now, we will also define t as a power series of s, so that

$$t = t_1 s + t_2 s^2 + t_3 s^3 + \ldots$$

where t_1, t_2 and t_3 are expressions to be determined in terms of f_i and g_j. We want to find the relationship between s and t, but we will initially only use terms of degree three or less for reasons that will become apparent. (Indeed, as we are working with power series, we have to stop somewhere!) With this in mind, we can now express the equation as follows.

$$f_2\frac{s^3}{2} - f_2\frac{ts^2}{2} - f_2\frac{as^2}{2} - f_3\frac{as^3}{2} + g_1\frac{st}{2} + g_2\frac{st^2}{2}$$

$$- g_1\frac{t^2}{2} - g_2\frac{t^3}{2} - g_1\frac{at}{2} - g_2\frac{at^2}{2} - g_3\frac{at^3}{2}$$

$$= f_2\frac{s^3}{3} - g_1\frac{t^2}{2} - g_2\frac{t^3}{3}$$

We will now use the power series form of t in the equation. Here we have to remember that $(t_1 s + t_2 s^2 + t_3 s^3)^2$, or indeed $(t_1 s + t_2 s^2 + t_3 s^3)^3$ contains terms of degree greater than three, so we must only use the relevant terms. We have

$$f_2\frac{s^3}{2} - f_2\frac{t_1 s^3}{2} - f_2\frac{as^2}{2} - f_3\frac{as^3}{2} + g_1\frac{t_1 s^2}{2} + g_1\frac{t_2 s^3}{2} + g_2\frac{t_1^2 s^3}{2}$$

$$- g_1\frac{t_1^2 s^2}{2} - g_1\frac{2t_1 t_2 s^3}{2} - g_2\frac{t_1^3 s^3}{2} - g_1\frac{at_1 s}{2} - g_1\frac{at_2 s^2}{2} - g_1\frac{at_3 s^3}{2}$$

$$- g_2\frac{at_1^2 s^2}{2} - g_2\frac{2at_1 t_2 s^3}{2} - g_3\frac{at_1^3 s^3}{2}$$

$$= f_2\frac{s^3}{3} - g_1\frac{t_1^2 s^2}{2} - g_1\frac{2t_1 t_2 s^3}{2} - g_2\frac{t_1^3 s^3}{3}$$

We can now arrange this so that we may compare coefficients of the different powers of s. This will then enable us to find some expressions for t_1, t_2 and t_3.

$$\frac{g_1 at_1}{2}s + \left(\frac{g_1 t_1}{2} - \frac{af_2}{2} - \frac{g_1 t_1^2}{2} - \frac{ag_1 t_2}{2} - \frac{ag_2 t_1^2}{2}\right)s^2$$

$$+ \left(\frac{f_2}{2} - \frac{f_2 t_1}{2} - \frac{af_3}{2} + \frac{g_2 t_1^2}{2} - g_1 t_1 t_2 - \frac{g_2 t_1^3}{2} - \frac{ag_1 t_3}{2} - ag_2 t_1 t_2 - \frac{ag_3 t_1^3}{2}\right)s^3$$

$$= \left(-\frac{g_1 t_1^2}{2}\right)s^2 + \left(\frac{f_2}{3} - g_1 t_1 t_2 - \frac{g_2 t_1^3}{3}\right)s^3$$

22

By comparing coefficients, we can now find out some very interesting information. Since there is no degree one s term on the right hand side, we can see that $ag_1t_1 = 0$. So at least one of a, g_1 and t_1 is zero. If a were zero, this would mean that this part of the curve $f(x)$ would be at the origin, but, by our definition of $f(x)$, this is a local part of the curve that is somewhere away from the origin, since the other part of the curve, where s is the variable, *is* at the origin. So $a \neq 0$.

Meanwhile, if g_1 was zero then the power series $g(t)$ would begin with the term g_2t^2 which is a quadratic term, and would, in the same way as $f(s)$, model a quadratic curve rather than a straight line. So $g_1 \neq 0$, which means that $t_1 = 0$. This is in fact rather convenient, as we can now remove many of the terms from the equation. Comparing coefficients of s^2 now gives us

$$
\begin{aligned}
-\frac{af_2}{2} - \frac{ag_1t_2}{2} &= 0 \\
\Rightarrow \qquad f_2 + g_1t_2 &= 0 \\
\text{Hence} \qquad t_2 &= -\frac{f_2}{g_1}
\end{aligned}
$$

This is, of course, always possible, as we have said that g_1 is never zero, so we will always avoid any zero denominators here. Let us now compare coefficients of s^3 to find an expression for t_3. Remembering that $t_1 = 0$, we have

$$
\frac{f_2}{2} - \frac{af_3}{2} + \frac{g_1t_2}{2} - \frac{ag_1t_3}{2} = \frac{f_2}{3}
$$

and substituting in for t_2 we have

$$
\frac{f_2}{2} - \frac{af_3}{2} - \frac{f_2}{2} - \frac{ag_1t_3}{2} = \frac{f_2}{3}
$$

Rearranging gives us

$$
\begin{aligned}
\frac{ag_1t_3}{2} &= -\frac{af_3}{2} - \frac{f_2}{3} \\
\Rightarrow \qquad t_3 &= -\frac{f_3}{g_1} - \frac{2f_2}{3ag_1}
\end{aligned}
$$

Finally, this gives us

$$
t = -\frac{f_2}{g_1}s^2 + \left(-\frac{f_3}{g_1} - \frac{2f_2}{3ag_1} \right)s^3 + \cdots
$$

So we may express t in terms of constant terms α and β. We have

$$
t = -\alpha s^2 - \beta s^3 + \ldots
$$

where $\alpha = \frac{f_2}{g_1}$ and $\beta = \left(\frac{f_3}{g_1} + \frac{2f_2}{3ag_1}\right)$.

Taking t in this form provides us with a useful piece of information. Clearly, t begins with a quadratic term, and so behaves locally (near $x = a$) in a quadratic way. It also has a negative coefficient, which means that the path it draws will always be negative, since the first term is quadratic. That is, any value that is given to s will be squared so that it becomes positive, and then negated by the minus sign. Geometrically, this indicates that as the one endpoint of the chord moves along the curve $y = f(x)$ and passes through the origin, the other endpoint, the behaviour of which is defined by our expression for t, will go along $f(x)$ up to the x-axis at $x = a$ and then back down along the same path. The following term is cubic, and so will have a lot less effect on the path that this endpoint takes.

3.4   The Area Parallel

Now we understand more about how the chord behaves, our next task is to find the area parallel. As usual, we must consider the midpoints of the chord, as it is these midpoints that trace out the area parallel. The midpoints will be denoted by the general expression (m_x, m_y). The coordinates of the endpoints are $(t + a, g(t))$ and $(s, f(s))$. In order to find the midpoint we will take the average of the x-coordinates and the average of the y-coordinates of the endpoints. So starting with the x-coordinate, we have

$$m_x = \frac{t + a + s}{2}$$

We know $t = -\alpha s^2 - \beta s^3 + \ldots$, which when substituted in, gives us

$$m_x = \frac{1}{2}(a + s - \alpha s^2 - \beta s^3 + \ldots)$$

For the y-coordinate we will again take the average of the endpoints, so that

$$m_y = \frac{1}{2}(f(s) + g(t))$$

Using the power series in the same way as in Section 3.3, we can then expand out the expression for the y-coordinate.

$$
\begin{aligned}
m_y &= \frac{1}{2}(f_2 s^2 + f_3 s^3 + \ldots + g_1 t + g_2 t^2 + \ldots) \\
&= \frac{1}{2}\left(f_2 s^2 + f_3 s^3 + \ldots + g_1(-\alpha s^2 - \beta s^3) + \ldots\right) \\
&= \frac{1}{2}\left((f_2 - \alpha g_1)s^2 + (f_3 - \beta g_1)s^3 + \ldots\right) \\
&= \frac{1}{2}\left(\left(f_2 - \frac{f_2}{g_1}g_1\right)s^2 + \left(f_3 - \frac{f_3}{g_1}g_1 - \frac{2f_2}{3ag_1}g_1\right)s^3 + \cdots\right) \\
&= \frac{1}{2}\left(-\frac{2f_2}{3a}\right)s^3 + \cdots \\
&= -\frac{f_2}{3a}s^3 + \cdots
\end{aligned}
$$

So we now have the general coordinates of the midpoint

$$
\begin{aligned}
(m_x, m_y) &= \left(\frac{1}{2}a + \frac{1}{2}s + \cdots, \ -\frac{f_2}{3a}s^3 + \cdots\right) \\
&= \left(\frac{1}{2}a, 0\right) + \left(\frac{1}{2}s + \cdots, \ -\frac{f_2}{3a}s^3 + \cdots\right)
\end{aligned}
$$

We are assuming here, for now, that $\alpha \neq 0$, which of course implies that $f_2 \neq 0$. If we were to write (m_x, m_y) as

$$
(m_x, m_y) = \left(\frac{1}{2}s + s^2 u(s), \ -\frac{f_2}{3a}s^3 + s^4 v(s)\right)
$$

where $u(s)$ and $v(s)$ are power series of s, then we can define $\bar{s}$ as

$$
\bar{s} = \frac{1}{2}s(1 + su(s))
$$

Here, the map $s \to \bar{s}$ is a local diffeomorphism since

$$
\left.\frac{d\bar{s}}{ds}\right|_{s=0} = \frac{1}{2} \neq 0
$$

and therefore the inverse exists. Hence, $s = 2\bar{s} + \ldots$. Using this reparametrisation, we have

$$
\begin{aligned}
(m_x, m_y) &= \left(\bar{s}, \ -\frac{f_2}{3a}(2\bar{s})^3 + \cdots\right) \\
&= \left(\bar{s}, \ -\frac{8f_2}{3a}\bar{s}^3 + \cdots\right)
\end{aligned}
$$

25

which is the parametric equation for

$$y = -\frac{8f_2}{3a}x^3 + \cdots$$

It is important to point out that $a < 0$, as this was originally how we decided the curve $y = f(x)$ should behave, and $f_2 > 0$, since the local parabola at the origin is above (or on) the x-axis, and curves anti-clockwise. This ensures that the coefficient of x^3 in the Cartesian equation is positive. Hence, the area parallel is locally defined by a cubic curve, sloping from bottom left, to top right. This also proves that the inflection point is at $(\frac{1}{2}a, 0)$, which is of course expected as the characteristic points of the area parallel are at the midpoint of the chords, and this is indeed the midpoint of the chord that lies on the x-axis, with endpoints at $x = a$ and $x = 0$. Figure 11 illustrates.
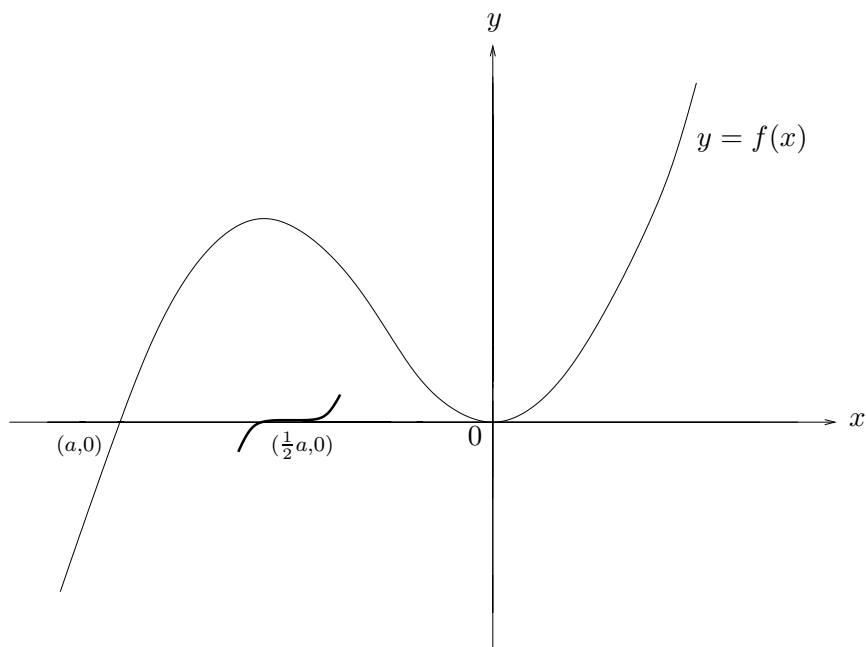
Figure 11: The area parallel is locally a cubic curve.

3.5 A Particular Case: $f(x) = x^2(1 + x)$

We will now illustrate these details in the form of an example. Let us take a function that has a zero at -1, so $a = -1$, and a double zero at the origin. The function $f(x) = x^2(1+x)$ satisfies these conditions. Multiplying out, we have $f(x) = x^2 + x^3$, which gives us coefficients $f_2 = 1 = f_3$, and all other f_i are zero.

26

We use the local variable t near $x = a = -1$, with $t = x - a = x + 1$, and so the curve in this area is defined by $y = g(t) = g(x + 1)$. Then $x = t - 1$ and so $y = t(t-1)^2$. Hence, $g(t) = t^3 - 2t^2 + t$, and so the coefficients are $g_1 = 1 = g_3$ and $g_2 = -2$, and all other g_j are zero.

To find the relationship between the local coordinates s and t, we will substitute in these coefficients. From Section 3.3, we have

$$t = -\frac{f_2}{g_1}s^2 + \left(-\frac{f_3}{g_1} - \frac{2f_2}{3ag_1}\right)s^3 + \cdots$$

and it can be easily checked that substituting in these values gives

$$t = -s^2 - \frac{5}{3}s^3 + \cdots$$

This of course shows how t behaves when s changes, and so we see how the endpoint of the chord near $x = -1$ behaves as the endpoint near the origin moves. Using this formation, we can plot some chords on the graph of $y = x^2(x + 1)$. With enough chords we can create an envelope which may give us some idea as to the shape of the local area parallel.

The general formula for the chord is

$$y = L(x) = \left(\frac{x - s}{s - t - a}\right)\left(f(s) - g(t)\right) + f(s)$$

as was discussed in Section 3.3. Clearly, $a = -1$, $f(s) = s^2(1+s)$ and $g(t) = t(t-1)^2$, so these can be inserted in the formula to get

$$y = \frac{(x - s)(s^2(1 + s) - t(t - 1)^2)}{s - t + 1} + s^2(1 + s)$$

We also have $t = -s^2 - \frac{5}{3}s^3 + \cdots$, which can also be inserted to give

$$y = \frac{(x - s)(s^2(1 + s) - (-s^2 - \frac{5}{3}s^3)(-s^2 - \frac{5}{3}s^3 - 1)^2)}{s - (-s^2 - \frac{5}{3}s^3) + 1} + s^2(1 + s)$$

and then using Maple to simplify this, we have

$$y = -\frac{(25s^5 + 30s^4 - 25xs^4 + 6s^3 - 30xs^3 + 6s^2 + 6xs^2 + 9s - 6xs - 18s - 9)s^2}{9}$$

This is now an equation involving x, y and s, and so we can substitute in any value for s, and we will be left with a linear equation involving x and y. Recall that s is

27

the local variable around the origin, so we will be taking values in this region, that is, taking a range something like $-0.3 \le s \le 0.3$.

We can then use Maple to draw a substantial amount of these chords within this range, which will give us an idea of the local behaviour of the area parallel. This is shown in Figure 13, over leaf.

We can now confirm that this is indeed the shape that the local area parallel will take by using the general formula for the locus of midpoints. From Section 3.4 we have

$$(m_x, m_y) = \left(\frac{1}{2}a, 0 \right) + \left(\frac{1}{2}s + \dots, -\frac{f_2}{3a}s^3 + \dots \right)$$

and substituting in the appropriate values gives

$$(m_x, m_y) = \left(-\frac{1}{2} + \frac{1}{2}s, \frac{1}{3}s^3 \right)$$

which gives us the parametric form of a cubic passing through the x-axis at $x = -\frac{1}{2}$, as wanted. This is illustrated in Figure 12.
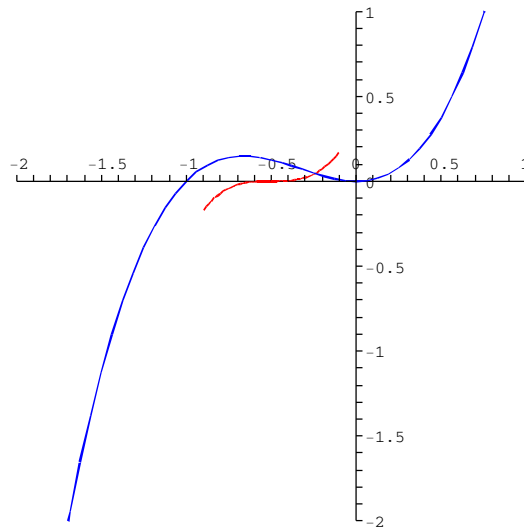
Figure 12: The curve plotted with a small section of the area parallel.

28
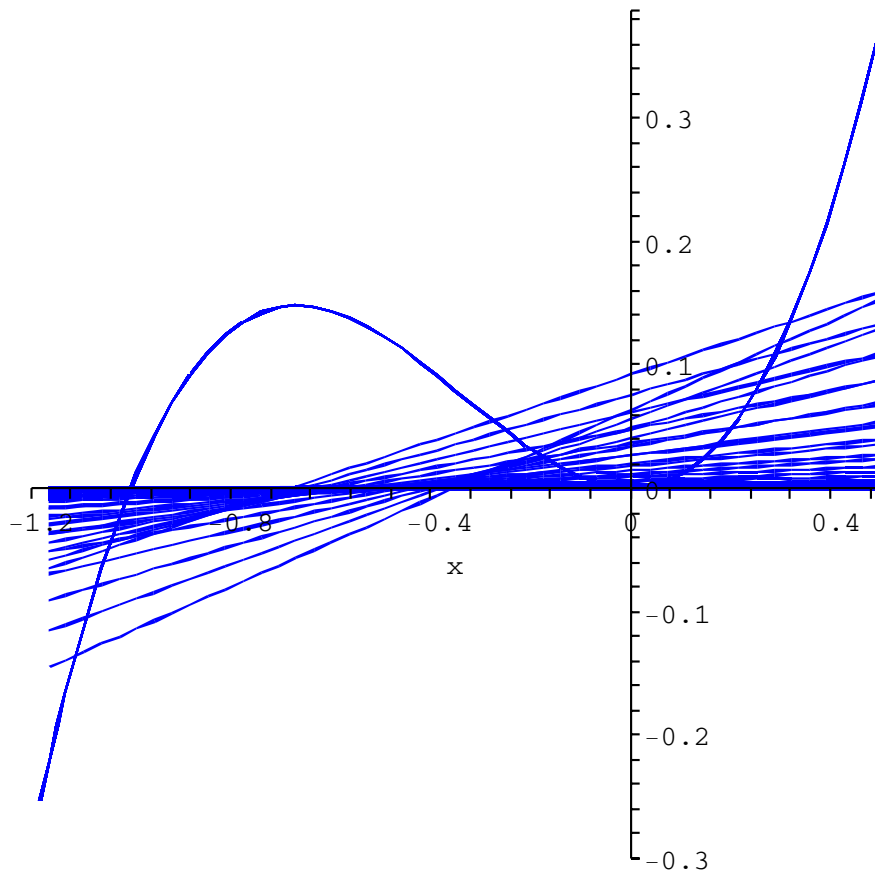
Figure 13: An envelope of chords tracing out the local area parallel.

3.6 Taking Coefficient $f_2 = 0$

With $f_2 \neq 0$, we have $t_1 = 0$, $t_2 = -\frac{f_2}{g_1}$ and $t_3 = -\frac{f_3}{g_1} - \frac{2f_2}{ag_1}$ as our coefficients of the series

$$t = t_1 s + t_2 s^2 + t_3 s^3 + \dots$$

If we were to set $f_2 = 0$, then we have $t_1 = 0 = t_2$ and $t_3 = -\frac{f_3}{g_1}$, and so there is only one non-zero coefficient. Therefore we must consider the coefficient t_4 which requires taking the fourth power of s into account in our calculations.

There will, however, only be a few extra terms involved, but we will see that this will change the shape of the area parallel. First of all we must find the *new* relationship between s and t. Recall from Section 3.3 that on the left we have

$$\frac{(s - t - a)(f(s) + g(t))}{2} = \frac{sf(s)}{2} - \frac{tf(s)}{2} - \frac{af(s)}{2} + \frac{sg(t)}{2} - \frac{tgt}{2} - \frac{ag(t)}{2}$$

and so the terms of degree four are

$$f_3 \frac{s^4}{2}, -f_3 \frac{ts^3}{2}, -f_4 \frac{as^4}{2}, g_3 \frac{st^3}{2} \text{ and } -g_4 \frac{t^4}{2}$$

Looking at the right hand side, we have

$$\sum_{n=2}^{\infty} f_n \frac{s^{n+1}}{n+1} - \sum_{n=1}^{\infty} g_n \frac{t^{n+1}}{n+1}$$

and so the terms of degree four are

$$f_3 \frac{s^4}{4} \text{ and } -g_3 \frac{t^4}{4}$$

We can then rewrite the required equation connecting s and t in the same way as in Section 3.3, but with the extra terms of degree four.

$$f_2 \frac{s^3}{2} + f_3 \frac{s^4}{2} - f_2 \frac{ts^2}{2} - f_3 \frac{ts^3}{2} - f_2 \frac{as^2}{2} - f_3 \frac{as^3}{2} - f_4 \frac{as^4}{2} + g_1 \frac{st}{2}$$

$$+ \quad g_2 \frac{st^2}{2} + g_3 \frac{st^3}{2} - g_1 \frac{t^2}{2} - g_2 \frac{t^3}{2} - g_1 \frac{at}{2} - g_2 \frac{at^2}{2} - g_3 \frac{at^3}{2} - g_4 \frac{t^4}{2}$$

$$= \quad f_2 \frac{s^3}{3} + f_3 \frac{s^4}{4} - g_1 \frac{t^2}{2} - g_2 \frac{t^3}{3} - g_3 \frac{t^4}{4}$$

Now we can use the power series $t = t_1 s + t_2 s^2 + t_3 s^3 + t_4 s^4$ in the equation. Also, it is useful to remember that $t_1 = 0$, so that we can remove many of the terms. We

get

$$f_2\frac{s^3}{2} + f_3\frac{s^4}{2} - f_2\frac{t_2s^4}{2} - f_2\frac{as^2}{2} - f_3\frac{as^3}{2} - f_4\frac{as^4}{2} + g_1\frac{t_2s^3}{2}$$

$$+ \quad g_1\frac{t_3s^4}{2} - g_1\frac{t_2^2s^2}{2} - g_1\frac{at_2s^2}{2} - g_1\frac{at_3s^3}{2} - g_1\frac{at_4s^4}{2} - g_2\frac{at_2^2s^4}{2}$$

$$= \quad f_2\frac{s^3}{3} + f_3\frac{s^4}{4} - g_1\frac{t_2^2s^4}{2}$$

We know from above that $t_2 = 0$ since we are setting $f_2 = 0$, and that $t_3 = -\frac{f_3}{g_1}$, for the same reason. We can now substitute these values into the equation, giving

$$f_3\frac{s^4}{2} - f_3\frac{as^3}{2} - f_4\frac{as^4}{2} - f_3\frac{s^4}{2} + f_3\frac{as^3}{2} - g_1\frac{at_4s^4}{2} = f_3\frac{s^4}{4}$$

Now, comparing coefficients of s^4 we have

$$-\frac{af_4}{2} - \frac{ag_1t_4}{2} = \frac{f_3}{4}$$

and rearranging gives the result

$$t_4 = -\frac{f_3}{2ag_1} - \frac{f_4}{g_1}$$

Finally, this gives us the connection between s and t to be

$$t = -\frac{f_3}{g_1}s^3 - \left(\frac{f_3}{2ag_1} + \frac{f_4}{g_1}\right)s^4 + \cdots$$

So we may express t in terms of constant terms γ and δ.

$$t = -\gamma s^3 - \delta s^4 + \ldots$$

where $\gamma = \frac{f_3}{g_1}$ and $\delta = \left(\frac{f_3}{2ag_1} + \frac{f_4}{g_1}\right)$.

Now in order to find an expression for the area parallel, we will use the same methods as in Section 3.4. The x-coordinate of the midpoint will be given by

$$
\begin{aligned}
m_x &= \frac{t + a + s}{2} \\
&= \frac{1}{2}(a + s - \gamma s^3 - \delta s^4 + \ldots)
\end{aligned}
$$

31

and the y-coordinate takes the form

$$
\begin{aligned}
m_y &= \frac{1}{2}(f(s) + g(t)) \\
&= \frac{1}{2}(f_3 s^3 + f_4 s^4 + \ldots + g_1 t + g_2 t^2 + \ldots) \\
&= \frac{1}{2}\left(f_3 s^3 + f_4 s^4 + \ldots + g_1(-\gamma s^3 - \delta s^4) + \ldots\right) \\
&= \frac{1}{2}\left((f_3 - \gamma g_1)s^3 + (f_4 - \delta g_1)s^4 + \ldots\right) \\
&= \frac{1}{2}\left(\left(f_3 - \frac{f_3}{g_1}g_1\right)s^3 + \left(f_4 - \frac{f_3}{2ag_1}g_1 - \frac{f_4}{g_1}g_1\right)s^4 + \cdots\right) \\
&= \frac{1}{2}\left(-\frac{f_3}{2a}\right)s^4 + \cdots \\
&= -\frac{f_3}{4a}s^4 + \cdots
\end{aligned}
$$

This gives us the general coordinates of the midpoint.

$$
\begin{aligned}
(m_x, m_y) &= \left(\frac{1}{2}a + \frac{1}{2}s + \cdots, -\frac{f_3}{4a}s^4 + \cdots\right) \\
&= \left(\frac{1}{2}a, 0\right) + \left(\frac{1}{2}s + \cdots, -\frac{f_3}{4a}s^4 + \cdots\right)
\end{aligned}
$$

We can then define $\bar{s}$, and use the same local diffeomorphism as in Section 3.4, so that $s = 2\bar{s} + \ldots$, giving

$$
\begin{aligned}
(m_x, m_y) &= \left(\bar{s}, -\frac{f_3}{4a}(2\bar{s})^4 + \cdots\right) \\
&= \left(\bar{s}, -\frac{4f_3}{a}\bar{s}^4 + \cdots\right)
\end{aligned}
$$

which clearly indicates that the Cartesian equation of the local area parallel is

$$
y = -\frac{4f_3}{a}x^4 + \cdots
$$

Clearly, a Cartesian equation of this form defines a quartic graph. Unlike the case where f_2 is non-zero, the sign of the coefficient of the principal term in the Cartesian equation is negative due to the fact that the constant a is negative, as we know, and f_3 is also negative. This is because the part of the curve $y = f(x)$ near the origin, where s is the local variable is locally a cubic, but sloping from top-left to bottom-right. That is, it follows the path of a cubic with negative coefficient. We therefore find that the local area parallel approximates the graph of a negative quartic. This is illustrated in Figure 14.

32
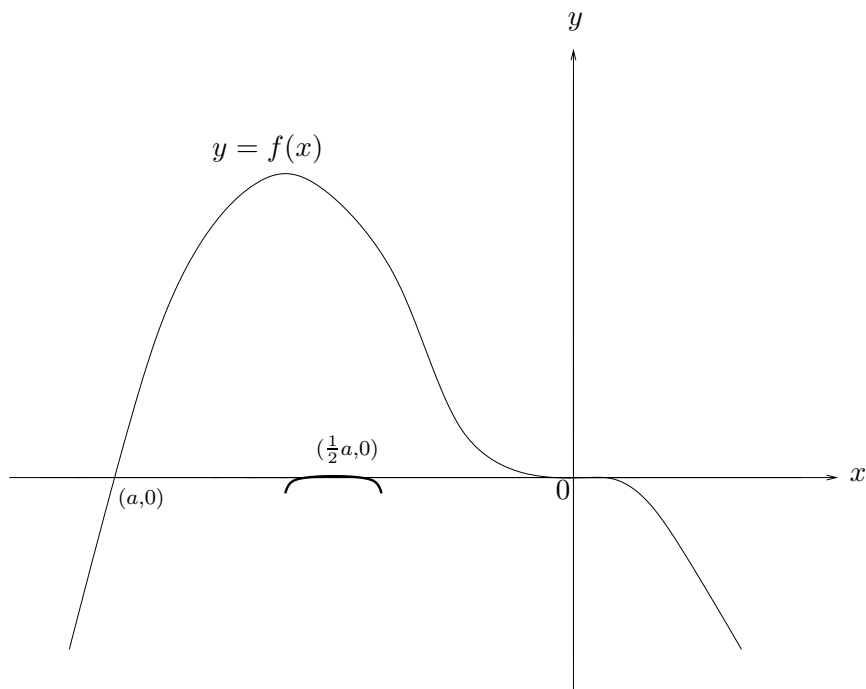
Figure 14: The area parallel locally approximates a quartic curve.

3.7 A particular Case: $f(x) = -x^3(1 + x)$

Let us now look at an example of the case where the coefficient $f_2 = 0$. We will once again set a at -1, but this time have a triple zero at the origin. This can be modelled by the function $f(x) = -x^3(1 + x) = -x^3 - x^4$, so that $f_3 = -1 = f_4$, and all other f_i are zero. It can be easily checked that the curve near $x = -1$ is given by $y = g(t) = t(t - 1)^3 = -t + 3t^2 - 3t^3 + t^4$ so that $g_1 = -1$, $g_2 = 3$, $g_3 = -3$ and $g_4 = 1$, and all other g_j are zero.

The connection between local variables s and t is then given by

$$t = -\frac{f_3}{g_1}s^3 - \left(\frac{f_3}{2ag_1} + \frac{f_4}{g_1}\right)s^4 + \cdots$$

from Section 3.6, and so substituting all the appropriate coefficients in, we have

$$t = -s^3 - \frac{1}{2}s^4 + \cdots$$

Once again, we have the information that describes how then endpoint near $x = -1$ behaves as its equivalent endpoint moves along the curve through the origin. So we can calculate the general expression for the chord, and use Maple to create an

33

envelope. Using $a = -1$, $f(s) = -s^3(s+1)$ and $g(t) = -t(t-1)^3$, we have

$$y = L(x) = \left(\frac{x-s}{s-t-a}\right)\left(f(s) - g(t)\right) + f(s)$$

$$= \frac{(x-s)(-s^3(1+s) + t(t-1)^3)}{s-t+1} - s^3(1+s)$$

$$= \frac{(x-s)(-s^3(1+s) - (-s^3 - \frac{1}{2}s^4)(-s^3 - \frac{1}{2}s^4 - 1)^2)}{s - (-s^3 - \frac{1}{2}s^4) + 1} - s^3(1+s)$$

which can again be simplified to give a linear equation involving x, y and s. Then, substituting in values of s in a range something similar to $-0.3 \le s \le 0.3$ will give an envelope of chords, tracing out the local area parallel. This is illustrated in Figure 16, over leaf.

We can once again confirm that the envelope shown in Figure 16 is indeed the accurate shape for the local area parallel by substituting the appropriate values into our general expression for the midpoint discussed in Section 3.6. We have

$$(m_x, m_y) = \left(-\frac{1}{2} + s, -\frac{3}{2}s^4\right)$$

which is the parametric form the quartic illustrated in Figure 15.
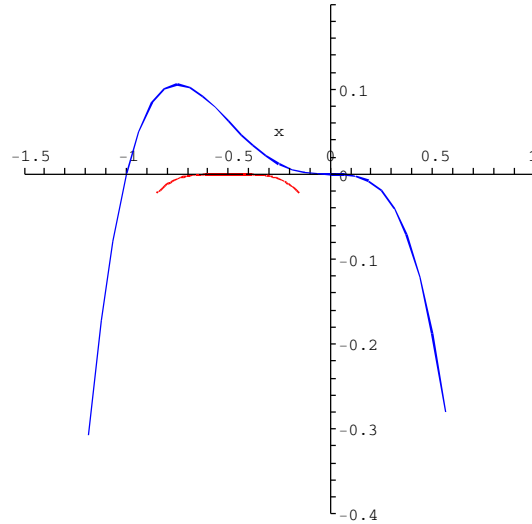
Figure 15: The curve plotted with a small section of the area parallel.

34
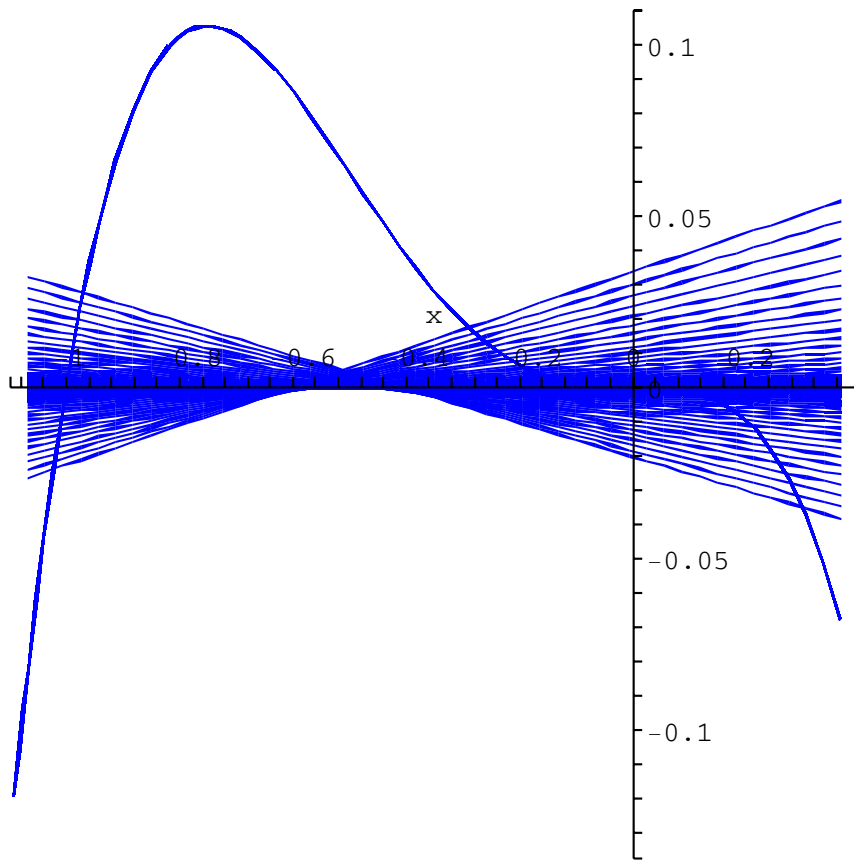
Figure 16: An envelope of chords tracing out the local area parallel.

3.8 The Zero Area Parallel

3.8.1 Zero Algebraic Area

As was discussed in Section 3.1, it is possible to have a negative algebraic area, if this area is below the relevant line. Subsequently, it is also possible to have a positive area and a negative area exactly the same, one above this line and one below, thus cancelling each other out. This is known as zero algebraic area, or in its shorter form, the zero area. Our task in this Section is to consider how the area parallel in such a situation will behave. This is known as the zero area parallel. See Figure 17 for further illustration.
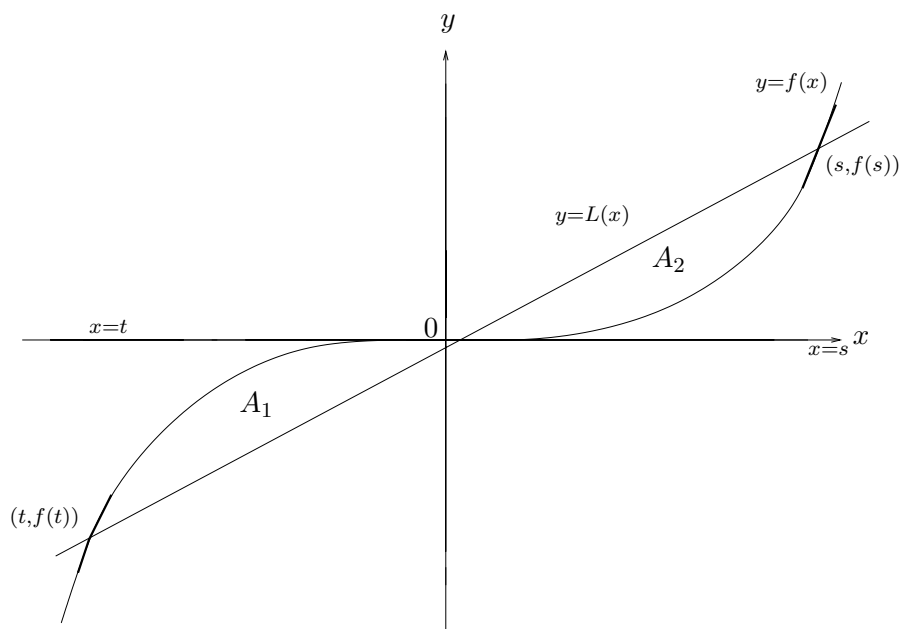
Figure 17: An example of zero area. The two areas cut off by the chord, one below (A_1) and one above (A_2) are equal.

As usual, the curve is $y = f(x)$ and the chord cutting off fixed area is $y = L(x)$. Clearly the fixed area is zero. We also have local variables s and t, between which we will be finding the connection, that is, the way one behaves as the other changes. In this situation, however, there is no need for any local redefinition of the curve, as there has been in previous cases, since the curve only has a zero at the origin. The important feature here is that $A_1 = A_2$.

3.8.2 The Relationship Between the Two Endpoints of the Chord

Any curve of this type will always contain an inflection, and takes the shape of a cubic. So when expressing the formula as a power series, it will start with a cubic term. It will therefore be of the form

$$f(x) = f_3 x^3 + f_4 x^4 + \ldots$$

Since the area cut off by the chord is now zero, we will have

$$\int_{x=t}^{x=s} (f(x) - L(x)) \, dx = 0$$

and so

$$\int_t^s f(x) dx = \int_t^s L(x) dx$$

giving us our connection between s and t. Similar to previous calculations, the line $y = L(x)$ will be of the form

$$y = \left(\frac{x-s}{s-t} \right) \left(f(s) - f(t) \right) + f(s)$$

and it is relatively simple to show that

$$\int_t^s L(x) dx = \frac{(s-t)(f(s) + f(t))}{2}$$

We must then consider the left hand side of the equation.

$$
\begin{aligned}
\int_t^s f(x) dx &= \int_t^s f_3 x^3 + f_4 x^4 + \ldots dx \\
&= \left[\frac{f_3 x^4}{4} + \frac{f_4 x^5}{5} + \cdots \right]_t^s \\
&= \frac{f_3 s^4}{4} + \frac{f_4 s^5}{5} + \cdots - \frac{f_3 t^4}{4} - \frac{f_4 t^5}{5} - \cdots \\
&= \frac{f_3(s^4 - t^4)}{4} + \frac{f_4(s^5 - t^5)}{5} + \cdots \\
&= \sum_{n=3}^{\infty} f_n \frac{(s^{n+1} - t^{n+1})}{n+1}
\end{aligned}
$$

The relationship can then be given in the form

$$\frac{f_3(s^4 - t^4)}{2} + \frac{2f_4(s^5 - t^5)}{5} + \cdots = (s-t)(f(s) + f(t))$$

37

Let us now consider the right hand side of this equation, using our power series approximations.

$$\begin{aligned}(s-t)(f(s)+f(t)) &= (s-t)(f_3 s^3 + f_4 s^4 + \ldots + f_3 t^3 + f_4 t^4 + \ldots) \\ &= f_3 s^4 + f_4 s^5 + \ldots + f_3 s t^3 + f_4 s t^4 + \ldots \\ &- f_3 s^3 t - f_4 s^4 t - \ldots - f_3 t^4 - f_4 t^5 - \ldots\end{aligned}$$

Now, as we have done in previous calculations, we will express t in terms of a power series, so that we may state our relationship between s and t in terms of increasing powers of s, where it is our task to find the coefficients thereof. As in Section 3.3, we have

$$t = t_1 s + t_2 s^3 + t_3 s^3 + \ldots$$

We can now substitute this power series into the equation. We first of all want to compare coefficients of s^4, so to avoid the risk of getting too confused we will express only the terms from this equation that will be required. We have

$$\frac{f_3 s^4}{2} - \frac{f_3 t_1^4 s^4}{2} = f_3 s^4 + f_3 t_1^3 s^4 - f_3 t_1 s^4 - f_3 t_1^4 s^4$$

We can divide through by f_3 since this coefficient is clearly non-zero, and then comparing coefficients of s^4 we have

$$\begin{aligned}1 - t_1^4 &= 2 - 2t_1 + 2t_1^3 - 2t_1^4 \\ \Rightarrow \quad 0 &= t_1^4 - 2t_1^3 + 2t_1 - 1 \\ \Rightarrow \quad 0 &= (t_1 - 1)^3(t_1 + 1)\end{aligned}$$

giving $t_1 = 1$ or $t_1 = -1$. If $t_1 = 1$ then the power series of t will start off $t = s$, and so this seems to suggest that the places at where the chord cuts the curve are in fact the same point. This in fact describes the trivial case where the chord is tangent to the curve at some particular point, and so the chord only meets the curve at one point. This is shown in Figure 18, over leaf.

The diagram shows that the chord is tangent to the curve. Strictly speaking, this chord *does* cut off zero area, but we are only interested in the case where a positive area is cancelled out by a negative area, and so we may assume that $t_1 \neq 1$. The other value, however, $t_1 = -1$ is of interest to us. This case implies that the power series starts of $t = -s$, and so the set up is originally symmetrical. This is a lot more valid, and so our discussion will be solely centred around this value for t_1.

Using the fact that $t_1 = -1$ we will now consider the coefficients of s^5 in order to devise an expression for t_2. Again, we will only use the relevant terms of the equation, so that we may simplify matters as much as possible. So the terms involving s^5 only
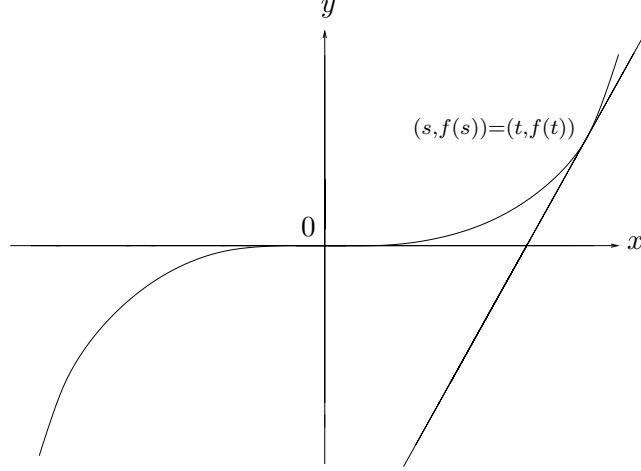
Figure 18: The chord is now tangent to the curve, and so only meets the curve in one place.

are now given. We have

$$
\frac{2f_4 s^5}{5} - \frac{2f_4 t_1^5 s^5}{5} - 2f_3 t_1^3 t_2 s^5 = f_4 s^5 + 3f_3 t_1^2 t_2 s^5 + f_4 t_1^4 s^5 - f_3 t_2 s^5
$$
$$
- f_4 t_1 s^5 - 4f_3 t_1^3 t_2 s^5 - f_4 t_1^5 s^5
$$

Taking $t_1 = -1$, this relationship can be simplified. Then, taking the coefficients of s^5 we have

$$
\frac{4f_4}{5} + 2f_3 t_2 = f_4 + 3f_3 t_2 + f_4 - f_3 t_2 + f_4 + 4f_3 t_2 + f_4
$$

Collecting up similar terms leaves us with

$$
\frac{4f_4}{5} = 4f_4 + 4f_3 t_2
$$
$$
\Rightarrow \quad 4f_3 t_2 = \frac{-16 f_4}{5}
$$
$$
\Rightarrow \quad t_2 = \frac{-4 f_4}{5 f_3}
$$

This will always exist since f_3 is non-zero. We can go on in a similar way to find t_3. As would be expected, this will involve the coefficients of s^6. We now have

$$
\frac{f_5 s^6}{3} - \frac{f_5 t_1^6 s^6}{3} - 2f_4 t_1^4 t_2 s^6 - 3f_3 t_1^2 t_2^2 s^6 - 2f_3 t_1^3 t_3 s^6
$$
$$
= f_5 s^6 + 3f_3 t_1 t_2^2 s^6 + 3f_3 t_1^2 t_3 s^6 + 4f_4 t_1^3 t_2 s^6 + f_5 t_1^5 s^6 - f_3 t_3 s^6
$$
$$
- f_4 t_2 s^6 - f_5 t_1 s^6 - 6f_3 t_1^2 t_2 s^6 - 4f_3 t_1^3 t_3 s^6 - 5f_4 t_1^4 t_2 s^6 - f_5 t_1 s^6
$$

39

Using the fact that $t_1 = -1$ and $t_2 = \frac{-4f_4}{5f_3}$, we can express the relationship in terms of t_3 only. We have

$$\frac{f_5}{3} - \frac{f_5 t_1^6}{3} + \frac{8f_4^2}{5f_3} - \frac{48f_4^2}{25f_3} + 2f_3 t_3 = f_5 - \frac{48f_4^2}{25f_3} + 3f_3 t_3 + \frac{16f_4^2}{5f_3} - f_5 - f_3 t_3$$

$$+ \frac{4f_4^2}{5f_3} + f_5 - \frac{96f_4^2}{25f_3} + 4f_3 t_3 + \frac{f_4^2}{f_3} - f_5$$

This can be greatly simplified by collecting up similar terms. We get

$$2f_3 t_3 - \frac{8f_4^2}{25f_3} = 6f_3 t_3 + \frac{56f_4^2}{25f_3}$$

$$\Rightarrow \quad 4f_3 t_3 = -\frac{64f_4^2}{25f_3}$$

$$\Rightarrow \quad t_3 = -\frac{16f_4^2}{25f_3^2}$$

We can see here that $t_3 = t_2^2$, which means that if the s^2 term in the power series disappears, then so does the s^3 term. We can now express our power series as

$$t = -s - \alpha s^2 - \alpha^2 s^3 + \ldots$$

where $\alpha = \frac{4f_4}{5f_3}$. This describes how the endpoint of the chord referring to the local part of the curve at which t is the local variable behaves as the endpoint at which s is the local variable changes.

3.8.3 The Locus of Midpoints of Such Chords

We will first of all recall the diagram from the start of the section, in order to familiarise ourselves with the way this model works. See Figure 19.

The diagram shows that the endpoints of the chord are given by $(s, f(s))$ and $(t, f(t))$. It is therefore quite clear that the general expression for the midpoint will be given by the average of these two coordinates, that is

$$(m_x, m_y) = \left(\frac{s+t}{2}, \frac{f(s) + f(t)}{2} \right)$$

Now, $f(s)$ and $f(t)$ can be expressed as power series, so we have

$$(m_x, m_y) = \left(\frac{s+t}{2}, \frac{f_3 s^3 + f_4 s^4 + \ldots + f_3 t^3 + f_4 t^4 + \ldots}{2} \right)$$

40
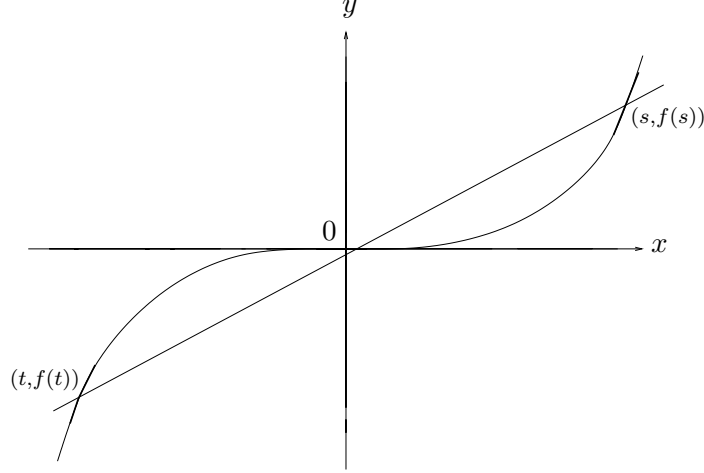
Figure 19: Recalling the diagram. The chord cuts the curve leaving zero algebraic area.

and we know that $t = -s - \alpha s^2 - \alpha^2 s^3 + \ldots$ where $\alpha = \frac{4f_4}{5f_3}$, giving

$$
\begin{aligned}
(m_x, m_y) &= \left(\frac{s - s - \alpha s^2 + \ldots}{2}, \frac{f_3 s^3 + f_4 s^4 - f_3 s^3 - 3\alpha f_3 s^4 + f_4 s^4 + \ldots}{2} \right) \\
&= \left(\frac{-\alpha s^2 + \ldots}{2}, f_4 s^4 - \frac{3\alpha f_3}{2} s^4 + \cdots \right) \\
&= \left(-\frac{2f_4}{5f_3} s^2 + \cdots, \left(f_4 - \frac{12 f_3 f_4}{10 f_3} \right) s^4 + \cdots \right) \\
&= \left(-\frac{2f_4}{5f_3} s^2 + \cdots, -\frac{f_4}{5} s^4 + \cdots \right)
\end{aligned}
$$

This calculation shows us that we have a cusp at the inflection point on the curve. We can improve the accuracy by calculating the coefficient of s^5 in the coordinate m_y. When fifth powers are included we have

$$
\begin{aligned}
m_y &= \frac{f_3 s^3 + f_4 s^4 + f_5 s^5 + \ldots + f_3 t^3 + f_4 t^4 + f_5 t^5 + \ldots}{2} \\
&= \frac{f_3 s^3 + f_4 s^4 + f_5 s^5 - f_3 s^3 - 3\alpha f_3 s^4 - 9\alpha^2 f_3 s^5 + f_4 s^4 + 4\alpha f_4 s^5 - f_5 s^5 + \ldots}{2} \\
&= \left(f_4 - \frac{3}{2}\alpha f_3 \right) s^4 + (2\alpha f_4 - 3\alpha^2 f_3) s^5 + \ldots \\
&= -\frac{f_4}{5} s^4 - \frac{8 f_4^2}{25 f_3} s^5 + \ldots
\end{aligned}
$$

41

which can be easily checked. This then gives us the extra term, and so we can now express a more accurate form of the general formula of the midpoint. We have

$$(m_x, m_y) = \left(-\frac{2f_4}{5f_3}s^2 + \cdots , -\frac{f_4}{5}s^4 - \frac{8f_4^2}{25f_3}s^5 + \cdots \right)$$

We find that the coefficient of s^5 disappears if and only if $f_4 = 0$, which is very interesting because if this was the case, then the coefficient of s^4 would also be zero. So, assuming f_4 is not zero, then we have, highly locally, a rhamphoid cusp. That is, for small values of s, we find there are two branches of a cusp curving in the same direction, meeting at the inflection point on the curve $y = f(x)$. In fact, the one branch is a slight deformation of the other, due to the high powers involved. This is illustrated in Figure 20, in order to give some idea of the shape of the local area parallel.
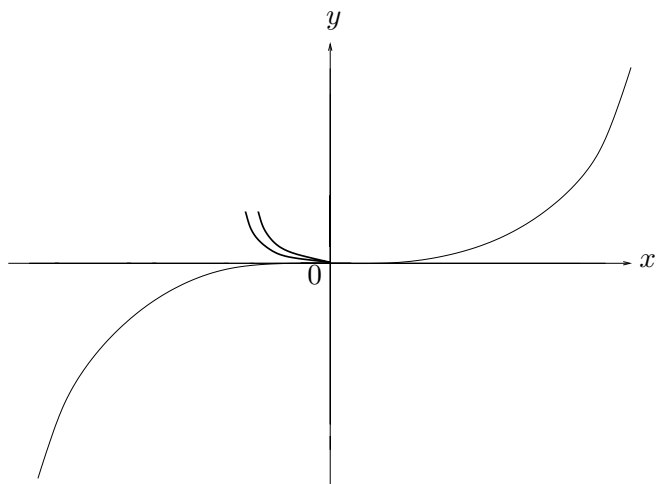
Figure 20: An example of how a local area parallel may look in this case. We have two branches meeting at the inflection point to form a rhamphoid cusp.

3.8.4   A Particular Case: $f(x) = x^3 + x^4 + x^5$

Let us now consider an example. We will take the function $f(x) = x^3 + x^4 + x^5$, so that $f_3 = f_4 = f_5 = 1$ and all other f_i are zero. Let us go straight on to look at the locus of midpoints of the chords. The previous section tells us that

$$(m_x, m_y) = \left(-\frac{2f_4}{5f_3}s^2 + \cdots , -\frac{f_4}{5}s^4 - \frac{8f_4^2}{25f_3}s^5 + \cdots \right)$$

If we substitute the coefficients f_i into this expression, then we have

$$(m_x, m_y) = \left(-\frac{2}{5}s^2 + \cdots, -\frac{1}{5}s^4 - \frac{8}{25}s^5 + \cdots \right)$$

and using maple, we can plot the local area parallel for a suitable range of s, such as $-0.6 \leq s \leq 0.5$. Maple returns the diagram shown in Figure 21.
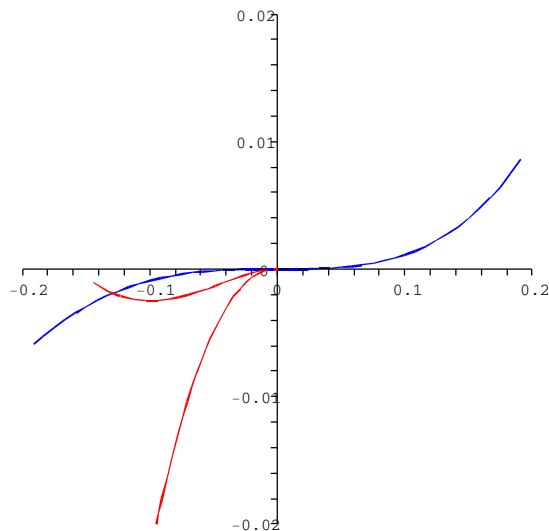
Figure 21: The curve plotted with a small section of the area parallel.

We can see in this diagram that it is only really within close proximity to the origin that the two branches of the rhamphoid cusp are curving in the same direction. Therefore, whilst our range of s shows the local behaviour, the more relevant parts of the curve will be described inside a smaller range such as $-0.3 \leq s \leq 0.25$. It is within this range that we observe a true rhamphoid cusp.

It is, according to an earlier part of this study, no surprise that we encounter a cusp in this particular case. As we have proven in Chapter 2, cuspidal formation corresponds with parallel tangent lines, and there are indeed limitless possibilities of parallel tangent lines in these types of curves.

3.9 Some Local Examples

In this final section, we will look at some examples of non-convex closed plane curves sketched with their global area parallels (with different areas cut off), and take a closer look at some of the local parts of these area parallels that have similar

43

characteristics to those described in the preceding sections of this chapter. This will help us to understand not only the local behaviour, but how each of these local parts of the curves in question relate to the global area parallel, and how they all connect up.

3.9.1   Examples of a Fixed Area Parallel

The curves we will analyse are shown in Figure 22. As can be seen, the left hand diagram contains four inflections, but the right hand diagram contains only two. The area parallels shown are non-zero area parallels, that is, they display the locus of midpoints of chords that cut off particular areas, each with some positive value.
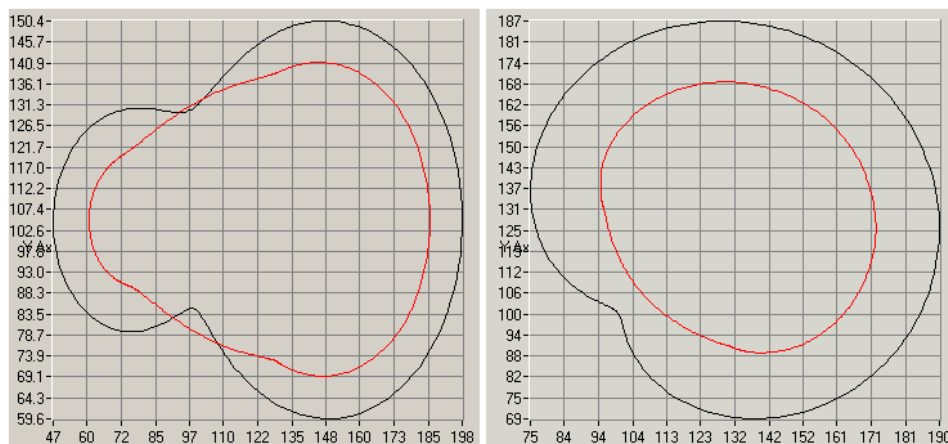
Figure 22: Two examples of non-convex closed plane curves with non-zero area parallels.

In the left hand diagram, we shall take one of the extreme points, and observe how the area parallel behaves locally. In order to clarify the shape of the local area parallel that refers to these parts of the curve, it will help us to zoom in somewhat to the local sections in question. This is discussed and illustrated over-leaf, in Figure 23. In this diagram, the local part of the curve near the extreme point behaves like a quadratic curve.

In the right hand diagram, the curve is similar, but we will see that the local part of the curve near the extreme point is more comparable with a cubic. This refers to the discussion in Section 3.6, where the coefficient of the quadratic term concerned is zero. This is discussed in Figure 24. These diagrams show us how seemingly similar curves can have area parallels with different characteristics.

44
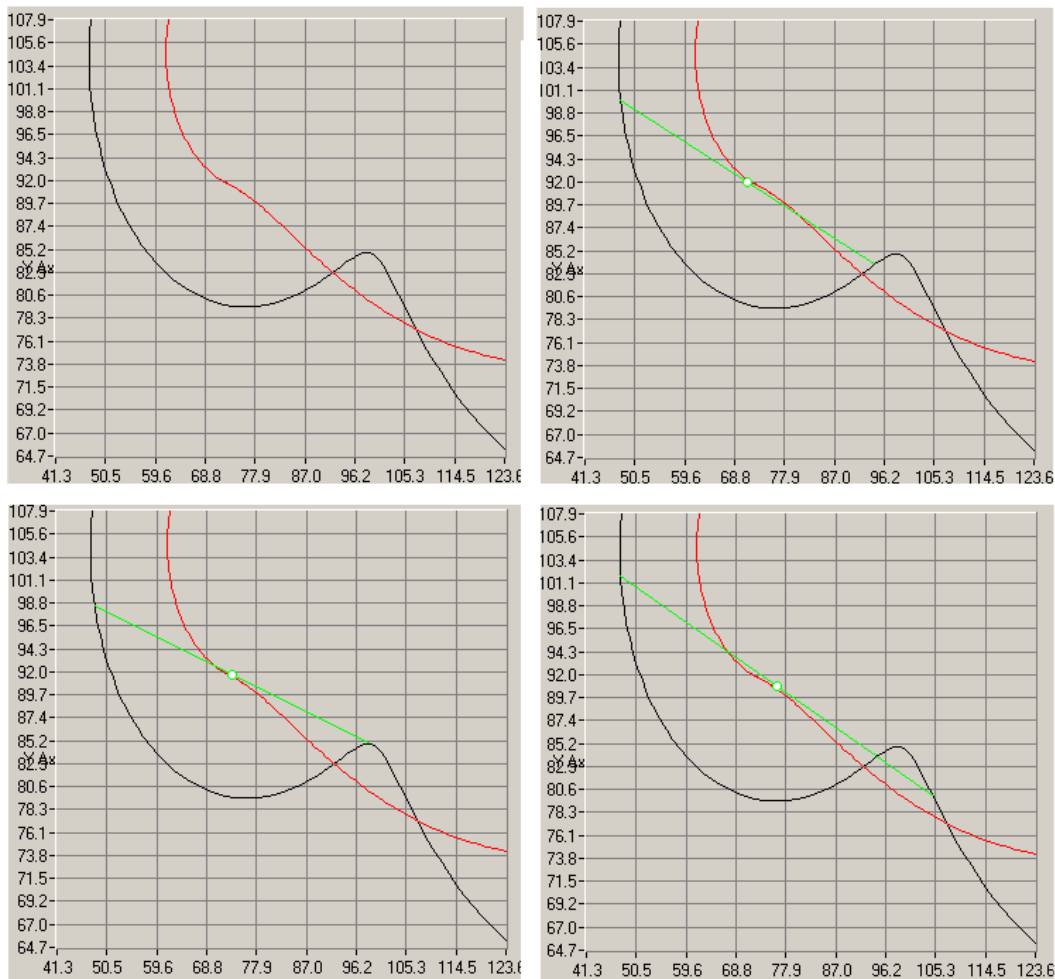
Figure 23: We have four interesting diagrams here. The top left diagram shows the local area that we are analysing. The other three show the chord in slightly different positions. The endpoint of the chord that is between the two inflections, which we will call e_1, moves along the curve, allowing the other endpoint, e_2, to move accordingly. We can see that there is a slight change in the direction in which the locus of such midpoints turn as the endpoint e_1 moves past the local extreme point. Indeed, we can see that these midpoints trace out a local inflection, and is therefore analogous to a cubic curve. Further analysis shows that the endpoint e_2 moves in one direction, and when e_2 moves past the local extreme point, it starts moving in the other direction along the same path. This curve corresponds to the discussion in Section 3.4.
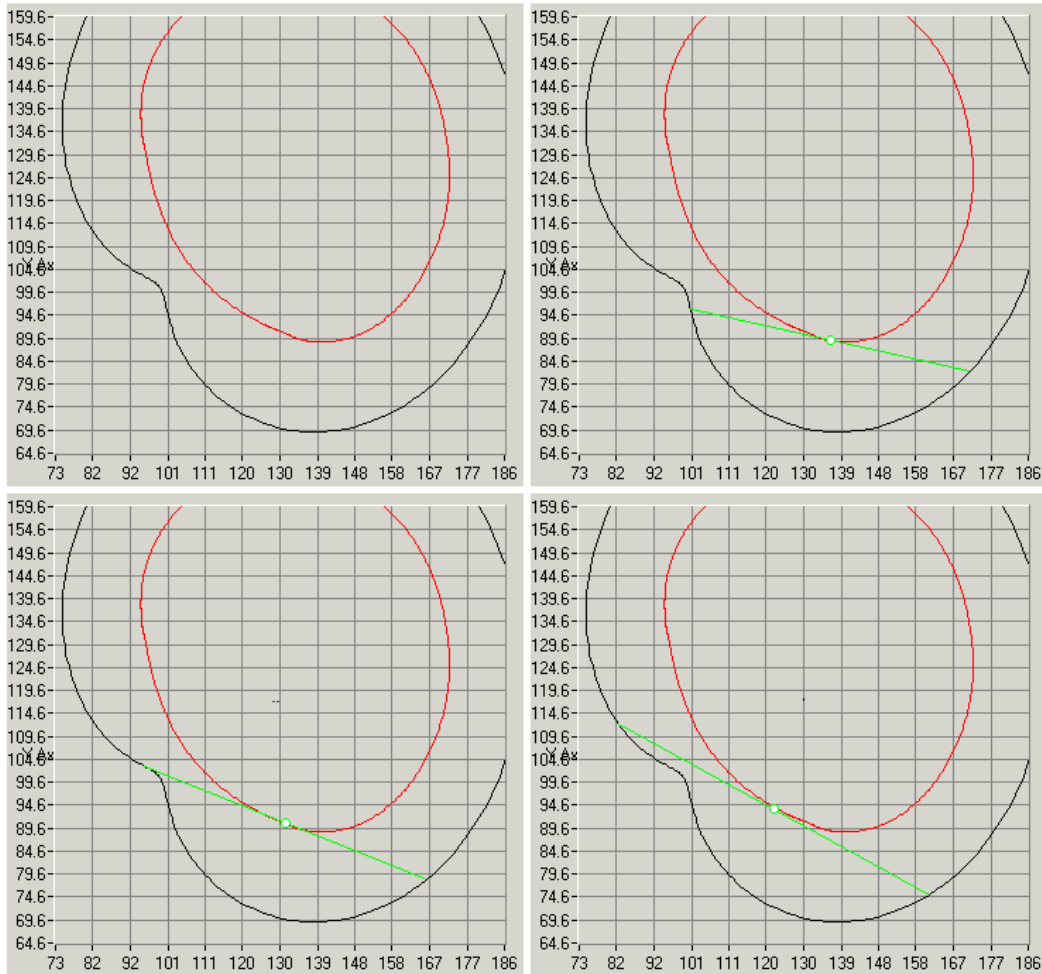
Figure 24: We now have four more interesting diagrams. The top left diagram once again shows the local area that we are analysing, and the other three show the chord in slightly different positions. In this instance, the area parallel is drawn, so that our corresponding chord has three point contact with the curve at the local extreme point. This means that the endpoint e_1 goes through an inflection point rather than a local minimum or maximum. This is shown in the bottom left diagram. If the chord were to continue beyond its intersection with the curve, then it would have three point contact at the inflection point. We can see that, in this case, there is no change in the direction in which the locus of the midpoints of the chords are turning indicating, as expected, that the local area parallel is analogous to a quartic curve. Indeed, the way the curve is drawn corresponds to the discussion in Section 3.6.

46

We find that, in all cases, the points of interest are at the non-convex parts of the curve. The area parallel at the convex parts of the curve behave as is expected, and therefore do not require any further discussion.

3.9.2 Examples of a Zero Area Parallel

The next example shows a zero area parallel. In fact, due to complications when drawing zero area parallels, we cannot be as accurate as we would like, so instead, we shall draw area parallels where the chord cuts off a tiny area, so we can get the gist of the general behaviour. We find that these pictures are good enough to show us what is required. For simplicity, we will continue to describe these area parallels as zero area parallels, as it is the local behaviour around the inflection points of the curve that are of interest in this study.

Figure 25 shows an example of a curve plotted with its zero area parallel. We find that there are indeed cusps near the local extrema.
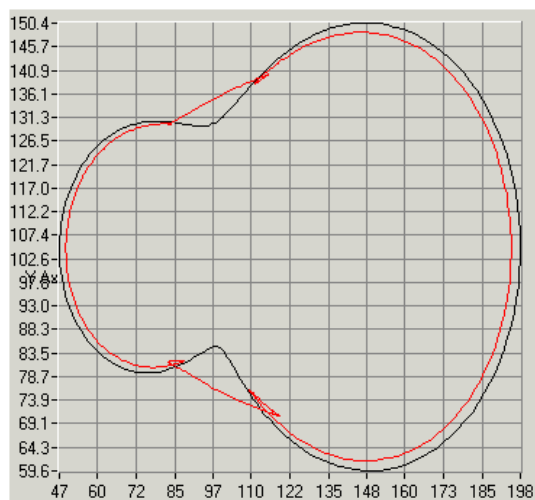
Figure 25: The curve here is the same as the curve in the left hand diagram in Figure 22. The zero area parallel is also shown.

We shall now zoom in to the lower of the two extrema. Figure 26, over leaf, shows the local behaviour of the area parallel near this extreme point.

This enlargement confirms much of the theory in Section 3.8. We find that there are indeed rhamphoid cusps which seem to exist near the inflection points. Obviously, this diagram shows the area parallel with chords cutting off a tiny area, but if we had a precise diagram of the zero area parallel, the two branches of each rhamphoid cusp would meet *at* the inflection points. We can also see from the diagram that the more extreme inflection point corresponds to a smaller rhamphoid cusp, with less of a deformation between the two branches.

47
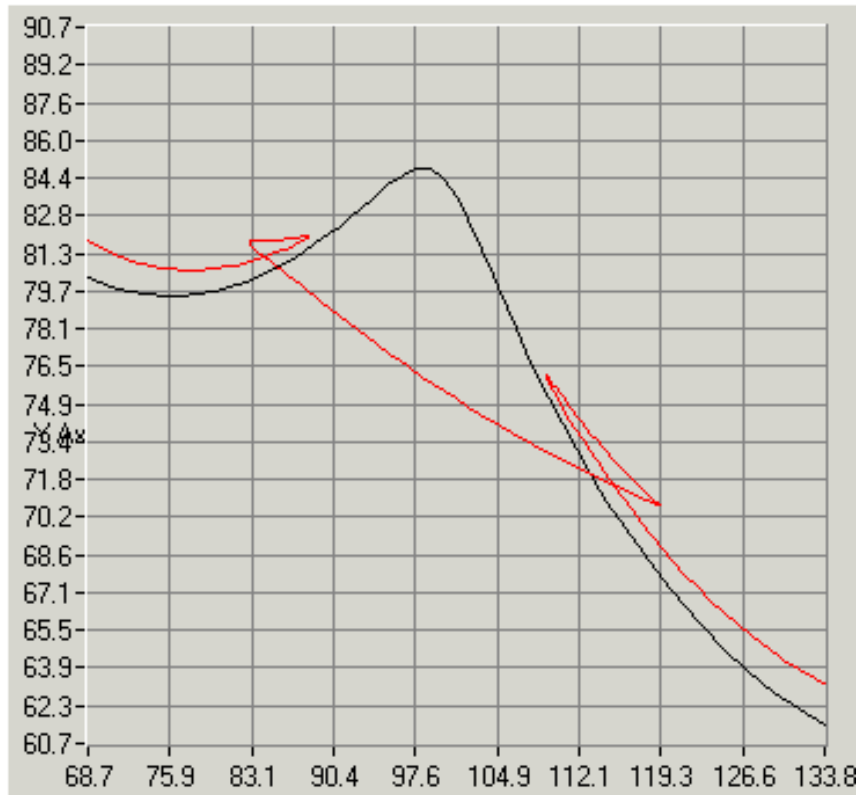
Figure 26: The curve and area parallel, having enlarged one of the areas of interest.

4 Quadrilaterals

4.1 An Affine Transformation

We know, from the discussion in the Mini-Dissertation, that any three points in the plane can be affinely transformed onto any other three points in the plane. It is due to this feature that, when discussing triangles in the Mini-Dissertation, we had the convenience of being able to affinely transform any arbitrary triangle onto any other triangle. After all, a triangle is made up of three points in the plane. This allowed us to constrict our study to that of equilateral triangles, and still produce general results involving the area parallel and the symmetry set.

In this section of the Dissertation, I would like to study the area parallels and symmetry sets of quadrilaterals. We have several more difficulties to consider. Firstly, unlike the case of triangles, we cannot affinely transform any arbitrary quadrilateral into one single quadrilateral. This is due to the fact that there are so many more considerations to be taken into account when specifying an arbitrary quadrilateral. We do, however, have some degrees of freedom that may help us specify *parts* of a quadrilateral that can be affinely transformed.

Using the same statement that three points in the plane can be affinely transformed onto any other three points, we may assume that two of the sides of the quadrilateral are the same length and perpendicular to each other. The other two sides are, however, completely arbitrary. Figure 27 shows some examples of the quadrilaterals we will be studying.
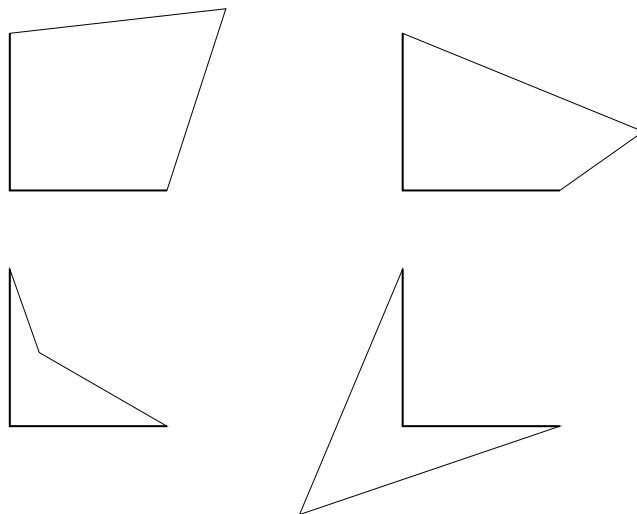
Figure 27: Some possible quadrilaterals. The dense lines show the affine levels of freedom.

Figure 27 leads us on to another difficulty that we did not encounter when looking at triangles. The two diagrams on the lower level are both non-convex. It

would be interesting to understand how the area parallel behaves with non-convex quadrilaterals, as we have done with non-convex curves.

4.2 Convex Quadrilaterals

We will begin with a discussion on convex quadrilaterals. In this section we will discuss how the area parallel behaves, giving numerous specific examples. The discussion will be connected with parallel tangent lines and cuspidal formation, and the self intersection points will be analysed in order to devise some details concerning the symmetry set. We find, in all cases, that it is helpful to imagine that the quadrilaterals in question have very slightly rounded vertices as well as slightly rounded sides, as this clarifies the existence of parallel tangents at the relevant parts of the figure. After all, finding tangent lines at corners can be rather problematic.

4.2.1 The Area Parallel

We will begin by taking some examples of quadrilaterals, and presenting them graphically with their area parallels. From these we will single out the points of interest, and discuss them in detail. First of all, however, it is worth mentioning something about the shape of the area parallel. Let us recall the definition of the area parallel. The area parallel is the locus of midpoints of chords which enclose a fixed area within a closed structure.

When discussing triangles in the Mini-Dissertation, we found that the area parallel was made up of six different branches of hyperbolae. In fact, it is no surprise that the area parallel of a quadrilateral is also made up of different branches of hyperbolae, for the simple reason that any quadrilateral can be turned into a triangle by simply extending the sides! This rather curious phenomenon is illustrated in Figure 28.
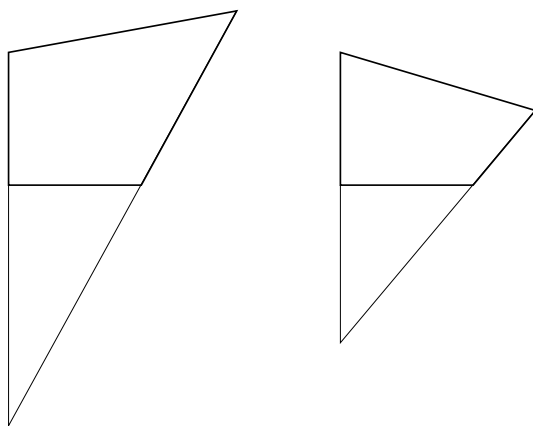
Figure 28: Any quadrilateral can be turned into a triangle by simply extending the sides.

50

The diagram shows two examples of how a quadrilateral can be compared to a triangle. It is then no surprise that the area parallel of a quadrilateral is formed in the same way as that of a triangle. The number of branches in the area parallel of a quadrilateral, however, will vary, unlike the case of a triangle, as we will come to realise shortly.

Let us now, as promised, take a look at a quadrilateral illustrated with its area parallel. Figure 29 shows a quadrilateral, with two different area parallels. That is, a different area is cut off by the envelope of chords in each diagram.
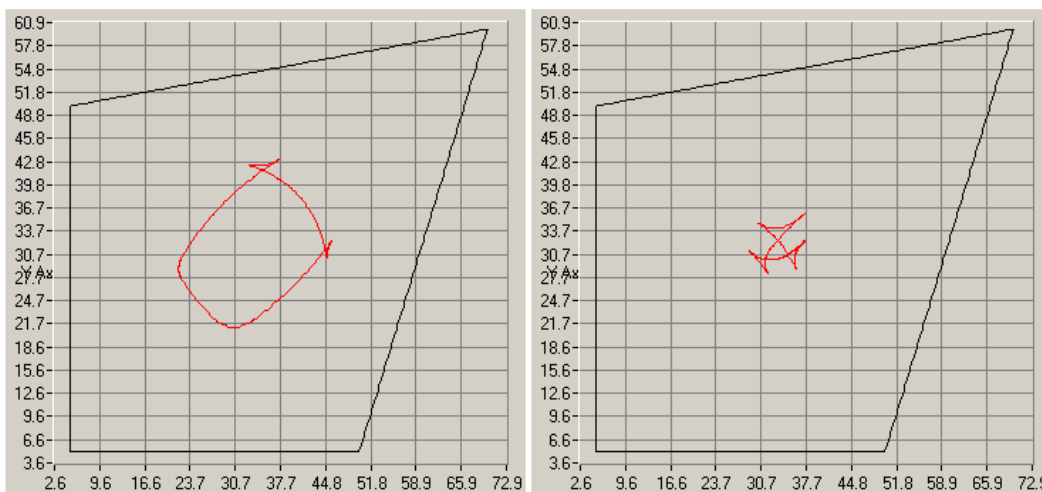
Figure 29: A quadrilateral with two different area parallels.

We can see clearly that the left hand diagram refers to chords cutting off a smaller area than in the right hand diagram. We also notice that, unlike the triangle, there appears to be a higher number of cusps when a larger area is cut off. Indeed, in the left hand diagram there are four cusps, but in the right hand diagram there are six, and so there is some value of fixed area at which we experience cuspidal formation. (As we know by now, cusps always appear in pairs.)

This opens up several points of interest. We can ask ourselves three questions. Firstly, at what value of area do the two new cusps first appear? Secondly, how does the existence of parallel tangent lines, as was described in the chapter discussing convex plane curves, correspond to the number of cusps present in a given area parallel? Finally, can we perhaps find area parallels with eight cusps?

Before addressing the question of cuspidal formation, let us try to understand more about the cusps. We know already that the cusps are the endpoints of branches of hyperbolae, but we don't really know, at this point, how they relate to the chord moving around the quadrilateral. In order to understand more, the reader is referred to Figure 30, which shows the different stages in order of the chord moving anti-clockwise around the quadrilateral.
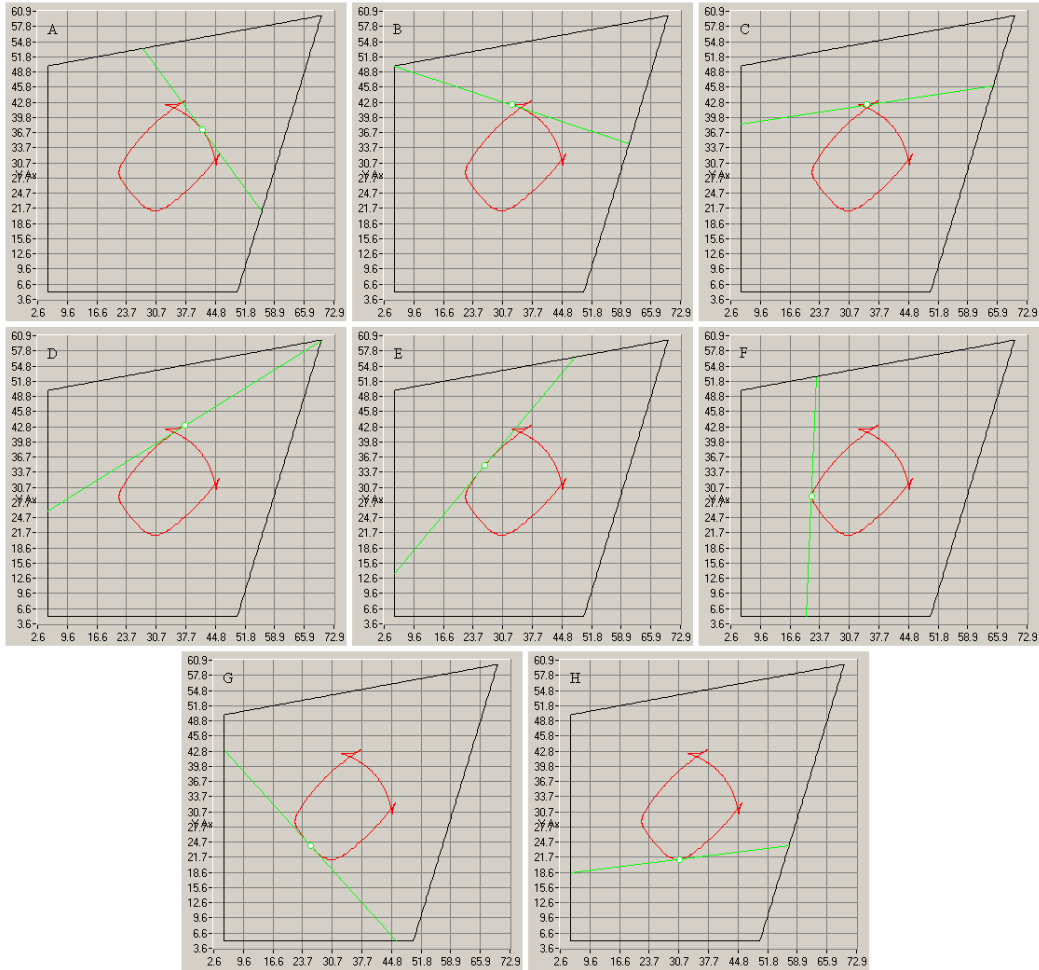
Figure 30: The chord moves around the quadrilateral. We can see how the area parallel is formed.

For referral purposes, let us define the coordinates of the vertices of the quadrilateral to be $(0,0)$, $(1,0)$, $(0,1)$ and (a,b). We can see from Figure 30 that when an endpoint of the chord passes through the vertices $(1,0)$, $(0,1)$ or (a,b), a cusp is formed on the area parallel. However, when either endpoint passes through $(0,0)$, we do not observe a cusp on the area parallel, but a continually (although locally) smooth curve.

The reason for this is that the area cut off by the chord in this particular example is small enough that one endpoint can be on the vertical side of the quadrilateral at the same time as the other endpoint being on the horizontal side. This is shown in diagram G of Figure 30. It is the fact that the endpoints continually exist on the two adjacent sides passing through the origin in this section of the quadrilateral that allows a smooth curve to be drawn.

This part of the area parallel, however, is no different to the rest. It still consists of three different branches of hyperbolae joined together, but they are joined with continuously turning tangents. That is, we find that at this part of the area parallel, the branches at the connection point curve in the same direction, whereas branches that join together to form cusps experience a change in direction of curvature at the connection. This is illustrated in Figure 31.
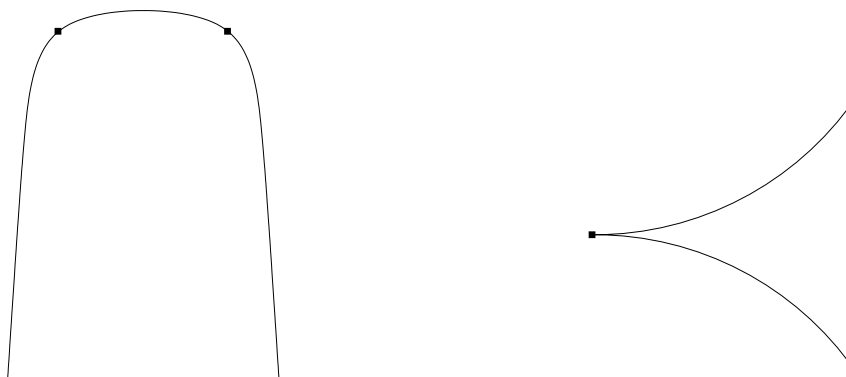
Figure 31: The branches of hyperbolae may join together to form a continuous curve, or a cusp.

In the left hand diagram we can see that it is possible to draw tangent lines at the connection points (clearly marked), but it is not possible to draw a tangent line at the connection point in the left hand diagram, as this is a cusp. This information must be considered in response to the second of our questions.

Meanwhile, however, let us take a look at an example of an area parallel with six cusps, similar to the right hand diagram of Figure 29.
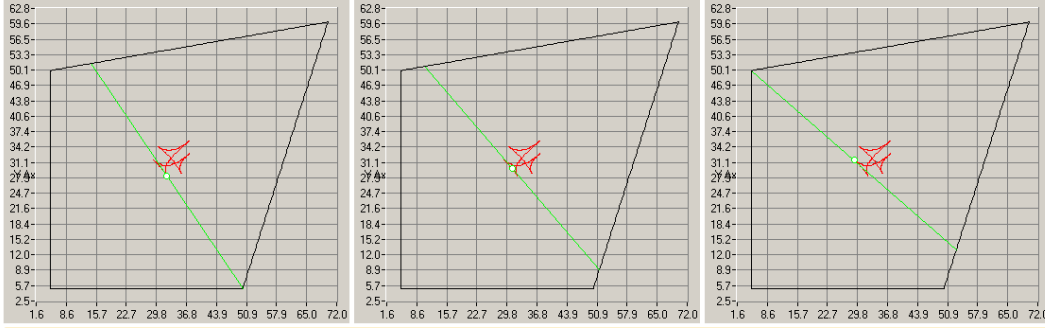
Figure 32: An area parallel with six cusps.

We can see in Figure 32 that there are two new cusps. Obviously, the cusps refer to the endpoints of the chord going past vertices, but the important feature here is the fact that, as can be seen in the middle diagram particularly, the endpoints of the chord are on the two diagonal sides of the quadrilateral at the same time. That is, they are on adjacent sides of the quadrilateral passing through the vertex (a, b), *not* the origin.

So what has changed? We know that the area cut off by the chord has increased, and we also know that the cusps appear when the endpoints can no longer be on the vertical *and* the horizontal sides of the quadrilateral simultaneously. This pretty much answers our first question. From this information, it is relatively clear that the cusps form when the area enclosed by the chord is the same as the area of the triangle with vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$. This is illustrated in Figure 33.
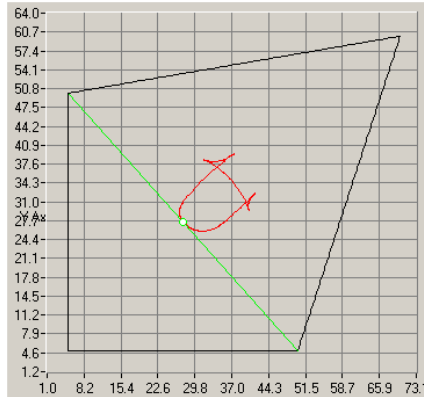
Figure 33: An area parallel at the point of cuspidal formation.

In this diagram, we can see that the area cut off by the chord is the same as the area of the triangle mentioned above. The midpoint of this chord is therefore at the exact point of cuspidal formation on the area parallel.

Let us now move on to discuss the second of our three questions, concerning parallel tangent lines. In the chapter on convex plane curves we proved that if the first and second derivatives of an area function are both zero, but the third derivative is non-zero, then we have an A_2 singularity, and we therefore observe cusps in the plane. This refers to the parts of the curves where the tangent lines at the two endpoints of the chord have the same gradient, and are therefore parallel.

As has been mentioned, once the vertices of a quadrilateral have been rounded, and we imagine that the sides are slightly bulging, we are suddenly dealing with a smooth closed plane curve.

We find that none of the cusps actually corresponds with the origin. As is stated above, the two new cusps form due to the fact that the endpoints of the chord exist on the two sides of the quadrilateral that meet at the point (a, b), and therefore correspond to the vertex (a, b). This is supported by the fact that there are no possible tangent lines at the origin that are parallel to any of the sides of the quadrilateral. The point (a, b), however, has two possible tangent lines that are parallel to the other sides. Also, the vertices $(1, 0)$ and $(0, 1)$ each have one tangent line parallel to one of the sides. This is shown in the diagram in Figure 34.
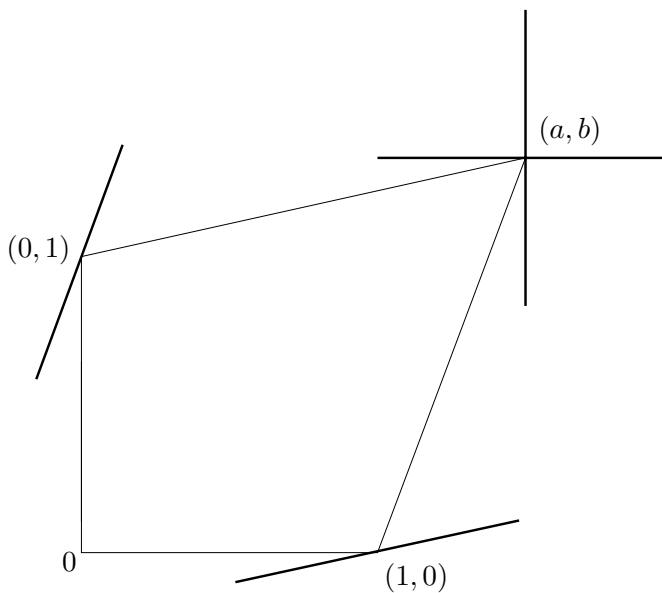
Figure 34: The dense black lines show the tangent lines that can be parallel to a side of the quadrilateral. It is best to imagine that the vertices of the quadrilateral are very slightly rounded, and that the sides are slightly bulging.

We can see in the diagram that the each of the dense black lines is parallel to one of the four sides of the quadrilateral, which helps explain why certain vertices correspond with the existence of cusps. As is mentioned above, the origin has no proper tangent lines parallel to any of the sides, and so does not correspond to the existence of any of the cusps.

In order to answer the third of our questions, referring to whether or not there exist area parallels containing eight cusps, let us first of all study the most degenerate case possible, that of the square, as there is a possibility that the area parallel may have eight cusps, due to its symmetry. Figure 35 shows the square with two different area parallels. That is, the envelope of chords separates a different area in each diagram.
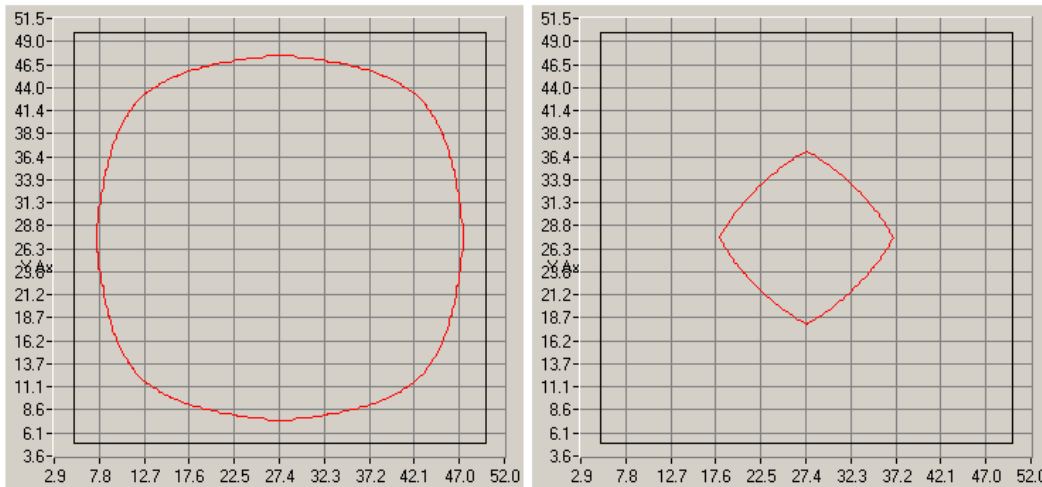
Figure 35: The most degenerate example of the quadrilateral is the square.

It is clear that the left hand diagram displays the envelope of chords cutting off a much smaller area than the right hand diagram. It is also clear that as this area increases, there seem to be four parts of the area parallel that become more and more angular. These parts of the curve are on the brink of turning into cusps, but never quite get there. They exist when the chord, which is moving around the square, goes through a vertex. In fact, the one end of the chord goes through one vertex, and the midpoint of this chord *stays* at its corresponding point on the area parallel, until the other end of the chord goes through its adjacent vertex. The midpoint actually stays at the angular part of the area parallel.

Alternatively, this can be explained by referring to the sides of the square. If the endpoints of the chord are on opposite sides of the square, then this refers to the angular parts of the area parallel. However, as they move along perpendicular sides of the square, the curved part of the area parallel is traced out. This is expressed in the diagrams in Figure 36.
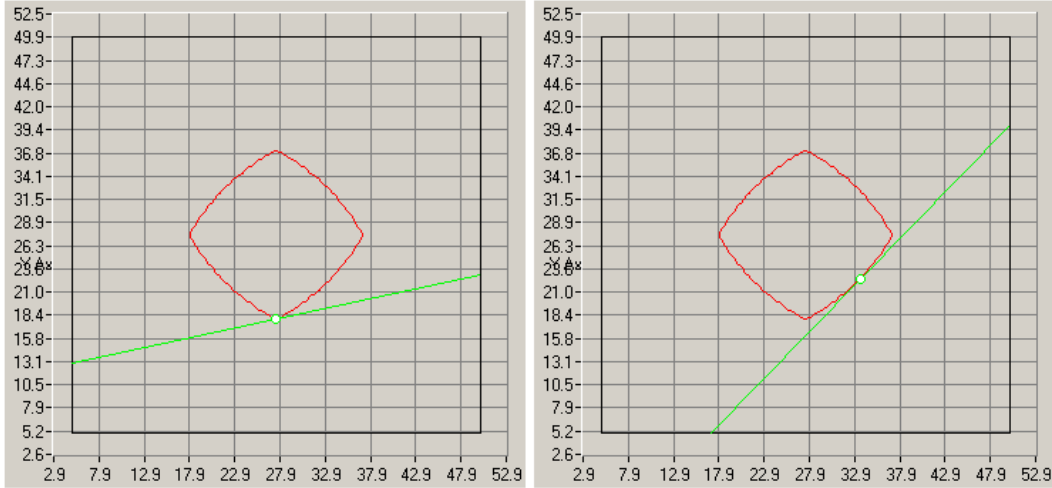
Figure 36: Two of the chords cutting off the same area.

In Figure 36, the chord is moving anti clockwise around the square. In the left hand diagram, we can see that the endpoints are on opposite sides of the square. We will call the left hand endpoint e_1 and the right hand endpoint e_2. As the chord moves around the square, the midpoint will stay at the same point, until e_1 moves through a vertex. The point where the midpoint is stationary is the angular part of the area parallel. Once e_1 has moved past the vertex, the two endpoints will then be present on two sides of the square that are perpendicular to each other. This is shown in the right hand diagram. The midpoint of the chord will then be in motion, tracing out one of the four curved sections of the area parallel, until e_2 goes through the next vertex. When this happens, the two endpoints are once again on opposite sides of the square, which means the midpoint of the chord will be stationary once more at the next angular part of the area parallel, and the whole process continues.

An important feature of this situation is that when the endpoints are on opposite sides of the square the midpoint stays at exactly the same point, even though the chord is still in motion. This shows that, despite the fact that the area parallel appears angular at this point, it does not quite turn into a pair of cusps. In fact, the area parallel of the square does not have eight cusps, but strictly speaking it does not have any!

So having looked at the most likely example, we can possibly assume that there are no area parallels of this type with eight cusps. But using the theory we have concerning parallel tangent lines, let us attempt to prove this.

In order to prove this, we shall look at a general quadrilateral, with angles α, β, γ and δ, as shown in the diagram in Figure 37. Clearly, $\alpha + \beta + \gamma + \delta = 2\pi$. For purposes of reference, we will call the vertex with angle α vertex A, angle β vertex B, and so on.

57
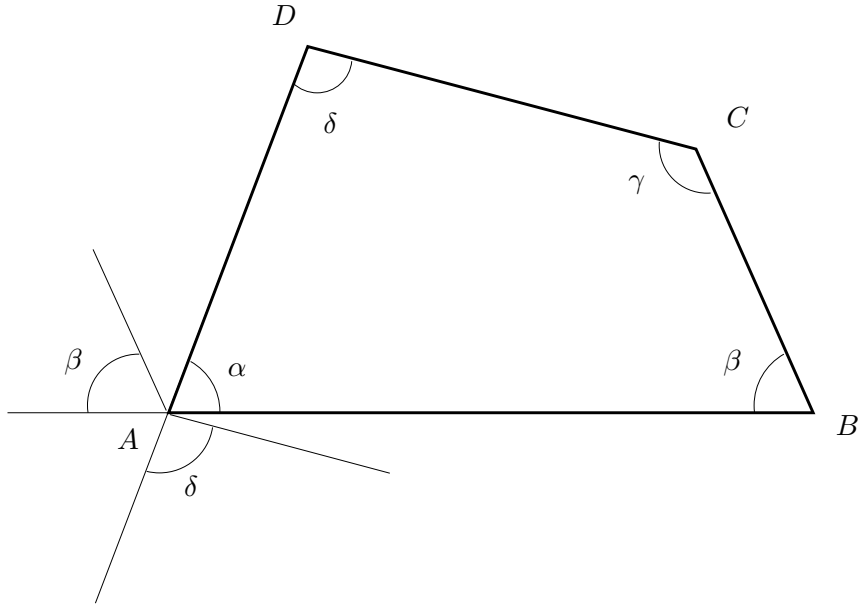
Figure 37: A general quadrilateral. Interior angles add up to 2π.

We will primarily look at vertex A. It is quite clear that this vertex will have two tangent lines that are each parallel to one of the sides of the quadrilateral. One will be parallel to side BC and the other parallel to side CD. We can also see from Figure 37 that since there is a parallel tangent line with line BC, then the angles at vertices A and B must add up to something less than π. That is to say that two cusps exist if and only if $\alpha + \beta < \pi$, due to the fact stated above that parallel tangent lines imply cusps.

By contrast, we can see that there are no possible tangent lines at vertex C that will be parallel to any of the sides. This is due to the fact that $\beta + \gamma > \pi$ and $\gamma + \delta > \pi$.

So, in order for there to be eight cusps, we require the sum of any two adjacent angles to be less than π. That is $\alpha + \beta < \pi$, $\alpha + \delta < \pi$, $\beta + \gamma < \pi$ and $\gamma + \delta < \pi$. Adding these together, this implies that

$$
\begin{aligned}
2(\alpha + \beta + \gamma + \delta) &< 4\pi \\
\Rightarrow \quad \alpha + \beta + \gamma + \delta &< 2\pi
\end{aligned}
$$

But we know that with *any* quadrilateral $\alpha + \beta + \gamma + \delta = 2\pi$, which is a contradiction. Hence, there does not exist an area parallel of a quadrilateral with eight cusps. We therefore come to the conclusion that the only cuspidal formation occurs from four cusps to six.

4.2.2 The Symmetry Set

Let us first recall the definition of the symmetry set. The symmetry set is the locus of self intersections of the area parallel. In order to observe how the symmetry set behaves we will set up a general quadrilateral on a set of Cartesian axes. The two fixed sides of the quadrilateral will remain where they are, and for simplicity we will give them each length one along their respective axes, meeting at the origin. The fourth vertex of the quadrilateral will have coordinates (a, b) with $a > b$. Figure 38 illustrates.
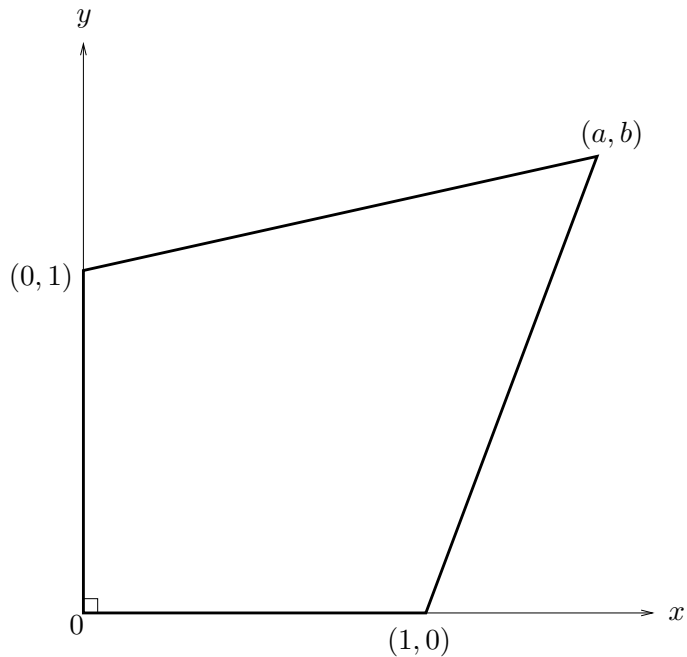
Figure 38: A general quadrilateral.

When finding the symmetry set of the equilateral triangle in the Mini-Dissertation, we used different sized rectangles within the triangle. We have seen that the characteristic points of the area parallel are always at the midpoint of the chord cutting off fixed area. So, in the case of the equilateral triangle, when two chords cutting off the same area meet at the midpoints, we have the two diagonals of a rectangle. So the symmetry set is traced out by the centres of all the rectangles that will fit into the triangle.

The reason why the symmetry set of a triangle is traced out by the centres of rectangles is simply down to the symmetry of the triangle. When considering quadrilaterals, however, the symmetry set will be traced out by the centres of parallelograms. Indeed, rectangles are rather degenerate examples of parallelograms.

So let us now add to Figure 38, by incorporating a parallelogram inside the quadrilateral. There is, in fact more than one way of doing this. In all cases there

will be one vertex of the parallelogram on each of the two diagonal sides of the quadrilateral. For the first case that we will consider, we will also have one vertex on each of the fixed sides as well. As these vertices are both on the axes, we will call them $(0, p)$ and $(q, 0)$. Clearly, $0 < p < 1$ and similarly, $0 < q < 1$. For now, we will call the other two vertices (u_1, v_1) and (u_2, v_2). Figure 39 illustrates further.
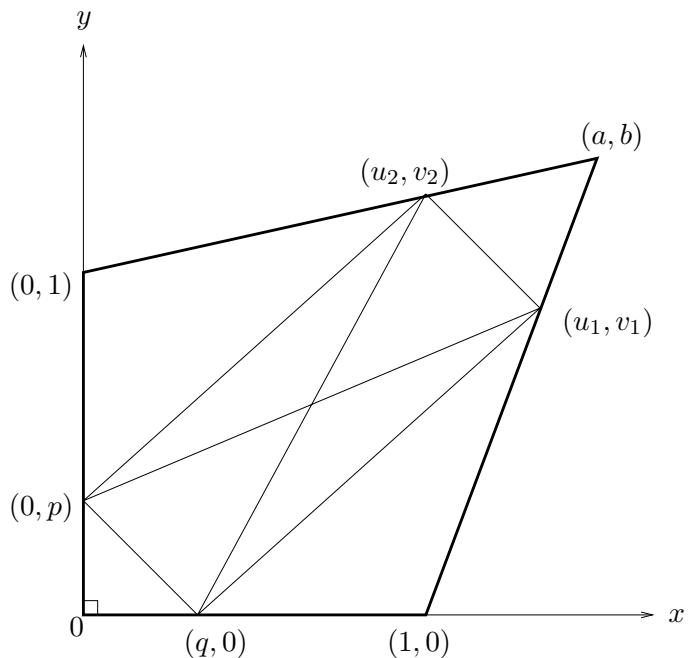
Figure 39: A general quadrilateral, with a parallelogram incorporated.

Let us now define two scalars λ and μ. We will restrict them so that $0 < \lambda < 1$ and $0 < \mu < 1$, with $\lambda \neq \mu$, for reasons that will become apparent. We will use these scalars to find the coordinates of (u_1, v_1) and (u_2, v_2), in terms of a and b. For (u_1, v_1), this point will have an x-coordinate greater than 1 but a y-coordinate between 0 and b. It will therefore take the form

$$(1, 0) + \mu(a - 1, b)$$

On the other hand, (u_2, v_2) has an x-coordinate between 0 and a and a y-coordinate greater than 1, and so will be of the form

$$(0, 1) + \lambda(a, b - 1)$$

We can see now that $\lambda \neq \mu$, since we are avoiding the degenerate case where $a = b$.

As we have said, the idea of drawing in this parallelogram is due to the fact that its two diagonals cut off the same area, but with respect to different vertices. It is therefore important to point out that the quadrilateral enclosed by the points

60

$(0,0)$, $(1,0)$, (u_1, v_1) and $(0,p)$ has the same area as the quadrilateral enclosed by the points $(0,0)$, $(0,1)$, (u_2, v_2) and $(q,0)$.

We would like to see how the midpoint of the diagonals of the parallelogram changes as the parallelogram changes position. We will call the midpoint (m_x, m_y). First of all, however, we can see from Figure 39 that the two areas mentioned above have some area in common which we can eliminate. The area in common is the quadrilateral enclosed by coordinates $(0,0)$, $(0,p)$, (u_1, v_1) and $(q,0)$, which is the same area as the quadrilateral enclosed by $(0,0)$, $(0,p)$, (u_2, v_2) and $(q,0)$. The reader is referred once again to Figure 39 for clarity. This leaves us with two areas, A_1 and A_2, as shown in the diagram in Figure 40. Clearly, $A_1 = A_2$.
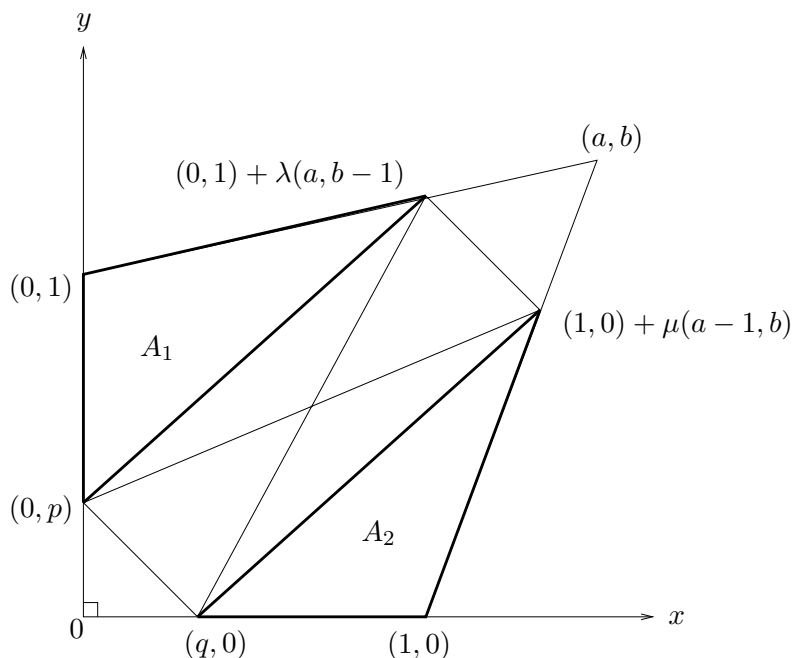
Figure 40: The areas enclosed by the dense black lines are equal.

So we have three conditions that must be met. Firstly, the midpoints must be the same, giving us two conditions (one for each of the chords), and in addition, the areas must be the same, that is $A_1 = A_2$.

Let us now find the coordinates of the midpoint. As usual, the midpoint is the average of the endpoints of the chord. So from one chord we have

$$(m_x, m_y) = \frac{1}{2}(1 + \mu(a-1), p + \mu b)$$

and from the other chord

$$(m_x, m_y) = \frac{1}{2}(q + \lambda a, 1 + \lambda(b-1))$$

61

Hence, we have

$$\begin{array}{rcl} 1 + \mu(a-1) & = & q + \lambda a \\ p + \mu b & = & 1 + \lambda(b-1) \end{array} \Rightarrow \left\{ \begin{array}{rcl} 1 - q & = & \lambda a - \mu(a-1) \\ 1 - p & = & \mu b - \lambda(b-1) \end{array} \right.$$

Now we can use the fact that the areas are the same.

$$\begin{array}{rcl} A_1 & = & \frac{1}{2}(1-p)\lambda a \\ A_2 & = & \frac{1}{2}(1-q)\mu b \end{array} \Rightarrow a(1-p)\lambda = b(1-q)\mu$$

We can now substitute in for $(1-p)$ and $(1-q)$. We have

$$\begin{array}{rcl} a(\mu b - \lambda(b-1))\lambda & = & b(\lambda a - \mu(a-1))\mu \\ a(\mu b - \lambda b + \lambda)\lambda & = & b(\lambda a - \mu a + \mu)\mu \\ \lambda\mu ab - \lambda^2 ab + \lambda^2 a & = & \lambda\mu ab - \mu^2 ab + \mu^2 b \\ \lambda^2 a(1-b) & = & \mu^2 b(1-a) \\ \dfrac{\lambda^2}{\mu^2} & = & \dfrac{b(a-1)}{a(b-1)} \end{array}$$

$$\text{Hence} \qquad \frac{\lambda}{\mu} = \sqrt{\frac{b(a-1)}{a(b-1)}}$$

Since λ and μ are always positive, then the right hand side will be the positive square root. Also, since a and b are constants, for simplicity we will redefine this right hand side as follows.

$$c = \sqrt{\frac{b(a-1)}{a(b-1)}}$$

So $\lambda = \mu c$. We can substitute this back in to the expressions for p and q, giving

$$\begin{array}{rcl} p & = & 1 + \mu c(b-1) - \mu b \\ & = & 1 + \mu bc - \mu c - \mu b \\ & = & 1 + \mu(bc - b - c) \\ q & = & 1 + \mu(a-1) - \mu ca \\ & = & 1 + \mu a - \mu - \mu ac \\ & = & 1 + \mu(a - ac - 1) \end{array}$$

Since a, b and c are constants, we now have p and q as functions of μ. As p and q move between 0 and 1 along their corresponding axes, the midpoint of the diagonals of the parallelogram moves. So we can now use these expressions for p and q to find the coordinates of the midpoint. We have

$$(m_x, m_y) = \frac{1}{2}(1 + \mu(a-1), 1 + \mu(bc - b - c) + \mu b)$$

62

Hence

$$m_x(\mu) = \frac{1}{2}(1 + \mu(a - 1))$$

$$m_y(\mu) = \frac{1}{2}(1 + \mu(bc - c))$$

Here, we can see that when $\mu = 0$ we have $(m_x, m_y) = (\frac{1}{2}, \frac{1}{2})$, proving that the endpoint of this section of the symmetry set is at $(\frac{1}{2}, \frac{1}{2})$. Now, if we differentiate these expressions, we get

$$m_x'(\mu) = \frac{1}{2}(a - 1)$$

$$m_y'(\mu) = \frac{1}{2}(bc - c)$$

both of which are constants. So the midpoint moves both horizontally and vertically with a constant rate of change, which proves that the midpoint traces a straight line. Hence, this section of the symmetry set is a straight line.

We can quite easily calculate where this straight line begins, by substituting $\mu = 0$ into the coordinates of the midpoint. Clearly, we have $(m_x, m_y) = (\frac{1}{2}, \frac{1}{2})$, meaning that this branch of the symmetry set starts on the diagonal of the triangle with vertices $(0, 0)$, $(1, 0)$ and $(0, 1)$. This is expected, as we proved in the last section that the area parallel starts to experience cuspidal formation when the area cut off by the chord is equal to the area of this triangle.

This is only a valid proof for the case where there is one vertex of the parallelogram on each of the axes as shown in Figure 39. Obviously, we may have a different situation once the vertex of the parallelogram, which is currently given by $(q, 0)$ moves around clockwise past the origin. So we must now look at a second case, where both of these vertices are on the y-axis. This case is illustrated in Figure 41. The vertices of the parallelogram remain the same as in the previous diagrams.

Once again, the diagonals of the parallelogram cut off the same area with respect to different vertices. But as before there is a segment of this area common to both which can be eliminated. The diagram in Figure 42 illustrates the two areas that are equal, once the area common to both has been eliminated. As before $A_1 = A_2$, and the same conditions must be met. That is, the midpoints of the diagonals of the parallelogram must be the same.

As before, we will find the general coordinates of the midpoint, in order to see how the symmetry set behaves once the vertices of the parallelogram have moved around clockwise. But first we need to calculate the areas A_1 and A_2 in terms of their respective variables. It is simple enough to find the area A_1. We have

$$A_1 = \frac{1}{2}(1 - q)\lambda a$$

63
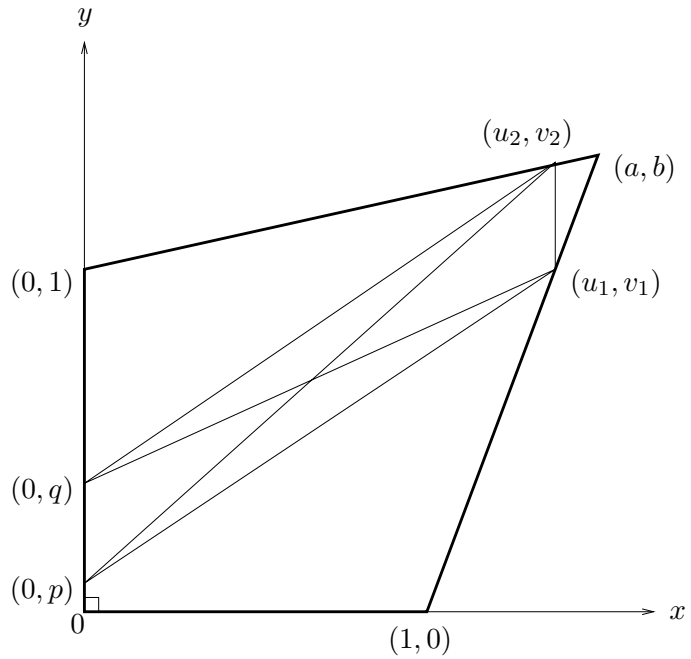
Figure 41: A general quadrilateral, with a parallelogram incorporated. Now the vertices of the parallelogram have moved around clockwise.
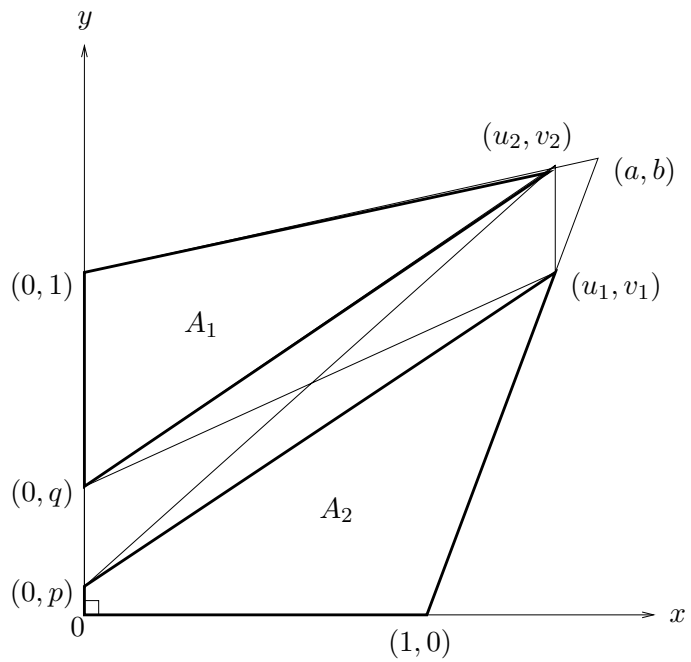
Figure 42: We can see the equal areas once the parallelogram has moved beyond the vertex.

64

But A_2 is a little more tricky, as we are dealing with the area of a quadrilateral, rather than a triangle. In fact, it can be split into two triangles, one with base p and height λa and the other with base 1 and height μb. We then have

$$A_2 = \frac{1}{2}(\lambda ap + \mu b)$$

Hence, this tells us that

$$(1-q)\lambda a \;=\; \lambda ap + \mu b$$
$$\Rightarrow \quad 1-q \;=\; p + \frac{\mu b}{\lambda a}$$

From the diagrams, we can see that the midpoint is given by

$$(m_x, m_y) \;=\; \frac{1}{2}(1 + \mu(a-1), q + \mu b)$$
$$=\; \frac{1}{2}(\lambda a, 1 + p + \lambda(b-1))$$

then, we can equate both the x-coordinates and the y-coordinates of these two expressions. Firstly, from the x-coordinates we have

$$\lambda a \;=\; 1 + \mu(a-1)$$
$$\Rightarrow \quad \lambda \;=\; \frac{1 + \mu(a-1)}{a}$$

and from the y-coordinates we have

$$q + \mu b \;=\; 1 + p + \lambda(b-1)$$
$$\Rightarrow \quad 1 - q \;=\; \mu b - p - \lambda(b-1)$$

Now we can use our expression for λ and substitute in, so that we get an equation involving just μ. We have

$$1 - q = \mu b - p - (b-1)\left(\frac{1 + \mu(a-1)}{a}\right)$$

and using the fact above that $A_1 = A_2$, we can substitute in for $1 - q$, giving

$$\mu b - p - (b-1)\left(\frac{1 + \mu(a-1)}{a}\right) = p + \frac{\mu b}{\lambda a}$$

and so, using our expression for λ once again, we have, having simplified the right hand side

$$\mu b - (b-1)\left(\frac{1 + \mu(a-1)}{a}\right) = 2p + \frac{\mu b}{1 + \mu(a-1)}$$

65

and after some more simplification, we can express p in terms of parameter μ and constants a and b. We have

$$p(\mu) = \mu b \left(\frac{1}{2} - \frac{1}{2 + 2\mu(a-1)} \right) - (b-1) \left(\frac{1 + \mu(a-1)}{2a} \right)$$

We now have our expression for p in terms of μ. We must then find an expression for the midpoint of the diagonals of the parallelogram, which we find is a lot more complicated than in the previous calculation, where one of the the vertices is on the x-axis. The x-coordinate of the midpoint is fairly straightforward, but the y-coordinate requires some rather tricky algebra. We have for the x-coordinate

$$m_x = \frac{\lambda a}{2} = \frac{1 + \mu(a-1)}{2}$$

and for the y-coordinate

$$
\begin{aligned}
m_y &= 1 + p + \lambda(b-1) \\
&= 1 + \mu b \left(\frac{1}{2} - \frac{1}{2 + 2\mu(a-1)} \right) - (b-1)\left(\frac{1+\mu(a-1)}{2a} \right) \\
&\qquad\qquad\qquad\qquad\qquad +(b-1)\left(\frac{1+\mu(a-1)}{a} \right) \\
&= 1 + \mu b \left(\frac{1}{2} - \frac{1}{2 + 2\mu(a-1)} \right) + (b-1)\left(\frac{1+\mu(a-1)}{2a} \right)
\end{aligned}
$$

It can then be checked in Maple that multiplying out gives

$$m_y = \frac{2a + b - 1 + (2 - 2b - 4a + 2ab + 2a^2)\mu + (2a + b - 1 + 2a^2 b - 3ab - a^2)\mu^2}{4a(1 + \mu a - \mu)}$$

We can now assign x, say to the x-coordinate of the midpoint, giving, once rearranged

$$\mu = \frac{2x - 1}{a - 1}$$

and using maple to substitute this expression for μ into the y-coordinate of the midpoint, we have

$$m_y = \frac{ab + (4a^2 - 4a - 4ab)x + (4 - 4a - 4b + 8ab)x^2}{8a(a-1)x}$$

Looking at the numerator of this expression, we have no y^2 term, which is interesting, since we know that, in general $ax^2 + bxy + cy^2 + $ lower order terms $= 0$ is a hyperbola

66

if and only if $b^2 - 4ac > 0$. Since there is no y^2 term, then c is 0 and so this expression models a hyperbola due to the simple fact that b^2 is always positive for real numbers.

It would be useful, now that we know that this part of the symmetry set takes the form of a hyperbola, to find the endpoints of this hyperbola, so that we can show that the two parts of the symmetry set are connected smoothly to one another. To do this, we will take our expressions for p, m_x and m_y, and substitute in the value $p = 0$ for the one endpoint, and $\mu = 1$ for the other.

It is clear from the diagrams that precede that as μ increases, one of the vertices of the parallelogram moves from $(1, 0)$ up to (a, b), we can see that at $(1, 0)$, $\mu = 0$ and at (a, b), $\mu = 1$. So let us first of all concentrate on the endpoint that corresponds with the vertex of the parallelogram coinciding with the vertex (a, b). That is, where $\mu = 1$. If $\mu = 1$ then

$$m_x = \frac{a}{2}$$

and

$$
\begin{aligned}
m_y &= \frac{2a + b - 1 + 2 - 2b - 4a + 2ab + 2a^2 + 2a + b - 1 + 2a^2b - 3ab - a^2}{4a(1 + a - 1)} \\
&= \frac{2a^2 + 2a^2b - ab}{4a^2} \\
&= \frac{2a + 2ab - b}{4}
\end{aligned}
$$

This tells us that the complete endpoint of this branch of the symmetry set is at the point

$$\left(\frac{a}{2}, \frac{2a + 2ab - b}{4} \right)$$

Now let's take a look at the other endpoint. This occurs where $p = 0$. Using Maple it can be shown that the value for μ when $p = 0$ is

$$\mu = \frac{(a - 1)(b - 1) + \sqrt{ab(a - 1)(b - 1)}}{(a - 1)(a + b - 1)}$$

This must take the positive square root, since both $(a - 1)$ and $(b - 1)$ are positive, and $(a + b - 1)$ is certainly positive. Let us concentrate on the coordinates of this endpoint. Further calculations in Maple show that the x-coordinate is given by

$$m_x = \frac{ab + \sqrt{ab(a - 1)(b - 1)}}{2(a + b - 1)}$$

and the y-coordinate takes the form

$$m_y = \frac{a^2b - 2ab^2 + 2ab^3 - b^3 + 2b^2 - b + (2b^2 + a - b - 1)\sqrt{ab(a - 1)(b - 1)}}{2(a + b - 1)(ab + \sqrt{ab(a - 1)(b - 1)})}$$

67

It would be a good idea now to check that these coordinates are the same as the coordinates of the endpoint of the section of the symmetry set that corresponds to our previous calculations, where one of the vertices of the parallelogram was on the x-axis. This will prove that, although this branch of the symmetry set is made up of two sections, it is still continuous. So let us recall the general coordinates of the midpoint for this case. We have

$$m_x(\mu) = \frac{1}{2}(1 + \mu(a - 1))$$

$$m_y(\mu) = \frac{1}{2}(1 + \mu(bc - c))$$

The vertex of the parallelogram that was on the x-axis had coordinates $(q, 0)$, so we also need to recall our expression for q. We have

$$q = 1 + \mu(a - ac - 1) \text{ where } c = \sqrt{\frac{b(a-1)}{a(b-1)}}$$

The endpoint of this section of the symmetry set is clearly where $q = 0$, so we have

$$0 = 1 + \mu\left(a - ac - 1\right)$$

If we then solve for μ and substitute into the x-coordinate of the midpoint, we have, after some simplification,

$$
\begin{aligned}
2m_x &= \frac{a\sqrt{b(a-1)}}{a\sqrt{b(a-1)} - (a-1)\sqrt{a(b-1)}} \\
&= \frac{a\sqrt{b(a-1)}}{a\sqrt{b(a-1)} - (a-1)\sqrt{a(b-1)}} \times \frac{a\sqrt{b(a-1)} + (a-1)\sqrt{a(b-1)}}{a\sqrt{b(a-1)} + (a-1)\sqrt{a(b-1)}} \\
&= \frac{a(ab(a-1) + (a-1)\sqrt{ab(a-1)(b-1)}}{a(a-1)(a+b-1)} \\
&= \frac{ab + \sqrt{ab(a-1)(b-1)}}{a+b-1}
\end{aligned}
$$

and so we finally have

$$m_x = \frac{ab + \sqrt{ab(a-1)(b-1)}}{2(a+b-1)}$$

as wanted. This proves that the x-coordinates of the endpoints are the same. The interested reader may wish to do the same check for the y-coordinates, by using similar methods.

In summary, we find that this branch of the symmetry set starts of in a straight line, but then turns into a branch of a hyperbola. We have also shown that the endpoints of this branch are at $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{a}{2}, \frac{2a+2ab-b}{4})$.

4.3 Non-Convex Quadrilaterals

Now that we understand the general behaviour of both the area parallel and the symmetry set of convex quadrilaterals, it should come of no surprise to the reader that a study of non-convex quadrilaterals is imminent. Indeed, having looked in detail at the subject of closed plane curves, we know that there are certain features connected with non-convex shapes that are not accessible with convex shapes. The area of interest that comes to mind most readily is that of the zero area parallel, so the bulk of this section will be concerned with this.

In the same way as convex quadrilaterals, we have certain degrees of freedom connected with affine transformations. Once again, any three points in the plane can be affinely transformed onto any other three points in the plane, meaning that two of the sides of the quadrilaterals under investigation in this section will be the same as with the convex quadrilaterals. These two sides are the two perpendicular sides of unit length that meet at the origin. Hence, the quadrilaterals shown in Figure 43 will have similar features to one another.
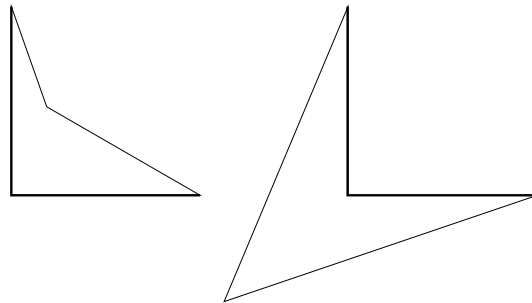
Figure 43: One quadrilateral can be affinely transformed onto the other.

Let us consider a quadrilateral with vertices $(0,0)$, $(1,0)$, $(0,1)$ and (a,b). Since we are now discussing non-convex quadrilaterals, we require a and b to be less than a half, and to avoid generality, we will set $a \neq b$. There are three possibilities available to us when considering the zero area parallel. These are illustrated in Figure 44.
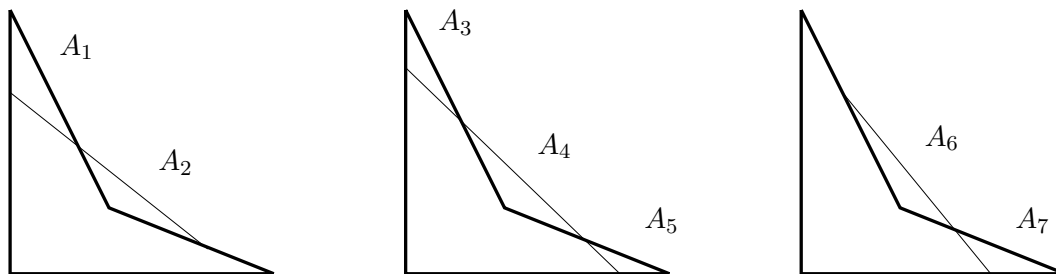
Figure 44: Three possibilities for the zero area parallel.

The A_i are the various areas cut off by the chords in each of the diagrams. Clearly, $A_1 = A_2$, $A_3 + A_5 = A_4$ and $A_6 = A_7$, so that we have zero algebraic area in each case. The chord can move around the quadrilateral as long as the area cut off remains zero. We will look at each of the left hand, and the centre diagrams only, since the right hand diagram is very similar to the left hand diagram, and will produce the same results.

So let's consider the first case. The chord has endpoints on the side joining $(0,0)$ with $(0,1)$ which we will call P, and on the side joining $(1,0)$ with (a,b), which we will call Q. The chord also cuts the line connecting $(0,1)$ with (a,b), and we shall call this point R. This is displayed in the diagram in Figure 45.
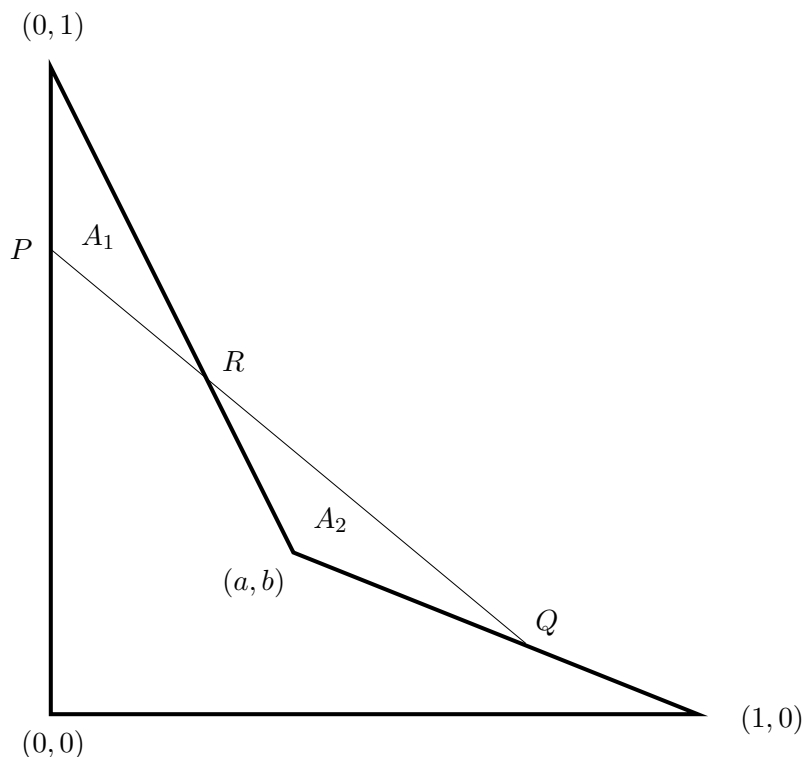
Figure 45: Case one: The chord cuts just one side of the quadrilateral.

Let us consider the points P, Q and R. Clearly P is on the y-axis, and so can be given coordinates $(0,p)$, where p is a constant between 0 and 1. Q and R are more complicated, and we require some more information about the sides of the quadrilateral, so for now we will assign them coordinates (q_1, q_2) and (r_1, r_2), respectively.

The side that connects $(0,1)$ with (a,b) has gradient $\frac{b-1}{a}$ and cuts the y-axis at

$(0, 1)$. It therefore has equation

$$y = \frac{b-1}{a}x + 1$$

For reasons that will become apparent very shortly, we can rearrange this to the form

$$a = ay + (1-b)x$$

The side connecting $(1, 0)$ with (a, b) has gradient $\frac{b}{a-1}$ and cuts the y-axis at $(0, \frac{b}{1-a})$, and so has equation

$$y = \frac{b}{a-1}x + \frac{b}{1-a}$$

In order to model the movement of the chord, we will once again take parameters λ and μ, with $0 < \lambda, \mu < 1$. The coordinates of Q are then given by

$$\lambda(a, b) + (1-\lambda)(1, 0)$$

Hence

$$(q_1, q_2) = (\lambda a + (1-\lambda), \lambda b)$$

Now, using the other parameter, we can derive the coordinates of R. We have

$$\begin{aligned}(r_1, r_2) &= \mu(q_1, q_2) + (1-\mu)(0, p) \\ &= (\mu q_1, \mu q_2 + (1-\mu)p)\end{aligned}$$

We can now use our rearrangement of the equation of the side that R rests on, and substitute in the coordinates (r_1, r_2) for x and y. We have

$$a = a(\mu q_2 + (1-\mu)p) + (1-b)\mu q_1$$

and after some more rearranging, we have

$$\begin{aligned}a(1-p) &= \mu(aq_2 - ap + (1-b)q_1)) \\ &= \mu(ab\lambda - ap + (1-b)(a\lambda + 1 - \lambda)) \\ &= \mu(ab\lambda - ap + a\lambda + 1 - \lambda - ab\lambda - b(1-\lambda)) \\ &= \mu(\lambda(a+b-1) + 1 - ap - b)\end{aligned}$$

We can then express μ in terms of λ and p. We get

$$\mu = \frac{a(1-p)}{\lambda(a+b-1) + 1 - ap - b}$$

Let us now discuss the area conditions. Obviously, for the zero area parallel, we require that $A_1 = A_2$. The area A_1 is quite simple to work out. It has base p and

71
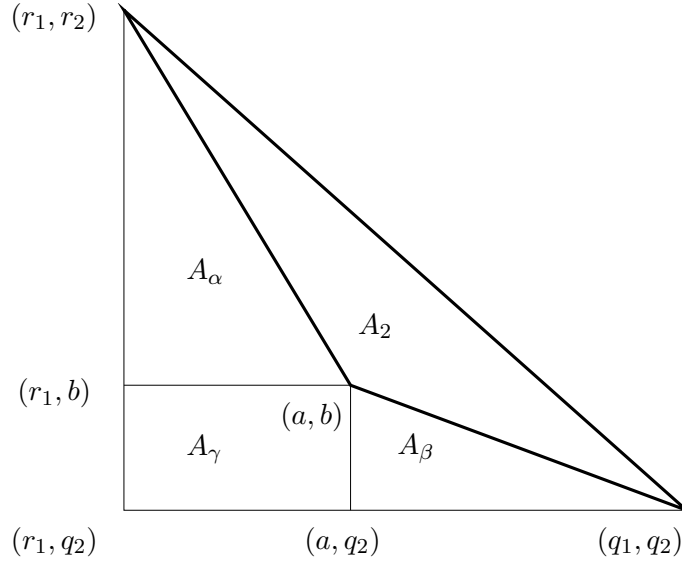
Figure 46: Method of calculating area A_2.

height r_1. Hence $A_1 = \frac{1}{2}(1 - p)r_1$. The area A_2 is a little more tricky. We have to take a triangle, the hypotenuse of which is the line QR, and delete the relevant areas. This is shown in Figure 46.

Using this diagram we can work out A_2. If we call the area of the whole triangle A_τ, we can see that $A_2 = A_\tau - A_\alpha - A_\beta - A_\gamma$. We will then consider each section separately. We have

$$A_\tau = \frac{1}{2}(q_1 - r_1)(r_2 - q_2)$$

$$A_\alpha = \frac{1}{2}(a - r_1)(r_2 - b)$$

$$A_\beta = \frac{1}{2}(q_1 - a)(b - q_2)$$

$$A_\gamma = (a - r_1)(b - q_2)$$

and since $A_1 = A_2$, we have

$$(1 - p)r_1 = (q_1 - r_1)(r_2 - q_2) - (a - r_1)(r_2 - b) - (q_1 - a)(b - q_2) - 2(a - r_1)(b - q_2)$$

We now use Maple to substitute in the following expressions.

$$q_1 = \lambda a + (1 - \lambda)$$
$$q_2 = \lambda b$$
$$r_1 = \mu q_1$$
$$r_2 = \mu q_2 + (1 - \mu)p$$

where

$$\mu = \frac{a(1-p)}{\lambda(a+b-1)+1-ap-b}$$

We find that the left hand side of this equation simplifies down to give

$$A_1 = \frac{(1-p)^2 a(\lambda a + 1 - \lambda)}{2(\lambda(a+b-1)+1-ap-b)}$$

and the right hand side simplifies to

$$A_2 = \frac{(\lambda-1)^2(a+b-1)(b+ap-p)}{2(\lambda a + \lambda b - \lambda + 1 - ap - b)}$$

The left hand side is then equated to the right hand side, and following some simplification and rearrangement, we have

$$0 = \lambda b + \lambda ap - p\lambda - a - b + p$$

It is then fairly easy to derive an expression for p in terms of λ. We have

$$p(\lambda) = \frac{b(\lambda - 1) - a}{\lambda(1-a) - 1}$$

We can then derive an expression for the general midpoint of the chord, giving us the parametric form of this branch of the area parallel. Following some simplification in the y-coordinate, we have

$$
\begin{aligned}
(m_x, m_y) &= \left(\frac{q_1}{2}, \frac{p+q_2}{2} \right) \\
&= \left(\frac{\lambda a + (1-\lambda)}{2}, \frac{a+b+\lambda^2 b(a-1)}{2(\lambda(a-1)+1)} \right)
\end{aligned}
$$

We must now find out how this branch of the area parallel behaves. We would expect, as has always been the case, that this models a hyperbola. In order to prove this, we will redefine the parametric expression in a more general form, as this will make things a little easier. We can express the midpoint as

$$(m_x, m_y) = \left(\alpha\lambda + \beta, \frac{\gamma\lambda^2 + \delta}{\zeta\lambda + \eta} \right)$$

Then, the x-coordinate has the form

$$x = \alpha\lambda + \beta \qquad \Rightarrow \qquad \lambda = \frac{x - \beta}{\alpha}$$

73

and y can be expressed as

$$y = \frac{\gamma \left(\frac{x-\beta}{\alpha}\right)^2 + \delta}{\zeta \left(\frac{x-\beta}{\alpha}\right) + \eta}$$

which has the general form

$$\frac{Ax^2 + Bx + C}{Dx + E}$$

and so

$$Dxy + Ey = Ax^2 + Bx + C$$

We can see that this does not contain any y^2 terms. Now that we have the general form, we can use Maple to convert our specific example. We find that the Cartesian form of our parametric equation is

$$0 = 4bx^2 + 4(1-a)xy - 4bx + a^2 + ab - a$$
$$\Rightarrow \quad 0 = 4x(bx + (1-a)y) - 4bx + a^2 + ab - a$$

We can see that the quadratic part of this clearly has real factors, and so we have proven that this branch of the area parallel is hyperbolic.

Let us now discuss the endpoints of the branch. In order to do that, we need to know more about the endpoints of the chord, specifically, the point P. We need to know how this point moves along the y-axis, and whether or not there are any changes of direction. To do this, we will take our expression for p and differentiate with respect to λ. Firstly, however, it is important to point out that if $\lambda = 0$ then $p = \frac{-b-a}{-1} = a + b$, and if $\lambda = 1$ then $p = \frac{-a}{-a} = 1$, which means that P moves up from $p = a + b$ to $p = 1$. Now, differentiating we have

$$\begin{aligned}
\frac{\partial p}{\partial \lambda} &= \frac{(\lambda(1-a)-1)b - (b(\lambda-1)-a)(1-a)}{(\lambda(1-a)-1)^2} \\
&= \frac{\lambda b - \lambda ab - b - \lambda b + b + a + \lambda ab - ab - a^2}{(\lambda(1-a)-1)^2} \\
&= \frac{a(1-b-a)}{(\lambda(1-a)-1)^2}
\end{aligned}$$

which contains only constants in the numerator. Furthermore, looking at the numerator, $a < \frac{1}{2}$ and $b < \frac{1}{2}$, which means $1 - b - a > 0$, and so the numerator is always positive. The denominator is also always positive, which makes the derivative positive. This proves there are no changes of direction in the movement of the endpoint P, and therefore it moves steadily up the y-axis from $p = a + b$ up to $p = 1$ as the chord moves around the quadrilateral.

74

We can now discuss the endpoints of the first branch of the area parallel. As we have said, when $\lambda = 0$, $p = a + b$, and when $\lambda = 1$, $p = 1$. Where $\lambda = 0$, we will call the endpoint e_1, and at $\lambda = 1$, we will use e_2 to denote the corresponding endpoint. We have

$$e_1 = (m_x, m_y) = \left(\frac{1}{2}, \frac{a+b}{2} \right)$$

$$e_2 = (m_x, m_y) = \left(\frac{a}{2}, \frac{a+b+b(a-1)}{2a} \right)$$

$$= \left(\frac{a}{2}, \frac{b+1}{2} \right)$$

It is relatively clear that the endpoint e_2 is the midpoint of the line connecting the vertices $(0, 1)$ and (a, b). This is expected, as a trivial zero area will occur when the chord comes into coincidence with this side of the quadrilateral. In fact, once the chord has moved onto this side of the quadrilateral, it can not go any further around the quadrilateral and still cut off zero area, so it goes back down, meaning that the midpoints will trace out the same path. This implies that the endpoint e_2 is a real endpoint, and does not join onto any other branches.

We will see however that the endpoint e_1 *does* join onto another branch. Once the endpoint of the chord Q goes through the vertex $(1, 0)$, the situation then turns into something more remenissant of the centre diagram in Figure 44. So let us now take a look at this diagram in detail.
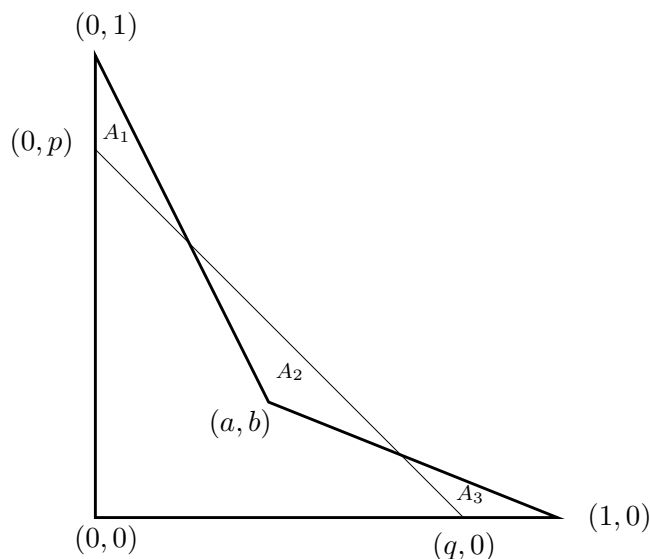

Figure 47: Case two: The chord now cuts two sides of the quadrilateral.

As can be seen from Figure 47, the point P still has coordinates $(0, p)$, but since Q is now on the x-axis, it makes sense to assign coordinates $(q, 0)$ to the point.

For zero area we now require $A_1 + A_3 = A_2$. For this case, however, we shall take a slightly different route to find our conditions required for equal area. Whilst it is true that the sum of the three (algebraic) areas has to equal zero, it is also true that the area of the quadrilateral has to equal the area of the triangle enclosed by vertices $(0, 0)$, $(q, 0)$ and $(0, p)$. It should be clear from the diagram that this condition implies that the sum of the areas cut off by the chord is zero.

Now that we know this, the algebra is very simple. The triangle has base q and height p and so the area is given by $\frac{1}{2}pq$. The quadrilateral is a trapezium plus a triangle. The trapezium has parallel sides length 1 and b, and height a, and so the area is $\frac{a(b+1)}{2}$. The triangle has base $1 - a$ and height b, giving area $\frac{b(1-a)}{2}$. Hence the quadrilateral has area

$$\frac{a(b+1)}{2} + \frac{b(1-a)}{2} = \frac{a+b}{2}$$

So our area condition is simply
$$a + b = pq$$

It is also clear from the diagram that the midpoint of the chord is given by

$$
\begin{aligned}
(m_x, m_y) &= \left(\frac{q}{2}, \frac{p}{2}\right) \\
&= \left(\frac{a+b}{2p}, \frac{p}{2}\right)
\end{aligned}
$$

If we then say $y = \frac{p}{2}$ and $x = \frac{a+b}{2p}$, then quite clearly $x = \frac{a+b}{4y}$ and so $4xy = a + b$, which defines a hyperbola since the quadratic part once again has real factors. In this expression, there is no x^2 term or y^2 term. So our second branch also models a hyperbola.

Let us now look at the endpoints of our second branch. Obviously, when Q is at the vertex $(1, 0)$, then $q = 1$, and therefore we have $p = a + b$ at the endpoint of the branch, which we will call e_3. Hence the midpoint at e_3 is given by

$$(m_x, m_y) = \left(\frac{1}{2}, \frac{a+b}{2}\right)$$

This is the same as e_1, proving that the two branches of hyperbolae are connected at this point. This is of no surprise. After all, the endpoint of the chord Q moves around continuously, so we expect that the locus of midpoints will also move from one branch to the next without any discontinuity.

With regards the other endpoint of this second branch, e_4, the chord intersects two of the sides of the quadrilateral, until the point P reaches the vertex $(0,1)$. At this point, $p = 1$ and so $q = a + b$. Then the midpoint at e_4 takes the form

$$(m_x, m_y) = \left(\frac{a+b}{2}, \frac{1}{2} \right)$$

which is a reflection in the line $y = x$ of the endpoint e_3.

Having found that the endpoints of the two branches coincide, that is $e_1 = e_3$, we would now like to find out whether the connection is smooth or is by means of a cusp. The way we will do this is by finding out whether the tangent vectors of the two branches are parallel or not. To do this we must differentiate, and then find the value of the first derivative at the endpoint.

It is clear from the diagram and the theory that precedes that the endpoint e_1 is where the first branch starts, and so $\lambda = 0$. Now, differentiating the parametric form of the hyperbola with respect to λ gives

$$(m'_x, m'_y) = \left(\frac{a-1}{2}, \frac{(a-1)(\lambda^2 ab - \lambda^2 b + 2\lambda b - a - b)}{2(\lambda a - \lambda + 1)} \right)$$

and then substituting in the value $\lambda = 0$ gives

$$(m'_x, m'_y) = \left(\frac{a-1}{2}, -\frac{(a+b)(a-1)}{2} \right)$$

We can now divide the x and y coordinates by $\frac{a-1}{2}$ to find a parallel tangent vector, but we must remember that since $a < \frac{1}{2}$, then $\frac{a-1}{2} < 0$, and so the direction of the tangent vector reverses. We have

$$(m'_x, m'_y) \parallel (1, -(a+b))$$

Meanwhile, let us consider the second branch. Here the parameter is p, and we can check that the second branch begins with $p = a + b$. If we differentiate the parametric form of the second branch with respect to p, we have

$$(m'_x, m'_y) = \left(-\frac{a+b}{2p^2}, \frac{1}{2} \right)$$

and substituting in our value of p, we have

$$
\begin{aligned}
(m'_x, m'_y) &= \left(-\frac{a+b}{2(a+b)^2}, \frac{1}{2} \right) \\
&= \left(-\frac{1}{2(a+b)}, \frac{1}{2} \right)
\end{aligned}
$$

77

Hence $(m'_x, m'_y) \,||\, (-1, a + b)$.

We can see that the tangent vector for the first branch is the negative of the tangent vector for the second branch. This means they meet at the common endpoint with a cusp, but the two branches of hyperbolae are curving in opposite directions. The third branch, as we said acts like the first due to the symmetry of the figure. The area parallel will therefore look something like the diagram in Figure 48.
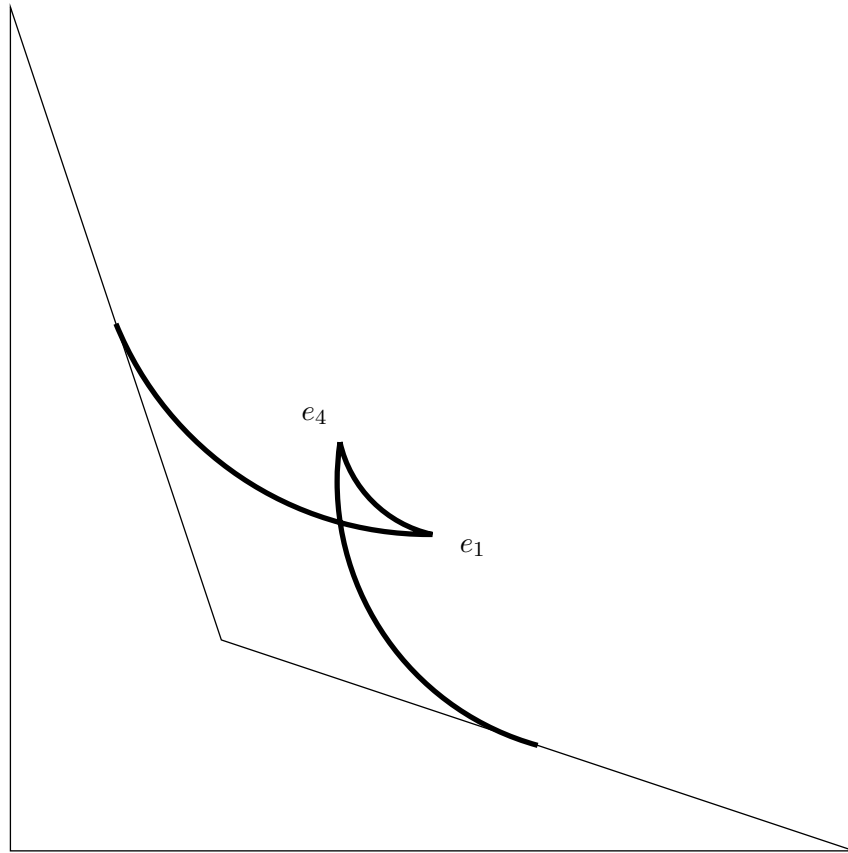

Figure 48: The bold lines trace the three branches of the area parallel.

So, we now have the area parallel, which, for the first time in this study, has proper endpoints. These are at the midpoints of two of the sides of the quadrilateral.

5 Concluding Remarks

In this final section we shall summarise the results that we have found through the course of this study. We began with a continuation from the mini-dissertation on the subject of closed convex plane curves, where an area function was defined. This area function found the area cut off by a chord, and depended on the angle that the chord made with the horizontal pole. We found that this pole is the midpoint of the chord, if and only if the first derivative of the area function is zero.

When classifying this area function, we considered the tangent lines at the two endpoints of the chord. We proved that there is a minimum of area if and only if these tangent lines meet above the horizontal, by considering the various cases separately. We also found that these two tangent lines are parallel if and only if the first *and* second derivatives of the area function are simultaneously zero, and that this is analogous to an A_2 singularity.

Having finished the section of convex curves, we moved onto non-convex curves, which encompassed extra complications connected with negative algebraic areas. A curve, which contained an inflection was defined, and local coordinates were defined near the zeros. As with the convex curves, a chord was also defined, and moved around the curve cutting of a fixed area. Using power series approximations, we found the relationship between the two endpoints of the chord, and also the locus of midpoints, which defines, as we know from the mini-dissertation, the area parallel.

We found that, for a cubic curve with a double zero at the origin and another zero elsewhere, the area parallel locally follows the path of a cubic curve with its inflection point half way between the zero points. In terms of power series, this curve begins with a cubic term, and may also involve higher powers, which will have lesser an effect the higher these powers get. Meanwhile, for a curve with a triple zero at the origin and another zero elsewhere, we found that the power series of the area parallel starts with a quartic term, and so, locally, modelled a quartic curve, with its maximum point half way between the zero points.

The non-convex case includes a rather important area of study that does not exist with convex curves. That is the zero area parallel, where the sum of the areas above a chord is equal to the sum of those below. That is, the algebraic area is zero. When discussing this possibility, we redefined our curve so that it only had a zero at the origin. It therefore began with a cubic term and included higher order powers. Similar methods were used to find the relationship between the two endpoints of a chord cutting off zero area, and an expression for the area parallel was found. This took the form of a rhamphoid cusp that coincided with the inflection point of the curve.

We then moved onto a discussion of quadrilaterals both convex and non-convex. With regards affine transformations, in both cases we found that we had fewer levels of freedom than we had with the case of a triangle. This is due to the fact that any three points in the plane can be affinely transformed onto any other three points. We therefore noted in the mini-dissertation that any triangle can be affinely transformed onto an equilateral triangle, but this is obviously not the case with quadrilaterals.

Nevertheless, with the convex case we managed to find some general results concerning the area parallel and symmetry set. We discovered that the area parallel can have four cusps or six cusps, depending on size of the area cut of by the chord. In fact this cusp appeared when the chord passed through two opposite vertices of the quadrilateral simultaneously. We also found that, as with the convex curves, parallel tangent lines imply the existence of cusps. We then proved that no quadrilateral can have any area parallel containing eight cusps, although the square comes the closest (and yet the furthest!).

For the symmetry set, we managed to prove that there are two sections to one of the branches, the first following a straight line, and then continuing onto the second which takes the form of a hyperbola. We did this by noting that the diagonals of a parallelogram incorporated inside the quadrilateral cut off the same area but with respect to different vertices of the quadrilateral. This is analogous to the case of the equilateral triangle in the mini-dissertation, which incorporated a rectangle. In both cases, the meeting point of the diagonals trace the self intersection points of the area parallel, and so define the symmetry set.

The final section dealt with the case of non-convex quadrilaterals. In this section we looked at the zero area parallel of the quadrilateral, and found that the area parallel had three branches that all took a hyperbolic form. We calculated the endpoints and proved that the branches were connected to one another. We also found that the zero area parallel had proper endpoints, which is the only time in this study that we encounter such a situation.

A lot more cases can be discussed with regards quadrilaterals, particularly the non-convex case. Further study may include a report into the fixed area parallel, both for a positive fixed area and a negative fixed algebraic area, as this would contain some fairly substantial material. An investigation concerning the symmetry set would also be of interest for future work, not only with the quadrilateral case, but with non-convex curves. Finally, it would be interesting to see how this kind of geometry works in three dimensions, and I'm sure many detailed reports on that particular subject will be composed in the future.

References

[1] Bruce and Giblin. "Curves and Singularities" (1992), various chapters.

[2] Niethammer, Betelu, Sapiro, Tannenbaum, Giblin. "Area Based Medial Axis of Planar Curves" (2004).

[3] Estrozi, Giblin. "Affine Area Parallels and Symmetry Sets" (2003).

[4] Betelu, Sapiro, Tannenbaum, Giblin. "Noise-Resistant Affine Skeletons of Planar Curves" (1999).

[5] Estrozi. "Affine Area Symmetry Sets 9" Computer Program (2003).