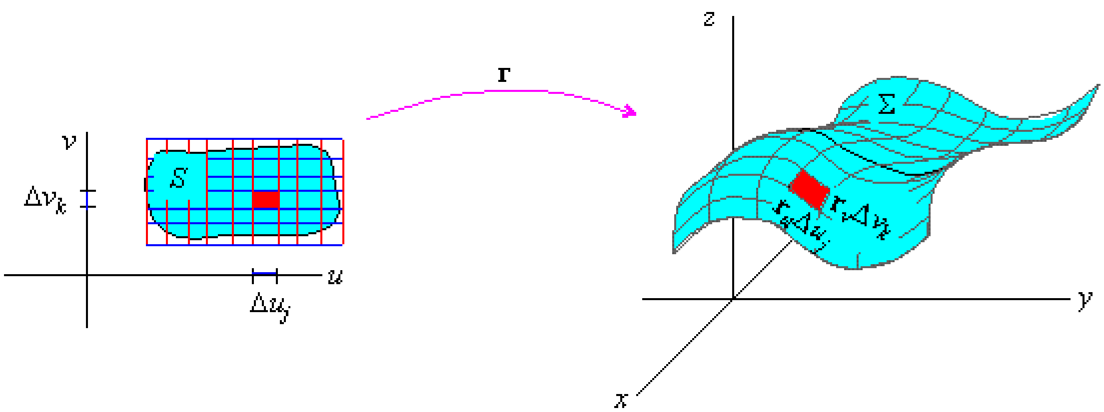
Comments on Colley, Section 7.1

We defined curves in terms of parametrizations, and in principle we would like to define surfaces likewise. However, there are some complications because we want to have a concept of surface which is broad enough to include most if not all the objects which are generally called surfaces in ordinary contexts.

One simple way to define a parametrized surface in coordinate **3 –** space is by means of a **3 –** dimensional vector valued function $\mathbf{X}(u, v)$ of two variables.



(Source: http://math.etsu.edu/MultiCalc/Chap5/Chap5-5/index.htm)

This definition includes the graphs of some function $f(u, v)$ of two variables, in which case the parametrization is given by

$$\mathbf{X}(u, v) \;=\; \big(u, v, f(u, v)\big)$$

and in order to define tangent planes we need to assume that f has continuous partial derivatives. In order to avoid complicated discussions about defining partial derivatives at boundary points of regions, we also need to stipulate that the function f is defined on an open set as defined in the previous course. More generally, the parametric equations for a smoothly parametrized surface are assumed to have the form

$$\mathbf{X}(u, v) \;=\; \big(x(u, v), y(u, v), z(u, v)\big)$$

where, as before, the coordinate functions x, y, z are assumed to have continuous partial derivatives over some open set. However, in order to obtain a decent notion of tangent plane we need an additional condition, just as we needed to assume the tangent vectors to parametrized curves were nonzero in order to define tangent lines. The condition may be stated as follows: Define the vector valued partial derivative functions

$$\frac{\partial X}{\partial u} = \left(\frac{\partial x}{\partial u}, \frac{\partial y}{\partial u}, \frac{\partial z}{\partial u}\right) \quad , \quad \frac{\partial X}{\partial v} = \left(\frac{\partial x}{\partial v}, \frac{\partial y}{\partial v}, \frac{\partial z}{\partial v}\right) .$$

Not only do we want these two partial derivatives to be nonzero everywhere, but we also want them to be linearly independent everywhere, so that neither is a nonzero scalar multiple of the other. This is conveniently summarized by the ***cross product condition***

$$\frac{\partial X}{\partial u} \times \frac{\partial X}{\partial v} \neq \mathbf{0} .$$

This condition automatically holds for the graph of a function, in which case the cross product is given by $(-f_u, -f_v, 1)$; the reader should check this directly using the definitions of partial derivatives and cross products. If we compare this with Theorem 2.3.3 on page 112 of the text, we see that this cross product is an upward normal vector to the tangent plane at a given point $\mathbf{p} = (u_0, v_0, f(u_0, v_0))$. Finally, for the sake of simplicity it is extremely useful to eliminate surfaces with self – intersections, so our default assumption will be that the parametrization is ***one – to – one***: Namely, ***the parametrization function* X *takes different points in its domain to distinct points in space***.

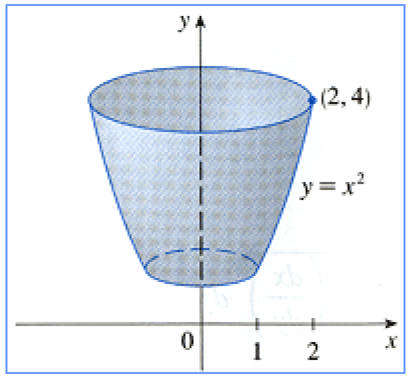
More generally, the tangent plane to a parametrized surface at some point (u_0, v_0) is defined to be the plane through the surface point $\mathbf{X}(u_0, v_0)$ for which the ***standard normal vector*** $\mathbf{N}(u_0, v_0)$ is equal to the cross product

$$\frac{\partial X}{\partial u} \times \frac{\partial X}{\partial v} .$$

In other words, the equation of the tangent plane to **X** at **p** is given by the set of all vectors **w** satisfying the equation $\mathbf{N}(u_0, v_0) \cdot (\mathbf{w} - \mathbf{p}) = 0.$

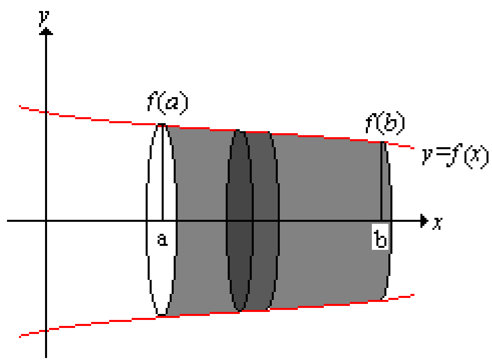
Here are some basic examples beyond graphs of functions.

Surfaces of revolution. Suppose that we are given a curve which is the graph of a smooth (continuously differentiable) function $y = f(x)$, where f is always positive valued and (say) $0 \leq a < x < b$. We can use such a curve to define surfaces of revolution about either the x – axis or the y – axis.



Revolution about the y – axis

(Source: http://curvebank.calstatela.edu/arearev/arearev.htm)



Revolution about the x – axis

(Source: http://www.mathresources.com/products/mathresource/maa/surface_of_revolution.html)

Here is a link to an animated graphic:

http://curvebank.calstatela.edu/arearev/rev3cont.gif

In the first case (around the y – axis) the most straightforward paremetrization is given by

$$\mathbf{X}(u,\, v) \;=\; (u\cos v,\; f(u),\; u\sin v)$$

and in the second (around the x – axis) the most straightforward parametrization is given by

$$\mathbf{X}(u,\, v) \;=\; (u,\; f(u)\cos v,\; f(u)\sin v).$$

The coordinate functions satisfy the smoothness condition, and the cross products of the partials of $\mathbf{X}$ with respect to u and v are given by

(1) $\mathbf{N}(u,\, v) \;=\; (u f'(u)\cos v,\; -u,\; u f'(u)\sin v)$ in the y – axis case,

(2) $\mathbf{N}(u,\, v) \;=\; (f(u)f'(u),\; -f(u)\cos v,\; -f(u)\sin v)$ in the x – axis case.

These formulas show that the lengths of the normal vectors are equal to $\boldsymbol{u}$ and $\boldsymbol{f(u)}$ times the square root of $\mathbf{1 + f'(u)^2}$ respectively**;** in each case the length of the cross product is given as a product of two positive numbers and hence is positive.

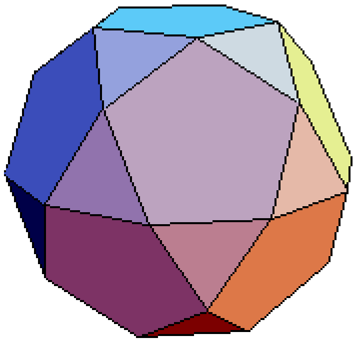
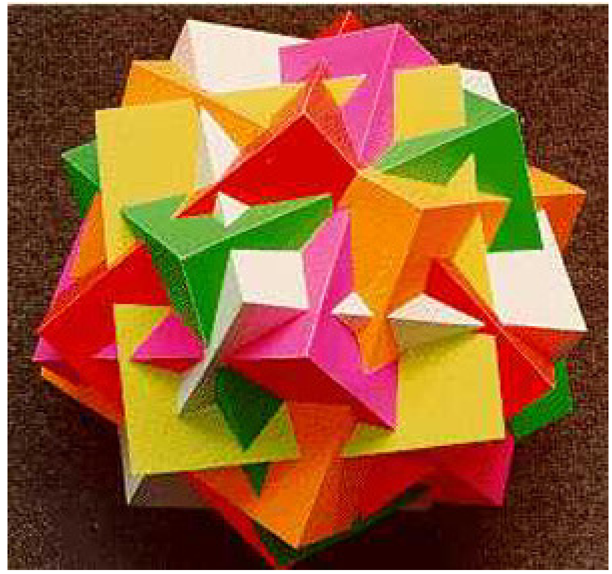
Examples of surfaces of revolution about the $\boldsymbol{x}$ – axis include a ***sphere*** of radius 1 and center (1, 0, 0) with the poles on the $\boldsymbol{x}$ – axis removed, in which case the curves are given by $\boldsymbol{y} = \mathbf{SQRT(1 - (}x\mathbf{ - 1)^2)}$ where $\mathbf{1 < \boldsymbol{x} < 2}$, ***right circular cylinders***, in which case the curves are given by $\boldsymbol{y} = \boldsymbol{r}$ for some positive constant $\boldsymbol{r}$ (and the $\boldsymbol{x}$ – limits are arbitrary nonnegative numbers), and ***right circular cones*** with the top vertices removed, in which case the curves are given by $\boldsymbol{y} = \boldsymbol{m}\boldsymbol{x}$ for some positive constant $\boldsymbol{m}$ (and the $\boldsymbol{x}$ – limits are $\mathbf{0}$ and some positive value $\boldsymbol{h}$).

More generally, if $\mathbf{\Gamma}$ is a parametrized regular smooth curve in the first quadrant of the coordinate plane, it is possible to extend the preceding constructions to obtain surfaces of revolution with respect to the $\boldsymbol{x}$ – and $\boldsymbol{y}$ – axes. However, we shall pass on describing such generalizations explicitly and move to another crucial family of surfaces which arise naturally.

Piecewise smooth surfaces

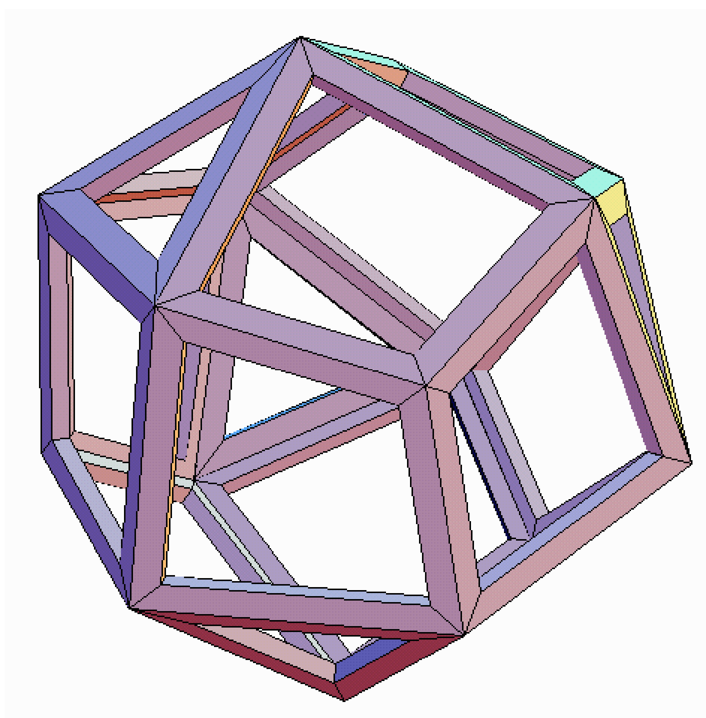
There are still other examples beyond the smooth surfaces considered just far; just as it is often convenient to consider piecewise smooth curves, we also want to consider piecewise smooth surfaces in many situations. One major complication is that it is definitely more difficult to put the pieces of such a surface together than it is to do the same thing for curves, and because of this we have to set up the definitions carefully.

Recall that an ***elementary region*** is one which is bounded by a pair of vertical lines $\boldsymbol{x} = \boldsymbol{a}, \boldsymbol{x} = \boldsymbol{b},$ and the graphs of two functions $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}),\ \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}).$ This concept figures in the definition of piecewise smooth surfaces which appears in Definition 7.1.3 on page 413 of the course text. Intuitively, one expects that a cube should be an example of a piecewise smooth surface, and a detailed verification of this point is given following the definition on pages 413 – 414 of the course text. Here are some further pictures showing piecewise smooth surfaces such that each piece is a flat portion of some plane; in classical geometry these are called ***polyhedra.***

(Sources: http://cset.mnsu.edu/mathstat/images/polyhedron.gif http://whyfiles.org/coolimages/images/csi/cubes.jpg)

And here is an example of a "surface with holes" to illustrate the possibilities even further.



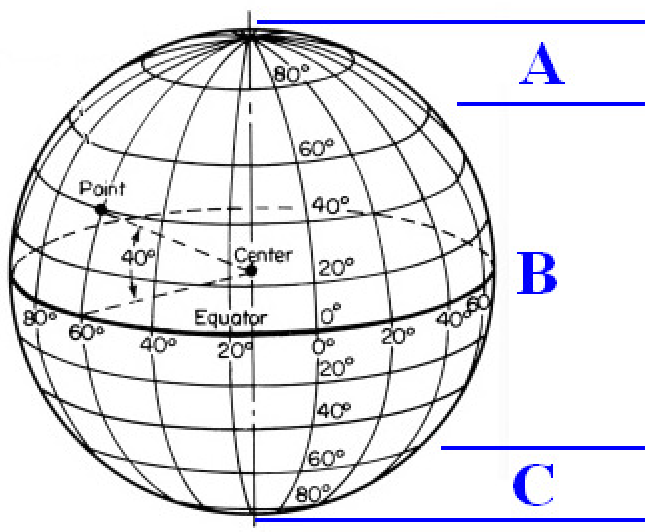
(Source: http://www.georgehart.com/cccg/Image75.gif)

The concept of piecewise smooth surface is also useful because it allows one to include many smooth – looking examples, such as spheres, which are at best awkward to describe using single parametrizations. In particular, the standard unit sphere with defining equation $x^2 + y^2 + z^2 = 1$ can be described using three parametrized pieces as follows:

(1) The ***upper third*** of the northern hemisphere above 60 degrees north latitude, defined by the parametrization $A(u, v) = (u, v, \text{SQRT}(1 - u^2 - v^2))$, where $u^2 + v^2 \le 1/4$.

(2) The ***equatorial – temperate band*** of the sphere between 60 degrees north and south latitudes, defined by the spherical coordinate parametrization $B(u, v) = (\cos u \sin v, \sin u \sin v, \cos v)$, where $0 \le u \le 2\pi$ and $-\pi/3 \le v \le \pi/3$.

(3) The ***lower third*** of the southern hemisphere below 60 degrees south latitude, defined by the parametrization $C(u, v) = (u, v, -\text{SQRT}(1 - u^2 - v^2))$, where $u^2 + v^2 \le 1/4$.

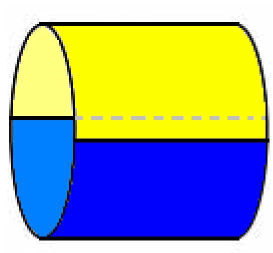


(Adapted from http://www.literacynet.org/sciencelincs/showcase/drifters/images/globe.jpg)

Note that in each case the parametrization is actually definable on thickenings of the given domains; specifically, in the first and third cases the parametrizations can be extended to $u^2 + v^2 < 1$, and in the second case the parametrizations can be extended to all values of u and v such that $-\pi/2 < v < \pi/2$.

Strictly speaking, the second example probably should be split into two pieces, one over the eastern hemisphere (defined by the condition $y \ge 0$) and one over the western hemisphere (defined by the condition $y \le 0$), in order to ensure that the parametrization maps are one – to – one.

One can proceed similarly with ***solids of revolution*** as described earlier; we shall assume that the continuously differentiable function f extends to a function with similar properties on an open interval containing the closed interval $[a, b]$. Then one can view the surface of revolution as being given by two parametrized pieces, one of which is the restriction of the standard parametrization to all (u, v) such that $a \leq u \leq b$ and $0 \leq v \leq \pi$, and the other of which is the restriction of the standard parametrization to all (u, v) such that $a \leq u \leq b$ and $\pi \leq v \leq 2\pi$. An illustration of this in the case of an ordinary circular cylinder is given below.



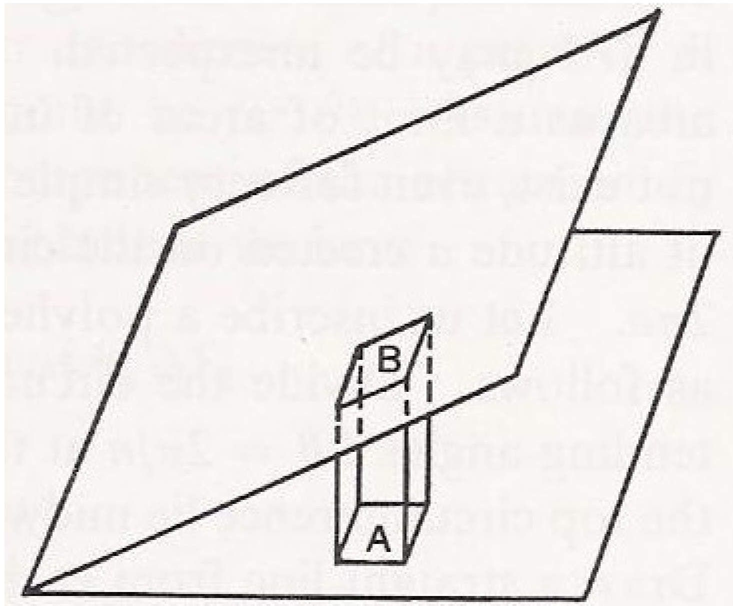
As suggested by this picture, the surface of revolution may be viewed as a pair of identical pieces (one yellow, one blue) that are glued together along certain boundary curves. In practice, we usually take a single parametrization such that $0 \leq v \leq 2\pi$ because such a parametrization is one – to – one almost everywhere.

Surface area

It is fairly straightforward to discuss arc length of curves and to derive integral formulas for arc lengths, and in many respects the theory of surface area can be viewed as a **2 –** dimensional analog of the theory of arc length. However, there are also some very important differences, and there are several phenomena that are at least somewhat surprising.

We shall begin the discussion with surfaces which are graphs of scalar – valued functions $z = f(x, y)$ of two variables, and in fact we shall begin with the simplest possible functions, which are linear functions $z = ax + by + c$; in such cases the graphs are simply ***planes,*** and thus we can analyze the situation using basic vector geometry

<u>CLAIM.</u> If we are given a small solid rectangular region A in the xy – plane with length L and width W, and B is the piece of the plane which lies above A, then it follows that B is a solid planar region bounded by a parallelogram, and it turns out that the area of B is equal to $L \cdot W \cdot \mathbf{SQRT}(1 + a^2 + b^2)$.



(Source for the drawing: *Widder*, Advanced Calculus, p. 203)

Why does this expression give the area? The original rectangular region A is the set of all points for which $p \leq x \leq p + L$ and $q \leq y \leq q + W$ for some p and q. The vertices of the parallelogram B are then given by the vectors

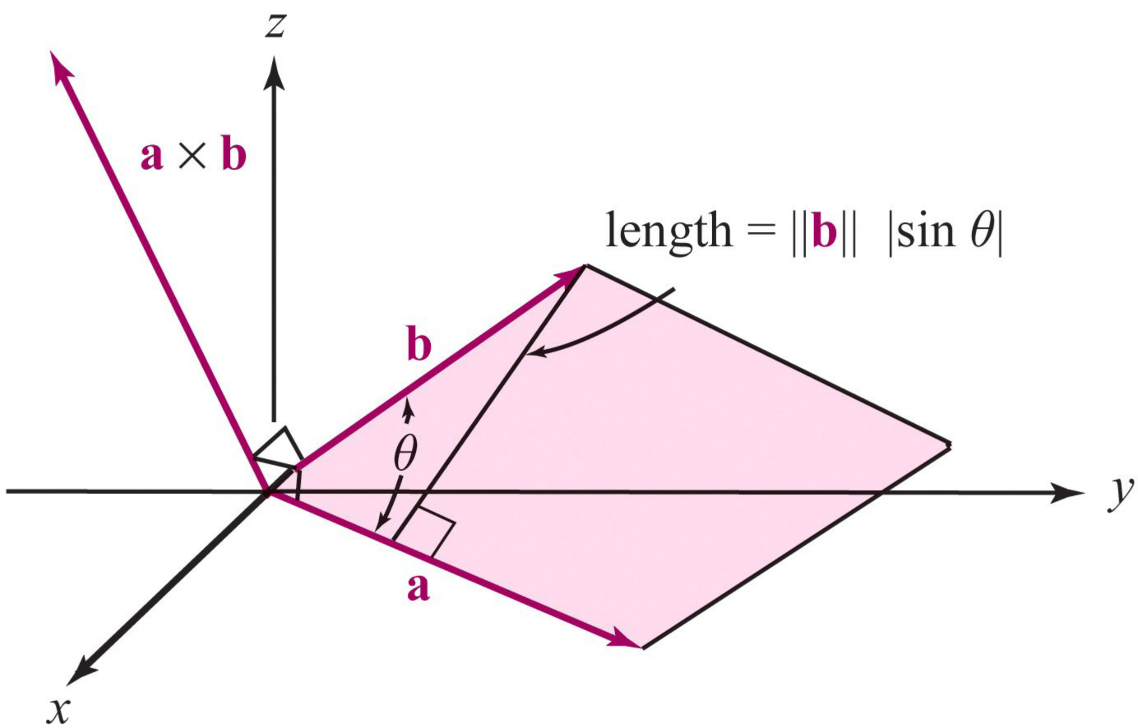
$$(p, q, ap + bq + c), \ (p + L, q, ap + aL + bq + c),$$
$$(p, q + W, ap + bq + bW + c), \ (p + L, q, ap + aL + bq + bW + c)$$

and therefore two adjacent sides of the parallelogram B correspond to the vectors $\mathbf{U} = L(1, 0, a)$ and $\mathbf{V} = W(0, 1, b)$. Thus the area of the solid parallelogram **B** is equal to the area of the parallelogram determined by the vectors **U** and **V.**

It turns out that this area is given by the length of the cross product $\mathbf{U} \times \mathbf{V}$. More generally, if $\mathbf{a}$ and $\mathbf{b}$ are any two nonzero vectors in 3 – space such that one is not a scalar multiple of the other, then the area of the parallelogram determined by $\mathbf{a}$ and $\mathbf{b}$ is equal to the length of the cross product $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}|\,|\mathbf{b}| \sin\theta$, where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$. An illustration taken from

http://stat.asu.edu/~eric/mat272/lect0403.pdf

is given below, and the cited file also has additional information of the derivation of the area formula.

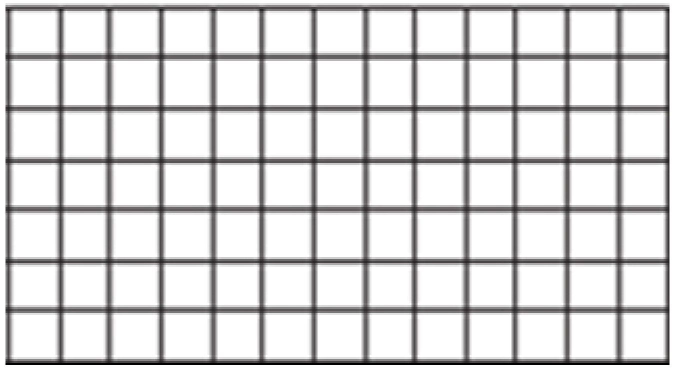


Returning to the issue of computing the area bounded by the parallelogram B, we may use the formula to conclude that this area is the length of $\mathbf{U} \times \mathbf{V}$**,** and direct calculuation implies that this length is equal to $L \cdot W \cdot \mathbf{SQRT}(\mathbf{1} + a^2 + b^2)$**.** For later reference, we note that the square root term is the length of the standard normal vector $\mathbf{N}(x, y)$ for the standard parametrization of the plane:

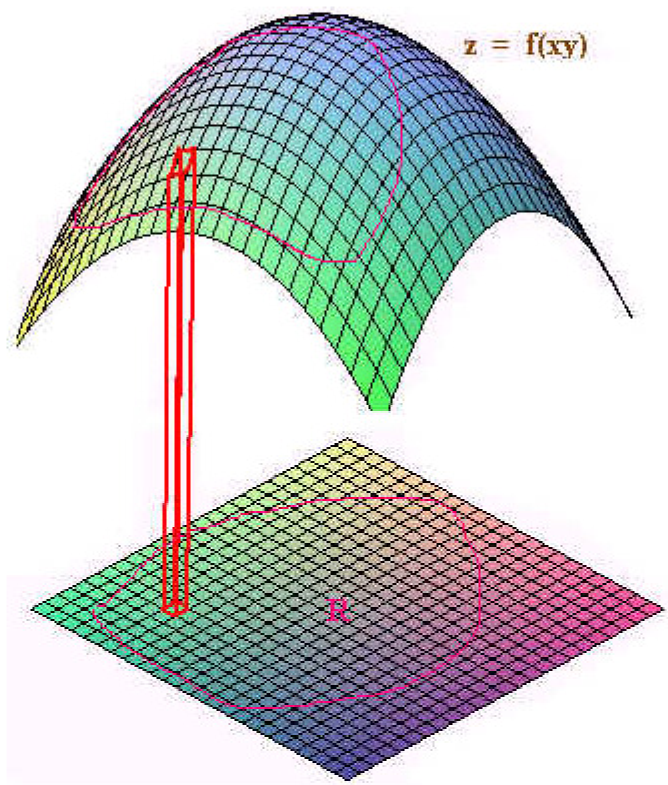
$$\mathbf{X}(x, y) = (x, y, ax + by + c)$$

(Recall that **N** is the cross product of the partial derivatives of **X** with respect to the two variables**.)**

<u>Surface areas of more general graphs.</u> Suppose now that we are given a surface which is the graph of an arbitrary function $z = f(x, y)$ with continuous partial derivatives; for the sake of simplicity, assume for the time being that the function is defined on a solid rectangular region whose lower left corner is (a, c) and whose upper right corner is (b, d)**.** Assume further that we have decomposed this region into smaller nonoverlapping rectangular regions in a standard grid pattern like the following drawing.



If we take a fine enough grid, then for each small rectangle $\boldsymbol{A}$ it is reasonable to expect that the area of the piece of the graph over $\boldsymbol{A}$ is approximately equal to the area of a piece of some tangent plane over $\boldsymbol{A}$**,** as in the picture below.



(Source: http://www.ltcconline.net/greenl/courses/202/multipleIntegration/surfaceArea.htm)

More precisely, choose some point $(\boldsymbol{p_A}, \boldsymbol{q_A})$ in $\boldsymbol{A}$**,** and let $\boldsymbol{T_A}$ be the tangent plane to the graph at $\big(\boldsymbol{p_A}, \boldsymbol{q_A}, \boldsymbol{f}(\boldsymbol{p_A}, \boldsymbol{q_A})\big)$; then the approximate value we want is the area of the portion of $\boldsymbol{T_A}$ which lies over $\boldsymbol{A}$, and by the preceding discussion this area is equal to

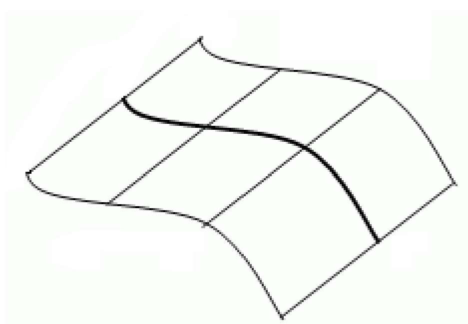
$$\mathbf{SQRT}\left(1 + f_{1st}(p_A, q_A)^2 + f_{2nd}(p_A, q_A)^2\right) \cdot \Delta x_A \cdot \Delta y_A$$

where f_{1st} and f_{2nd} denote the partial derivatives with respect to the first and second variables.

The total estimate for the surface area will be the sum of all such terms with A running through all the small squares in the grid. As usual, if we take finer and finer grids, we expect to get better and better approximations to the surface area, and we also expect that the actual surface area should be a limit of such approximations. This turns out to be the case, but we shall not try to prove it here (the most efficient approach requires some fairly sophisticated tools which are beyond the scope of this course). In fact, we can also carry out a similar analysis if the function is defined on some "reasonable" set D of the usual type(s), and the result is the following standard **surface area formula**:

$$S = \iint_D \sqrt{\left(\frac{\partial z}{\partial x}\right)^2 + \left(\frac{\partial z}{\partial y}\right)^2 + 1}\, dx\, dy.$$

It is usually worthwhile to check that new, general formulas yield the answers one has already obtained in special cases. By construction, this is true if the function is linear. Suppose next that we have a cylindrical surface for which the function f can be rewritten as a function $g(x)$, so that its value depends only upon the first variable and not the second.

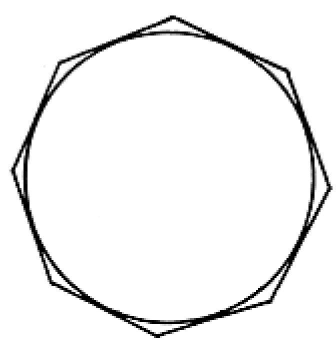


In this case the partial derivative of f with respect to the second variable is zero, and if the region D is rectangular, and if the lower left and upper right corners then the formula implies that the surface area equal to the product $L(d-c)$, where L is the length of the graph curve defined by $z = g(x)$ on the plane $y = c$. This formula is consistent with physical experiments in which one takes a sheet of paper (or metal) and bends it into a cylindrical surface. Finally, if we are given a surface of revolution obtained by revolving the graph of some function $z = g(x)$ around the $z-$ axis, then the surface of revolution is the graph of $z = g(x^2 + y^2)$. In this case the surface area formula yields a double integral over the ring (or annulus) shaped region $a \le x^2 + y^2 \le b$; if we compute the partial derivatives of z with respect to x and y, and then convert the integral to polar coordinates, the result will be the usual integral formula for computing surface areas by the shell method from single variable calculus (verify this!). Near the end of this document we shall give a second argument to show that our formula for surface area yields the same answer as the one from single variable calculus for surfaces of revolution about the y – axis.

Comparison with the arc length definition. In the usual definition of arc length, one approximates a curve by broken lines which are inscribed in the curve.

(Source: http://en.wikipedia.org/wiki/Arc_length)

As noted on page 204 of Widder, ***Advanced Calculus***, this definition of surface area "*may be unexpected. It might seem more natural to consider the area as a limit of areas of inscribed polyhedral*." However, the subsequent discussion in Widder explains why such an approach does not work; specifically, "*the … limit need not exist, even for very simple surfaces*." The **1 –** dimensional analog of this standard approach to surface area does not involve approximating small pieces of the curve by secant lines joining a pairs of nearby points on the curve, but rather by taking the lengths of small pieces of tangent lines. In several respects the approach generalizes the approximation to the circumference of a circle by means of circumscribed polygons.



More generally, if one considers approximations to curves by means of pieces of tangent lines, then it turns out that the approximations become better as one refines the decomposition of the interval on which the parametrization is defined, and a suitably defined limit is equal to the same integral formula obtained for the arc length using inscribed broken line approximations.

Surface areas for general parametrizations. Pages 414 – 416 of the course text discusses surface areas for parametrizations in the general case, for which

$$\mathbf{X}(u, v) \;=\; \big(x(u, v), y(u, v), z(u, v)\big)$$

where the coordinate functions have sufficiently many continuous partial derivatives and the cross product of the partial derivative vectors

$$\mathbf{N}(u, v) \;=\; \mathbf{X}_{1st}(u, v) \times \mathbf{X}_{2nd}(u, v)$$

is nonzero ($\mathbf{X}_{1st}$ and $\mathbf{X}_{2nd}$ denote partial derivatives with respect to the appropriate variables). One standard version of the area formula is given by

$$\textit{Surface area} \;=\; \iint_D |\mathbf{X}_{1st}(u,v) \times \mathbf{X}_{2nd}(u,v)|\, dA \;=\; \iint_D |\mathbf{N}(u,v)|\, dA$$

and another extremely useful version of the area formula is stated as item **(8)** on page 416 in terms of Jacobians. Specifically, since the integrand is given by

$$\mathbf{N}(u,v) \;=\; \frac{\partial \mathbf{X}}{\partial u} \times \frac{\partial \mathbf{X}}{\partial v} \;=\; \left(\frac{\partial(y,z)}{\partial(u,v)}, \frac{\partial(z,x)}{\partial(u,v)}, \frac{\partial(x,y)}{\partial(u,v)}\right)$$

it follows that one can express the integrand in the following form:

$$|\mathbf{N}(u,v)| \;=\; \sqrt{\left(\frac{\partial(y,z)}{\partial(u,v)}\right)^2 + \left(\frac{\partial(z,x)}{\partial(u,v)}\right)^2 + \left(\frac{\partial(x,y)}{\partial(u,v)}\right)^2}$$

Interpretation for surfaces of revolution. Single variable calculus courses often give formulas for the surface areas of surfaces of revolution, so we should really check that our surface area reduces to the previously known ones in such cases. For surfaces of revolution with respect to the y – axis, a previously derived formula shows that the integrand $|\mathbf{N}(u, v)|$ is equal to $u \cdot \mathbf{SQRT}(1 + f'(u)^2)$. Since the parametrization is defined for u between a and b and v between $\mathbf{0}$ and $\mathbf{2\pi}$, it follows that the surface area is equal to

$$\int_0^{2\pi} \int_a^b u\sqrt{1 + f'(u)^2}\, du\, dv \;=\; 2\pi \int_a^b u\sqrt{1 + f'(u)^2}\, du\,.$$

Likewise, for surfaces of revolution with respect to the x – axis, the corresponding formula shows that the integrand $|\mathbf{N}(u, v)|$ is equal to $f(u) \cdot \mathbf{SQRT}(1 + f'(u)^2)$; the parametrization is again defined for u between a and b and v between $\mathbf{0}$ and $\mathbf{2\pi}$, so in this case the surface area is equal to

$$\int_0^{2\pi} \int_a^b f(u)\sqrt{1 + f'(u)^2}\, du\, dv \quad = \quad 2\pi \int_a^b f(u)\sqrt{1 + f'(u)^2}\, du.$$

In each case, the expression on the right hand side equals the surface area formula from single variable calculus.

Final remarks

Extension to piecewise smooth surfaces. Suppose now that we are given a piecewise smooth surface, with smooth parametrizations defined over (open sets containing) the regular regions $D_1, D_2, \ldots, D_k$. In such cases the total surface area is given by the sum of the surface areas of the smooth pieces.

$$\textit{Total surface area} \quad = \quad \sum_{i=1}^{k} \iint_{D_i} |\mathbf{N}(u,v)|\, dA$$

This formula is probably not surprising, but it is included for the sake of completeness.

Independence of parametrizations. By construction, all surface area formulas depend upon choosing parametrizations. It is possible to show that the values obtained for surface areas do not depend upon choices of parametrizations, but this requires concepts beyond the scope of this course.

Level surfaces. The unit sphere is a special case of a ***bounded nonsingular level surface in 3 – space***, which is the set $\mathbf{S}$ of all (x, y, z) such that $f(x, y, z) = 0$ for some function f with continuous partial derivatives such that $\mathbf{S}$ is ***bounded*** (it lies inside some large disk or cube) and the following ***nonsingularity condition*** is satisfied:

For all points (x, y, z) *in* $\mathbf{S}$, *the gradient* $\nabla f(x, y, z)$ *is nonzero.*

One can check directly that these conditions apply to the sphere defined by the equation $x^2 + y^2 + z^2 - 1 = 0$. More generally, every bounded nonsingular level surface in **3** – space admits a decomposition into a piecewise smooth surface. However, a rigorous proof of this fact requires graduate level mathematics.