Conic Sections Beyond $\mathbb{R}^2$

Mzuri S. Handlin

May 14, 2013

Contents

1 Introduction

As with many powerful concepts, the basic idea of a conic section is simple. Slice a cone with a plane in any direction and what you have is a conic section, or conic; it is straightforward enough that the concept is discussed in many high school geometry classes. Their study goes back at least to 200 BC, when Apollonius of Perga studied them extensively [1]. We can begin to see the power of these simple curves by noticing the diverse range of fields in which they appear. Kepler noted that the planets move in elliptical orbits. Parabolic reflectors focus incoming light to a single point, making them useful both as components of powerful telescopes and as tools for collecting solar energy. Hyperbolas are used in the process of trilateration, the determination of a location based on the difference in arrival time of synchronized signals; this is what makes GPS possible. All of these are conic sections. Even more intriguing than the different places in which conics show up are the connections between these occurrences. It is no coincidence that ellipses, parabolas, and hyperbolas all show up when discussing the orbital motion of two bodies. An elliptic orbit occurs when neither object has reached escape velocity. A parabolic orbit occurs at exactly escape velocity. A hyperbolic orbit occurs when one object has exceeded escape velocity. Fundamentally, this is because a parabola is the limit between an increasingly "eccentric" ellipse and a decreasingly "eccentric" hyperbola.

By studying connections such as these, mathematicians made interesting new observations beyond the work of the ancient Greeks. Dandelin spheres were invented to facilitate the proof of important geometric properties of conic sections. When constructed in perspective geometry, the three distinct conics were found to collapse into a single type of object. It is in this vein of considering old ideas in a new light that we wish to explore conic sections in this paper. In particular, we will examine how the ideas of conic sections generalize to higher dimensions. Curves become surfaces and hypersurfaces, some ideas break down, and some new connections arise. In order to make these connections, we will utilize some of the tools of differential geometry. Ultimately, we hope to convince the reader that deep results can be obtained by studying the "simple" conic sections.

2 One Dimensional Conic Sections

In order to begin generalizing conic sections, we must first have a solid foundation in their basics. We will discuss multiple definitions of conics, both algebraic and geometric. We will also explore some important properties of conics which will prove to be useful later on. In particular, we will see that it is possible to classify a conic using only the coefficients of its implicit equation.

2.1 Geometric Definitions

We begin with the fundamental definition:

Definition 1. *A conic section is the curve resulting from the intersection of a plane and a cone.*

Some terminology is required to flush out this definition. The *principal axis* of a cone is the line passing through the vertex and perpendicular to what we would call the base if we cut off the cone instead of allowing it to extend to infinity. For simplicity we will assume that the vertex of the cone lies at the origin and the principal axis is the z-axis. Unless specified otherwise, from now on when we refer to a cone we will mean a *circular* cone, which means that cross sections perpendicular to the principal axis are circles. An example of a non-circular cone is an elliptic cone, which has elliptical cross sections perpendicular to the principal axis. We will see these surfaces again when we begin discussing generalizing conic sections to higher dimensions. Finally, a *generating line* for a cone is a line in $\mathbb{R}^3$ other than the x, y, or z axes passing through the origin which, when revolved about the z-axis, sweeps out the surface of the cone. The (circular) cone is thus a surface of revolution. This, too, we will discuss in more detail later. An example of a cone with the various properties just described is shown in Figure 1. We can now make our geometric definition more

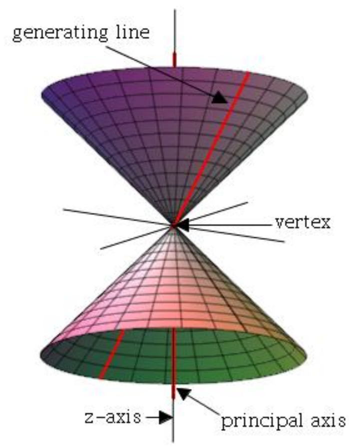


Figure 1: A cone with circular cross sections, principal axis aligned with the z-axis, and vertex at the origin

explicit by identifying different ways that a plane and cone can intersect and the different classes of curves that these intersections result in:

Ellipses The plane intersects the cone in a closed curve.

Hyperbolas The plane intersects both halves of the cone.

Parabolas The plane is parallel to (but does not contain) a generating line of the cone (notice that it passes through only one half of the cone).

Degenerate curves The plane passes through the vertex of the cone in one of the following ways:

- **Intersecting lines** The plane passes through the vertex and two distinct generating lines of the cone.
- **Single line** The plane contains exactly one generating line of the cone.
- **Point** The plane intersects the cone only at its vertex.

Examples of non-degenerate conics generated by the intersection of a plane and cone are shown in Figure 2.1. The degenerate curves are somewhat unusual in that we don't normally see them referred to as conic sections. They are, however, most certainly sections of a cone. They are each limiting cases of one of the more familiar conics. The intersecting lines are the limit of a hyperbola which approaches its asymptotes, the point is the limit of a circle with zero radius, and the single line is the limit of an increasingly flat parabola. We call them degenerate because some of their features (curvature, dimension) have collapsed. We will see that degenerate conics also consist only of degree one and lower terms, whereas the non-degenerate conics have degree two terms in their

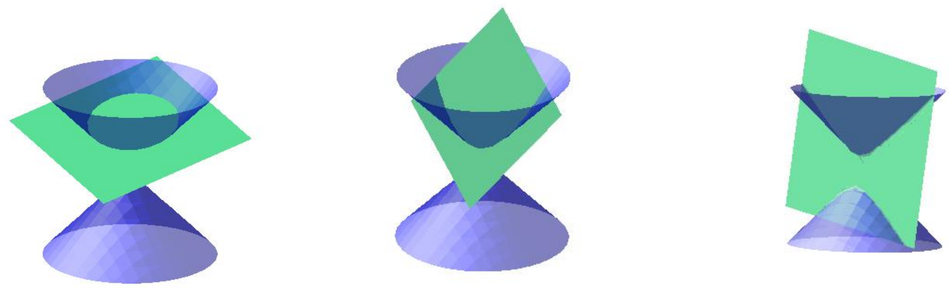


Figure 2: Generating conic sections (an ellipse, parabola, and hyperbola respectively)

equations, which gives us a more concrete definition of what degenerate means: *a degenerate conic section is one whose equation does not have the highest possible degree.* What we mean by a conic section's equation will be explained shortly (Section 2.2).

We might also include two further classes of degenerate conic sections: two parallel lines and the null set. These occur when we allow the cone to degenerate into a cylinder by pulling its vertex out to infinity. The parallel lines thus occur when the plane lies parallel to the principal axis and between two generating lines. The null set occurs when the plane never intersects the cone: it is parallel to the principal axis but lies outside of the cylinder. However, by assuming that our vertex lies at the origin we are precluding these limiting cases. In particular, this means that the null set is not an option: there must be an intersection between a plane and a (non-degenerate) cone.

Notice that we can generate all three classes of conics by starting with a plane whose intersection produces a circle, then rotating the plane about a line contained in the plane which does not pass through the z-axis and in the direction which does not cause the plane to pass through the origin before hitting the other half of the cone. Our circle becomes an increasingly "squished" ellipse (we will formalize this with the idea of eccentricity shortly) until the plane is parallel to a generating line, producing a parabola. Continuing past this orientation causes the plane to intersect the other half of the cone, producing a hyperbola. Notice that if we allow the line which the plane rotates around to intersect with the z-axis, then we get the degenerate cases. These curves are thus in some sense determined by how "steep" the plane is with respect to the cone.

In order to consider this statement more formally, we introduce three new ideas. A *focus* and *directrix* are a point and a line in a plane, respectively. The *eccentricity* of a particular point with respect to a particular focus and directrix (where all three objects are contained in the same plane) is the ratio of the distance from the point to the focus to the perpendicular distance from the point to the directrix. If P is a point, F is a focus and l is a directrix, we define $\overline{PF}$ to be the distance from P to F and $d_{min}(Pl)$ to be the minimum distance from P to l. Then the eccentricity e is given by

$$e = \frac{\overline{PF}}{d_{min}(Pl)}$$

We will begin to see how the different conics are related to each other beyond their common definition of intersection by exploring this parameter of eccentricity. We begin with the following new definition for the conics:

Definition 2. *A conic section is the set of all points in a plane with the same eccentricity with respect to a particular focus and directrix.*

This leads to the following classifications:

Ellipses Conic sections with $0 \leq e < 1$. Circles are the special case of $e = 0$.

Parabolas Conic sections with $e = 1$.

Hyperbolas Conic sections with $e > 1$.

In Section 2.2.1, we will show that our two geometric definitions are equivalent.

We may be used to thinking of eccentricity as a measure of how far an ellipse has deformed from a perfect circle, but with this definition in mind we can also think of it as the slope of the plane with respect to the $xy-$plane and a particular generating line of the cone. Parallel to the $xy-$plane (no tilt) or $e = 0$ produces a circle. Slope between zero and that of the generating line (which we will call 1) gives an ellipse. Slope equal to that of the generating line ($e = 1$) gives a parabola. Slope greater than the generating line, or $e > 1$, produces a hyperbola. Notice that we have not offered any proof that our two competing definitions give us the same curves. We will return to this problem after we have developed a necessary algebraic picture of the conics.

2.2 Algebraic Definitions

We will now examine ways of representing conics algebraically. To start, consider the following well known formulas.

- Ellipse:
$$\frac{x^2}{a} + \frac{y^2}{b} = c^2 \tag{1}$$
- Hyperbola:
$$\frac{x^2}{a} - \frac{y^2}{b} = c^2 \tag{2}$$
- Parabola:
$$y = ax^2 \tag{3}$$

Most introductory textbooks (e.g. [4]) find these by using the distance formula and the definition of a conic section as the set of points with a constant ratio between the distance from a focus and a directrix. We will not give the derivation of these formulas, but use them as motivation to establish a general algebraic definition.

It is important to note that equations (1) through (3) do not give us all possible ellipses, parabolas, and hyperbolas. Instead, they are the conic equations in *standard form.* This means that the principal axes are aligned with the x and y axes and the center of the conic is at the origin. If we want to consider any conic section, we must allow the center to be at any point and the principal axes to be arbitrarily rotated. To move the center to (x_0, y_0), we can simply substitute $x - x_0$ and $y - y_0$ for x and y respectively in equations (1) through (3). This gives us the following:

- Ellipse:

$$\begin{aligned}
\frac{(x-x_0)^2}{a} + \frac{(y-y_0)^2}{b} &= c^2 \\
\frac{1}{a}(x^2 - 2xx_0 + x_0^2) + \frac{1}{b}(y^2 - 2yy_0 + y_0^2) &= c^2 \\
\frac{1}{a}x^2 + \frac{1}{b}y^2 - \frac{2x_0}{a}x - \frac{2y_0}{b}y + (\frac{x_0^2}{a} + \frac{y_0^2}{b} + c^2) &= 0
\end{aligned} \tag{4}$$

- Hyperbola:

$$\begin{aligned}
\frac{(x-x_0)^2}{a} - \frac{(y-y_0)^2}{b} &= c^2 \\
\frac{1}{a}(x^2 - 2xx_0 + x_0^2) - \frac{1}{b}(y^2 - 2yy_0 + y_0^2) &= c^2 \\
\frac{1}{a}x^2 - \frac{1}{b}y^2 - \frac{2x_0}{a}x + \frac{2y_0}{a}y + (\frac{x_0^2}{a} - \frac{y_0^2}{b} + c^2) &= 0
\end{aligned} \tag{5}$$

- Parabola:

$$\begin{aligned}
(y - y_0) &= a(x - x_0)^2 \\
(y^2 - 2yy_0 + y_0^2) &= a(x^2 - 2xx_0 + x_0^2) \\
(-a)x^2 + y^2 + (2ax_0)x + (-2y_0)y + (-x_0^2 + y_0^2) &= 0
\end{aligned} \tag{6}$$

Next we want to consider a conic whose principle axes have been arbitrarily rotated by an angle θ. Recall (perhaps from linear algebra) that if a point (x, y) is rotated by an angle θ about the origin, then the coordinates of the rotated point (x_r, y_r) are given by

$$\begin{aligned}
x_r &= x\cos\theta - y\sin\theta \\
y_r &= x\sin\theta + y\cos\theta
\end{aligned}$$

Thus we can simply substitute x_r and y_r for x and y in equations (4) through (6). This gives us

- Ellipse:

$$\begin{aligned}
0 = &\frac{1}{a}x_r^2 + \frac{1}{b}y_r^2 - \frac{2x_0}{a}x_r - \frac{2y_0}{b}y_r + (\frac{x_0^2}{a} + \frac{y_0^2}{b} + c^2) \\
&\frac{1}{a}(x\cos\theta - y\sin\theta)^2 + \frac{1}{b}(x\sin\theta + y\cos\theta)^2 - \frac{2x_0}{a}(x\cos\theta - y\sin\theta) \\
&- \frac{2y_0}{b}(x\sin\theta + y\cos\theta) + (\frac{x_0^2}{a} + \frac{y_0^2}{b} + c^2) \\
= &(\frac{\cos^2\theta}{a} + \frac{\sin^2\theta}{b})x^2 + (\frac{\sin^2\theta}{a} + \frac{\cos^2\theta}{b})y^2 + (\frac{-2\cos\theta\sin\theta}{a} + \frac{2\cos\theta\sin\theta}{b})xy \\
&+ (\frac{-2x_0\cos\theta}{a} - \frac{-2y_0\sin\theta}{b})x + (\frac{2x_0\sin\theta}{a} + \frac{-2y_0\cos\theta}{b})y + (\frac{x_0^2}{a} + \frac{y_0^2}{b} + c^2)
\end{aligned}$$

- Hyperbola:

$$\begin{aligned}
0 &= \frac{1}{a}x_r^2 - \frac{1}{b}y_r^2 - \frac{2x_0}{a}x_r + \frac{2y_0}{a}y_r + (\frac{x_0^2}{a} - \frac{y_0^2}{b} + c^2) \\
&\quad \frac{1}{a}(x\cos\theta - y\sin\theta)^2 - \frac{1}{b}(x\sin\theta + y\cos\theta)^2 - \frac{2x_0}{a}(x\cos\theta - y\sin\theta) \\
&\quad + \frac{2y_0}{b}(x\sin\theta + y\cos\theta) + (\frac{x_0^2}{a} + \frac{y_0^2}{b} + c^2) \\
&= (\frac{\cos^2\theta}{a} - \frac{\sin^2\theta}{b})x^2 + (\frac{\sin^2\theta}{a} - \frac{\cos^2\theta}{b})y^2 + (\frac{-2\cos\theta\sin\theta}{a} - \frac{2\cos\theta\sin\theta}{b})xy \\
&\quad + (\frac{-2x_0\cos\theta}{a} + \frac{-2y_0\sin\theta}{b})x + (\frac{2x_0\sin\theta}{a} + \frac{2y_0\cos\theta}{b})y + (\frac{x_0^2}{a} + \frac{y_0^2}{b} + c^2)
\end{aligned}$$

- Parabola:

$$\begin{aligned}
0 &= (-a)x_r^2 + y_r^2 + (2ax_0)x_r + (-2y_0)y_r + (-x_0^2 + y_0^2) \\
&\quad - a(x\cos\theta - y\sin\theta)^2 + (x\sin\theta + y\cos\theta)^2 + 2ax_0(x\cos\theta - y\sin\theta) \\
&\quad + (-2y_0)(x\sin\theta + y\cos\theta) + (-x_0^2 + y_0^2) \\
&= (-a\cos^2\theta + \sin^2\theta)x^2 + (-a\sin^2\theta + \cos^2\theta)y^2 + (2a\cos\theta\sin\theta)xy \\
&\quad + (2ax_0\cos\theta - 2y_0\sin\theta)x + (-2ax_0\sin\theta - 2y_0\cos\theta)y + (-x_0^2 + y_0^2)
\end{aligned}$$

Although they all appear fairly messy, notice that all three of these equations have the same general form:

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0$$

They differ only in the choice of coefficients. We therefore make the following algebraic definition of a conic section:

Definition 3. *A* conic section *is the set of points* (x, y) *satisfying the implicit formula*

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0 \tag{7}$$

Because this formula gives any second degree polynomial in two variables, we may also refer to these curves as *quadric curves* or *one-dimensional quadrics*. The significance of this terminology will become apparent when we begin discussing generalizations of conic sections.

2.2.1 Polar Coordinates

Now we are equipped to develop an algebraic picture of the important concept of eccentricity introduced in Section 2.1. We will accomplish this by finding an equation for the conic sections in polar coordinates.

Consider a conic section whose focus lies at the origin and whose directrix is the line $x = k$ for some constant k. This situation is shown in Figure 3. Notice that although the specific conic section shown in this figure is an ellipse, we are not using any of the particular properties of the ellipse, only the common properties defining a generic conic. Thus, the following theorem, whose proof we are adopting from [2], holds for all one-dimensional conic sections.

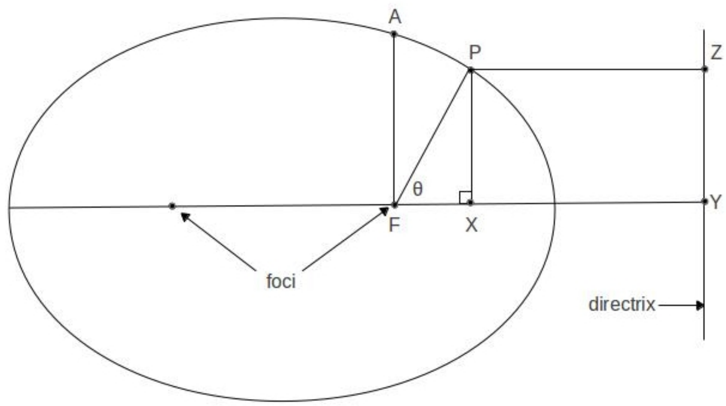


Figure 3: Important points in a conic section

Theorem 1. *Consider a one-dimensional conic section with at least one focal point lying at the origin. Then the equation for this conic in polar coordinates is*

$$r = \frac{l}{1 + e\cos\theta} \tag{8}$$

where r is the distance from the origin, θ is the counterclockwise angle from the positive x axis, and l is a constant.

Proof. Notice that triangle FXP in Figure 3 is a right triangle. Therefore we have the relationship

$$FX = r\cos\theta$$

Furthermore, we know that

$$PZ = FY - FX$$

Because F lies at the origin and the point Y lies at $(k, 0)$ (since it lies at the intersection of the directrix and the x-axis), the distance $FY = k$. Thus we have

$$PZ = k - r\cos\theta \tag{9}$$

Recalling the definition for the eccentricity e, we notice that

$$\begin{aligned} PZ &= FP/e \\ &= r/e \end{aligned} \tag{10}$$

Thus, by combining equations (9) and (10) and solving for r, we get

$$r = \frac{ek}{1 + e\cos\theta}$$

□

This would be a fine place to stop, but in this form we have some difficulties. Namely, for the circle we know that $e = 0$, which seems to give us the polar equation for a circle of $r = 0$, which is not true. Fortunately, it is known that $ek = l$ where l is a parameter known as the semi-latus rectum; it is the distance parallel to the directrix from a focus to a point on the conic. It is indicated in Figure 3 by the line segment FA. Thus, by making this substitution, we remove the ambiguity of $e = 0$ and arrive at equation (8) as desired.

This theorem has the useful property that it makes possible an immediate classification of a conic section from its equation in polar form. Recall that we have the following three cases for the eccentricity:

1. $0 \leq e < 1$ indicates an ellipse
2. $1 = e$ indicates a parabola
3. $1 < e$ indicates a hyperbola

Thus we only need to look to e in the denominator of equation (8). Moreover, we are now in a position to verify this property of conics being classified by eccentricity. We can rearrange equation (8) to get

$$r = l - er\cos\theta$$

and convert from polar to rectangular coordinates to get

$$\sqrt{x^2 + y^2} = l - ex$$

and square both sides to get

$$x^2 + y^2 = l^2 - 2ex + e^2x^2$$

which we can rearrange to get

$$(1 - e^2)x^2 + 2ex + y^2 = l^2 \tag{11}$$

Notice that this is a conic section, although we are not yet certain which type. There are three cases:

1. $0 \leq e < 1$

   We can complete the square in x in equation (11) to get

$$\begin{aligned}(1 - e^2)\left(x^2 + \frac{2e}{1 - e^2}x\right) + y^2 &= l^2 \\ (1 - e^2)\left[\left(x + \frac{e}{1 - e^2}\right)^2 - \frac{e^2}{(1 - e^2)^2}\right] + y^2 &= \\ (1 - e^2)\left(x + \frac{e}{1 - e^2}\right)^2 + y^2 &= l^2 + \frac{e^2}{1 - e^2} \qquad (12)\end{aligned}$$

   Since all of the coefficients are positive, we recognize this as a ellipse. Moreover, if $e = 0$, we have

$$x^2 + y^2 = l^2$$

   which is the equation of a circle of radius l.

2. $1 = e$

   The x^2 term in equation (11) disappears and we get

$$y^2 = l^2 + 2ex \tag{13}$$

   which we recognize as a parabola.

3. $1 < e$

   As in case 1, we can complete the square in x in equation (11) to obtain equation (12) again, but this time the coefficients on the x and y terms have opposite signs. We recognize this as a hyperbola.

Thus our polar and rectangular equations and their properties are consistent. More importantly, our intersection definition and eccentricity definition are consistent.

2.3 Classifying Conic Sections

We will continue this process of comparing our definitions of conics by developing an algebraic means of classifying a conic from its implicit equation. In other words, we want to check that an arbitrary conic section can be given by an appropriate choice of constants in

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

and moreover that we can identify which type of conic we have from these constants. A similar classification from the geometric picture of intersecting cones and planes was already given above. Before giving a general theorem for quickly identifying which class of conic is represented by a particular equation, we will give some examples to show that this identification is not always immediately obvious.

Consider the equation

$$5x^2 + y^2 + y - 8 = 0 \tag{14}$$

which is in the form of equation (7). Using Maple, it is relatively simple to plot the set of solution points to this equation. We see in Figure 4 that it appears to be an ellipse, but this is not immediately obvious from the form of the equation. We can rearrange to arrive at the standard

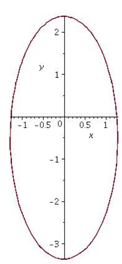


Figure 4: Plot of equation (14)

form for the equation of an ellipse. First, move the constant term to the right side

$$5x^2 + y^2 + y = 8$$

Next, complete the squares for the x and y terms. In this example, x is already done.

$$5x^2 + y^2 + y + \frac{1}{4} = 8 + \frac{1}{4}$$
$$\frac{x^2}{5} + (y + \frac{1}{2})^2 = \frac{33}{4}$$

Which is the desired form. It is more difficult to perform such rearrangements in equations such as

$$5x^2 - \frac{3}{2}xy + y^2 + y - 8 = 0 \tag{15}$$

due to the xy cross term. However, we can still get a plot, and from Figure 5 it is clear that this is once again an ellipse. In fact, it appears to be very similar to the ellipse from Figure 4, simply rotated.

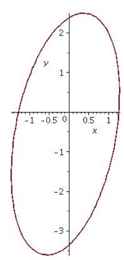


Figure 5: Plot of equation (15)

Recall the equation

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \tag{16}$$

The quantity $B^2 - 4AC$ is the *discriminant* of equation (7). It can help classify conics using the following well known theorem, for which we are primarily following [3].

Theorem 2. *A conic section can be classified by its discriminant as follows:*

1. *If $B^2 - 4AC < 0$, the equation represents an ellipse, a point, or else has no graph.*

2. *If $B^2 - 4AC = 0$, the equation represents a parabola, a line, or else has no graph.*

3. *If $B^2 - 4AC > 0$, the equation represents a hyperbola, a pair of intersecting lines, or else has no graph.*

Proof. We begin by noting that the discriminant is invariant under rotation. In other words, if we take our conic section and rotate without deforming it, the discriminant is unaffected. Recall that if a point (x, y) is subject to a rotation of an angle θ about the origin, then the coordinates of the rotated point (x_r, y_r) are given by

$$x_r = x\cos\theta - y\sin\theta$$
$$y_r = x\sin\theta + y\cos\theta$$

If we substitute x_r for x and y_r for y in equation (7) and rearrange our terms, we get:

$$\begin{aligned}
0 &= Ax_r^2 + Bx_ry_r + Cy_r^2 + Dx_r + Ey_r + F \\
&= A(x\cos\theta - y\sin\theta)^2 + B(x\cos\theta - y\sin\theta)(x\sin\theta + y\cos\theta) + C(x\sin\theta + y\cos\theta)^2 \\
&\quad + D(x\cos\theta - y\sin\theta) + E(x\sin\theta + y\cos\theta) + F \\
&= x^2(A\cos^2\theta + B\cos\theta\sin\theta + C\sin^2\theta) + xy(-2A\cos\theta\sin\theta + B\cos^2\theta - B\sin^2\theta + 2C\cos\theta\sin\theta) \\
&\quad + y^2(A\sin^2\theta - B\sin\theta\cos\theta + C\cos^2\theta) + x(D\cos\theta + E\sin\theta) + y(-D\sin\theta + E\cos\theta) + F \\
&= A'x^2 + B'xy + C'y^2 + D'x + E'y + F
\end{aligned}$$

where the new coefficients are constants in terms of the old coefficients and θ. Notice that although it is slightly messier, this is the same general form that we started with, a fact which is highlighted by the last line. The discriminant of this rotated equation is thus given by

$$\begin{aligned}
B'^2 - 4A'C' &= (-2A\cos\theta\sin\theta + B\cos^2\theta - B\sin^2\theta + 2C\cos\theta\sin\theta)^2 \\
&\quad - 4(A\cos^2\theta + B\cos\theta\sin\theta + C\sin^2\theta)(A\sin^2\theta - B\sin\theta\cos\theta + C\cos^2\theta) \\
&= -8AC\cos^2\theta\sin^2\theta + B^2\cos^4\theta + 2B^2\cos^2\theta\sin^2\theta + B^2\sin^4\theta - 4AC\cos^4\theta - 4AC\sin^4\theta \\
&= \cos^2\theta\sin^2\theta(2B^2 - 8AC) + B^2(\cos^4\theta\sin^4\theta) - 4AC(\sin^4\theta + cos^4\theta) \\
&= (B^2 - 4AC)(2\cos^2\theta\sin^2\theta + \cos^4\theta + \sin^4\theta) \\
&= (B^2 - 4AC)(\cos^2\theta(2\sin^2\theta + \cos^2\theta) + \sin^4\theta) \\
&= (B^2 - 4AC)(\cos^2\theta(1 + \sin^2\theta) + \sin^4\theta) \\
&= (B^2 - 4AC)(\cos^2\theta + \sin^2\theta\cos^2\theta + \sin^4\theta) \\
&= (B^2 - 4AC)(\cos^2\theta + \sin^2\theta(\cos^2\theta + \sin^2\theta)) \\
&= (B^2 - 4AC)(\cos^2\theta + \sin^2\theta) \\
&= (B^2 - 4AC) \qquad (17)
\end{aligned}$$

Therefore we see that the discriminant is invariant under rotation, as we sought to show. Before we move on, it is important to realize that this property of invariance under rotation has significance outside of the context of this proof, and we will see it again. Continuing with the proof, note that after rotation, the coefficient on the xy term becomes

$$B' = -2A\cos\theta\sin\theta + B\cos^2\theta - B\sin^2\theta + 2C\sin\theta\cos\theta \qquad (18)$$

Our goal is to eliminate the xy cross-term which makes classifying different conics difficult. Thus, we wish to find a θ such that $B' = 0$. With this in mind, we set this B' equal to zero and solve for θ to get

$$\theta = -\arctan\left(\frac{(A-C) \pm \sqrt{(A-C)^2 + B^2)}}{B}\right)$$

Recall that the domain of the arctan function is all of $\mathbb{R}$ and note that the term in the square root is positive. Thus the only place we might have difficulty is if $B = 0$. However, if this is the case, then we had no xy cross term in equation (16) to begin with and we don't need to rotate the conic. Therefore we can assume that our conic does not have a cross term and is therefore given by:

$$A'x^2 + C'y^2 + D'x + E'y + F = 0 \qquad (19)$$

with discriminant

$$B^2 - 4AC = -4A'C' \tag{20}$$

for some new constants A' and C'. There are then three cases.

1. Let $B^2 - 4AC = -4A'C' < 0$. It follows that $A' \neq 0 \neq C'$, and therefore we can complete the square in the $A'x^2 + D'x$ and $C'y^2 + E'y$ terms in equation (19) to get

$$\begin{aligned} 0 &= A'x^2 + C'y^2 + D'x + E'y + F \\ -F &= \left[A'\left(x + \frac{D'}{2A'}\right)^2 - A'\left(\frac{D'}{2A'}\right)^2\right] + \left[C'\left(y + \frac{E'}{2C'}\right)^2 - C'\left(\frac{E'}{2C'}\right)^2\right] \\ \frac{-4FA'C' + D'^2C' + E'^2A'}{4A'C'} &= A'\left(x + \frac{D'}{2A'}\right)^2 + C'\left(y + \frac{E'}{2C'}\right)^2 \\ k &= \end{aligned}$$

   for some constant k. Since $-4A'C' < 0$, both A' and C' must have the same sign. Without loss of generality, we will assume they are both positive. If this were not the case, we could simply divide equation (19) through by -1. Thus if $k > 0$, this is the general equation for an ellipse. If $k = 0$, only one point satisfies the equation, namely $(-D'/2A', -E'/2C')$. If $k < 0$, there is no (real) solution and thus no graph.

2. Let $B^2 - 4AC = -4A'C' > 0$. Following the same derivation as in the first case but noting that A' and C' must have the opposite sign, we see that if $k = 0$, we get a pair of lines intersecting at the point found in case 1. If $k \neq 0$, we get a hyperbola.

3. Let $B^2 - 4AC = -4A'C' = 0$. At least one of A', C' must be zero. Without loss of generality, choose C'. Then equation (19) becomes

$$A'x^2 + D'x + E'y + F' = 0$$

   When $A' \neq 0 \neq E'$, we recognize this as a parabola. If $A' \neq 0 = E'$, there are no real solutions. If $A' = 0$ and E' or $D' \neq 0$, this is a line. If $A' > 0$ and $F' > 0$ and $D' = E' = 0$, there is no (real) solution and thus no graph.

This completes the proof □

Notice that in each of the three cases at the end of our proof, we found one imaginary solution, one non-degenerate solution, and one degenerate solution. Moreover, each degenerate solution is paired with its non-degenerate partner. In other words, an ellipse degenerates into a point, and we get both of these for $B^2 - 4AC < 0$. A hyperbola degenerates into a pair of intersecting lines, and we get both of these for $B^2 - 4AC > 0$. A parabola degenerates into a line, and we get both of these for $B^2 - 4AC = 0$. It therefore seems that the discriminant is not just an algebraic trick, but has actual geometric meaning.

3 Generalizing Conic Sections

With an understanding of the different types of conic sections, how to identify them, and some of their properties, we can begin discussing generalizations. In particular, we will generalize the one-dimensional conic sections we have already studied to two-dimensional surfaces. We will examine two different methods: geometric and algebraic. Both require simple modifications to our definitions of conic sections, but the results of these changes will prove to be non-trivial.

3.1 Algebraic Generalization

The main algebraic definition which we gave for a conic section is the set of points in $\mathbb{R}^2$ satisfying the implicit equation

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0 \tag{21}$$

Generalizing this definition essentially requires adding a third variable. Thus we might define the two-dimensional conic sections as the set of points (x, y, z) in $\mathbb{R}^3$ satisfying the implicit equation

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0 \tag{22}$$

for some constants A through J. However, for reasons that will become apparent in Section 4, we will avoid reusing the term "conic." Instead, because equation (22) is a general quadratic equation in three variables, surfaces defined by equation (22) will be called *quadric surfaces* or *quadrics*. Before we can move on to the geometric generalizations, we must briefly examine these quadric surfaces. We begin with several examples.

One of the simplest quadric surfaces is the unit sphere (see Figure 6),

$$x^2 + y^2 + z^2 = 1 \tag{23}$$

which we arrive at from equation (22) by letting $A = B = C = F = 1$ and setting all other coefficients to zero.

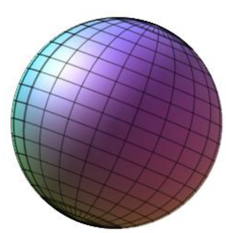


Figure 6: Unit sphere corresponding to equation (23)

A slightly more unusual quadric is given by

$$x^2 + y^2 - z^2 = 2 \tag{24}$$

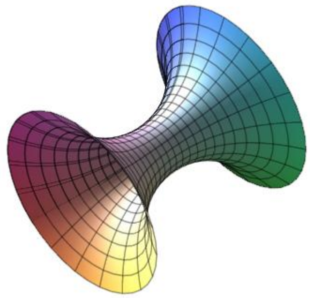


Figure 7: Hyperboloid of one sheet described by equation (24)

where we have set $A = B = 1$, $C = -1$, $J = -2$, and all other coefficients equal to zero (see Figure 7). It appears, and later we will prove this, that one could arrive at the same surface by revolving a hyperbola about the axis between its branches. Moreover, notice that the hyperboloid of one sheet has cross sections of ellipses (circles in particular, for this case) in one direction and hyperbolas in a perpendicular direction. This type of quadric surface is referred to as a *hyperboloid of one sheet.* To understand this name, it helps to see the *hyperboloid of two sheets* as well. An example of a hyperboloid of two sheets is

$$x^2 - y^2 - z^2 = 2 \tag{25}$$

Notice that we have simply switched the sign on the y term in equation (24). By looking at Figure 8, we can see that this surface also has ellipses as one cross section and hyperbolas as another, hence the term hyperboloid, but in this case it is in two distinct pieces, or sheets. In contrast, the surface in Figure 7 is in only one piece, or one sheet.

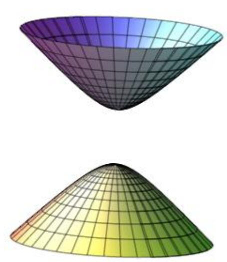


Figure 8: Hyperboloid of two sheets described by equation (25)

3.1.1 Classifying Quadric Surfaces

We wish to develop a means of identifying the type of quadric surface directly from the coefficients of its implicit form, similar to the method we developed for identifying a conic section. This will later allow us to begin comparing the algebraic and geometric generalizations. A basic familiarity with concepts from linear algebra is assumed, such as matrix multiplication and properties of matrices such as rank, determinant, and eigenvalues.

We begin by altering the presentation of the quadric equation by adding a factor of 2 to the cross terms

$$Ax^2 + By^2 + Cz^2 + 2Dxy + 2Exz + 2Fyz + Gx + Hy + Iz + J = 0 \tag{26}$$

This simplifies some of our later work, but does not actually affect the types of surfaces we can obtain. We can use linear algebra to rewrite equation (26) in the form

$$\boldsymbol{x}^T q\boldsymbol{x} + L\boldsymbol{x} + J = 0 \tag{27}$$

where q, L, and $\boldsymbol{x}$ are given by

$$q = \begin{bmatrix} A & D & E \\ D & B & F \\ E & F & C \end{bmatrix} \qquad L = \begin{bmatrix} G & H & I \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

By adopting the form of equation (27), we have separated the equation into three parts containing the coefficients of terms of degree two, one, and zero. These terms can be combined into a single matrix:

$$Q = \begin{bmatrix} A & D & E & G \\ D & B & F & H \\ E & F & C & I \\ G & H & I & J \end{bmatrix} = \begin{bmatrix} q & L^T \\ L & J \end{bmatrix} \tag{28}$$

Any quadric surface can be completely classified using properties of the matrix Q. In particular, we must know the rank of Q and of its sub-matrix q, the sign of the determinant of Q, and the signs of the eigenvalues of q. We therefore adopt the following notation from [6].

$$\rho_3 = \text{rank} q$$
$$\rho_4 = \text{rank} Q$$
$$\Delta = \det Q$$

Recall from linear algebra that the rank of a matrix is the number of linearly independent columns and the determinant is a scalar value associated with square matrices. Methods for determining these values should be familiar and can be found in detail in [5]. Finally, the eigenvalues λ of a matrix are given by the roots of the characteristic equation:

$$\det(q - \lambda I) = 0$$

where I is the identity matrix. Again, we are not concerned with the magnitudes of the eigenvalues, merely their signs. Therefore the following notation is again adopted from [6]:

$$\Lambda = \begin{cases} 1 & \text{if the signs of the nonzero eigenvalues are all the same} \\ 0 & \text{otherwise} \end{cases}$$

Table 1 classifying the quadric surfaces is from [6] and [7]. We present it without proof. The equations provided are the surfaces in standard form, meaning that they have been rotated and translated so that their principal axes are aligned with the familiar axes of $\mathbb{R}^3$ and they are centered at the origin.

Surface	Equation	ρ_3	ρ_4	sgn(Δ)	Λ
ellipsoid (real)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$	3	4	-	1
ellipsoid (imaginary)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=-1$	3	4	+	1
elliptic cone (real)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=0$	3	3		0
elliptic cone (imaginary)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=0$	3	3		1
elliptic cylinder (real)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$	2	3		1
elliptic cylinder (imaginary)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=-1$	2	3		1
elliptic paraboloid	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=z$	2	4	-	1
hyperbolic paraboloid	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=z$	2	4	+	0
hyperbolic cylinder	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=-1$	2	3		0
hyperboloid of one sheet	$\frac{x^2}{a^2}+\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$	3	4	+	0
hyperboloid of two sheets	$\frac{x^2}{a^2}-\frac{y^2}{b^2}-\frac{z^2}{c^2}=1$	3	4	-	0
parabolic cylinder	$x^2+cz=0$	1	3		
intersecting planes (real)	$\frac{x^2}{a^2}-\frac{y^2}{b^2}=0$	2	2		0
intersecting planes (imaginary)	$\frac{x^2}{a^2}+\frac{y^2}{b^2}=0$	2	2		1
parallel planes (real)	$x^2=a^2$	1	2		
parallel planes (imaginary)	$x^2=-a^2$	1	2		
coincident planes	$x^2=0$	1	1		

Table 1: Classification of quadric surfaces from parameters of their matrix equation

Notice that there are several surfaces with two types, real and imaginary. This simply denotes the type of points (real or imaginary) which satisfy the given equation. Note that there are two sets of entries which contain ambiguity: the elliptic cylinders and the parallel planes. In particular, Table 1 does not distinguish between the real and imaginary versions of these surfaces, which is not a trivial difference. The following test from [7], which we will give without proof, allows exactly this distinction:

Theorem 3. *The graph of equation (26) is imaginary if and only if $\rho_4 > 1$ and all of the eigenvalues of Q have the same sign.*

We are now equipped with a means of classifying any given two dimensional quadric surface. Before we move on, we will give an example of using this classification. Consider the surface implicitly defined by

$$9x^2 + y^2 - 3z^2 - 2xy - xz + 5x + 5y + 12 = 0 \tag{29}$$

From equation (28), we can see that this surface has associated with it the matrix

$$Q = \begin{bmatrix} 9 & -1 & -1/2 & 5/2 \\ -1 & 1 & 0 & 5/2 \\ -1/2 & 0 & -3 & 0 \\ 5/2 & 5/2 & 0 & 12 \end{bmatrix}$$

A few elementary row operations on Q gives us the row-equivalent matrix

$$\begin{bmatrix} 1 & 0 & 6 & 0 \\ 0 & 2 & 12 & 5 \\ 0 & 0 & -7 & 10 \\ 0 & 0 & 0 & -1039/14 \end{bmatrix}$$

which has four pivot columns. Thus the dimension of the column space is four, so the rank of Q is four. Furthermore, the submatrix consisting of the first three rows and columns–which is row equivalent to q–has three pivot columns, so q has rank three. Thus we have

$$\rho_3 = 3, \qquad \rho_4 = 4$$

Next we need to find the sign of the determinant of Q. By using cofactor expansion on the last row of Q (before the row operations), we get

$$\begin{aligned} \Delta = \det Q &= -(5/2)(0 - (-3)(-1 \cdot (5/2) - (5/2) \cdot 1) + 0) \\ &\quad + (5/2)((-1/2)((-1/2)(5/2) - 0) - (-3)(9(5/2) - (-1)(5/2)) + 0) \\ &\quad - 0 + 12((-1/2)(0 - 1(-1/2)) - 0 + (-3)(9 \cdot 1 - (-1 \cdot -1))) \\ &= -1031/16 \end{aligned}$$

Finally, we need the signs of the eigenvalues of q, which we get from the characteristic equation:

$$\begin{aligned} 0 &= \det(q - \lambda I) \\ &= \begin{vmatrix} 9-\lambda & -1 & -1/2 \\ -1 & 1-\lambda & 0 \\ -1/2 & 0 & -3-\lambda \end{vmatrix} \\ &= -1/2 \cdot (0 - (-1/2) \cdot (1-\lambda)) - 0 + (-3-\lambda) \cdot ((9-\lambda) \cdot (1-\lambda) - 1) \\ &= -\lambda^3 + 7\lambda^2 + (153/4)\lambda + 95/4 \end{aligned}$$

This cubic has three real roots of approximately -3.0, -0.7, and 10.8. Again, we don't care about the exact answers, just that two of these roots are negative and one is positive. Thus–because the eigenvalues do not all have the same sign–we have

$$\Lambda = 0$$

All together, the values we have calculated are

$$\rho_3 = 3 \qquad \rho_4 = 4, \qquad \Delta = -, \qquad \Lambda = 0$$

Consulting our table, we see that these four values correspond to a hyperboloid of two sheets. If we look at Figure 9, we see that this answer makes sense.

We will utilize this method of characterizing quadric surfaces when we begin comparing conic surfaces and quadric surfaces. First we must formally introduce the conic surfaces.

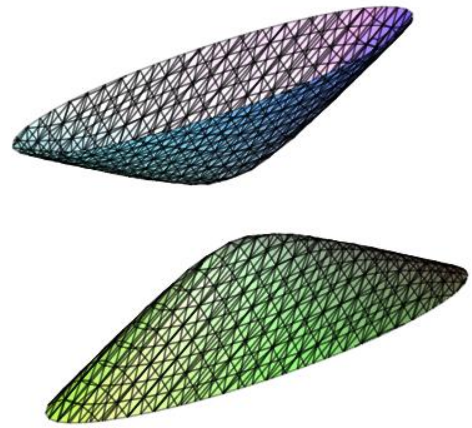


Figure 9: Plot of $9x^2 + y^2 - 3z^2 - 2xy - xz + 5x + 5y + 12 = 0$

3.2 Geometric Generalization

The conic surfaces are the result of generalizing our main geometric definition of the one-dimensional conic sections. To generalize the geometric definition of a conic section to higher dimensions, we want to maintain the idea of intersecting planes and cones. In particular, we will consider the intersections of the three dimensional (i.e. embedded in $\mathbb{R}^4$) analogues of the cone and plane. The generalization of a surface from $\mathbb{R}^3$ to higher dimensions is called a *hypersurface*, which we define formally as follows.

Definition 4. *A* hypersurface *in $\mathbb{R}^n$ is a subset M of $\mathbb{R}^n$ such that the neighborhood of each point in M is contained in the image of a one-to-one regular mapping $\boldsymbol{x} : D \to \mathbb{R}^n$ where D is an open subset of $\mathbb{R}^{n-1}$.*

For a more detailed explanation of the terminology in this definition, see Section 5.1.3, which discusses the formal definition for two-dimensional surfaces in particular. Essentially, a hypersurface in $\mathbb{R}^n$ is a deformation of the space $\mathbb{R}^{n-1}$. By this we mean that we can take $\mathbb{R}^{n-1}$ and bend, stretch, and cut it to obtain our hypersurface. We think of surfaces as being formed from manipulations of the xy plane, or $\mathbb{R}^2$. The dimension of a hypersurface is thus $n-1$, or one less than the dimension of the ambient space. The three dimensional analogues of the plane and cone would thus be called the *hyperplane* and *hypercone* in $\mathbb{R}^4$. Unless noted otherwise, for the remainder of this paper we will assume that our hypersurfaces are in $\mathbb{R}^4$, thus when we refer to the hyperplane and hypercone, we are specifically referring to the appropriate hypersurfaces in $\mathbb{R}^4$.

The hyperplane is implicitly defined as follows:

$$g(x, y, z, w) = a_1x + a_2y + a_3z + a_4w - b = 0$$

We can simplify this by eliminating one of the coefficients. If $a_4 \neq 0$, we are done. Otherwise, we can divide through by a_4 and rename our coefficients to get:

$$g(x, y, z, w) = a_1x + a_2y + a_3z + w - b = 0 \tag{30}$$

The hypercone is implicitly defined by:

$$h(x, y, z, w) = c_1^2x^2 + c_2^2y^2 + c_3^2z^2 - c_4^2w^2 = 0$$

Notice that this looks very similar to the regular cone, except that we have added another positive squared term. We will be working with a specific, simpler hypercone, just as we did for the conic sections:

$$h(x, y, z, w) = x^2 + y^2 + z^2 - w^2 = 0 \tag{31}$$

Before we move on, we will attempt to develop some sort of understanding of what these hypersurfaces "look" like. The difficulty, of course, is that they are embedded in four-dimensional space, which is somewhat difficult to visualize. For a hyperplane, this embedding is trivial: $\mathbb{R}^3$ is essentially an example of a hyperplane in $\mathbb{R}^4$, and any others are given by rotations and translations of this basic hyperplane. A similar trick is not possible with hypercones. The solution is to change one of the four spatial dimensions into something which we can visualize simultaneously with three spacial dimensions. We will refer to the four principal directions in $\mathbb{R}^4$ as x, y, z, and w where w is the direction perpendicular to what we think of as $\mathbb{R}^3$. This is the dimension which we will reassign. One choice is color, which is how many complex functions are visualized. Each point in $\mathbb{R}^3$ is assigned a different color on a given scale, with different color values corresponding to different w values. Another choice is time, which is what we will use.

In particular, we will imagine what we, as three-dimensional observers, would see as we pass a hypercone through $\mathbb{R}^3$. To see how this would work, we first start with the simpler example of a two-dimensional cone in $\mathbb{R}^3$. Consider the situation shown Figure 10. A two-dimensional inhabitant of the indicated plane would not be able to see the cone. However, if we translate this cone up in the z direction with constant speed, as time passes this flatlander would see a circle of steadily increasing size. Thus she can visualize the cone by replacing the z dimension with time and considering the intersections that she sees as time passes. Some of these intersections are shown in Figure 11 to demonstrate how this would be represented.

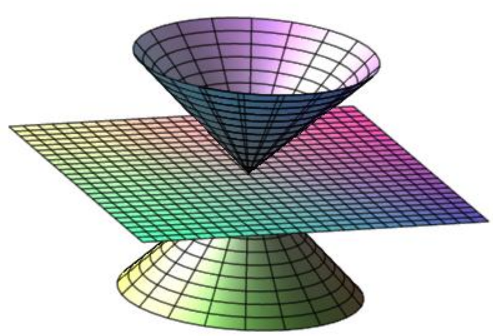


Figure 10: A cone with a plane through its vertex perpendicular to its principal axis

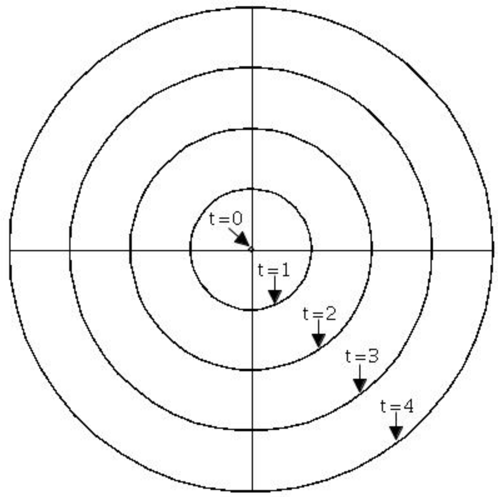


Figure 11: Visualizing a cone from a plane

Similarly, we will consider ourselves as inhabitants of xyz space in $\mathbb{R}^4$. We want to visualize the hypercone whose principal axis is the w axis and whose vertex lies at the origin. As the hypercone begins translating in the w direction, we see a sphere of increasing size centered around the origin. Three discrete intersections are shown in Figure 12 corresponding to three different times. There

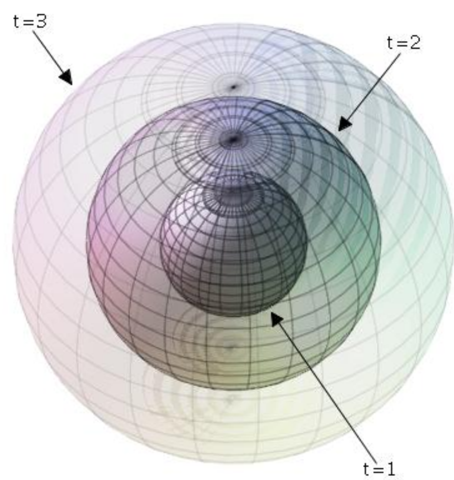


Figure 12: Visualizing a hypercone

are a few things to note here. First, this hypercone is symmetric under any rotations in xyz space, since it only ever has spherical cross-sections in that space. This is analogous to a cone, which we know has rotational symmetry about its principal axis. Second, note that we have just found that spheres are the simplest class of surfaces given by the intersection of a hyperplane and hypercone, and we know that spheres are also quadric surfaces. Finally, the reason that we chose the time visualization rather than the color visualization is partly because it is somewhat easier to visualize, but also because it has an interesting application: the hypercone with one dimension represented by time is an important object in general relativity. For a more detailed look at the following discussion, see [8].

In particular, if we imagine a point in space emitting a pulse of light uniformly in all directions at some time which we call $t = 0$, then the location of that light as time passes–the wavefront–is a series of concentric spheres. In other words, it is a hypercone. For this particular example, we will refer to it as a *light cone*. A visualization of such a light cone is shown in Figure 13, which is taken from [9]. We must recognize that the plane labelled "hyperspace of the present" is actually

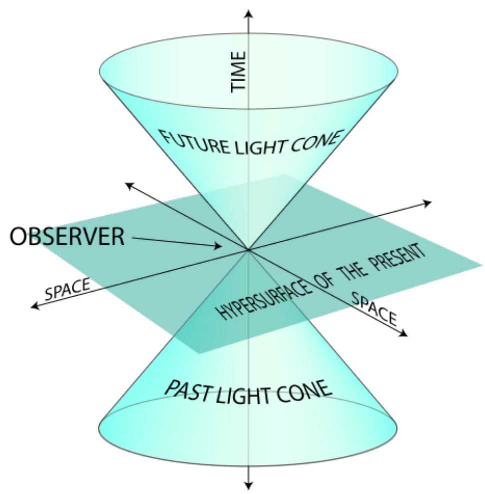


Figure 13: Hypercone representing the location of a point-source of light in space-time

a snapshot of the universe at time $t = 0$. In other words, it is not actually a plane as the picture might suggest, but a three-dimensional hyperplane. The cone which is being represented as two-dimensional is thus actually a three-dimensional hypercone, as we just described. The significance of this picture, and thus of the light cone, is that events within the boundary of this cone can be *causally connected*. An event in space-time is essentially a particular spatial location at a particular time. Two events are causally connected if one can reach the other with the passage of time. Thus two events with the same time coordinate cannot be causally connected. This is represented in Figure 13 by the fact that only a single point is in the light cone at time $t = 0$. Furthermore, if two events are spatially two light-years apart, then they cannot be causally connected before two years pass, because general relativity requires that nothing travels faster than the speed of light. Thus points which lie outside the boundary of the light cone are too far apart at that particular time coordinate to be causally connected with the event at the origin. In other words, the light emitted from the origin at time $t = 0$ has not had time to reach the points on the outside of the cone at that time. Any spatial point can eventually be reached, but there is a boundary on how quickly it can be reached.

With this example of the importance of hypercones to motivate us, and some sense of what a hypercone looks like, we will begin considering the intersections of hyperplanes and hypercones.

We start by solving for w in equation (30) and substituting it into equation (31). This gives us

$$x^2 + y^2 + z^2 = (b - (a_1 x + a_2 y + a_3 z))^2$$

We can expand and rearrange this equation to get

$$(a_1^2 - 1)x^2 + (a_2^2 - 1)y^2 + (a_3^2 - 1)z^2 + 2a_1 a_2 xy + 2a_1 a_3 xz + 2a_2 a_3 yz - 2ba_1 x - 2ba_2 y - 2ba_3 z + b^2 \quad (32)$$

This is certainly a two dimensional quadric surface. However, it is not the surface of intersection. The surface implicitly described by equation (32) is essentially the parameter space for the surface of intersection. To find the actual intersection, we would have to plug these points back into a parametrization for the hyperplane (or hypercone) parametrized by x,y, and z. For the hyperplane, such a parametrization is given by

$$\boldsymbol{x} = (x, y, z, b - (a_1 x + a_2 y + a_3 z))$$

The problem is that if we were to use this approach, we would end up with a two-dimensional surface, as we would expect, but it would be embedded in $\mathbb{R}^4$, thus we would have difficulty visualizing it, which should be a simple task. One way around this is to take the surface of intersection in $\mathbb{R}^4$ and perform the appropriate rotations and translations to embed it solely in $\mathbb{R}^3$. We know that such an embedding is possible because the surface of intersection necessarily lies entirely in the hyperplane, which is equivalent to $\mathbb{R}^3$ after a particular rotation and translation. However, this has proven to be an impractically difficult solution. Fortunately, it is still useful to have a description of our parameter space, since constraints on our parameter space will necessarily place constraints on our solution space.

Now that we have developed our two competing generalizations of the one-dimensional conic sections, we can begin comparing them.

4 Comparing Generalizations of Conic Sections

In order to keep our two generalizations separate, recall that we will refer to *conic surfaces* as the surfaces given by the intersection of a hyperplane and hypercone, and we will refer to *quadric surfaces* as the set of points given by the implicit equation

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0$$

Thus, again, conic surfaces are the geometric generalization and quadric surfaces are the algebraic generalization. The question we wish to examine is this: *is the set of all quadric surfaces identical to the set of all conic surfaces?*.

4.1 Some Quadric Surfaces may not be Conic Surfaces

In order to begin exploring this question, recall the following equation for the parameter space of the intersection of an arbitrary hyperplane and a hypercone with its vertex at the origin and principal axis aligned with the w axis:

$$(a_1^2 - 1)x^2 + (a_2^2 - 1)y^2 + (a_3^2 - 1)z^2 + 2a_1 a_2 xy + 2a_1 a_3 xz + 2a_2 a_3 yz - 2ba_1 x - 2ba_2 y - 2ba_3 z + b^2 \quad (33)$$

We can see that these are quadric surfaces. Although we are aware that these are not the actual surfaces of intersection, it is useful to know what kind of constraints exist on the types of parameter surfaces we can get. Therefore, we turn to the classification tools developed in Section 3.1.1. If we put the coefficients of equation (33) into the matrix in equation (28), we get the matrix

$$Q = \begin{bmatrix} a_1^2 - 1 & a_1 a_2 & a_1 a_3 & -ba_1 \\ a_1 a_2 & a_2^2 - 1 & a_2 a_3 & -ba_2 \\ a_1 a_3 & a_2 a_3 & a_3^2 - 1 & -ba_3 \\ -ba_1 & -ba_2 & -ba_3 & b^2 \end{bmatrix}$$

Without plugging in specific values, we can get some interesting information from this matrix (the following calculations were primarily done in Mathematica 9). For example, it turns out that the determinant is

$$\det Q = -b^2$$

Thus we immediately see that none of the surfaces with the requirement that

$$\Delta = \operatorname{sgn}(\det Q) = +$$

are possible. Even if $\det Q = 0$, this means that the last row and column of Q become zero, so $\rho_4 \leq 3$. Thus the surfaces requiring knowledge of $\operatorname{sgn}\det Q$ in the table are not possible, since they all require $\rho_4 = 4$. In particular, this means that we can't get the hyperbolic paraboloid or the hyperboloid of one sheet.

The eigenvalues of the sub-matrix q are

$$\lambda = -1, -1, -1 + a_1^2 + a_2^2 + a_3^2$$

We can clearly get all negative or simultaneously positive and negative eigenvalues, giving us both choices of $\Lambda = 1$ or $\Lambda = 0$. Let us look at the particular case of $a_1 = 1$ and $a_2 = a_3 = b = 0$. Geometrically, this corresponds to the hyperplane perpendicular to the x_1 axis at the origin. The matrix Q then becomes

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and the eigenvalues of q become

$$\lambda = -1, -1, 0$$

Thus, since $\rho_3 = 2 = \rho_4$, we have intersecting imaginary planes. If instead of setting $b = 0$ we set $b = 1$, we get $\rho_3 = 2$ and $\rho_4 = 4$. The eigenvalues of q remain unchanged since the submatrix does not have any dependence on b. Thus we still have $\Lambda = 1$, and we have already shown that $\operatorname{sgn}\det Q < 0$, so this corresponds to an elliptic paraboloid.

It is impossible to make $\rho_3 = 1$. We can see this by noting that having the rank of q equal to one corresponds to all three rows of q being identical up to a constant multiple. Scaling out this constant so that each row is identical (which does not change the rank of the matrix), we see from comparing entries in the first column of q that $a_1 a_2 = a_1 a_3$, or $a_2 = a_3$. Comparing entries in the last column of q then tells us that $a_2 a_3 = a_3^2 = a_3^2 - 1$, which is a contradiction. Thus $\rho_3 > 1$, and therefore it is impossible to get a parabolic cylinder. It may be true that it is also impossible to make $\rho_3 = 2$ while simultaneously making $\rho_4 = 3$, which would any type of cylinder impossible. However, we have not proven this: it would be an important question to investigate with more time.

We have explored limitations on the parameter space of the intersection of a hyperplane and hypercone. In particular, we have found that not all quadric surfaces are possible *in the parameter space*. What does this tell us about the solution space, about the actual intersection? Unfortunately, we have not yet developed a complete answer. However, we can draw on our understanding of the one-dimensional conics to make a reasonable conjecture. The only difference between curves in the parameter space and the actual curves of intersection is a simple dilation. An ellipse in the parameter space might have a different eccentricity from the same curve in the solution space, but it will still be an ellipse. This is essentially because the only difference between the plane of the solution space and the plane of the parameter space is a simple rotation and translation. Since the only difference between the xyz hyperplane that our parameter spaces are embedded in and the hyperplane that the solution space is embedded in is a rotation and translation, we might expect a similar effect. Thus we make the following conjecture:

Conjecture 1. *If a class of quadric surface cannot be obtained in the parameter space of the intersection of a hyperplane and hypercone, then it also cannot be obtained in the solution space.*

An immediate result of proving this conjecture would be proving that the parabolic cylinder, the hyperbolic paraboloid, and the hyperboloid of one sheet are not conic surfaces. It would therefore follow that not all quadric surfaces are conic surfaces. Therefore, proving or disproving this conjecture would be of particular interest for future directions of this project. One difficulty that might arise is that rotations in $\mathbb{R}^3$ are in some ways very different from rotations in $\mathbb{R}^4$, and therefore the linear algebra associated with changing between the parameter and solution spaces for the one-dimensional conic sections (i.e. the requisite rotations matrices) might be fundamentally different from that associated with the two-dimensional conic surfaces.

4.2 Non-Spherical Hypercones

Before we move on, we will consider a slight modification of the hypercone we have been investigating. We know that equation (31) is a specific case of a hypercone with spherical cross sections parallel to xyz space. Now let us look at a more general problem: consider a non-symmetric hypercone, similar to an elliptic cone in $\mathbb{R}^3$ with non-circular cross sections perpendicular to its principal axis. Such a surface would have the equation

$$h(x, y, z, w) = a_1^2x^2 + a_2^2y^2 + a_3^2z^2 - a_4^2w^2 = 0$$

Let us generalize this hypersurface one step further by allowing it to be translated in the w direction to get

$$h(x, y, z, w) = a_1^2x^2 + a_2^2y^2 + a_3^2z^2 - a_4^2(w - c)^2 = 0 \tag{34}$$

with $c \neq 0$. It is exceptionally easy to find the intersection of this hypercone with the hyperplane of xyz space. This hyperplane is simply given by

$$w = 0$$

Plugging this into equation (34), we get

$$
\begin{aligned}
0 &= a_1^2x^2 + a_2^2y^2 + a_3^2z^2 - a_4^2(w-c)^2 \\
&= a_1^2x^2 + a_2^2y^2 + a_3^2z^2 - a_4^2c^2
\end{aligned}
$$

This gives us all possible real ellipsoids, since the coefficients are all nonzero and independent. This is all it allows us to get, but this should be unsurprising since this is analogous to deforming the two dimensional circular cone to an elliptic cone, then lowering it to pass through the $xy-$plane: we could only ever get bounded figures and are thus restricted to ellipses. Similarly, deforming the hypercone and lowering it into $xyz-$space will give bounded quadrics, which are ellipsoids.

There are two reasons we are discussing these elliptic hypercones. First, it gives us a slightly better idea of the geometry of hypercones in general and in particular of their connection to two-dimensional cones. Second, it suggests and begins an interesting extension of this project, which is considering the intersections of all quadric surfaces rather than simply spherical hypercones and hyperplanes. This is another way of generalizing our geometric definition of the conic sections, and it is important to realize that many possible (reasonable) generalizations exist which we have not explored.

4.3 All Conic Surfaces are Quadric Surfaces

There is one last question that we must consider: which conic surfaces are also quadric surfaces? This questions turns out to be simpler than its inverse which we have been exploring so far. To answer it, we will consider the intersection of the hyperplane $w = 0$ with a general hypercone which has been arbitrarily rotated and translated in $\mathbb{R}^4$. Notice that this gives us all possible conic surfaces, because this is identical to fixing a particular hypercone and passing arbitrary hyperplanes throuh it. The general hypercone is given by

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gxw + Hyw + Izw + Jx + Ky + Lz + M - Pw^2 + Qw = 0$$

where the coefficients are not necessarily independent. Thus the desired intersection is given by

$$
\begin{aligned}
0 &= Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx(0) + Hy(0) + Iz(0) \\
&\quad + Jx + Ky + Lz + M - P(0) + Q(0) \\
&= Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J
\end{aligned}
$$

which is clearly a two-dimensional quadric surface. This does not necessarily mean that we get all quadric surfaces, since the coefficients may have interdependence, but it does guarantee that we *only* get quadric surfaces. In other words, all intersections of hyperplanes and hypercones are quadric surfaces, thus *all conic surfaces are quadric surfaces.*

5 Differential Geometry of Quadric Surfaces

We now know that all conic surfaces are quadric surfaces. Therefore any information we can obtain about the more general class of quadric surfaces will give us insight into conic surfaces as well. This is particularly useful, since quadric surfaces are somewhat easier to study; unlike the conic surfaces where we have to deal with the problem of switching between parameter spaces and solution spaces, we have a concrete formula immediately available for the quadric surfaces:

$$Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0$$

Using this formula, we will begin using the tools of differential geometry to study quadric surfaces. After developing a few fundamental tools, we will investigate two questions in particular:

- Which quadric surfaces are surfaces of revolution?
- What are the constraints on the curvature of a quadric surface?

The first question is straightforward to understand, although non-trivial to answer. The second will require the development of a fair amount of background material to understand the question, let alone answer it. However, answering this question will prove to be one of the most powerful and original results of this paper.

5.1 Preliminaries

5.1.1 Higher Dimensional Derivatives

At several points, we will find ourselves needing a definition of the derivative which applies both to vector valued and multivariable functions. In order to introduce this definition, we require some notation. Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$ and let $a \in \mathbb{R}^n$. Then there exist m unique functions $f_i : \mathbb{R}^n \to \mathbb{R}$ such that $f(a) = (f_1(a), f_2(a), \ldots, f_m(a))$. The j^{th} partial derivative of the i^{th} component of f is denoted by $D_j f_i(a)$.

Example 1. Let $f : \mathbb{R}^2 \to \mathbb{R}^3$ be given by $f(x, y) = (x^2, y^2, xy^3)$. Then we have

$$\begin{aligned}
D_1 f_1(x, y) &= 2x \\
D_1 f_2(x, y) &= 0 \\
D_1 f_3(x, y) &= y^3 \\
D_2 f_1(x, y) &= 0 \\
D_2 f_2(x, y) &= 2y \\
D_2 f_3(x, y) &= 3xy^2
\end{aligned}$$

We are now equipped to give a general definition of the derivative.

Definition 5. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable at a point $a \in \mathbb{R}^n$ if, for $1 < i < m, 1 < j < n$, $D_j f_i(x)$ exists and is continuous for every point x in some open set containing a. The derivative of f at a, denoted by $f'(a)$, is given by the $m \times n$ matrix*

$$f'(a) = \begin{bmatrix} D_1 f_1(a) & D_2 f_1(a) & \ldots & D_n f_1(a) \\ D_1 f_2(a) & D_2 f_2(a) & \ldots & D_n f_2(a) \\ \vdots & \vdots & \ddots & \vdots \\ D_1 f_m(a) & D_2 f_m(a) & \ldots & D_n f_m(a) \end{bmatrix}$$

This matrix is often called the *Jacobian* of f and will be important in our discussion of the Gauss curvature of quadric surfaces in Section 5.3.1. Understanding this generalization of the derivative might be easier if we notice that the rows of $f'(a)$ are simply the gradients of the component functions. In symbols,

$$f'(a) = \begin{bmatrix} \nabla f^1(a) \\ \nabla f^2(a) \\ \vdots \\ \nabla f^m(a) \end{bmatrix}$$

Example 2. In order to illustrate this tool, we will find the derivative of the function in example 1. The necessary partial derivatives were already found, thus we have

$$\begin{aligned} f'(a) &= \begin{bmatrix} D_1 f_1(x,y) & D_2 f_1(x,y) \\ D_1 f_2(x,y) & D_2 f_2(x,y) \\ D_1 f_3(x,y) & D_2 f_3(x,y) \end{bmatrix} \\ &= \begin{bmatrix} 2x & 0 \\ 0 & 2y \\ y^3 & 3xy^2 \end{bmatrix} \end{aligned}$$

5.1.2 Curves

We must now introduce a few basic ideas about curves. Since we will want to apply these ideas to curves on quadric surfaces, we will restrict ourselves to $\mathbb{R}^3$. We can then make the following definition:

Definition 6. *A* curve *in $\mathbb{R}^3$ is a continuous mapping $\boldsymbol{\alpha} : I \to \mathbb{R}^3$ where I is an interval in $\mathbb{R}$.*

We will often represent curves parametrically as follows:

$$\boldsymbol{\alpha}(t) = (x(t), y(t), z(t))$$

where $x(t)$, $y(t)$, and $z(t)$ are the *coordinate functions.* If we can replace the codomain in Definition 6 with $\mathbb{R}^2$, then $\boldsymbol{\alpha}$ is a plane curve. If $I = [a, b]$, $\boldsymbol{\alpha}(a) = \boldsymbol{\alpha}(b)$, and the n^{th} derivatives of $\boldsymbol{\alpha}(a)$ and $\boldsymbol{\alpha}(b)$ are equal for all n, then $\boldsymbol{\alpha}$ is a closed curve. Notice that all conic sections are plane curves by the simple fact that they are sections of a plane.

Example 3. Consider the ellipse in standard form parametrized by

$$\boldsymbol{\alpha}(t) : [0, 2\pi] \to \mathbb{R}^2, \qquad \boldsymbol{\alpha}(t) = (a \sin\theta, b\cos\theta)$$

We will show that this is a closed curve. First, to verify that the parametrization is valid, we simply note that the coordinate functions satisfy the implicit equation for an ellipse in standard form:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

Next, we check that the endpoints are equal:

$$\begin{aligned}\boldsymbol{\alpha}(0) &= (0, b) \\ &= (a\sin(2\pi), b\cos(2\pi)) \\ &= \boldsymbol{\alpha}(2\pi)\end{aligned}$$

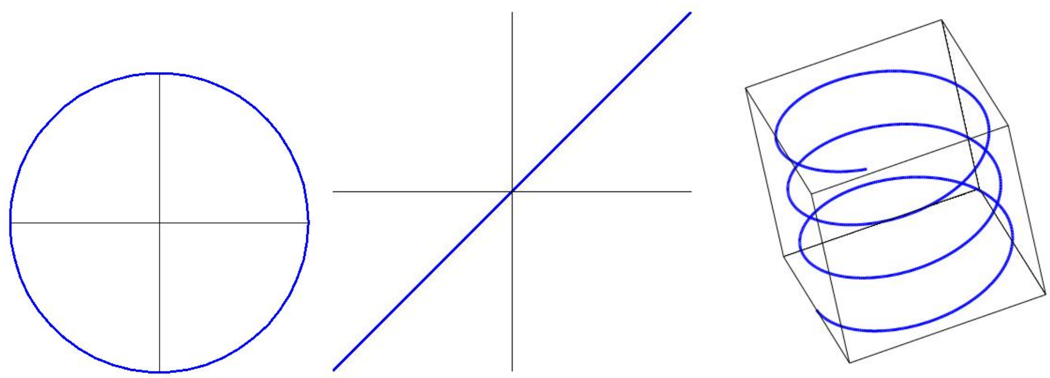
Finally, note that the n^{th} derivative of $\boldsymbol{\alpha}(t)$ is given by

$$\boldsymbol{\alpha}^{(n)}(t) = \begin{cases} (a\sin\theta, b\cos\theta) & n \equiv 0 \mod 4 \\ (a\cos\theta, -b\sin\theta) & n \equiv 1 \mod 4 \\ (-a\sin\theta, -b\cos\theta) & n \equiv 2 \mod 4 \\ (-a\cos\theta, b\sin\theta) & n \equiv 3 \mod 4 \end{cases}$$

Without checking each case individually, we can see that the values of the each of these derivatives will also be equal at the two endpoints. Therefore, an ellipse in standard form is a closed curve.

Since the only difference between a standard ellipse and a general ellipse is a rotation and translation, it in fact follows that all ellipses are closed curves. Furthermore, although we will not formally derive it, it is easy to see that none of the other conic sections have points given by two distinct parameter values. Therefore ellipses are the only closed conic sections, which agrees with our previous knowledge of the conic sections.

Figure 5.1.2 illustrate some of our new definitions. It is important to realize that all curves can



A circle is a closed plane curve — A line is a non-closed plane curve — A helix is a non-closed, non-plane curve

Figure 14: Three different types of curves

be parametrized in more than one way. For example, consider the following parametrization of the line segment from $(0, 0, 0)$ to $(1, 0, 0)$:

$$\boldsymbol{\alpha}(t) = (t, 0, 0) \qquad 0 \le t \le 1$$

Clearly, the same line segment is given by

$$\boldsymbol{\beta}(t) = (2t, 0, 0) \qquad 0 \le t \le 1/2$$

In many cases, the differences seem less trivial. We will discuss this issue further in Section 5.2.

It will be useful to talk about vectors tangent to curves. We make the following definition:

Definition 7. *The tangent vector to a curve* $\boldsymbol{\alpha}(t) = (\alpha_1(t), \alpha_2(t), \alpha_3(t))$ *at a point* $\boldsymbol{\alpha}(t_0)$ *is denoted* $\boldsymbol{\alpha}'(t)$ *and given by*

$$\boldsymbol{\alpha}'(t_0) = \left(\frac{d\alpha_1}{dt}(t_0), \frac{d\alpha_2}{dt}(t_0), \frac{d\alpha_3}{dt}(t_0) \right)$$

Notice that this definition is consistent with Definition 5, justifying our use of the notation $\boldsymbol{\alpha}'(t)$. Intuitively, if we think of a curve as a particle moving in $\mathbb{R}^3$ such that its position is given by $\boldsymbol{\alpha}(t)$ at time t, the tangent vector at time t_0 is the direction in which the particle is moving at that moment with magnitude equal to the speed at which the particle is moving in that direction. Because of this picture, the tangent vector is sometimes called the velocity vector of a curve.

We require one last property to complete our preliminary discussion of curves:

Definition 8. *A* regular *curve* $\boldsymbol{\alpha}$ *is one for which* $\boldsymbol{\alpha}'(t) \neq 0$ *for all* t *in the domain of* $\boldsymbol{\alpha}$.

Note that in our picture of $\boldsymbol{\alpha}(t)$ as describing the motion of a particle, $\boldsymbol{\alpha}(t)$ being regular essentially means that the particle never stops moving, ie its velocity never goes to zero.

5.1.3 Surfaces

Now that we are equipped with some basic ideas about curves, we can begin discussing surfaces, starting with the formal definition:

Definition 9. *A* surface *in* $\mathbb{R}^3$ *is a subset* M *of* $\mathbb{R}^3$ *such that the neighborhood of each point in* M *is contained in the image of a one-to-one regular mapping* $\boldsymbol{x} : D \to \mathbb{R}^3$ *where* D *is an open subset of* $\mathbb{R}^2$.

There are some technical points to understand in this definition. First, we say that a set S is an *open set* in $\mathbb{R}^n$ if for every $x \in S$, there exists an $\epsilon > 0$ such that for all $y \in \mathbb{R}^n$ where $|x - y| < \epsilon$, $y \in S$. Intuitively, we might say that all of the points "surrounding" elements of the set are also in the set, but we should not rely too heavily on this idea for any actual work since it is not well defined. This leads to the next idea of a *neighborhood.* A neighborhood of a point p in M is a subset of M which contains an *open set* containing p. A disk which contains its boundary (a circle) is not an open set, as any point on the boundary of a disk cannot have a neighborhood of points in that disk. However, points on the interior of the disk do have a neighborhood of points in the disk. Thus implicit in our definition of a surface is the fact that every point must have a neighborhood, thus we cannot have any problematic edges where a point might be in M but not have a neighborhood in M. This is why we require that the domain D be an open set.

Notice that we have required that the points in the neighborhood of any point on a surface can be given by two independent coordinates, and that their is an invertible mapping from $\mathbb{R}^2$ into these neighborhood points. This essentially means that a surface as we have defined it is *locally Euclidean*, which we can think as the surface "looking like" $\mathbb{R}^2$ on a small, or local, scale. This fits our intuitive idea of a surface as a deformation of a plane. If we return momentarily to the hypersurfaces in $\mathbb{R}^4$ from Section 3.2, notice that the neighborhood of each point on a four-dimensional hypersurface is locally Euclidean in the sense that the hypersurface locally "looks like" $\mathbb{R}^3$.

If we hold one parameter constant in our domain D, we get a function of a single parameter, which is simply a curve. We call the curve with constant parameter u the *v-parameter curve* and the curve with constant v the *u-parameter curve.* We can find tangent vectors to these parameter curves by taking derivatives just as we did during our discussion of curves. Thus we denote the derivative of the u-parameter curve by $\boldsymbol{x}_u$ and the derivative of the v-parameter curve by $\boldsymbol{x}_v$. Note that these are simply the partial derivatives of $\boldsymbol{x}$. We generalize the idea of regularity from curves with the following definition:

Definition 10. *A surface $\boldsymbol{x}(u, v) : D \to \mathbb{R}^3$ is regular if $\boldsymbol{x}_u \times \boldsymbol{x}_v \neq 0$ for all $(u, v) \in D$.*

Notice that geometrically, this requires that the u and v parameter curves are never parallel. If they were, then our mapping would "pinch" and become a curve at those points.

Finally, we want to generalize the idea of the tangent vector of a curve (ie velocity vectors) to surfaces:

Definition 11. *The* tangent plane $T_P(M)$ *of a regular surface M at a point $P \in M$ is the set of all vectors tangent to M at P.*

From our discussion of tangent vectors earlier, we know that $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ are in $T_P(M)$. In fact, since we have required that our surface is regular, we know that $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ are linearly independent. We can use this fact to justify calling the set of all tangent vectors at a point a *tangent plane*. Let M be a surface given by $\boldsymbol{x}(u, v) = (x_1(u, v), x_2(u, v), x_3(u, v))$ and let $\boldsymbol{\alpha}(t)$ be a curve lying on M. Because $\boldsymbol{\alpha}$ lies on M, there exist functions $u(t)$ and $v(t)$ such that $\boldsymbol{\alpha}(t) = \boldsymbol{x}(u(t), v(t))$. Thus by application of the chain rule, we see that the tangent vector of $\boldsymbol{\alpha}$ is given by

$$\boldsymbol{\alpha}'(t) = \boldsymbol{x}_u u'(t) + \boldsymbol{x}_v v'(t)$$

Thus any tangent vector on the surface can be written as a linear combination of $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$, so $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ form a basis for $T_P(M)$. Therefore $T_P(M)$ is a two dimensional vector space with its origin at the point $(u(t), v(t))$ on the surface M. In other words, $T_P(M)$ is a plane in $\mathbb{R}_3$. An example is shown in Figure 15, which shows the tangent plane of a point on a sphere. The two red lines on the plane indicate the spanning vectors $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$. Furthermore, because $\boldsymbol{x}_u$ and $\boldsymbol{x}_v$ are linearly

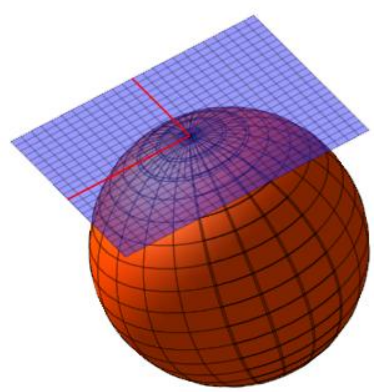


Figure 15: Tangent plane of a point on a sphere

independent vectors in the tangent plane, their cross product is perpendicular to the tangent plane and so we call it a normal vector to the surface. In general, we prefer to work with unit vectors, so let

$$\boldsymbol{U} = \frac{\boldsymbol{x}_u \times \boldsymbol{x}_v}{|\boldsymbol{x}_u \times \boldsymbol{x}_v|}$$

be the *unit normal vector*.

5.2 Surfaces of Revolution

Recall from calculus that a surface of revolution is generated by revolving a curve about an axis. We will formalize this slightly in a moment, but first consider the surface given by the equation

$$x^2 + 2z = 0 \tag{35}$$

By simply looking at the plot of this surface in Figure 16, we intuitively believe that it is not a

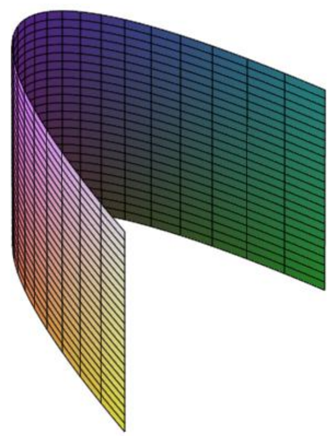


Figure 16: Parabolic cylinder given by equation (35)

surface of revolution. One possible explanation as to why is that there is no line through which we can take perpendicular cross sections which are only circles. In this section, we will formalize this explanation, which will help give us insight into the relationship between quadric surfaces and surfaces of revolution. We begin with a definition.

Definition 12. *A surface of revolution is formed by revolving a plane curve about an axis in that plane.*

By using this basic definition, we can derive the more explicit characterization of surfaces of revolution given in the following lemma due to [10]:

Lemma 1. *All surfaces of revolution where the axis of revolution is the $x-$axis can be parametrized to have the form*

$$\boldsymbol{x} = (g(u), h(u)\cos v, h(u)\sin v) \tag{36}$$

The constraint that the $x - axis$ be the axis of revolution turns out not to be as limiting as it may seem, because we can rotate and translate any surface of revolution so that this is true.

Proof. Let our plane curve be defined parametrically by

$$\boldsymbol{\alpha}(u) = (g(u), h(u), 0) \tag{37}$$

where g and h are arbitrary functions. We will take our plane to be the xy plane and our axis of rotation to be the x axis. We can always rotate and translate an arbitrary plane curve in $\mathbb{R}$ to make this true. We will define our surface of revolution parametrically by $\boldsymbol{x}(u, v) = (l(u, v), m(u, v), n(u, v))$. Thus in order to find the surface of revolution, we must find expressions for l, m, and n in terms of g, h, u, and our new parameter v. Let v be the angle from the xy plane through which a point on our generating curve has been rotated. Since we are rotating about the x axis, the x coordinate of a point on our generating curve does not change, thus $l(u, v) = g(u)$. It follows that $m(u, v) = h(u)\cos v$ and $n(u, v) = h(u)\sin v$. A diagram of this situation is shown in Figure 17.

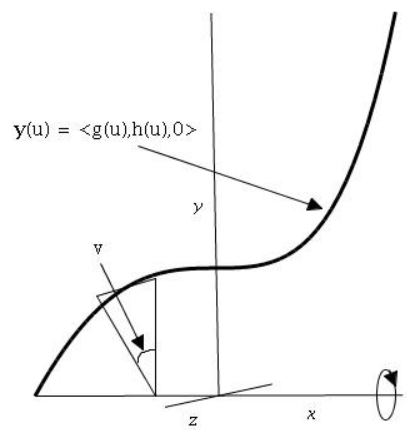


Figure 17: Parametrizing a surface of revolution

Thus our surface of revolution is given by

$$\boldsymbol{x} = (g(u), h(u)\cos v, h(u)\sin v)$$

□

We will use this equation to prove the following theorem which was suggested at the beginning of the section:

Theorem 4. *A surface is a surface of revolution if and only if it has circular cross sections perpendicular to some line (the axis of revolution).*

Proof. Consider an arbitrary cross section of $\boldsymbol{x}$ from equation (36) perpendicular to the x axis at some point $(g(u), 0, 0)$ along the x axis. In order to specify that this cross section does not change in the x direction, we require that the first coordinate of each of the points in this cross section is constant. In other words. $g(u) = c_1$ for some constant c_1. Since $g(u)$ is not constant in general, this requires that u be constant. Thus $h(u) = c_2$ for some constant c_2. Our parametrization then simplifies to

$$\boldsymbol{x} = (c_1, c_2 \cos v, c_2 \sin v)$$

which we recognize as the equation of a circle with radius c_2.

Now suppose that we have a surface which has circular cross sections along a particular line. We wish to show that this is a surface of revolution, i.e. that it has the form of equation (36). We can rotate and translate our surface such that this line with perpendicular circular cross sections is the $x-$axis. This surface can be parametrized by

$$\boldsymbol{x} = (g(u,v), i(u,v), k(u,v)) \tag{38}$$

for some coordinate functions g, h, and k (i.e. one-to-one regular mappings from an open subset of $\mathbb{R}^2$ into $\mathbb{R}$). However, we require that when g is held constant, this parametrization simply traces out a circle. Thus equation (38) becomes

$$\boldsymbol{x} = (g(u,v), h(u,v)\cos v, h(u,v)\sin v)$$

for some functions g and h such that h is constant whenever g is constant. In order for this to be true, g and h must depend only on u, thus we get

$$\boldsymbol{x} = (g(u), h(u)\cos v, h(u)\sin v)$$

which is the desired form of a surface of revolution. □

Recall that these curves generated by holding one parameter constant and allowing the other to vary are called *parameter curves*. The circular cross sections we just investigated are v-parameter curves because they were generated by holding u constant and allowing v to vary in equation (36). Note that the u parameter curves of equation (36) are the original generating curve, equation (37), rotated by the angle v in the $yz-$plane.

Example 4. Recall the hyperboloid of one sheet:

$$x^2 + y^2 - z^2 = 1 \tag{39}$$

Looking at Figure 18, we can be fairly certain that it has circular cross sections perpendicular to the x axis, and thus we suspect that it is a surface of revolution. In order to verify this, we first require a parametrization of the surface. One possibility is

$$\boldsymbol{x}(u,v) = (\cosh u \cos v, \cosh u \sin v, \sinh v)$$

We can check that this is a correct parametrization by plugging in the values for x, y, and z and recalling the hyperbolic trig identity $\cosh^2 x - \sinh^2 x = 1$.

$$\begin{aligned}\cosh^2 u \cos^2 v + \cosh^2 u \sin^2 v - \sinh^2 v &= \cosh^2(\cos^2 v + \sin^2 v) = \sinh^2 u \\ &= \cosh^2 u(1) - \sinh^2 u \\ &= 1\end{aligned}$$

Thus, because it can be parametrized in the form of equation (36), the hyperboloid of one sheet is a surface of revolution.

Now that we are equipped with a more rigorous characterization of surfaces of revolution, we can revisit the surface from equation (35).

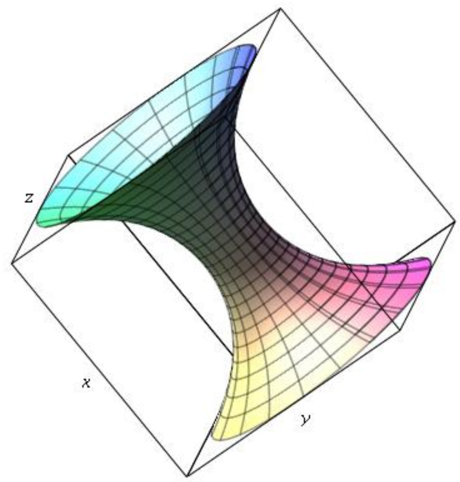


Figure 18: Hyperboloid of one sheet described by equation (39)

Example 5. We can parametrize the parabolic cylinder relatively easily:

$$\boldsymbol{x}(u, v) = (u, u^2, v)$$

We see that this does not have the required form for a surface of revolution of equation (36). If it could be reparametrized to have the correct form, then it is a surface of revolution. However, we propose that such a reparametrization is impossible. In order to prove this, we need to understand what it means for two parametrizations to be equivalent. Recall that this question was first raised in Section 5.1.2. Intuitively, we want two parametrizations to be equivalent if they trace out the same surface as both sets of parameters range over all possible values. This is guaranteed by the following definition:

Definition 13. *Consider the parametrizations* $\boldsymbol{\alpha} : A \to \mathbb{R}^3$ *and* $\boldsymbol{\beta} : B \to \mathbb{R}^3$ *with* A *and* B *both open domains in* $\mathbb{R}^2$*. We say that* $\boldsymbol{\alpha}$ *and* $\boldsymbol{\beta}$ *are* equivalent *if there exists a bijection* $\phi : A \to B$ *such that the following are true for* $(u, v) \in A$*:*

1. *The first two derivatives of* ϕ *exist and are continuous.*
2. $\phi'(u, v) \neq 0$
3. $\boldsymbol{\beta}(\phi(u, v)) = \boldsymbol{\alpha}(u, v)$

The first two requirements assure us that when moving between the input domains D_1 and D_2, no jumps are introduced that might turn a well-behaved manifold into a structure with different geometric properties. The last requirement ensures that we are not altering the actual parametrizations $\boldsymbol{\alpha}$ or $\boldsymbol{\beta}$, we are simply modifying their input values. In other words, the two parametrizations may trace out the surface differently–in different directions or with different "speeds"–based on their input values, but these input values can be altered so that the parametrizations overlap perfectly, making it clear that they ultimately give the same surface. In essence, for two parametrizations to give the same surface, they must be equal component-wise up to a difference in parameter, since vectors are equal exactly when their components are equal.

Returning to the example of the parabolic cylinder, let $\boldsymbol{\alpha} : A \to \mathbb{R}^2$ and $\boldsymbol{\beta} : B \to \mathbb{R}^2$ be given by

$$\begin{aligned}\boldsymbol{\alpha}(u, v) &= (g(v)\cos u, g(v)\sin u, h(v))\\ \boldsymbol{\beta}(x, y) &= (x, x^2, y)\end{aligned}$$

Recall Definition 13, which tells us that in order for two parametrizations to define the same surface, they must be equal component-wise up to a change in parameter values. Assume by way of contradiction that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are equivalent parametrizations. Then we can choose u and v such that the following hold simultaneously:

$$g(v)\cos u = x \tag{40}$$

$$g(v)\sin u = x^2 \tag{41}$$

$$h(v) = y$$

By comparing equations (40) and (41), we see that

$$x = \frac{x^2}{x} = \frac{g(v)\sin u}{g(v)\cos u} = \tan u$$

We can plug this back into equation (40) to get

$$\begin{aligned}g(v)\cos u &= \tan u\\ g(v) &= \tan u \sec u\end{aligned}$$

Since we have functions of two independent variables equal to each other, both functions must be constant. However, $\tan u \sec u$ is not constant, which is a contradiction, thus $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are not equivalent. Therefore, the parabolic cylinder is not a surface of revolution.

5.3 Curvature

One fundamental property of a curve which we will examine is how sharply it bends at a given point. We will call this property *curvature* and define it as follows:

Definition 14. *The* curvature *of a point on a curve is given by*

$$k(t) = \frac{|\boldsymbol{\alpha}'(t) \times \boldsymbol{\alpha}''(t)|}{|\boldsymbol{\alpha}'(t)|^3} \tag{42}$$

Recall that the magnitude of a vector $\boldsymbol{v} = (a, b, c)$ is given by $|\boldsymbol{v}| = \sqrt{a^2 + b^2 + c^2}$. Unfortunately, this definition does not immediately lend itself to an intuitive understanding. However, we may consider the special case when $|\boldsymbol{\alpha}'(t)| = 1$ for all t. This is called a *unit-speed curve*. The curvature then becomes

$$k(t) = |\boldsymbol{\alpha}''(t)|$$

and we see that this is simply a measure of magnitude of the rate of change of the velocity vector of the curve. Returning to our picture of parametrized curves as paths of particles, we can see that if the velocity of the particle changes, it is either changing speed or direction. However, we have defined this particle to have unit speed, thus it must be changing direction. Therefore, a larger curvature corresponds to a more rapid change in direction, which causes the path to bend more sharply. This may seem like a contrived example, but it turns out that any regular curve can be reparametrized to have unit speed [10], so our understanding of this simple case applies in general. We can obtain a more explicit geometric understanding of curvature by first finding the curvature of a circle.

Example 6. We can parametrize a circle of radius r lying in the xy plane and centered at the origin by $\boldsymbol{\alpha}(t) = (r\cos t, r\sin t, 0)$. Thus we have

$$\begin{aligned}
\boldsymbol{\alpha}'(t) &= (-r\sin t, r\cos t, 0)\\
\boldsymbol{\alpha}''(t) &= (-r\cos t, -r\sin t, 0)\\
\boldsymbol{\alpha}'(t) \times \boldsymbol{\alpha}''(t) &= (0, 0, r^2\sin^2 t + r^2\cos^2 t)\\
&= (0, 0, r^2)\\
|\boldsymbol{\alpha}'(t)| &= \sqrt{r^2\sin^2 t + r^2\cos^2 t}\\
&= r\\
|\boldsymbol{\alpha}'(t) \times \boldsymbol{\alpha}''(t)| &= r^2
\end{aligned}$$

Plugging this into equation (42) gives us

$$k = \frac{r^2}{r^3} = \frac{1}{r}$$

Thus the curvature of a circle is simply the inverse of its radius.

This simple picture can be generalized to any curve with the idea of the *osculating circle*. If a curve $\boldsymbol{\alpha}$ has curvature $k(t_0)$ at some point $\boldsymbol{\alpha}(t_0)$, then the osculating circle is the circle tangent to $\boldsymbol{\alpha}(t_0)$ with radius $1/k(t_0)$. It is the circle that "fits in" the curve at that point. Thus at places where a curve does not bend very much and looks more like a straight line, the osculating circle is very large, thus the inverse of its radius is small, so that point on the curve has a small curvature. At places where a curve bends very sharply, a much smaller circle can fit in that bend and thus it has a very large curvature. To test this idea, we will look at an example of another conic section.

Example 7. Consider the parabola parametrized by $\boldsymbol{\alpha}(t) = (t, t^2, 0)$. We will show that, consistent with what we would expect from the idea of the osculating circle, the curvature of this parabola is largest at the origin and decreases as we approach positive or negative infinity. As before, we have

$$\begin{aligned}
\boldsymbol{\alpha}'(t) &= (1, 2t, 0)\\
\boldsymbol{\alpha}''(t) &= (0, 2, 0)\\
\boldsymbol{\alpha}'(t) \times \boldsymbol{\alpha}''(t) &= (0, 0, 2)\\
|\boldsymbol{\alpha}'(t)| &= \sqrt{1 + 4t^2}\\
|\boldsymbol{\alpha}'(t) \times \boldsymbol{\alpha}''(t)| &= 2\\
k(t) &= \frac{2}{(1 + 4t^2)^{3/2}}
\end{aligned}$$

Thus we see that $k(0) = 2$ is the largest value that the curvature can take on, and $\lim_{t \to \pm\infty} k(t) = 0$. In other words, farther from the origin, this parabola looks more like a straight line.

We can begin generalizing the idea of curvature from curves to surfaces with the following definition:

Definition 15. *The* normal curvature *of a surface M at a point P in a direction $\boldsymbol{u}$, denoted $\kappa(\boldsymbol{u})$), is plus or minus the curvature of the curve resulting from the intersection of M and the plane spanned by the unit normal vector and $\boldsymbol{u}$ at the point P. The sign is positive if the surface is curving towards $\boldsymbol{U}$ and negative otherwise.*

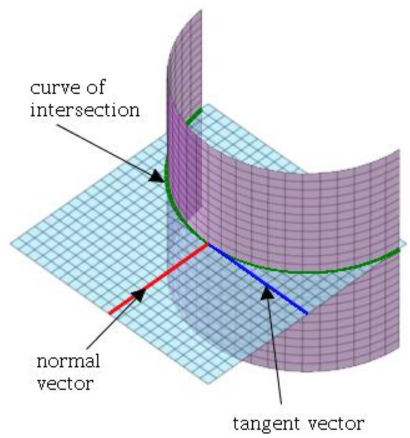


Figure 19: Finding the normal curvature of a point on a cylinder

An example of finding the normal curvature of a point on a cylinder is shown in Figure 19. In it, the red line is $\boldsymbol{U}$, blue is $\boldsymbol{u}$, the plane contains both $\boldsymbol{U}$ and $\boldsymbol{u}$, and the green curve which we will call $\boldsymbol{\alpha}$ is the intersection of the cylinder and plane. The value of κ at this point is thus the curvature of $\boldsymbol{\alpha}$, which we know is the inverse of the radius of the circle which "fits" the green curve. Conveniently, in this simple case $\boldsymbol{\alpha}$ is a circle, thus κ is simply one over the radius of the circular cross section of the cylinder. In general, the curves of intersection are often more complicated.

One of the most important but difficult to deal with components of the normal curvature is its sign. We said that κ is positive if $\boldsymbol{\alpha}$ is bending towards $\boldsymbol{U}$ at P and negative otherwise. For the situation in Figure 19, κ is negative. It is useful to understand what this sign means, but in practice, proving which case is occurring can be tedious. Fortunately, it turns out that we will not have to. Notice that in general, κ depends on our choice of $\boldsymbol{u}$. If we stand on a surface and face one direction, it may curve sharply upwards, but if we turn 30° it may bend gently downwards. We say that the *principle curvatures* are the maximum and minimum possible values of the normal curvature at a given point and the principle vectors are the choices of direction for which we obtain these extreme values. We denote the principle curvatures by κ_{max} and κ_{min} and the principle directions by $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$ respectively.

Example 8. Consider the parabolic cylinder parametrized by

$$\boldsymbol{x}(u, v) = (u, u^2, v) \tag{43}$$

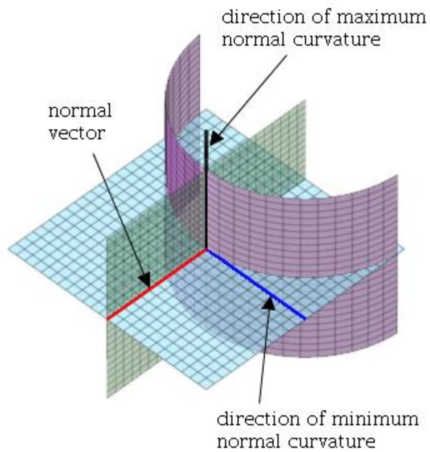


Figure 20: Finding the principle curvatures of a point on the cylinder in equation (43)

We want examine the normal curvatures of this surface at the origin. Two tangent vectors for M are $\boldsymbol{x}_u = (1, 2u, 0)$ and $\boldsymbol{x}_v = (0, 0, 1)$. At the origin, these become $\boldsymbol{x}_u(0) = (1, 0, 0)$ and $\boldsymbol{x}_v(0) = (0, 0, 1)$. We will find the normal curvature at the origin in these two tangent directions. To do so, we must first find the normal vector:

$$\begin{aligned}
\boldsymbol{U}(0, 0) &= \frac{\boldsymbol{x}_u(0) \times \boldsymbol{x}_v(0)}{|\boldsymbol{x}_u(0) \times \boldsymbol{x}_v(0)|} \\
&= \frac{(0, -1, 0)}{1} \\
&= (0, -1, 0)
\end{aligned}$$

Notice that the plane containing $\boldsymbol{U}$ and $\boldsymbol{x}_u(0)$ is the xy plane, and thus the curve of intersection is a parabola. We already know from our calculations above that the curvature of a parabola parametrized by $\boldsymbol{\alpha}(t) = (t, t^2, 0)$ at the origin is 2. However, note that our surface is bending away from $\boldsymbol{U}$, thus $\kappa(1, 0, 0) = -2$. Similarly, the plane spanned by $\boldsymbol{U}$ and $\boldsymbol{x}_v(0)$ is the yz plane, and the curve of intersection is a line. We have stated previously that the curvature of a line is 0, but we will briefly prove it now. A line can be parametrized by $\boldsymbol{\beta}(t) = (at, bt, ct)$ for some constants a, b, and c. Thus $\boldsymbol{\beta}'(t) = (a, b, c)$ and $\boldsymbol{\beta}''(t) = (0, 0, 0)$. Therefore the curvature is given by

$$\kappa(0, 0, 1) = \frac{0}{(a^2 + b^2 + c^2)^{3/2}} = 0$$

Note that because the cylinder is bending away from $\boldsymbol{U}$ in every direction other than along this line, $\kappa = 0$ is the maximum normal curvature at the origin.

5.3.1 Gauss Curvature of Quadric Surfaces

With the tools we have developed, the definition of the Gauss curvature is simple, although not immediately intuitive:

Definition 16. *The* Gauss curvature *of a surface at a given point is the product of the principal curvatures of the surface at that point.*

We will denote Gauss curvature by K_G. Sometimes we will write $K_G(P)$ for some point P on the surface to indicate that the Gauss curvature depends on the location on the surface. Usually, however, the location we are discussing will be clear from context.

Example 9. Recall the parabolic cylinder discussed in the previous section. We found that the maximum normal curvature was $\kappa_{max} = 0$. Thus the Gauss curvature at this point is given by

$$K_G = \kappa_{min} \cdot 0 = 0$$

We did not have to find the the minimum normal curvature to do this computation, since no matter what its value is, we will have $K_G = 0$.

Note that K_G can be positive, negative or zero depending on the sign of the principal curvatures. We think of zero as a third type of sign rather than talking about zero being positive or negative. The sign of K_G has important geometric significance and will be the subject of the rest of this paper. There are three cases:

- If $K_G > 0$, both principle curvatures are positive, thus the surface is either bending towards the normal vector $\boldsymbol{U}$ in every direction or bending away in every direction. An example of positive Gauss curvature is the point on the sphere in Figure 21.

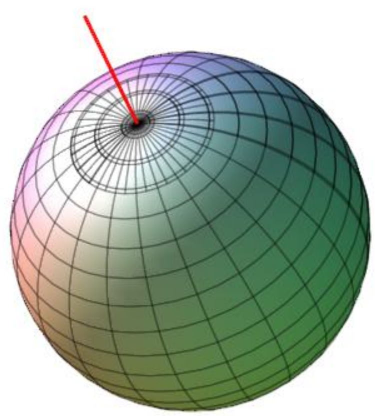


Figure 21: A point (indicated by the red line) with positive Gauss curvature

- If $K_G = 0$, at least one principal curvature has a value of 0, thus in that direction the surface is neither bending towards nor away from $\boldsymbol{U}$; it is flat. We saw that this was the case for the point at the origin for the parabolic cylinder in Figure 20.

- If $K_G < 0$, one principal curvature is negative and the other is positive, thus in one direction the surface is bending away from $\boldsymbol{U}$ and in another it is bending towards $\boldsymbol{U}$. This would be the case for the point on the hyperbolic paraboloid in Figure 22.

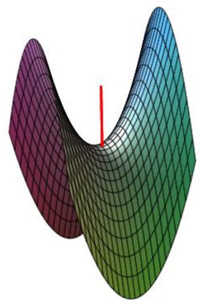


Figure 22: A point (indicated by the red line) with negative Gauss curvature

In more familiar terms, a point with positive Gauss curvature looks like a bowl, a point with negative Gauss curvature looks like a plane, and a point with negative Gauss curvature looks like a saddle. Using these ideas, we can classify points on a surface as follows:

1. If $K_G(P) > 0$ then P is an elliptic point.
2. If $K_G(P) < 0$ then P is a hyperbolic point (or saddle point).
3. If $K_G(P) = 0$ and exactly one of the principle curvatures is nonzero then P is a parabolic point.

We will refer to this classification of a point on a surface as the *curvature type* of the point.

Before we present our primary result, we will briefly discuss Gauss's well known *Theorem Egregium*, which we state here without proof.

Theorem 5 (Theorem Egregium)**.** *The Gauss curvature of a surface is invariant under local isometries.*

We do not want to get too bogged down in details, but we will give a sufficient definition of a local isometry from [11]:

Definition 17. *A local isometry between two surfaces M_1 and M_2 is a smooth (ie differentiable of all orders) map $f : M_1 \to M_2$ such that for any curve $\boldsymbol{\alpha} \in M_1$, its image curve $f \circ \boldsymbol{\alpha} \in M_2$ has the same length as $\boldsymbol{\alpha}$.*

In other words, a local isometry is essentially a map which preserves distances. Two simple cases are of interest to us: rotations and translations. From the Theorem Egregium and the definition of a local isometry, we know that the Gauss curvature of a surface is invariant under rigid rotations and translations of that surface. This seems reasonable since we would suspect that these operations do not deform the surface or affect how it bends. Another important consequence of this theorem is that the Gauss curvature can be calculated directly from equations defining the surface, rather than having to go through calculating two different normal curvatures first. The definition of Gauss curvature as the product of principle curvatures provides an important geometric picture of what the Gauss curvature means, but in practice it is usually calculated by other means, as we will see shortly.

We are now equipped to make the following claim:

Theorem 6. *Every point on a quadric surface has the same curvature type.*

Before we prove this theorem, motivation might be helpful. In particular, we will consider an example of a surface which does not have the property that all of its points have the same curvature type. There are countless examples, one of them being the torus. There are three points indicated by their normal vectors in Figure 23. Red corresponds to an elliptic point, blue to a hyperbolic point, and black to a parabolic point. The green curves are simple the parameter curves corresponding to the principle directions of the three points to help see why they have the indicated curvature types. For instance, the black vector sits on one circle bending away from it, thus corresponding to negative normal curvature, and on one circle which lies in a plane perpendicular to the normal, thus corresponding to zero normal curvature (it is neither bending away nor towards the normal). Since exactly one of the principle curvatures at this point is zero, the black point is parabolic. Similar arguments explain the classification of the other two points. This example shows why Theorem 6 is

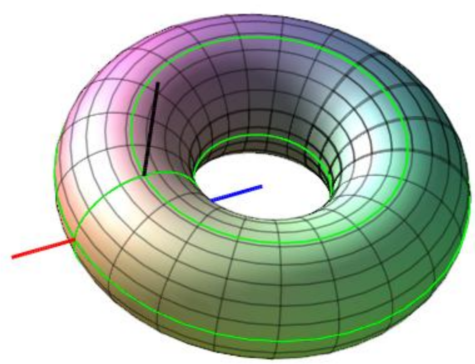


Figure 23: Torus containing points with all three curvature types (red = elliptic, black = parabolic, blue = hyperbolic)

non-trivial. Unlike many other surfaces we encounter, classifying the curvature types of all points on a quadric surface merely requires that we classify one point. We will now prove this.

Proof. Consider an implicit surface given by $f(x, y, z) = 0$ with Gauss curvature $K_G(x, y, z)$. We will use the following important formula from [12] without proof:

$$K_G = \frac{\begin{vmatrix} D(D(f)) & D(f)^T \\ D(f) & 0 \end{vmatrix}}{|D(f)|^4} \tag{44}$$

where D is the derivative operator and $D(f)^T$ is the transpose of the derivative of f. Notice that, as indicated earlier, we have sidestepped the issue of principle curvatures entirely.

Recall that a quadric surface is implicitly defined by

$$f(x,y,z) = Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + Gx + Hy + Iz + J = 0 \tag{45}$$

Thus we have

$$\begin{aligned}
\nabla f &= (f_x, f_y, f_z) \\
&= (2Ax + Dy + Ez + G, 2By + Dx + Fz + H, 2Cz + Ex + Fy + I) \\
D(D(f)) &= D(\nabla f) \\
&= \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix} \\
&= \begin{bmatrix} 2A & D & E \\ D & 2B & F \\ E & F & 2C \end{bmatrix}
\end{aligned}$$

Plugging these values into equation (44), we get

$$K_G = \frac{\begin{vmatrix} D(D(f)) & \nabla f^T \\ \nabla f & 0 \end{vmatrix}}{|\nabla f|^4}$$

$$= \frac{\begin{vmatrix} 2A & D & E & f_x \\ D & 2B & F & f_y \\ E & F & 2C & f_z \\ f_x & f_y & f_z & 0 \end{vmatrix}}{((2Ax + Dy + Ez + G)^2 + (2By + Dx + Fz + H)^2 + (2Cz + Ex + Fy + I)^2)^{4/2}} \tag{46}$$

(47)

We will use cofactor expansion on the last row of the numerator to evaluate the determinant:

$$\begin{aligned}
\begin{vmatrix} 2A & D & E & f_x \\ D & 2B & F & f_y \\ E & F & 2C & f_z \\ f_x & f_y & f_z & 0 \end{vmatrix} &= -f_x \begin{vmatrix} D & E & f_x \\ 2B & F & f_y \\ F & 2C & f_z \end{vmatrix} + f_y \begin{vmatrix} 2A & E & f_x \\ D & F & f_y \\ E & 2C & f_z \end{vmatrix} - f_z \begin{vmatrix} 2A & D & f_x \\ D & 2B & f_y \\ E & F & f_z \end{vmatrix} + 0 \\
&= -f_x(D(Ff_z - 2Cf_y) - E(2Bf_z - Ff_y) + f_x(4BC - F^2)) \\
&\quad + f_y(2A(Ff_z - 2Cf_y) - E(Df_z - Ef_y) + f_x(2DC - EF)) \\
&\quad - f_z(2A(2Bf_z - Ff_y) - D(Df_z - Ef_y) + f_x(DF - 2EB))
\end{aligned}$$

Replacing f_x, f_y, and f_z with their expressions and substituting this back into equation (47) gives us the following formula for the Gauss curvature of a quadric surface:

$$\begin{aligned}K_G = [&-(2Ax+Dy+Ez+G)(D(F(2Cz+Ex+Fy+I)-2C(2By+Dx+Fz+H))\\&-E(2B(2Cz+Ex+Fy+I)-F(2By+Dx+Fz+H))+(2Ax+Dy+Ez+G)(4BC-F^2))\\&+(2By+Dx+Fz+H)(2A(F(2Cz+Ex+Fy+I)-2C(2By+Dx+Fz+H))\\&-E(D(2Cz+Ex+Fy+I)-E(2By+Dx+Fz+H))+(2Ax+Dy+Ez+G)(2DC-EF))\\&-(2Cz+Ex+Fy+I)(2A(2B(2Cz+Ex+Fy+I)-F(2By+Dx+Fz+H))\\&-D(D(2Cz+Ex+Fy+I)-E(2By+Dx+Fz+H))+(2Ax+Dy+Ez+G)(DF-2EB))]\\&/((2Ax+Dy+Ez+G)^2+(2By+Dx+Fz+H)^2+(2Cz+Ex+Fy+I)^2)^2)\end{aligned} \tag{48}$$

Although it is a useful formula to have for a computer, manipulating this equation would be horrifying. We must make some simplifications. Recall that we have previously stated that, just as in the case of the conic sections, the cross terms in equation (45) can be removed by an appropriate rotation. Furthermore, the Theorem Egregium tells us that Gauss curvature is invariant under rotations of the surface. In other words, we can choose an appropriate rotation to make $D = E = F = 0$ *without changing the Gauss curvature*. This gives us

$$f(x, y, z) = A'x^2 + B'y^2 + C'z^2 + G'x + H'y + I'z + J' = 0 \tag{49}$$

for some new constants $A', B', C', G', H', I', J'$. For the sake of clarity, we will omit the primes, but it is important to realize that the constants A through J after the rotation are not necessarily the same as they were before. Applying this to equation (48) then gives us

$$\begin{aligned}K_G &= \frac{-f_x(0-0+4f_xBC)+f_y(-4ACf_y-0+0)-f_z(4ABf_z-0)-0+0)}{(f_x^2+f_y^2+f_z^2)^2}\\&= \frac{-4(BCf_x^2+ACf_y^2+ABf_z^2)}{(f_x^2+f_y^2+f_z^2)^2}\\&= \frac{-4(BC(2Ax+G)^2+AC(2By+H)^2+AB(2Cz+I)^2)}{(f_x^2+f_y^2+f_z^2)^2}\end{aligned}$$

Notice that the sign of the denominator is always positive. Furthermore, for a regular surface, the denominator cannot be zero. This will not be much of a hindrance; one of the only non-regular points on a quadric surface which we must avoid is the tip of a cone. Thus we must only prove that if the numerator–which we will denote $N(K_G)$–is positive, negative, or zero at some point on the surface, then it has the same type of value at every point on the surface. There are three cases:

1. If two or more of A, B, or C are zero, then the numerator goes to zero, thus for every point on M, $K_G = 0$.

2. If exactly one of A, B, or C is zero, then two of the terms in the numerator go to zero. Without loss of generality, choose $A = 0$. Then the numerator becomes

   $$N(K_G) = -4BC(0 \cdot x + G)^2 = -4BCG^2$$

   Since this is only a function of constants, clearly the sign cannot change.

3. If none of A, B, or C are zero, then we can complete the square for three terms in equation (49) and then translate the surface to remove the degree one terms. We know from the Theorem Egregium that this does not change the Gauss curvature. This gives us

$$f(x,y,z) = A''x^2 + B''y^2 + C''z^2 + J'' = 0$$

Again, we will drop the primes and simply recall that the constants may have changed, thus

$$f(x,y,z) = Ax^2 + By^2 + Cz^2 = -J$$

The numerator of the Gauss curvature then becomes

$$\begin{aligned} N(K_G) &= -16(A^2BCx^2 + B^2ACy^2 + C^2ABz^2) \\ &= -16ABC(Ax^2 + By^2 + Cz^2) \\ &= -16ABC(-J) \\ &= 16ABCJ \end{aligned}$$

Once again, the sign of the Gauss curvature depends only on constants, thus it does not change.

This completes the proof. □

We should note that this does not necessarily mean that the Gauss curvature itself is constant on a quadric surface. In the case where $K_G = 0$ this is true, but if $K_G \neq 0$ it may be the case that the magnitude of the denominator of K_G in equation (47) changes even though the overall sign does not. As an example, consider the ellipsoid shown in Figure 24. The point indicated by

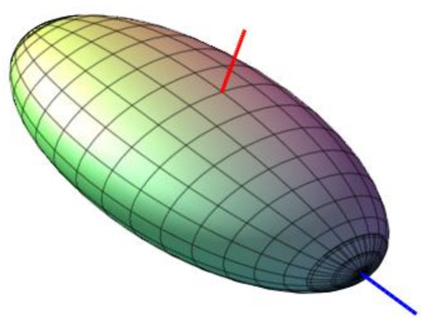


Figure 24: Two points on an ellipsoid with different values of positive Gauss curvature

the blue normal vector has a Gauss curvature which is larger than the point indicated by the red normal vector, although both are positive.

6 Conclusion

We have developed and applied a great deal of machinery to the problem of generalizing conic sections to higher dimensions. Our introductory discussion of conic sections not only gave us important algebraic and geometric definitions to work with, it saw the complete classification of conic sections from basic algebraic properties, pointing the way for our work with generalized conic sections, and both an intuitive and formal definition of the important property of eccentricity. With this setup, we explored two possible generalizations of conic sections into three dimensions. The algebraic generalization introduced the quadric surfaces and a means of classifying them. The geometric generalization was somewhat more difficult, requiring us to work with objects that we cannot directly visualize and ultimately leaving us with a conjecture rather than an answer to the question of whether all quadric surfaces are conic surfaces. This would be an excellent area for further exploration, likely requiring the use of more powerful tools from linear algebra. Finally, we proposed and proved the theorem that all quadric surfaces have a single type of Gauss curvature. Due to its originality and relative depth, this is the most significant part of this paper.

Apart from proving or disproving the conjecture on the parameter spaces of conic surfaces, there are many possibilities for continuing the project. First is extending the questions of this paper to arbitrary dimensions. Second is considering ways of generalizing the conic sections other than dimensionality, such as considering conic sections in the complex plane or the intersections of quadric surfaces other than planes and cones. Finally, there are the many questions raised by the brief discussion of hypercones in the context of general relativity. In particular, what can our exploration of the sections of hypercones tell us about the paths of light in space-time?

With the results we have developed and the new questions raised, it should be clear that the conic sections remain now as powerful and interesting as they were to Apollonius of Perga.

References

[1] J.J. O'Connor, E.F. Robertson. *Apollonius of Perga*. MacTutor History of Mathematics. http://www-history.mcs.st-and.ac.uk/Mathematicians/Apollonius.html. January 1999.

[2] David Arnold. *The Conic Sections in Polar Coordinates*. http://online.redwoods.cc.ca.us/instruct/darnold/MULTCALC/PolarConics/conics.htm January 26, 1999.

[3] Howard Anton, Irl Bivens, Stephen Davis. *Calculus: Early Transcendentals Combined*. Somerset, New Jersey, 8th Edition, 2005.

[4] James Stewart. *Essential Calculus: Early Transcendentals*. http://www.stewartcalculus.com /data/ESSENTIAL%20CALCULUS%20Early%20Transcendentals/upfiles/ess-reviewofconics.pdf

[5] D. C. Lay, *Linear Algebra and its Applications*, 3rd ed., Addison-Wesley, 2006.

[6] Beyer, W. H. *CRC Standard Mathematical Tables*, 28th ed. Boca Raton, FL: CRC Press, pp. 210-211, 1987.

[7] John M. H. Olmsted. *Matrices and Quadric Surfaces*. National Mathematics Magazine, Vol. 19, No. 6 (Mar., 1945), pp. 267-275. http://www.jstor.org/stable/3030043

[8] http://physics.syr.edu/courses/modules/LIGHTCONE/minkowski.html

[9] http://en.wikipedia.org/wiki/File:World_line.svg#file

[10] John Oprea. *Differential Geometry and Its Applications*. 2nd Ed. The Mathematical Association of America. 2007.

[11] Jim Belk. *Pullbacks, Isometries, and Conformal Maps*. http://math.bard.edu/belk/math352/Outline%20-%20Pullbacks%20and%20Isometries%20Revised.pdf. Bard College. 2011.

[12] Michael Spivak. *Calculus on Manifolds*. W. A. Benjamin, Inc. 1965.