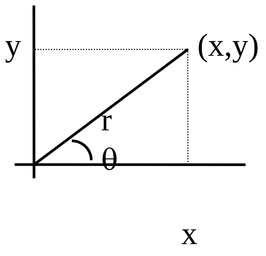
Essential Mathematics for Computer Graphics

- Trigonometry
- Polar Coordinates
- 3-D Coordinate Systems
- Parametric Representations
- Points and Vectors
- Matrices
- Analytic Geometry

This handout provides a compendium of the math that we will be using this quarter. It does not provide a complete presentation of the topics, but instead just gives the “facts” that we will need. We will not cover most of this in class, so you may want to keep this nearby as a handy reference.

Trigonometry

The trigonometric functions of a general angle: Let θ be an angle with its vertex at the origin and its initial side coincident with the x-axis, and let (x,y) be any point in the Cartesian plane distinct from the origin, and on the terminal side of the angle. The six trigonometric functions are as follows:



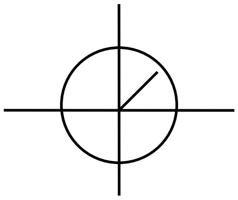
$\sin\theta = y / r$	$\cos\theta = x / r$	$\tan\theta = y / x$
$\csc\theta = r / y$	$\sec\theta = r / x$	$\cot\theta = x / y$

Angles can be specified in degrees or radians. One radian is defined as the measure of the central angle subtended by an arc of a circle equal to the radius of the circle.

$1 \text{ radian} = 180/\pi = 57.296°$ $\qquad 1° = \pi/180 \text{ radian} = 0.017453 \text{ radian}$

Polar Coordinates

We are used to the standard Cartesian coordinate system of graphing, but there are many other methods. One common, and very useful method is the polar coordinate method. In this system, the origin is at the center of the plane and one specifies coordinates by giving an angle θ and a radius r:



If you need to translate polar coordinates to Cartesian coordinates, use these formulas:

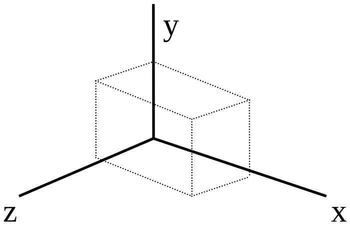
$x = r \cos \theta$ $\qquad$ $y = r \sin \theta$

Going from Cartesian to polar:

$r = \mathrm{sqrt}(x^2 + y^2)$ $\qquad$ $\theta = \tan^{-1}(y/x)$

3-D Coordinate Systems

In a 3-dim Cartesian system, we specify 3 values to identify a point (x,y,z):



This is a right-handed system, which is most often used in graphics, i.e., the thumb of the right hand points down the z-axis if we imagine grabbing the z-axis with the fingers of the right hand curling from the positive x-axis toward the positive y-axis.

Parametric Representations

There are two ways to visualize a curve: as a line "frozen" in space or as a path of a particle as it moves along the curve. The first view leads us to describe the curve according to a function that defines the x- and y-coordinates of the curve. This produces an *implicit form* as in the equation $F(x,y) = 0$. For example, the equation for a circle is $x^2 + y^2 - r^2 = 0$. Or, the function might be in *explicit form* which specifies the y-coordinate for each value of x: $y = f(x)$. For example, the equation for a line is $y = mx + b$.

A *parametric form* is the other way to view a curve. It suggests the movement of a point through time. The *parameter* is the value t (time) and is used to distinguish one point on the curve from another. The path of the particle traveling along the curve is fixed by two functions: x() and y() and we speak of (x(t), y(t)) as the position of the particle at time t. The curve is all the points visited by the particle over some interval, e.g., from 0 to 1.

Lines, line segments and rays all share the same parametric representation. (A line is infinite in length and is defined by two points through which it passes; a line segment is defined by two endpoints extending only from one endpoint to another; a ray starts at an endpoint, passes through another point and is infinite in one direction.) All points on a line segment from a = (ax, ay) to b = (bx, by) are represented by the parametric form:

$$x(t) = ax + (bx - ax)t$$
$$y(t) = ay + (by - ay)t$$

as t varies from 0 to 1. When $t = 0$, the point (x(t), y(t)) is at point a. As t increases toward 1, the point moves in a straight line toward b. It is midway between the two points when $t = 1/2$, and in general, it is a fraction f of the way from a to b at $t = f$.

The ray that starts at a and passes through b is also defined by these equations but t is allowed to take on any nonnegative value. The ray passes through b at $t = 1$ but then continues forever along the same path. The direction of the ray (or line segment) is given by:

$$\text{slope} = (by - ay) / (bx - ax)$$

The line defined by points a and b is also defined by these equations but now all real values of t are permitted. Thus, line segments, rays and lines differ parametrically only in the values of t that are relevant:

line segment: $0 \leq t \leq 1$
ray: $0 \leq t < \infty$
line: $-\infty < t < \infty$

A circle with radius R centered at the origin of the Cartesian plane has the following parametric representation:

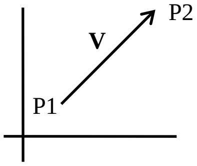
$$x(t) = R * \cos(2\pi t)$$
$$y(t) = R * \sin(2\pi t)$$

for t between 0 and 1. If you try graphing this with equidistant values of t, you'll end up with a circle.

Points and Vectors

Basics

A *point* is a position specified with coordinate values in some reference frame. A *vector* is the difference between two point positions:



$$\begin{aligned}\mathbf{V} &= P2 - P1\\ &= (x2 - x1, y2 - y1)\\ &= (Vx, Vy)\end{aligned}$$

(Vx, Vy) are called the Cartesian *components* or *elements*, and are the projections of **V** onto the x and y axes. We can describe a vector as a directed line segment (starting an the *initial* point and ending with the *terminal point*) that has two properties: magnitude (length) and direction. To calculate magnitude, we use a variation of the Pythagorean Theorem:

$$|\ \mathbf{V}\ | = \mathrm{sqrt}(Vx^2 + Vy^2)$$

The direction is given in terms of the angular displacement with the x axis:

$$\alpha = \tan^{-1} (Vy / Vx)$$

In a 3-dim Cartesian space, magnitude is defined by:

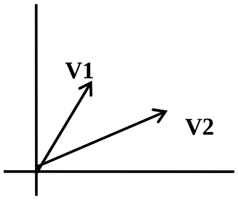
$$|\ \mathbf{V}\ | = \mathrm{sqrt}(Vx^2 + Vy^2 + Vz^2)$$

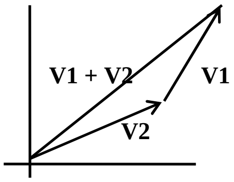
Vector direction in 3-D is given with direction angles, α (x-axis), β (y-axis), and γ (z-axis):

$$\begin{aligned}\cos\ \alpha &= Vx / |\ \mathbf{V}\ |\\ \cos\ \beta &= Vy / |\ \mathbf{V}\ |\\ \cos\ \gamma &= Vz / |\ \mathbf{V}\ |\end{aligned}$$

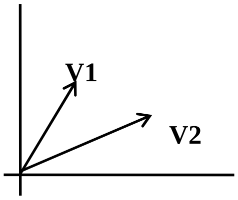
Addition and Subtraction of Vectors

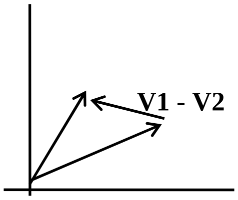
Sum of two vectors: **V1** + **V2** = (V1x + V2x, V1y + V2y, V1z + V2z). Geometrically, the sum of two vectors means placing the start position of one vector at the tip of the other and then drawing the resulting summation vector (or view **V1** and **V2** in the first diagram as two adjacent sides of a parallelogram; **V1** + **V2** = the diagonal of the parallelogram).





Subtraction is defined as the other diagonal of the parallelogram formed by the two vectors:





Scalar Multiplication

Scalar multiplication of a vector and a constant: a**V** = (aVx, aVy, aVz). For example, if a = 2, each component is doubled. If we scale a vector with a negative scalar, we reverse its direction.

It is often convenient to scale a vector so that the result has a length equal to 1. This is called *normalizing* the vector and the result is known as the *unit vector*. To form the normalized versions $\mathbf{u}_{\mathbf{a}}$ of a vector **a**, we scale the vector **a** with the value 1 / |**a**|:

$$\mathbf{u}_{\mathbf{a}} = \mathbf{a} / |\mathbf{a}| \qquad \text{giving} \qquad |\, \mathbf{u}_{\mathbf{a}} \,| = 1$$

We sometimes refer to a unit vector as a *direction* since its magnitude = 1. Note that any vector can be written as its magnitude times its direction, thus if $\mathbf{u}_{\mathbf{a}}$ is a normalized version of **a** then: $\mathbf{a} = |\mathbf{a}|\, \mathbf{u}_{\mathbf{a}}$

Standard Unit Vectors

The special vectors **i**, **j** (and **k** in 3-d) are called *standard unit vectors* and are defined as follows:

$$\mathbf{i} = (1,0,0) \qquad \mathbf{j} = (0,1,0) \qquad \mathbf{k} = (0,0,1)$$

These vectors allow us to obtain an alternate way of denoting vectors in 2-D or 3-D. In 2-D:

$$\text{if } \mathbf{a} = (a1, a2) \text{ then } \mathbf{a} = (a1,0) + (0, a2) = a1(1,0) + a2(0,1) = a1\mathbf{i} + a2\mathbf{j}$$

This final sum is called a *linear combination* of **i** and **j**.

Linear Combinations

Now that we know how to add and scale vectors, it is useful to define more about linear combinations of m vectors **v1**, **v2**, ..., **vm** which is a vector of the form:

$$\mathbf{w} = a1\mathbf{v1} + a2\mathbf{v2} + \ldots + am\mathbf{vm}$$

where a1, a2, ..., am are scalars. For example, the linear combination of 2(3,4) + 3/2(2,4) gives the vector (9,14). Thus, all we are doing is scaling or “weighting” them by some values and then adding the weighted versions.

A special class of linear combinations has an important place in graphics: *the convex combinations*. These are linear combinations for which the coefficients are nonnegative and add up to one. $\mathbf{w} = a1\mathbf{v1} + a2\mathbf{v2} + \ldots + am\mathbf{vm}$ will be a convex combination if all the nonnegative scalars ai add up to 1, and each ai >= 0. These coefficients are said to make a partition of unity meaning that a unit amount of material is partitioned into pieces.

A very important convex combination is

$$\mathbf{p}(t) = \mathbf{a}(1-t) + \mathbf{b}t$$

in which **a** and **b** are vectors and t is between 0 and 1 (inclusive). The scalars are (1-t) and t which clearly add up to 1. We call **p**(t) a *vector function* of t meaning that it defines a vector whose length and direction vary with t. At t = 0, **p**(t) = vector **a**; at t = 1, **p**(t) = vector **b**. If t increases beyond 1, the ray is still defined, but we no longer call this a convex combination. This convex combination is often written in the form:

$$\mathbf{p}(t) = \mathbf{a} + (\mathbf{b} - \mathbf{a})t$$
$$\mathbf{p}(t) = \mathbf{a} + \mathbf{c}t \quad \text{where } \mathbf{c} = \mathbf{b} - \mathbf{a}$$

This is still a convex combination if t varies between 0 and 1, but it is not so obvious because of how we re-arranged things. Now t appears as a scalar on the vector **c**, and the term “**c**t” acts as an offset or displacement on the vector **a**. As t increases, an increasing amount of offset **c** is added to **a**.

Multiplication of Vectors: Dot Product

There are two ways to perform multiplication of vectors, one resulting in another vector, the other a scalar value. To get the scalar value

$$\mathbf{V}1 \bullet \mathbf{V}2 = |\,\mathbf{V}1\,|\,|\,\mathbf{V}2\,| \cos\theta \quad \text{where } 0 \le \theta \le \pi \text{ and } \theta \text{ is the angle between the two vectors}$$

This is the *scalar* or *dot* or *inner product* of two vectors. We can also calculate dot product this way:

$$\mathbf{V}1 \bullet \mathbf{V}2 = V1xV2x + V1yV2y \ (+\ V1zV2z)$$

What does this single resulting value mean? The dot product is often said to measure the “match” or “similarity” between two vectors. Imagine two vectors moving in space like the hands of a clock. If we hold their lengths constant, we see by the first equation above that the dot product is proportional to the cosine of the angle between them. For example, if the angle = 0, the cosine is at its max and the dot product is also at its max. As the two vectors move farther apart, the cosine decreases and the dot product decreases. It is its most negative at 180° when the vectors point in opposite directions. Thus, the closer the vectors are, the larger the dot product; the more they point in the opposite direction, the more negative their dot product. This can be summarized as follows:

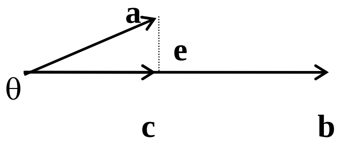
The angle between two vectors is:

a) less than 90° if $\mathbf{a} \bullet \mathbf{b} > 0$
b) exactly 90° if $\mathbf{a} \bullet \mathbf{b} = 0$
c) greater than 90° **if** $\mathbf{a} \bullet \mathbf{b} < 0$

Dot products have the following properties:

a) symmetry: $\mathbf{a} \bullet \mathbf{b} = \mathbf{b} \bullet \mathbf{a}$
b) linearity: $(\mathbf{a} + \mathbf{c}) \bullet \mathbf{b} = \mathbf{a} \bullet \mathbf{b} + \mathbf{c} \bullet \mathbf{b}$
c) homogeneity: $(s\mathbf{a}) \bullet \mathbf{b} = s(\mathbf{a} \bullet \mathbf{b})$ where s is a scalar.
d) $|\mathbf{b}|^2 = \mathbf{b} \bullet \mathbf{b}$

A useful application of the dot product arises when we discuss *resolving* a vector into two components: one component in the direction of a second given vector, and the other that is perpendicular to the second vector.



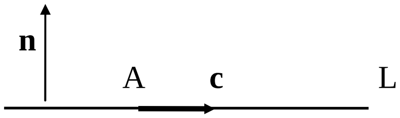
Vector **a** is shown here resolved into a first component **c** along the vector **b**, and a second component **e** that is perpendicular to **b**. By vector addition, we see that **e** = **a** - **c** which is basically what the idea of resolving a vector is all about. The vector **c** is called *the perpendicular projection* of **a** onto **b**; it has the same direction as **b**, but a different length. We know that |**c**| = |**a**| cos θ from basic trig. Using this equation and the definition for dot product **a** • **b** = |**a**| |**b**| cos θ, we get:

$$
\begin{aligned}
|\mathbf{c}| &= |\mathbf{a}| \left[(\mathbf{a} \cdot \mathbf{b}) / (|\mathbf{a}|\,|\mathbf{b}|) \right] \\
&= (\mathbf{a} \cdot \mathbf{b}) / |\mathbf{b}| \\
&= \mathbf{a} \cdot \mathbf{u}_b \qquad \text{(remember: } \mathbf{b} / |\mathbf{b}| = \mathbf{u}_b\text{)}
\end{aligned}
$$

Thus, the length of **c** depends on the length of **a** (which we would expect), but not the length of **b**. To form the actual vector **c**, we can attach the direction of **b** to it: $\mathbf{c} = |\mathbf{c}|\,\mathbf{u}_b$. Using the equation above, we get $\mathbf{c} = (\mathbf{a} \cdot \mathbf{u}_b)\,\mathbf{u}_b$.

Point-Normal Form

Consider a line L which passes through point A = (Ax, Ay) in direction **c** = (cx, cy).



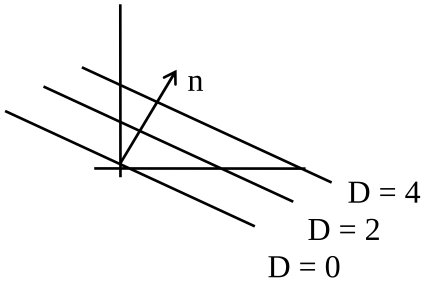
Sometimes it is useful to find the *normal* direction for line L (**n** = (nx, ny)), which is in the perpendicular direction. We know that the dot product of two vectors perpendicular to one another = 0, so **c** • **n** = 0.

To obtain an equation for line L, consider some arbitrary point R = (x,y) on L. The vector R - A (the difference of two points is a vector) must be perpendicular to **n** and so **n** • (R - A) = 0. We might be tempted to do something like: **n** • R = **n** • A but we really can't form dot products between a vector and a point.

To transform this equation we need to find a way of representing R and A as vectors not points. Try this: replace R with the vector **r** whose initial point is at the origin; replace A with the vector **a** whose initial point is the origin also. This makes everything we do dependent on our choice of origin but the equation of any line is dependent on where the origin is, so there really is no loss of generality. With this replacement, we can simply write:

$$\mathbf{n} \bullet \mathbf{r} = D \text{ where } D = \mathbf{n} \bullet \mathbf{a} = nx * Ax + ny * Ay$$

This is the *point-normal form* for a line. This equation shows that all points on a straight line, i.e., all vectors from the origin to those points on the line, share the same dot product value with the normal direction, i.e., all these vectors share the same projection onto the normal **n**. The value D reports the “position” of the line, meaning D is altered by shifting the line to a position parallel to itself.



In this diagram, a set of lines **n** • R = D having the same normal (**n** = (1,2)) is shown. As D increases, the lines shift in parallel in the direction of **n**.

Point-Normal Form for a Plane

Planes can also be represented in point normal form. A plane is completely specified by giving a single point S = (sx, sy, sz) that lies within it, and the normal direction to the plane. Just as the normal vector orients a line in 2-D, the normal to a plane orients it in space. This normal, **n** = (nx, ny, nz), is perpendicular to any line lying in the plane. For any arbitrary point in the plane R = (x, y, z) form the vector from R to S which must be perpendicular to **n**: **n** • (R - S) = 0. Replace R and S with **r** and **s** as we did above (these are vectors from the origin to R and S respectively) and use linearity (see properties of dot product above) to get: **n** • **r** = D where D = **n** • **s**. This is the point normal form for a plane. It’s the same as the one we defined for a line but it operates on 3-D vectors not 2-D. The basic premise, however, is the same: All points in a plane have the same dot product with the normal, i.e., they all have the same projection onto **n**.

Recall that the equation of a plane is traditionally written as Ax + By + Cz = D where the coefficients A, B, C, D distinguish one plane from another. By calculating out the dot product using the components of the vectors in the equation **n** • **r** = D, we see that the point normal form is actually nx*x + ny*y + nz*z = D. Thus, (A,B,C) of the traditional equation gives the normal direction to the plane.

Multiplication of vectors: Cross Product

The other way of multiplying vectors gives another vector as a result:

$\mathbf{V}1 \times \mathbf{V}2 = \mathbf{u}\,|\,\mathbf{V}1\,|\,|\,\mathbf{V}2\,|\sin\theta$ where $0 \leq \theta \leq \pi$ and θ is the angle between the two vectors

This is the *vector* or *cross product* of two vectors. **u** is a *unit vector* of magnitude 1 that is perpendicular to both **V**1 and **V**2.

We can also calculate cross product this way:

$$\mathbf{V}1 \times \mathbf{V}2 = (V1yV2z - V1zV2y, V1zV2x - V1xV2z, V1xV2y - V1yV2x)$$

The cross product of two vectors is a vector that is perpendicular to the plane defined by the two vectors and with magnitude equal to the area of the parallelogram formed by the two vectors.

Finding the Normal to a Plane

The classic application of the cross product is to find the normal vector to a plane. The equation **n** • **r** = D completely specifies a plane once its normal and parameter D are known. But if we only know that the plane passes through three specific points, how can **n** and D be computed?

We know from geometry that three points specify a plane as long as they do not lie in a straight line (i.e., they are *noncollinear*). To find the normal vector, build two vectors **a** = P2 - P1 and **b** = P3 - P1. Their cross product must be normal to every line in the plane and is thus the desired vector **n**. Any scalar multiple of this cross product is also a normal (its just at a different point in the plane). **b** x **a** is also a normal but it points in the opposite direction. For whichever choice of **n** that is made, use one of the three points P1, P2 or P3 as the **r** element in **n** • **r** = D to compute D. This then gives the point normal form for the plane.

Matrices

A *matrix* is a rectangular array of quantities, called the elements of a matrix. We identify matrices according to the number of rows X number of columns.

2 X 3 matrix $\begin{vmatrix} 1 & 3 & 8 \\ 4 & 2 & 9 \end{vmatrix}$

When number of rows = number of columns, we have a *square* matrix. A matrix with a single row or column represents a vector, and a larger matrix can be viewed as a collection of row or column vectors.

$$\mathbf{V} = [V_x \quad V_y \quad V_z]$$

To multiply a matrix **A** by a scalar value s, we just multiply each element of the matrix by s. Matrix addition is defined only for matrices with the same number of rows and columns and is simply the sum of corresponding elements.

We can multiply an m X n matrix **A** by a p X q matrix **B** only if n = p. We obtain an m X q matrix **C** whose elements are calculated as in the following example:

$$\begin{vmatrix} 2 & 3 \\ 4 & 5 \\ 1 & 4 \end{vmatrix} \quad X \quad \begin{vmatrix} 7 & 8 \\ 9 & 3 \end{vmatrix} \quad =$$

$$\begin{vmatrix} 2*7 + 3*9 & 2*8 + 3*3 \\ 4*7 + 5*9 & 4*8 + 5*3 \\ 1*7 + 4*9 & 1*8 + 4*3 \end{vmatrix}$$

Vector multiplication in matrix notation produces the same result as the inner product providing the first vector is a row vector and the second is a column vector.

The *transpose* of a matrix $\mathbf{A}^T$ is obtained by interchanging rows and columns.

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{vmatrix} \quad \text{transposes to} \quad \begin{vmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{vmatrix}$$

The matrix B is the *inverse* of a matrix A if AB = I where I is the identity matrix, being all 0's except for 1's on the diagonal.

Analytic Geometry

Plane geometry includes the study of figures like lines, circles, triangles, etc. all of which lie in a plane. Theorems are proved deductively from certain postulates. In *analytic* geometry, plane geometric figures are investigated by introducing a coordinate system and then using equations and formulas of various types. If the study of analytic geometry were to be summarized with one statement, this would be appropriate: "Given an equation, find its graph, and conversely, given a graph, find its equation." This is relevant to graphics because many of our primitives can be defined using equations, and drawn by graphing coordinates.

The basic figures one investigates in analytic geometry are the conic sections, i.e., the figures one obtains by intersection of a double-napped right circular cone with a plane. We get either a parabola, ellipse, circle or hyperbola. The equations for a circle are given in the "Parametric Representations" section above. Here are the others:

The equation of a parabola: $y^2 - 4ax = 0$

$x(t) = a * t^2$
$y(t) = 2a * t$

The equation of a hyperbola: $(x/a)^2 - (y/b)^2 = 1$
$x(t) = a * \sec(t)$
$y(t) = b * \tan(t)$

The equation of an ellipse: $(x/a)^2 + (y/b)^2 = 1$
$x(t) = a * \cos(2\pi t)$
$y(t) = b * \sin(2\pi t)$

The graph of a second-degree equation in x and y (i.e., in 2-D) gives us a conic section. In 3-D, the graph of a second-degree equation in x, y, and z gives us a *quadric* surface. Here are the equations of the six primary quadric surfaces:

a) Ellipsoid: $(x/a)^2 + (y/b)^2 + (z/c)^2 -1$

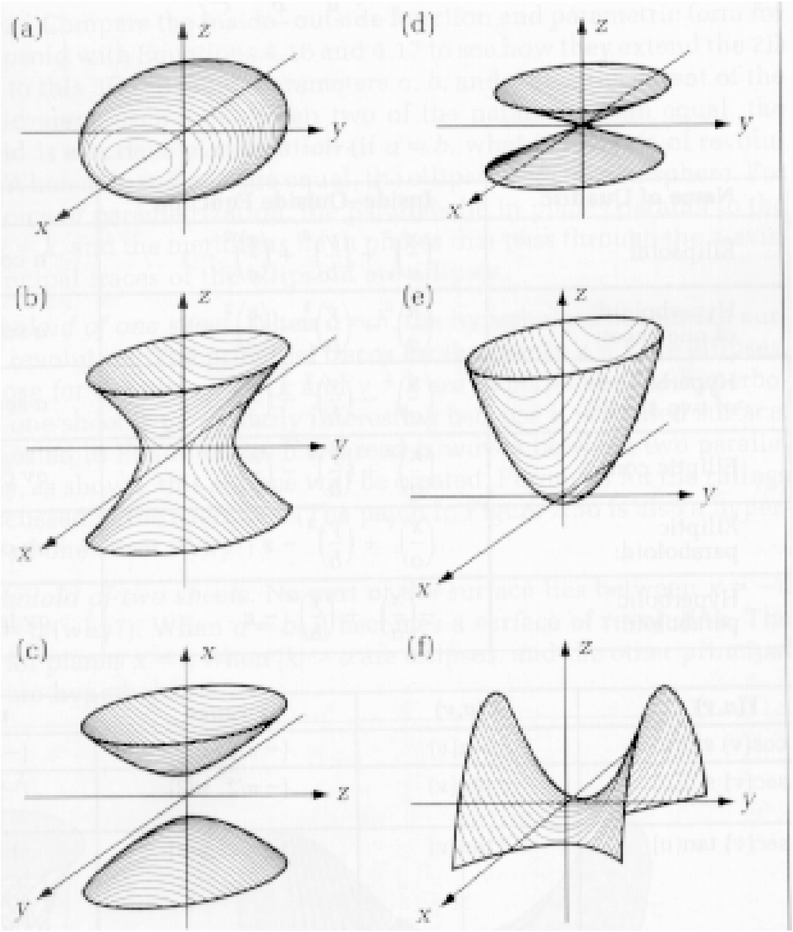
b) Hyperboloid of one sheet: $(x/a)^2 + (y/b)^2 - (z/c)^2 -1$

c) Hyperboloid of two sheets: $(x/a)^2 - (y/b)^2 - (z/c)^2 -1$

d) Elliptic cone: $(x/a)^2 + (y/b)^2 - (z/c)^2$

e) Elliptic paraboloid: $(x/a)^2 + (y/b)^2 - z$

f) Hyperbolic paraboloid: $-(x/a)^2 + (y/b)^2 - z$


(a)
z
y
x
(d)
z
y
x
(b)
z
y
x
(e)
z
y
x
(c)
x
z
y
(f)
z
y
x