Tilted Planes and Curvature in Three-Dimensional Space: Explorations of Partial Derivatives

By

Andrew Grossfield, Ph. D., P. E.

Abstract

Many engineering students encounter and algebraically manipulate partial derivatives in their fluids, thermodynamics or electromagnetic wave theory courses. However it is possible that unless these students were properly introduced to these symbols, they may lack the insight that could be obtained from a geometric or visual approach to the equations that contain these symbols.

We accept the approach that just as the direction of a curve at a point in two-dimensional space is described by the slope of the straight line tangent to the curve at that point, the orientation of a surface at a point in 3-dimensional space is determined by the orientation of the plane tangent to the surface at that point.

A straight line has only one direction described by the same slope everywhere along its length; a tilted plane has the same orientation everywhere but has many slopes at each point. In fact the slope in almost every direction leading away from a point is different. How do we conceive of these differing slopes and how can they be evaluated?

We are led to conclude that the derivative of a multivariable function is the gradient vector, and that it is wrong and misleading to define the gradient in terms of some kind of limiting process. This approach circumvents the need for all the unnecessary delta-epsilon arguments about obvious results, while providing visual insight into properties of multi-variable derivatives and differentials.

This paper provides the visual connection displaying the remarkably simple and beautiful relationships between the gradient, the directional derivatives and the partial derivatives. We find that the altitude above the horizontal coordinate plane varies sinusoidally with direction. The properties of multivariable derivatives can be easily grasped in terms of the properties of the orientation or tilt of planes in a 3-dimensional Cartesian coordinate system.

The spotlight is turned on the curvature or deviation from the tangent plane in terms of the classic second degree surfaces that prevails almost everywhere on well-behaved, that is, continuous and smooth (differentiable), warped surfaces. Here too the curvature is found to vary sinusoidally, only at twice the frequency and raised or lowered vertically. We see the significance of that wonderful intrinsic point property of surfaces, the Gaussian curvature and what it reveals about the differences between the curvature at the mountain passes and the curvature of the mountaintops and valleys.

This visual treatment of fundamental mathematical theory should serve as an introduction for precollege students of what lies ahead in their continuing study of mathematics.

Review

A fundamental development in analytic theory was the idea of Descartes that the graph of a curve could be constructed from an equation in two variables, say x and y. The variable x would be plotted along one axis and the variable y could be plotted along a perpendicular axis. A glance at the graph would indicate at each point, x, whether the value of y was positive or negative and also whether y was near or far from the horizontal axis. The glance would also indicate if y rises as x increases, and how the graph turns away or deviates from a tangent line.

A comment is required here about possible meanings of the variables, x, and y. In the following, it will be assumed that x and y represent distances measured in the same units. In this case the graph of the equation $y = x$ will be the straight line rising in the first quadrant at an angle of 45^o with the horizontal axis. Furthermore every straight line has a direction which can be measured by the angle the line makes with the horizontal. If the straight line has the form $y = mx + b$ then it can be shown that the value of m, called the slope of the line is equal to the tangent of the angle with the horizontal and will serve as a measure of the direction of the line

On the other hand say the vertical variable, y, represents distance and the horizontal variable, t, represents time, then it makes no sense to combine the vertical and horizontal into a single space; and there are no such concepts as either direction or angle in such a combined space. In this case if there is a linear relationship between y and t say $y = vt + y_0$, v cannot be treated as describing a direction but instead is called a “rate of change.” If y represents a distance and t represents time then v will be velocity. In this light all 1st degree equations graph as straight lines and have a constant rate of change.

Straight lines have a single direction or rate of change at all points but this is not true for curves. At all but a few points, well-behaved curves have a unique tangent line but the direction or rate of change of this tangent line varies from point to point and therefore is a function of either the horizontal or vertical coordinate of the point being examined. The development of the concept of differentiation enabled the computation of the direction or rate of change of these tangent lines.

In order to study the relationships between variables, it should be noticed that the same relationship can be expressed in different forms. Below are two tables of forms of equations relating two variables x and y. In the first table the equations are 1st degree and the graphs plot as straight lines. In the second table, the equations are not 1st degree and the graphs plot as curves.

Forms of straight lines in 2-dimensional x-y space

Implicit	$Ax + By = C$
Explicit	$y = mx + b$
Parametric	
	$x - x_0 = A_x (t - t_0)$
	$y - y_0 = A_y (t - t_0)$
Taylor Series	$y = y_0 + m (x - x_0)$

Forms of curves in 2-dimensional x-y space

Implicit	$F(x, y) = 0$
Explicit	$y = f(x)$
Parametric	
	$x = f(t)$
	$y = g(t)$

Taylor Series $y = y_0 + a_1 (x - x_0) + a_2 (x - x_0)^2 + a_3 (x - x_0)^3 + a_4 (x - x_0)^4 \ldots$

In the explicit form the equation is solved for the dependent variable, y. The slope of the line is the coefficient, m, of the independent variable, x. The equations of some curves cannot be easily solved for either of the dependent variable, y, or the independent variable, x. These curves might be expressed in a form called the implicit form, where x and y are knotted together.

The parametric form uses two equations and an intermediate variable, say t, as above called a parameter to describe the relationship between x and y. The slope of the straight line of y vs. x described above in parametric form is the ratio of the coefficients, $\frac{\Delta y}{\Delta x} = A_y/A_x$. The advantage of the parametric form is that it allows the curve $F(x, y) = 0$ to loop or double-back while $x(t)$ and $y(t)$ are single-valued functions.

The Taylor series describes a curve with a sum of possibly an infinite number of terms each of which is related to one of the derivatives of the function at a fixed point. For the Taylor series of the function $y = f(x)$, the coefficient $a_n = \frac{f^{(n)}}{n!}$ where $f^{(n)}$ is the n^{th} derivative of f(x) at the horizontal value x. As an example the Taylor series about the origin of an exponential function is $e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots, \ -\infty < x < \infty$

What is important to know about forms is that while a curve can be described in many forms, the properties of a curve belong to the curve and not any particular form. A form may display one property and conceal other properties. In quantitative analysis this leads to the computation drifting to a form that reveals the property being sought.

In the study of differential calculus, we are concerned with the local behavior of a curve; that is the behavior of the curve in a small interval surrounding a fixed point. There are three properties of major interest:

1) Position
2) Direction or rate of change
3) Deviation from the line of tangency, curvature, or rate of turn.

If the horizontal coordinate of our point of concern is the value x_0, the equation will provide the calculation of the corresponding vertical coordinate, y_0. The two coordinates (x_0, y_0) then locate the position of the point on the curve.

The direction or rate of change of the curve as determined by the slope of the tangent line is calculated from the derivative of the original function.

The last major property to be computed is the curvature or rate of turn. If the slope is increasing then the curve is turning up away from the tangent line. A decreasing slope indicates the curve is turning down below the tangent line. If the slope is increasing then the derivative of the slope is positive and vice versa. The derivative of the slope is the 2nd derivative of the original function, y(x). The 2nd derivative then can serve as a measure of the rate of turn. Another measure of the rate of turn is the reciprocal of the radius of the circle that best approximates the curve. This reciprocal is called the curvature of the curve and is usually represented by the Greek letter κ. A small circle accompanies a high rate of turn.

The previous paragraphs can be summed up with the following table:

1)	Position	y(x)
2)	Direction	y'(x)
3)	Rate of turn	y''(x)

If our curve is described by a Taylor series then the table can be extended to:

1)	Position	y(x)	y_0
2)	Direction	y'(x)	a_1
3)	Rate of turn	y''(x)	$2a_2$

Consider an application in physics where altitude is being treated as a function of time, t. Then the table would appear as:

1)	Height	y(t)
2)	Velocity	y'(t)
3)	Acceleration	y''(t)

I plan to extend these relationships between two variables in a 2-dimensional space to equations in three variables and their 3-dimensional representations.

Considering three variables and three dimensional space

In considering the geometry of three dimensions, we will be studying surfaces and curves. Surfaces in 3-dimensional x-y-z space can be described with a single equation which in implicit form is F(x, y, z) = 0. In explicit form, the equation of a surface is z = f(x, y). The explicit form requires the values of the two variables x and y in order to compute the value of z and locate the point, P(x, y, z). The explicit functional form treats the horizontal x-y plane as the domain of the function and the vertical z-axis as the range. Above each point in the x-y plane, a point on the surface at altitude z can be plotted. The implicit form provides more generality in allowing any one of the three variables to lie in the range while the other two variables make up the domain. The variables x, y and z will be treated as lying in a space whose axes are perpendicular and measured in equal units. Such a coordinate system is called orthonormal.

On the other hand, a space curve can be viewed as the intersection of two surfaces and so one of the representations of space curves will have to require two equations, each in possibly three variables.

Now how do we determine whether the object being studied is a surface or a space curve? It depends on the number of variables and the number of equations. Every additional variable increases by one the dimension of the space being considered. Every additional equation decreases by one the dimension of that space. The difference between the number of variables and the number of equations determines the dimension of our range of motion. This dimension is called the degree of freedom.

No. of variables	V
No. of equations	E

$V - E =$ degrees of freedom.

A point has no freedom. The intersection of 3 equations in 3 variables might consist of isolated points. A curve has one degree of freedom. From any particular point one can move only forward or backward. The coordinates of the points on a space curve can all be described as functions of one variable, say t for time or s for distance from an origin. The parametric form of a space curve is then;

$$x = f(t)$$
$$y = g(t)$$
$$z = h(t)\ .$$

The number of variables less the number of equations above, 4 – 3 yields one degree of freedom.

A surface has two degrees of freedom. Locating a point on the surface of the globe requires two coordinates: longitude and latitude. The parametric form of a surface in 3-space is then;

$$x = x(u, v)$$
$$y = y(u, v)$$
$$z = z(u, v)$$

The count of 5 variables minus 3 equations above leaves two degrees of freedom. One can depart from any point on the globe, except the poles, along lines of latitude or lines of longitude or some combination of both.

Straight lines and flat planes in space

Just as in the 2-dimensional space, straight lines have a position and a direction. If we were interested in the trajectory of a particle moving in space along a straight line as a function of time we would pick an initial point P(x, y, z) and specify a direction and our equations would enable us to compute where the particle was at every time, t. The equations for the position of the particle at every time can be described in parametric form extended to three dimensions:

$$x - x_0 = A_x\,(t - t_0)$$
$$y - y_0 = A_y\,(t - t_0)$$
$$z - z_0 = A_z\,(t - t_0)\ .$$

These three equations could be written as one vector equation: $\vec{\mathbf{P}} - \vec{\mathbf{P}_o} = \vec{\mathbf{M}}\,(t - t_0)$ where $\vec{\mathbf{P}_o}$ is a vector locating the position of the particle on the line at time t_0 and $\vec{\mathbf{M}} = [A_x, A_y, A_z]$ is a vector which indicates the direction of the line.

Now we want to examine the straight line when it is initially described as the intersection of two flat planes but first we need to recognize some facts usually found in geometry and solid geometry courses.

Facts from Plane and Solid Geometry

1) In a 2-dimensional plane, 2 straight lines are either parallel, coincide or intersect in a single point.
2) In 3-dimensional space, 2 planes are either parallel, coincide or intersect in a single line.
3) In 3-dimensional space, there can exist pairs of lines, described as “skew” that are not parallel and are not co-incident and in addition do not intersect.
4) Straight lines have a unique direction but the same cannot be said of planes. A tilted plane contains an infinite number of lines whose angle with the horizontal plane varies with the direction of displacement.
5) However, a plane in 3-space has a unique normal direction which is said to determine the “orientation” of the plane.
6) Every line in the plane which passes through the point where a normal line pierces the plane is perpendicular to the normal line.
7) In both 2 and 3-dimensional spaces, the equation of a line is determined by a point on the line and the direction of the line.
8) The equation of a plane is determined by a point on the plane and the direction of the normal to the plane.
9) The line of intersection of two planes must be perpendicular to each of the normal directions of the two planes.

Vectors and vector operations

Vector notation is a significant advance in mathematical writing. A single vector combines in one symbol a set of numbers or variables. In addition, a vector equation can combine a set of equations into a single equation. This provides for easier and faster reading and computation but requires some study and practice to acquire proficiency.

There are three vector operations which we will find useful in our study. One operation, called the dot product, is written as $\vec{\mathbf{A}} \circ \vec{\mathbf{B}}$ and yields a scalar (that is, a number and not a vector). The next operation, called the cross product, is written as $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$ and yields a vector which is perpendicular to the plane spanned by the vectors $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ and the third operation produces a scalar and is called the triple scalar product, and denoted by $\vec{\mathbf{A}} \times \vec{\mathbf{B}} \circ \vec{\mathbf{C}}$.

The dot product provides a means to compute the length of the projection of a line segment onto some intersecting line. In the ordinary 2-dimensional space the projection of a vector $\vec{\mathbf{A}}$ onto the horizontal axis is $|A| \cos(\varphi)$ where φ is the angle between $\vec{\mathbf{A}}$ and the horizontal axis.

In 3-dimensional space if we have a vector $\vec{\mathbf{A}}$ and a direction described by a unit vector $\vec{\mathbf{u}}$ the projection of $\vec{\mathbf{A}}$ in the direction $\vec{\mathbf{u}}$ is $\vec{\mathbf{A}}\circ\vec{\mathbf{u}} = |A| \cos(\varphi)$. We would like our dot product to be symmetric in the two vectors and so the dot product is defined as: $\vec{\mathbf{A}}\circ\vec{\mathbf{B}} = |A||B|\cos(\varphi)$. It can be proven that if the coordinates of the vectors are given in an orthonormal coordinate system; $\vec{\mathbf{A}} = [a_1, a_2, a_3]$ and $\vec{\mathbf{B}} = [b_1, b_2, b_3]$ then the dot product can be computed by adding the pairwise products of the coordinates:

$$\vec{\mathbf{A}}\circ\vec{\mathbf{B}} = |A||B| \cos(\varphi) = a_1 b_1 + a_2 b_2 + a_3 b_3 = \sum_{i=1}^{3} a_i b_i \qquad \text{Equation 1}$$

Imagine a vector whose tail is at the origin rotating in a plane around the origin. The projections on both the horizontal and vertical axes or in any direction are simple sinusoidal oscillations. In a plane in 3-dimensional space the projection on a fixed vector, $\vec{\mathbf{A}}$, of a rotating vector, $\vec{\mathbf{B}}$, will vary sinusoidally with the angle between the two vectors.

The cross product provides a means to compute the area of the parallelogram spanned by two vectors. In the ordinary 2-dimensional space, the area of the parallelogram spanned by the two vectors, $\vec{\mathbf{A}} = [a, b]$ and $\vec{\mathbf{B}} = [c, d]$, can be shown to equal the determinant of the matrix formed by the coordinates of the vectors:

$$\mathrm{Det}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

Similarly, in 3-dimensional space with an orthonormal basis, it can be shown that the area of the parallelogram spanned by two vectors with a common tail can also be described by a determinant. Say $\vec{\mathbf{A}} = [a_1, a_2, a_3]$ and $\vec{\mathbf{B}} = [b_1, b_2, b_3]$ The cross product $\vec{\mathbf{A}} \times \vec{\mathbf{B}}$ produces a directed vector in the plane perpendicular to the plane of $\vec{\mathbf{A}}$ and $\vec{\mathbf{B}}$ whose magnitude equals the area of the parallelogram spanned by the vectors. The direction of the cross product is determined by the right hand rule of physics. It should be noted that

$$\vec{\mathbf{B}} \times \vec{\mathbf{A}} = -\vec{\mathbf{A}} \times \vec{\mathbf{B}}$$

The cross product can be evaluated from the components of $\vec{\mathbf{A}} = [a_1, a_2, a_3]$ and $\vec{\mathbf{B}} = [b_1, b_2, b_3]$ by computing the determinant of the matrix

$$\vec{\mathbf{N}} = \vec{\mathbf{A}} \times \vec{\mathbf{B}} = \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = (a_2 b_3 - a_3 b_2)\,\mathbf{i} - (a_1 b_3 - a_3 b_1)\,\mathbf{j} + (a_1 b_2 - a_2 b_1)\,\mathbf{k}$$

Our last vector operation, the triple scalar product provides the volume of the parallelepiped spanned by three vectors, $\vec{\mathbf{A}} \times \vec{\mathbf{B}}\circ\vec{\mathbf{C}} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix}$

Now these basic facts and notations concerning lines and planes in 3-dimensional spaces have been noted we can begin to examine derivatives for surfaces in space.

Slopes in tilted planes in 3-dimensional space

A straight line in the 2-dimensional x-y plane has the same direction at every point. This direction is measured by its slope, the ratio of rise to run which has the same value between any two points on the line.

However, at a point in a tilted plane the slopes of lines through the point differ in almost every direction. The magnitude of the slope depends on the displacement direction. There is an uphill direction of maximum rise and a maximum fall in the opposite direction. There are two directions in which there is no rise or fall and there are intermediate directions in which a change in altitude may be any value between the maximum and the minimum. The slope in any particular direction is denoted by the symbol, $\frac{dz}{ds}$, and is called the directional derivative.

Just as the variables of the equations of a straight line in 2 and 3-dimensional spaces appear to the 1^{st} degree, the variables of an equation describing a flat plane in 3-dimensions also occur to the 1^{st} degree. Therefore the equation of a plane passing through the point, $P_0 = [\ x_0, y_0, z_0\ ]$ in explicit form will be:

$$z = A\,(x - x_0) + B(y - y_0) + z_0 = A\,x + By - A\,x_0 - By_0 + z_0$$

The slope in the x direction is called the partial derivative of z with respect to x and is denoted by the symbol, $\frac{\partial z}{\partial x}$ and correspondingly the partial derivative of z with respect to y is $\frac{\partial z}{\partial y}$. To evaluate $\frac{\partial z}{\partial x}$ differentiate z while holding y constant and to evaluate $\frac{\partial z}{\partial y}$, differentiate z while holding x constant. We find: $\frac{\partial z}{\partial x} = A$, the coefficient of x and $\frac{\partial z}{\partial y} = B$, the coefficient of y. The change in z due to a change, dx, parallel to the x-axis will be $\frac{\partial z}{\partial x}dx$ and a change in z due to a change, dy, in the y direction will be $\frac{\partial z}{\partial y}dy$.

Since the equation of a plane is linear the effect on z of varying both x and y in either order can be computed by adding the separate effects:

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy \qquad \text{Equation 2}$$

This variation in height above the x-y plane, dz, which is due to changes in both the x and y directions is called the total differential.

Let a differential displacement vector in the x-y plane be $\overrightarrow{\mathbf{dr}} = [dx, dy]$.

Then $$ds^2 = \overrightarrow{\mathbf{dr}}\circ\overrightarrow{\mathbf{dr}} = dx^2 + dy^2$$

and then the directional derivative is: $$\frac{dz}{ds} = \frac{\partial z}{\partial x}\frac{dx}{ds} + \frac{\partial z}{\partial y}\frac{dy}{ds}.$$

It appears from Equation 2 that the variation in height, dz can be viewed as a dot product between the vector displacement, $\overrightarrow{\mathbf{dr}}$ = [dx, dy], in the x-y plane and something new that here acts like a vector. This new object, [$\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}$], denoted by the symbol by ∇z, is called the gradient of z.

Now employing Equation 1, Equation 2 can be rewritten in the form:

$$dz = \nabla z \circ \overrightarrow{\mathbf{dr}} = |\nabla z||\overrightarrow{\mathbf{dr}}| \cos(\varphi). \qquad \text{Equation 2}$$

We note that a plane has the same orientation, the same gradient at every point. It has the same slopes in parallel directions. We should also note that the slopes in the coordinate directions which are the partial derivatives determine the gradient. The gradient is an invariant property of the surface which does not change even when the coordinate system is changed. See Figures 1, 2 and 3 below showing a plane in 3-space and the slopes in the coordinate and gradient directions.

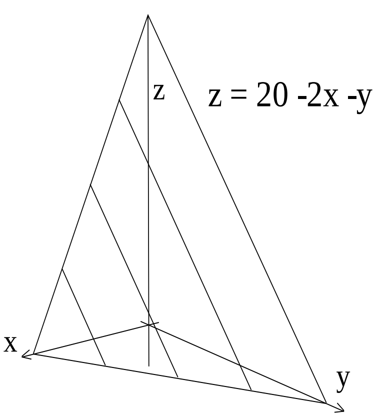


Figure 1 The straight lines in a plane when x is constant

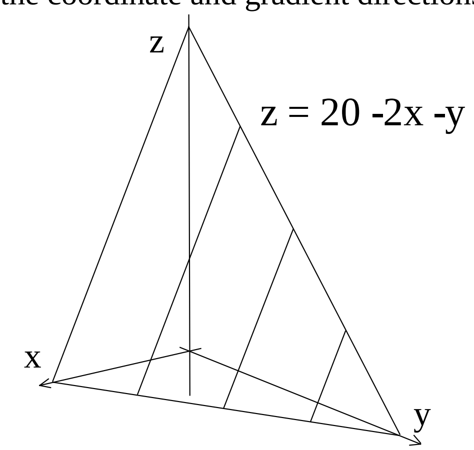


Figure 2 The straight lines in a plane when y is constant

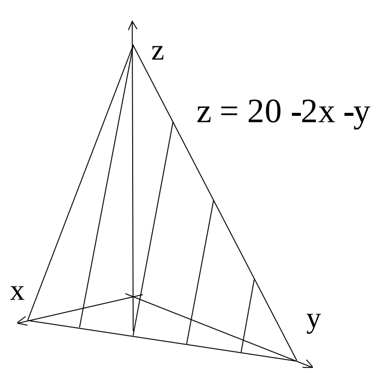


Figure 3 The gradient direction

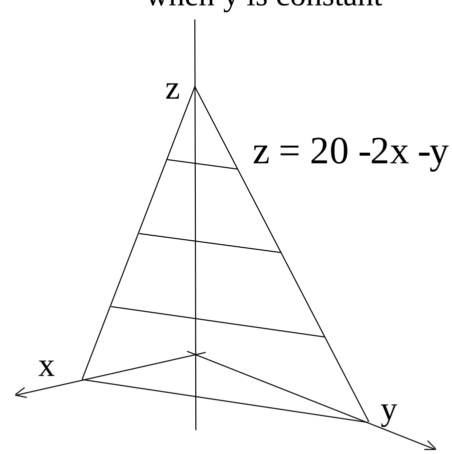


Figure 4 The level directions

As an example let us compute the properties of orientation for the plane shown in the above four figures described in explicit form by the equation: $z = 20 - 2x - y$.

If y is held constant then the slope in the x direction is $\frac{\partial z}{\partial x} = -2$ and similarly if x is held constant the slope in the y direction is $\frac{\partial z}{\partial y} = -1$.

The gradient is found to be $\nabla z = -2\mathbf{i} - \mathbf{j}$.

Note: The gradient is not a vector like the displacement vector (a gradient cannot be added to a displacement), but if a dot product is performed between the gradient and a displacement, a total differential can be computed. If the vectors and displacements in the original x-y-z space are written as column vectors then the gradients will be a row vectors.

Let a displacement in the x-y plane be $2\mathbf{i} + 2\mathbf{j} = 2\sqrt{2}e^{i\pi/4}$, then

$$dz = \left(\frac{\partial z}{\partial x}\mathbf{i} + \frac{\partial z}{\partial y}\mathbf{j}\right)\circ(2\mathbf{i} + 2\mathbf{j}) = (-2\mathbf{i} - \mathbf{j})\circ(2\mathbf{i} + 2\mathbf{j}) = (-2*2 - 1*2) = -6 \ .$$

The directional derivative at an angle of 45^o between the x and y directions will be:

$$\frac{dz}{ds} = \frac{-6}{2\sqrt{2}} = -1.5\sqrt{2} \ .$$

On implicit and explicit forms

The implicit form of an equation treats the three variables x, y and z equally. However, neither a range or domain is called out but the extent of the values may be limited as in the case of the sphere $x^2 + y^2 + z^2 - 25 = 0$ where all the points are limited to the inside of the cube $|x| \leq 5$, $|y| \leq 5$ and $|z| \leq 5$. It is also common as in this case that the implicit form might have more than one value for any of the variables

On the other hand, an explicit form of the same relationship $z = \sqrt{25 - x^2 - y^2}$ is a functional form with the range being along the vertical z axis and the domain lying in the horizontal x-y plane. If the operations in the function are single-valued then so will be the surface in the x-y-z space. In the next section we will examine the slopes of a tilted plane, $z = Ax + By + C$ above a circle in the x-y horizontal domain, and now we are positioned to observe something which is remarkable and wonderful.

Take any point, P_0 in the x-y plane and consider a point, P circling P_0 at a fixed distance, dr, as shown below in figure 5. Watch the motion of the point, Q, in the tilted plane directly above P. Say the motion of P starts when Q is highest. Q will be observed to descend monotonically to a minimum and then rise monotonically back to its starting height.

We can now derive the relationship between the height of the point Q and the angle that P makes in the x-y space. Since the gradient of a plane is constant and $|\overrightarrow{dr}|$ is constant, equation 2 above leads to the conclusion that if the angle, φ, of the direction is chosen to be 0° in the uphill direction, then dz can be shown to vary as a cosine. This means that the direction of the gradient is uphill; the opposite direction is downhill; at 90 ° there will be no change in z and at 45° the change in z will be 0.707 of the maximum change.

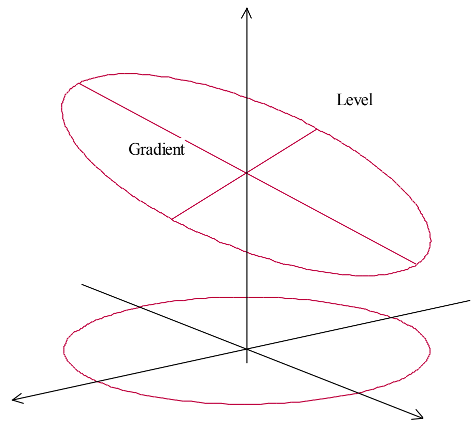


Figure 5 A tilted plane above a circle

While the tilted plane has many slopes depending on direction it has only one orientation which is described by its gradient. The gradient, $[\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}]$ of the explicit form of the plane indicates the uphill direction in the x-y horizontal domain.

In 3-dimensional space say the implicit equation of the plane is:

$$A(x - x_o) + B(y - y_0) + C(z - z_0) = 0. \qquad \text{Equation 3}$$

This states that the dot product of the vector [A, B, C] with the displacement on the plane, $[x - x_o, y - y_0, z - z_0]$ from the point P_o (x_o, y_o, z_o) is 0 which implies that the vector [A, B, C] is perpendicular to the plane in space. If the perpendicular or normal vector to a plane is called $\vec{\mathbf{N}} = [N_x, N_y, N_z]$ then equation 3 can be rewritten as

$$N_{x,} (x - x_o) + N_y (y - y_0) + N_z (z - z_0) = 0.$$

To sum up our findings so far:

A point, $P_0(x_0, y_0, z_0)$ and the direction vector, $[M_x, M_y, M_z] = \vec{M}$ determine the parametric and vector equations of a line in 3-dimensional space;

$$\begin{aligned} x - x_0 &= M_x (t - t_0) \\ y - y_0 &= M_y (t - t_0) \\ z - z_0 &= M_z (t - t_0). \end{aligned} \qquad \text{and} \qquad \vec{r}\mathbf{(t)} = \vec{M}\,\mathbf{(}t - t_0\mathbf{)} + \vec{P_o}$$

The vector $\vec{M}$ is the derivative $\frac{d\vec{r}(t)}{dt}$ of the straight line trajectory $\vec{r}\mathbf{(t)}$.

A point, $P_0(x_{0,} y_{0,} z_0)$, and the normal or orientation vector, $[N_x, N_y, N_z] = \vec{N}$ determine the implicit equation, $N_x (x - x_0) + N_y (y - y_0) + N_z(z - z_0) = 0$ of a flat plane in space which can be written as a vector equation, $\overrightarrow{P - P_o} \circ \vec{\mathbf{N}} = 0$.

The direction of the line of intersection of two planes

Let two planes H and K intersect in line L. Since L is in H it must be perpendicular to any normal, $N_{\mathbf{H}}$, to H that intersects L. Since L is in K it must be perpendicular to any normal, $N_{\mathbf{K}}$, to K that also intersects L. Therefore since L is perpendicular to both $N_{\mathbf{H}}$ and $N_{\mathbf{K}}$, L is parallel to the cross product of $N_{\mathbf{H}}$ and $N_{\mathbf{K}}$.

If the equation of H is $\qquad A(x - x_0) + B(y - y_0) + C(z - z_0) = 0$

and the equation of K is $\qquad D(x - x_0) + E(y - y_0) + F(z - z_0) = 0$

then the direction of the line of intersection, L, of the two planes is:

$$\left[\begin{bmatrix} B & C \\ E & F \end{bmatrix}, \; -\begin{bmatrix} A & C \\ D & F \end{bmatrix}, \; \begin{bmatrix} A & B \\ D & E \end{bmatrix}\right].$$

Now we are set to apply the properties of straight lines and flat planes to space curves and warped surfaces.

Differentials and the derivatives of space curves

Curves in space are continuous if the functions in their parametric form are continuous. Curves in space are differentiable if at every point they have a unique direction set by the tangent line. This means that the functions in the parametric forms are differentiable and that the derivatives of these functions are not all zero at the same point.

The derivative of a function of a single variable at a fixed point is the slope of the line tangent to the curve of the function. Let the horizontal displacement between two points on the line be represented by the variable, dx, and similarly let the vertical displacement between the two points be represented by the variable, dy. In the 2-dimensional x-y space the equation describing the relation between these two variables called differentials appear as in ordinary algebra:

$$dy = \frac{dy}{dx} * dx \qquad \text{Equation 4}$$

This derivative concept can be extended naturally to curves in space. In 3-dimensional space the position vector is

$$\vec{R}(\mathbf{s}) = x(\mathbf{s})\,\vec{\imath} + y(\mathbf{s})\,\vec{\jmath} + z(\mathbf{s})\,\vec{k} = \begin{bmatrix} x(s) \\ y(s) \\ z(s) \end{bmatrix} \quad \text{and the vector derivative is}$$

$$\frac{\mathbf{d\vec{R}(s)}}{\mathbf{ds}} = \frac{\mathbf{d}}{\mathbf{ds}}\mathbf{x(s)}\,\vec{\mathbf{i}} + \frac{\mathbf{d}}{\mathbf{ds}}\mathbf{y(s)}\,\vec{\mathbf{j}} + \frac{\mathbf{d}}{\mathbf{ds}}\mathbf{z(s)}\,\overrightarrow{\mathbf{k}} = \begin{bmatrix} \frac{\mathbf{dx(s)}}{\mathbf{ds}} \\ \frac{\mathbf{dy(s)}}{\mathbf{ds}} \\ \frac{\mathbf{dz(s)}}{\mathbf{ds}} \end{bmatrix} = \begin{bmatrix} x'(s) \\ y'(s) \\ z'(s) \end{bmatrix} \text{ and}$$

the vector differential in position is related to the scalar parameter differential by the vector equation

$$\overrightarrow{\mathbf{dR}} = \frac{\mathbf{d\vec{R}}}{\mathbf{ds}}\,\mathbf{d}s \qquad \text{Equation 5}$$

The derivative of the vector function at a point on a space curve is a vector with the same direction as the tangent vector. If the parametric equations represent the trajectory of a moving projectile and the parameter is time, then the derivative will be velocity (that is, a vector in the direction of the tangent vector whose magnitude is the speed of the projectile). As the functions x(s), y(s), and z(s) are single variable functions, these are ordinary derivatives.

Surfaces in space – continuity

We could begin a study of surfaces by noting some properties of the different classifications of surfaces: polynomial, rational, algebraic and transcendental. Surfaces described by polynomial functions are defined everywhere, single-valued, continuous and differentiable and rise to infinity in every direction. The surfaces of rational functions are also single-valued but are not defined at the zeros in their denominators. Algebraic surfaces described by implicit forms like the sphere may be multiple-valued and like the sphere may only be defined in restricted domains. It is important to be aware that in the regions where they are useful all of these surfaces are mostly continuous and mostly differentiable.

The function $z = \frac{x^2 - y^2}{x^2 + y^2}$ provides an example of discontinuous behavior of a function of two variables. We see at the origin the value of z is undefined. If this function is described in polar (r, φ) coordinates instead of Cartesian coordinates we observe it becomes a function of the single-variable, φ. That is z is constant on every ray of constant φ, but the value of z varies between ∓ 1 as φ varies.

$$z = \frac{r^2\cos^2\varphi - r^2\sin^2\varphi}{r^2\cos^2\varphi + r^2\sin^2\varphi} = \cos^2\varphi - \sin^2\varphi = \cos 2\varphi \qquad \text{Equation 6}$$

A closer inspection of equation 6 reveals that approaching the origin along the x-axis when $\varphi = 0$, z maintains the value 1. Approaching the origin from a direction of 45 degrees to the x-axis when $\varphi = 45^o$, z keeps the value 0. Approach the origin down along the y-axis when $\varphi = 90^o$, and z maintains the value –1. Since the values for z in every neighborhood of the origin do not cluster around a single value, the function z is discontinuous.

This surface can be thought of as the area swept out by the motion of a straight rod attached at a right angle to the z-axis but free to rotate about the z-axis and move upward along the z-axis maintaining its perpendicularity with the z-axis.

On the relationship between the tangent plane to a surface and the local behavior of the surface

We should now recognize some facts about the relationship between the tangent plane at a point, P, on a surface and the local behavior of the surface.

1) Both the tangent plane and the surface have the same orientation, that is the same normal direction at P.
2) The directional derivatives in every direction of both the surface and the tangent plane are identical including the partial derivatives, the gradient and the level directions.
3) The constant and the linear terms of the 2-dimensional Taylor series of the surface describe the tangent plane.

The definition of the derivative of a function of several variables

We observe from Equations 4 and 5 that the derivative of both an ordinary function of a scalar and the vector function of a scalar linearly relate differential displacements in the domain to the corresponding range displacements. It should be taken as a general definition that the derivative provides a linear relationship between displacements in the domain and their corresponding range displacements. Taking this definition of the derivative and applying Equation 2 leads to the conclusion that the derivative of a multi-variable function is the gradient.

Locating the interior extrema of multivariable smooth surfaces

We observe that for a function representing a smooth curve in a 2-dimensional space at the maximum or minimum points the tangent line is horizontal implying that the derivative of the function is zero. Similarly, for multivariable functions representing smooth surfaces, the tangent plane at the extreme points is horizontal. The gradient of a horizontal plane being zero implies that the gradient of a multivariable function must also be zero and that at an extreme point of a multivariable function z = f(x, y) both of the equations:

$$\frac{\partial z}{\partial x}(x, y) = 0 \quad \text{and} \quad \frac{\partial z}{\partial y}(x, y) = 0$$

must be satisfied. A second strategy for locating a maximum that is incremental is to pick a starting point and then follow the local gradients uphill.

An application to error analysis

Say a function of several variables w = F(x, y, z) has respective small errors in the variables x, y and z of values dx, dy and dz. Then to a first approximation the effects of these errors add. This means the error in w is approximately

$$dw = \frac{\partial w}{\partial x}dx + \frac{\partial w}{\partial y}dy + \frac{\partial w}{\partial z}dx .$$

On the curvature of curves in space

Envision the curve as the trajectory of a moving point. Use the distance from a fixed point on the curve, s, as a parameter in the description of the curve. Then in a fixed orthonormal coordinate system the position, direction and rate of turn of the point on the curve could be described respectively as vector functions of s:

$$\overrightarrow{\mathbf{R}(\mathbf{s})} = \begin{bmatrix} x(s) \\ y(s) \\ z(s) \end{bmatrix}, \qquad \overrightarrow{\frac{\mathbf{dR}}{\mathbf{ds}}}(\mathbf{s}) = \begin{bmatrix} x'(s) \\ y'(s) \\ z'(s) \end{bmatrix} \qquad \text{and} \qquad \overrightarrow{\frac{\mathbf{dR}}{\mathbf{ds}^2}}(\mathbf{s}) = \begin{bmatrix} x''(s) \\ y''(s) \\ z''(s) \end{bmatrix}$$

If the components of the trajectory are described by a Maclaurin series then

$$x(s) = a_0 + a_1 s + a_2 s^2 + \ldots \quad ,$$
$$y(s) = b_0 + b_1 s + b_2 s^2 + \ldots \quad \text{and}$$
$$z(s) = c_0 + c_1 s + c_2 s^2 + \ldots$$

If a_2, b_2 and c_2 are not all 0 then when s = 0, the position, direction and rate of turn of the point will be respectively the vector constant, the vector derivative and the vector 2nd derivative:

$$\overrightarrow{\mathbf{R}(\mathbf{0})} = \begin{bmatrix} a_0 \\ b_0 \\ c_0 \end{bmatrix}, \qquad \overrightarrow{\frac{\mathbf{dR}}{\mathbf{ds}}}(\mathbf{0}) = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} \qquad \text{and} \qquad \overrightarrow{\frac{\mathbf{d^2R}}{\mathbf{ds}^2}}(\mathbf{0}) = \begin{bmatrix} 2a_2 \\ 2b_2 \\ 2c_2 \end{bmatrix}$$

This approach provides the components of the vector derivatives in the directions along fixed coordinate axes.

The Frenet-Serret equations

There is another description of a space curve which robotics engineers will find valuable which was studied by Jean Frédéric Frenet and Joseph Alfred Serret near the middle of the 19th century. The Frenet-Serret frame is comprised of three orthonormal vectors which move along the curve. The first is the unit tangent vector, $\vec{\mathbf{T}}$, which points in the direction of motion. The second is the unit normal vector, $\vec{\mathbf{N}}$, which points in the direction of $\overrightarrow{\frac{\mathbf{dT}}{\mathbf{ds}}}$. The third vector in the frame is called the bi-normal and is denoted by $\vec{\mathbf{B}}$. $\vec{\mathbf{B}}$ is obtained from the cross product of $\vec{\mathbf{T}}$ and $\vec{\mathbf{N}}$.
The magnitude of $\overrightarrow{\frac{\mathbf{dT}}{\mathbf{ds}}}$, denoted by κ is a measure of the curvature or the departure from straight line motion. The curve is turning sharper when κ increases. If κ = 0 the motion is linear.
If the motion lies in a plane, which would be the $\vec{\mathbf{T}}$-$\vec{\mathbf{N}}$ plane then the Frenet-Serret description

reduces to $$\overrightarrow{\frac{\mathbf{dT}}{\mathbf{ds}}} = \kappa\, \vec{\mathbf{N}}, \quad \overrightarrow{\frac{\mathbf{dN}}{\mathbf{ds}}} = -\kappa\, \vec{\mathbf{T}} \quad \text{and} \quad \overrightarrow{\frac{\mathbf{dB}}{\mathbf{ds}}} = 0.$$

However, if $\overrightarrow{\frac{\mathbf{dB}}{\mathbf{ds}}}$ is not 0, then the magnitude of $\overrightarrow{\frac{\mathbf{dB}}{\mathbf{ds}}}$ which is called the torsion and is denoted by the symbol, τ, provides a measure of the rate of turning of the $\vec{\mathbf{T}}$-$\vec{\mathbf{N}}$ plane. The complete Frenet-Serret equations are:

$$\frac{d\vec{T}}{ds} = \qquad \kappa \vec{N}\ .$$

$$\frac{d\vec{N}}{ds} = \quad -\kappa \vec{T} \qquad\qquad \tau \vec{B}\ .$$

$$\frac{d\vec{B}}{ds} = \qquad -\tau \vec{N}\ .$$

The reader is encouraged to research the Frenet-Serret frame and to view images on the web.

On the curvature of surfaces in space

A study of the curvature at a point, P, on a surface should begin by examining the Taylor series form of the equation of the surface. If we subtract the constant and the linear terms of the series, the 2^{nd} and higher degree terms which contribute to the curvature remain. We know that near P the values of the higher degree terms of the surface z = F(x, y) , x^3 and x^4 etc., are much smaller than x^2. If the higher degree terms are dropped and only the 2^{nd}-degree terms are retained the Taylor series for z will appear as a quadratic form:

$$z = F(x,y) = A_Q\, x^2 + 2\, B_Q\, xy + C_Q\, y^2$$

Note: In the following paragraphs it will be necessary to distinguish the curvature of ordinary 2-dimensional parabolas from the curvature of surfaces represented by quadratic forms. The subscript Q will be used to distinguish the surface quadratic form parameters from the parameters A, B, and C which are conventionally used in the equations describing parabolas. The curvature at the vertex of an ordinary parabola is encapsulated in A the coefficient of the second degree term.

The quadratic form parameters A_Q, B_Q and C_Q determine the surface curvature at point P. If A_Q, B_Q and C_Q are allowed to have any values, only 17 possible different kinds of surfaces will be produced[1]. Five of the possibilities are imaginary. Three of the other possibilities have the form $(A_1x + B_1y) * (A_2x + B_2y)$ which are the products of either parallel, intersecting or coincident planes. The products of the planes are flat and have no curvature.

Nine surfaces remain which can be placed in three categories distinguished by nature of their curvatures:

> the elliptic cone and the elliptic, the parabolic and the hyperbolic cylinders;
> the paraboloid, the ellipsoid, and the hyperboloid of two sheets and
> the hyperbolic paraboloid, and the hyperboloid of one sheet.

A straight line in 2-space has a constant slope and a plane in 3-space has a constant gradient or orientation. In 2-space curvature is described as a change in direction or slope and related to the second derivative of position. In 3-space curvature will be seen to be described as a change in orientation or gradient, that is as the gradient of the gradient.

The derivative of a row vector

In this section the numerical subscripts 1, 2 and 3 will be used to represent the orthonormal coordinate axes. These subscripts will be used to represent the components of the vectors in our 3-space and the components of the differentials in our 3-space. If our row-vector is

$$\vec{V} = [\ A_1(x, y, z)\,\vec{\mathbf{i}} + A_2(x, y, z)\,\vec{\mathbf{j}} + A_3(x, y, z)\,\vec{\mathbf{k}}\ ]$$

then the gradient of A_1 is $\frac{\partial A_1}{\partial x}\vec{\mathbf{i}} + \frac{\partial A_1}{\partial y}\vec{\mathbf{j}} + \frac{\partial A_1}{\partial z}\vec{\mathbf{k}}$ or $[\frac{\partial A_1}{\partial x}, \frac{\partial A_1}{\partial y}, \frac{\partial A_1}{\partial z}]$ and the differential of A_1 is

$$dA_1 = [\ \frac{\partial A_1}{\partial x}, \frac{\partial A_1}{\partial y}, \frac{\partial A_1}{\partial z}\ ] \circ [\ dx, dy, dz\ ].$$

Repeating this process on the remaining two components yields

$$dA_1 = [\ \frac{\partial A_1}{\partial x}, \frac{\partial A_1}{\partial y}, \frac{\partial A_1}{\partial z}\ ] \circ [\ dx, dy, dz\ ].$$

$$dA_2 = [\ \frac{\partial A_2}{\partial x}, \frac{\partial A_2}{\partial y}, \frac{\partial A_2}{\partial z}\ ] \circ [\ dx, dy, dz\ ].$$

$$dA_3 = [\ \frac{\partial A_3}{\partial x}, \frac{\partial A_3}{\partial y}, \frac{\partial A_3}{\partial z}\ ] \circ [\ dx, dy, dz\ ].$$

Combining the row vectors to form a matrix leads to a 3×3 matrix equation where the matrix relates differential changes in the range vector components to differential domain displacements.

$$\begin{bmatrix} dA_1 \\ dA_2 \\ dA_3 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix}$$

Here we observe that this matrix is the derivative of the row vector. This derivative will be used to derive the shape operator.

The shape operator

In this section a clean notation will be used where the partial derivatives $\frac{\partial F}{\partial x}$ and $\frac{\partial F}{\partial y}$ are written using subscripts, F_x and F_y.

The original surface is described by the function $z = F(x, y)$. The gradient of the z is the row vector $\nabla z = [\ F_x\,\vec{\mathbf{i}} + F_y\,\vec{\mathbf{j}}\ ]$. The derivative of ∇z is the 2^{nd} derivative of the function $z = F(x, y)$. This 2^{nd} derivative is a matrix that contains fundamental information about the curvature of the surface and is composed of the 2^{nd} degree terms of the Taylor series of F(x, y) . This matrix is called the shape operator or Hessian and its determinant is called the Gaussian curvature, K, of the surface described by the function F. The Gaussian curvature can be shown to be independent of the orthonormal coordinate system used in the description of the horizontal plane. If the 2^{nd} degree terms of the Taylor series of F(x, y) are A, B and C then the Hessian matrix appears as:

$$\begin{vmatrix} F_{xx} & F_{xy} \\ F_{xy} & F_{yy} \end{vmatrix} = \begin{vmatrix} 2A & 2B \\ 2B & 2C \end{vmatrix} = 4\begin{vmatrix} A & B \\ B & C \end{vmatrix} .$$

The quadratic surface is seen to result from the matrix operations;

$$[x \quad y]\begin{vmatrix} F_{xx} & F_{xy} \\ F_{xy} & F_{yy} \end{vmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = F_{xx}\, x^2 + 2F_{xy}\, xy + F_{yy}\, y^2$$

It should be noted that while generally matrices describe linear operations this particular matrix describes a surface that is not linear in the vector components x and y.

The variation of the normal parabolic curvature with direction

If the origin of the orthonormal coordinate system is moved to the point P and the surface is rotated so that the tangent plane to the surface coincides with the x-y plane, the normal to the surface at P will coincide with the z-axis as shown in figure 6. If the surface is defined over the two variable x-y space and is of 2nd degree then the equation for the surface will have the form

$$z = F\,(x, y) = A_Q x^2 + 2B_Q\, xy + C_Q y^2$$

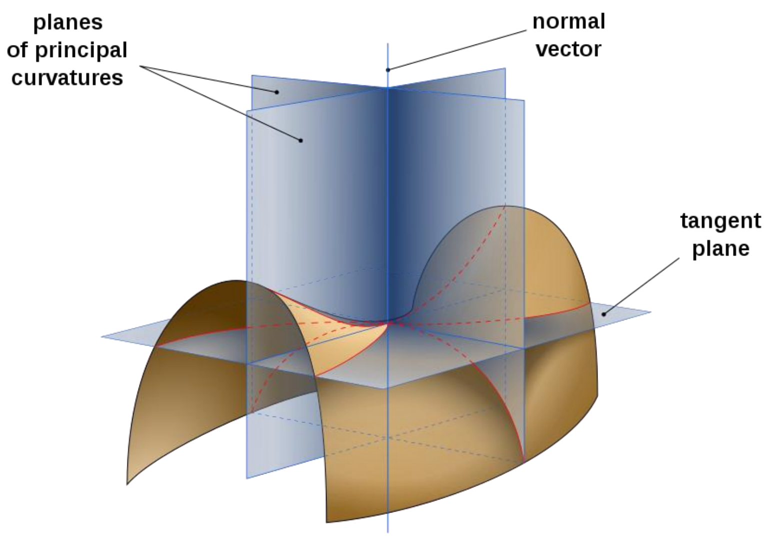


Figure 6 The principle planes of the hyperbolic paraboloid[16]. This figure shows that a surface with negative Gaussian curvature lies above the tangent plane in two opposing sectors of normal planes and below the tangent plane in the complimentary sectors.

If a plane with the equation y = mx is passed through the z-axis and rotated it will intersect the 2nd degree surface in the parabola $z = A_Q\, x^2 + 2B_Q\, m\, x^2 + C_Q\, m^2 x^2 = (A_Q + 2B_Q\, m + C_Q\, m^2)\, x^2$

The information about the curvature of the surface as a function of the direction of a normal plane is contained in the coefficient of x^2 but it is hidden. More can be learned if the coordinate system is changed to polar coordinates and the coefficient of x^2 is described in terms of the angle with the x-axis, θ. In polar coordinates,

$$x = r\cos(\theta)\ ;\ \ y = r\sin(\theta)$$

and the description of the curvature in the normal plane as a function of direction becomes

$$z = \{\ A\cos^2(\theta) + 2B\cos(\theta)\sin(\theta) + C\sin^2(\theta)\ \}\ r^2 = H(\theta)\ r^2$$

where $H(\theta)$ is the curvature of the normal parabola as a function of θ . $H(\theta)$ is positive if the parabola turns up above the x-y plane and negative if the parabola turns down below.

We can apply the trigonometric identities $\cos^2(\theta) = ½ + ½\cos(2\theta)$,
$\sin^2(\theta) = ½ – ½\cos(2\theta)$ and
$\cos(\theta)\sin(\theta) = ½\sin(2\theta)$ to obtain

$$H(\theta) = A\ \{½ + ½\cos(2\theta)\} + B\ ½\sin(2\theta) + C\ \{½ – ½\cos(2\theta)\}.$$
$$H(\theta) = ½(A + C) + ½(A – C)\cos(2\theta) + ½\ B\sin(2\theta).$$

It is a principle of trigonometry that the sum of sinusoids of the same frequency is another sinusoid of the same frequency shifted in phase, ignoring the exceptional case when the sum is 0. Therefore

$$H(\theta) = H_1 + H_2\cos(2\theta + \alpha). \qquad \text{Equation 6}$$

where H_1, H_2 and α are constants that depend on A_Q, B_Q and C_Q, the values that describe the surface. It is seen that $H(\theta)$ is a sinusoidal oscillation that is raised or lowered by an amount H_1.

Consider the following three cases:

Case 1,when $|H_1| > |H_2|$, $H(\theta)$ is either always positive or always negative as shown in figure 7.
Case 2,when $|H_1| < |H_2|$, $H(\theta)$ crosses the horizontal axis, being positive in some directions and otherwise negative as shown in figure 9.
And last in Case 3, $|H_1| = |H_2|$, $H(\theta)$ is tangent to the horizontal axis, exhibiting zero curvature in two opposing directions and otherwise either always positive or always negative as shown in figure 11.

The maximum and minimum values of $H(\theta)$ are called the principle curvatures, κ_1 and κ_2 of the surface at P. The directions corresponding to κ_1 and κ_2, which are 90° apart, are called the principal directions. Be careful in reading the following to distinguish the Gaussian curvature K from the principle curvatures κ_1 and κ_2 that represent the maximum and minimum parabolic curvatures. When the principle directions are used for the coordinate system, the off-diagonal entries, F_{xy} in the Hessian matrix become equal to 0. The Hessian matrix in this coordinate system becomes

$$\begin{vmatrix} F_{xx} & F_{xy} \\ F_{xy} & F_{yy} \end{vmatrix} = \begin{vmatrix} 2A & 2B \\ 2B & 2C \end{vmatrix} = \begin{vmatrix} \kappa 1 & 0 \\ 0 & \kappa 2 \end{vmatrix}\ .$$

In Case 1, H(θ) is always either above or below the horizontal axis as shown in Figure 7. Both the maximum and the minimum curvatures have the same sign and lie in the principle planes on the same side of the tangent plane. At P, the Gaussian curvature, $K = \kappa_1\kappa_2$ is positive. The 2nd degree surfaces that have positive curvature are the ellipsoids, the elliptical paraboloids and the hyperboloids of two sheets. A surface that has positive curvature everywhere like an ellipsoid or a paraboloid cannot contain a straight line. See figure 8.

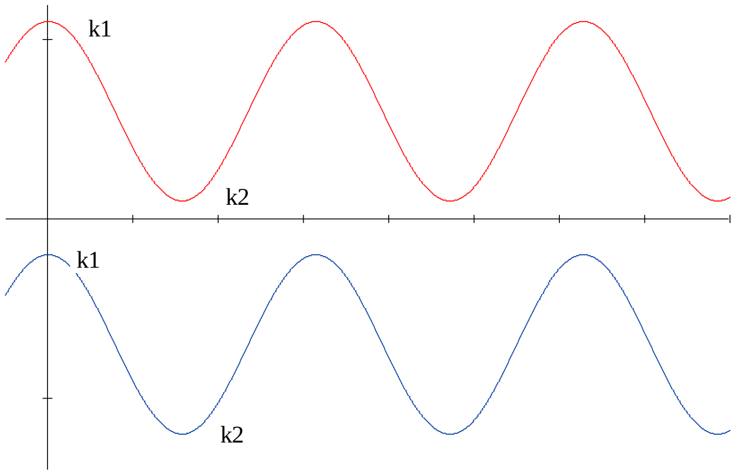
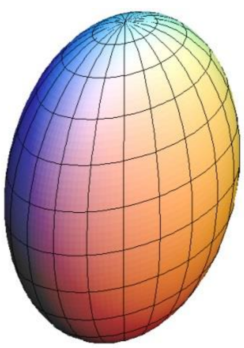
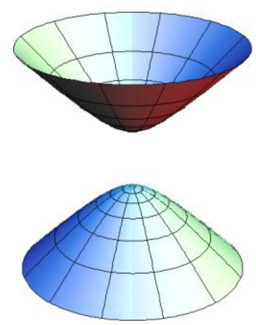


Figure 7 Case 1 Positive Gaussian curvature $|H_1| > |H_2|$



An Ellipsoid[17]

A Paraboloid[18]



An Hyperboloid of Two Sheets[19]

Figure 8 The three 2nd-degree surfaces above have positive Gaussian curvature.

In Case 2, H(θ) crosses the horizontal axis in two places as shown in figure 9. The maximum and the minimum curvatures have opposite signs and lie in the principle planes on opposite sides of the tangent plane. The Gaussian curvature is negative at P. The 2nd-degree surfaces that have negative curvatures are the hyperboloids of one sheet and the hyperbolic paraboloids. There are two values of θ for which H(θ) = 0. In these two directions the surface has zero parabolic curvature. These directions that contain straight lines are called the asymptotic directions.

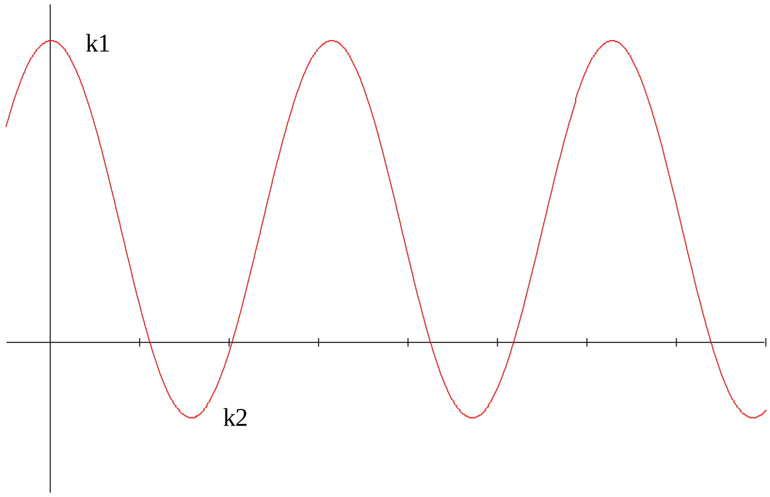


Figure 9 Case 2 The two surfaces that have negative Gaussian curvature $|H_1| < |H_2|$

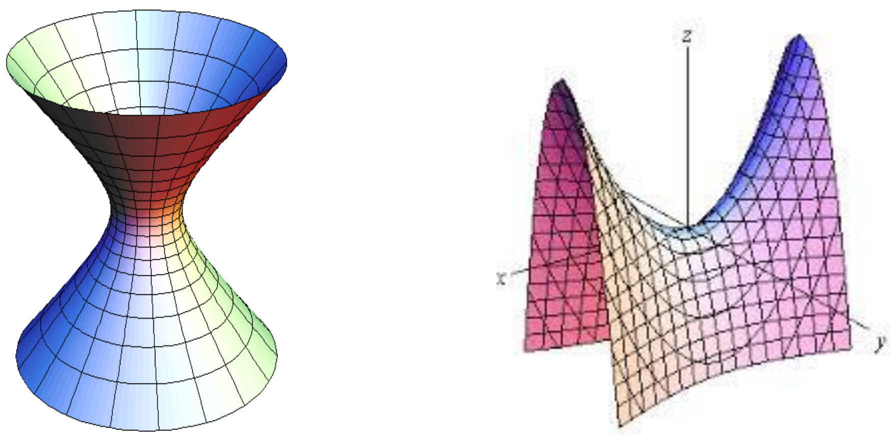


Figure 10 The hyperboloid of one sheet[20] and the saddle surface[21] have negative Gaussian curvatures.

In Case 3, $H(\theta)$ just touches the horizontal axis either at the maximum or at the minimum. Either κ_1 or κ_2 is 0 and the Gaussian curvature is 0. This behavior is exhibited by the cone and the parabolic, the elliptical, and the hyperbolic cylinders. Surfaces that have zero Gaussian curvature everywhere are described as developable, meaning that they can be laid flat on a table if you cut or bend them but do not stretch them. Because they have zero curvature a flat sheet of paper or a painting can be rolled up for easier transportation.

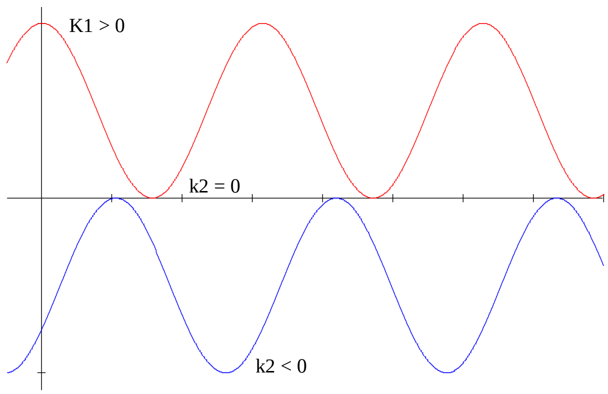


Figure 11 Case 3 Zero Gaussian curvature $| H_1 | = | H_2 |$

It can be shown that the Gaussian curvature of a surface does not change if the surface is only cut or bent but will change if it is stretched. Because of this Gaussian curvature property it is not possible to draw a flat undeformed map of the earth. A flat paper map has zero Gaussian curvature but the earth does not. The reader is referred to the fine images and discussion by Aatish Bhatia in *Wired Magazine* in his article[15] on curvature and strength.

It is possible to generate surfaces in space by moving a straight line. Such surfaces are called ruled surfaces. Examples of ruled surfaces are cones, cylinders, hyperboloids of one sheet and hyperbolic paraboloids. The hyperboloids of one sheet and hyperbolic paraboloids are doubly ruled but are not developable.

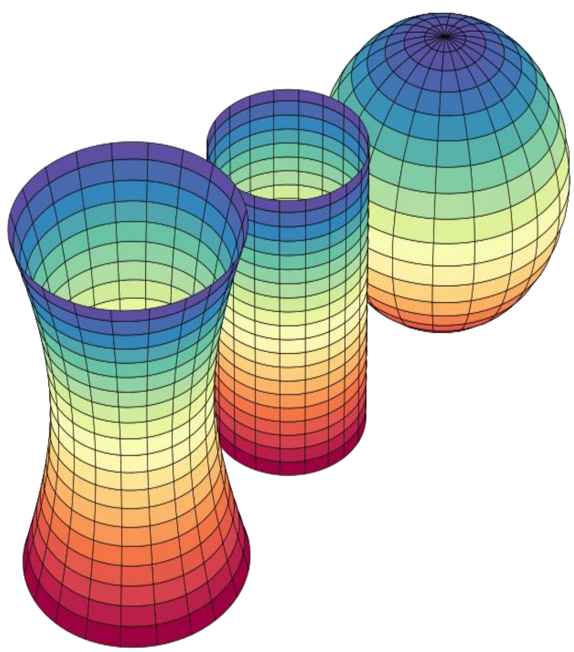


Figure 12 Examples of negative, zero and positive Gaussian Curvatures[22]

Curvature of higher degree surfaces

The surfaces of 2nd-degree are important because they serve as examples of the behavior of more complicated surfaces and because the 2^{nd}-degree terms swamp the higher degree terms. The points where all the 2^{nd}-degree terms vanish at the same time are very special.

Consider the following example. It can be seen in the figure 13 below that points on the outer region of the torus have positive Gaussian curvature and points on the inner region have negative Gaussian curvature. At the top and bottom of the torus there are two circles of points of zero Gaussian curvature which separate the two non-zero Gaussian curvature regions.

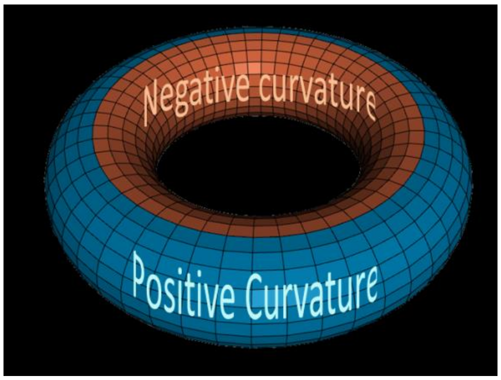


Figure 13 The torus[23]

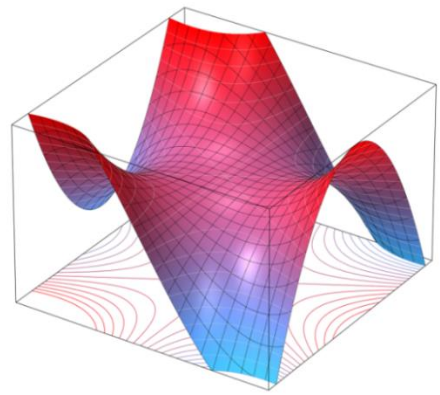


Figure 14 The monkey saddle[24]

Another example is the monkey saddle in figure 14 which has negative Gaussian curvature at every point except its center. At the center of the monkey saddle the Taylor series has 3^{rd}-degree terms but no 2^{nd}-degree terms. A plane rotated about the normal at the center intersects the surface in cubic curves except for the straight lines in 3 directions. At the center of the monkey saddle there are no principle directions or principle curvatures.

On mountain tops, valleys and passes

Applying our study to terrain, at the mountain tops and valleys the Gaussian curvature is positive and at the passes the Gaussian curvature is negative.

There are some remarkable properties of Gaussian curvature that are consequences of the Bertrand-Diquet-Puiseux theorem. Imagine that we examine the region inside a closed path of points that lie at a small constant distance, r, from a fixed point, P. If at P our surface has zero Gaussian curvature, the area will be πr^2 and the circumference will be $2\pi r$. If at P our surface has positive Gaussian curvature, the area will be less than πr^2 and the circumference will be less than $2\pi r$. And vice versa, if at P our surface has negative Gaussian curvature, the area will be greater than πr^2 and the circumference will be greater than $2\pi r$. If the space curves which consist of points on a surface at a constant distance from a fixed point are called circles then circles on hilltops and valleys have smaller circumferences and areas than circles located at the mountain passes.

Here is an example. On a sphere of radius R the area enclosed by all points less than a distance of $\pi/2$ R from a fixed point is ½ the total area of the sphere or $2\pi R^2$. On a flat plane the area enclosed by all points less than a distance of $\pi/2$ R from a fixed point is $\pi \{ \pi/2 \text{ R} \}^2$ which is equal to $\{ \pi^2/4\} \pi R^2$ or $2.47\ \pi R^2$. On the sphere the circumference of a circle at a distance of $\pi/2$ R from a fixed point is $2\pi R$ while on a flat plane the circumference of the circle of points at a distance of $\pi/2$ R from a fixed point is $2\pi \{\pi/2 \text{ R}\}$ which equals $\pi \{\pi \text{ R}\}$. It is found that in both cases the circumference and the area on the plane are greater than the circumference and the area on the sphere.

On partial differential equations

The visual approach can easily provide insight to the nature of the surfaces which represent solutions to Laplace's partial differential equation, $\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0$.

Laplace's partial differential equation describes the fundamental solutions in such disciplines as electrostatics, magnetic fields, fluid potentials and steady state thermodynamics. If Laplace's equation holds inside some region then

$$\frac{\partial^2 z}{\partial x^2} = -\frac{\partial^2 z}{\partial y^2}$$

and the Gaussian curvature is negative. This means that at every interior point, if in one direction the surface bends down then in the orthogonal direction the surface bends up. These points cannot be extreme. Since this equation prevails at every interior point, the extreme values of the surface cannot lie in the interior of the region and therefore must lie on the boundary.

Summary

It seems that this paper naturally divides into three parts. The 1st part treats the forms and notation for describing straight lines, flat planes, curves and surfaces and their derivatives in a 3-dimensional space described by a Cartesian orthonormal coordinate system.

The 2nd part treats the complexities of flat planes. While only one parameter is required to describe the direction of a straight line in a 2-dimensional space, the relationships between directional derivatives, partial derivatives and the gradient in a 3-dimensional space take some explaining that is not customarily available in conventional math texts.

The 3rd part treats the curvature of surfaces and curves in 3-dimensional space, which leads to the important surface descriptor, Gaussian curvature. The simple deviations at a point of a surface from its tangent plane can be described by nine 2nd-degree functions of two variables. Three surfaces, the paraboloid, the ellipsoid and the hyperboloid of two sheets have positive Gaussian curvature everywhere. Two others, the hyperboloid of one sheet and the hyperbolic paraboloid have negative Gaussian curvature. The three cylinders have zero Gaussian curvature as does the elliptic cone.

Differential single variable calculus can be compressed to exploring three concepts:

Position	The original function		$y = f(x)$
Direction	The derivative function	the slope of the tangent line	$y'(x)$
Turning	The 2nd derivative function	the deviation from the tangent line	$y''(x)$

This paper extends this progression to multivariable calculus:

Position	The original equation	$z = F(x, y)$
Orientation	The gradient of the tangent plane	$\nabla z = [z_x \;\; z_y] = [\frac{\partial z}{\partial x} \;\; \frac{\partial z}{\partial y}]$
Curvature	The shape operator	$\begin{vmatrix} z_{xx} & z_{xy} \\ z_{xy} & z_{yy} \end{vmatrix}$

We see that the three entries in the shape operator are related to the three coefficients of the 2^{nd}-degree Taylor polynomial which is the simplest warped approximation to the surface. And now we can also comprehend Gauss's great discovery that the determinant of the Hessian matrix or shape operator is an intrinsic characterization of curvature at a point on a surface reduced to a single number, K, the value of the appropriately named Gaussian curvature.

References

1. Aleksandrov, et al <u>Mathematics Its Content, Methods and Meaning</u> Vol II The MIT Press, Cambridge MA, 1963
2. Alsina, Claudi & Nelson, Roger B. <u>Math Made Visual</u> MAA, Washington, 2006
3. Dray, Tevian. <u>Using Differentials to Differentiate Trigonometric and Exponential Functions</u> The College Mathematics Journal, Vol. 44 No. 1 January 2013 pg's. 18, 19
4. Grossfield, Andrew <u>Mathematical Forms and Strategies</u> ASEE Annual Conference, 1999
5. Grossfield, Andrew <u>Are Functions Real?</u> ASEE Annual Conference, 2005
6. Grossfield, Andrew <u>Visual Analysis and the Composition of Functions</u> ASEE Annual Conference, 2009
7. Grossfield, Andrew <u>Introducing Calculus to the High School Curriculum: Curves, Branches and Functions</u> ASEE Annual Conference, 2013
8. Grossfield, Andrew <u>Visual Differential Calculus</u> ASEE Zone 1 Conference, 2014
9. Nelson, Roger B. <u>Proofs without Words: Exercises in Visual Thinking</u> MAA Washington, 1993
10. Nelson, Roger B. <u>Proofs without Words II: More Exercises in Visual Thinking</u> MAA Washington, 2000
11. Dawkins, Paul http://tutorial.math.lamar.edu
12. Parris, Rick WINPLOT, a general-purpose plotting utility, Peanut Software Figures not listed below were drawn with Winplot
13. Struik, Dirk J. *Lectures on Classical Differential Geometry* Dover Publications Inc. New York 1961
14. O'Neill, Barrett *Elementary Differential Geometry* Academic Press 1966
15. Aatish Bhatia Wired Magazine *How a 19th Century Math Genius Taught Us The Best Way To Hold A Pizza Slice* https://www.wired.com/2014/09/curvature-and-strength-empzeal/

16 Principle planes, retrieved from https://en.wikipedia.org/wiki/Gaussian_curvature#/media/File:Minimal_surface_curvature_planes-en.svg

17. Ellipsoid, retrieved from http://virtualmathmuseum.org/Surface/ellipsoid/ellipsoid.html
18. Paraboloid, retrieved from http://virtualmathmuseum.org/Surface/paraboloid/paraboloid.html
19. Hyperboloid of two sheets, retrieved from http://virtualmathmuseum.org/Surface/hyperboloid2/hyperboloid2.html
20. Hyperboloid of one sheet, retrieved from http://virtualmathmuseum.org/Surface/hyperboloid1/hyperboloid1.html
21. Hyperbolic paraboloid, retrieved from http://tutorial.math.lamar.edu/Classes/CalcIII/QuadricSurfaces.aspx
22. Gaussian curvatures 3 examples, retrieved from https://en.wikipedia.org/wiki/Gaussian_curvature#/media/File:Gaussian_curvature.svg
23. Torus, retrieved from https://en.wikipedia.org/wiki/Gaussian_curvature#/media/File:Torus_Positive_and_negative_curvature.png
24. Monkey saddle, retrieved from, http://virtualmathmuseum.org/Surface/monkey_saddle/monkey_saddle.html