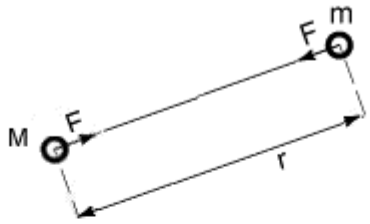
Gravity and the Earth
Newtonian Gravity and Earth Rotation Effects

James R. Clynch, 2003

I. Newtonian Gravity

A. Newtonian Gravitational Force

Newton published his law of gravitational attraction in 1687. This says that all things that have mass attach each other. The attraction is proportional to the product of the masses and inversely proportional to the distance between them.



$$\left|\vec{F}\right| = \frac{GMm}{r^2}$$

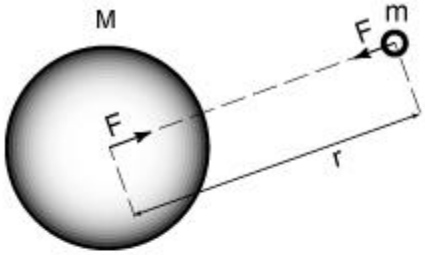
Newton's Law of Gravity

Two Masses, M and m, Attract Each Other With the Force, |F|.

The force is directed toward the other mass. Each attracts the other.

The value G is a constant that was first measured by Cavendish in 1798 using lead balls.

The two masses are assumed to be points. However it can be shown that if the masses are spherical symmetric the same equation holds using the center of symmetry as the location for the total mass of the object. The equation holds for extended masses in this special, but useful case.



$$\left|\vec{F}\right| = \frac{GMm}{r^2}$$

Newton's Law of Gravity For Extended Masses

The equation holds when the masses are spherically symmetric. The force acts as if all the mass is at the center.

This can be proved by looking at the mass as being made up of an infinite number of small points of volume dV and mass $dm = \rho dV$ where the density is ρ. The effects of all these points is added up. This is the integral of calculus. (See the basic physics note for a few details.) This procedure is required for the general case.

In the spherically symmetric case you get the result above, but only for points outside the mass. For points inside the mass, only the mass that is at a radius less than that of the test mass counts. So the force decreases as you go inside a spherical mass. (These results are just dependent on the inverse square law nature of the force. So they also hold for the electrical force.) The case of non-spherical Newtonian gravity will be taken up shortly.

B. Newtonian Gravitational Field

Newton's second law states that the force and acceleration are related through the mass. The mass occurs in both the equation for the gravitational force and the equation for the acceleration that force causes. Assuming that some small test mass, m is used, then the force on that mass is:

$$\vec{F} = \frac{GMm}{r^2}\hat{e}_r$$
$$\vec{F} = m\vec{a}$$
$$\vec{a} = \frac{GM}{r^2}\hat{e}_r$$
$$\vec{g} \equiv \vec{a}$$

where we have introduced the Newtonian gravitational acceleration g. (This is not what we call the gravitational acceleration of the earth, that is slightly different.) Notice that we can find the value of the acceleration of gravity from these formulas, at least in the spherical approximation:

$$g = \frac{GM_e}{r_e} \ .$$

This is the value on the surface of a spherical earth. It can be measured, for example with a falling pebble or a pendulum. Even if you have an idea of the size of the earth, and hence r_e , you still do not know the mass of the earth M_e, but only the product GM_e . This is why Cavendish said he had "weighted the earth" when he measured G.

There is an acceleration at each point in space due to the mass M, whether there is a test mass there or not. We say there is a **gravitational field**. This concept is very common in physics. Things produce effects even if there is nothing there to feel it.

C. Newtonian Gravitational Potential

There will be a force required to move a test mass away from our field source mass, M. This means that we have added potential energy to the test mass. Call this potential energy V. It will be a function of the position relative to the mass M. This potential energy must be related to the force through its spatial gradient. The rate of change of this potential energy with respect to distance is the same force as the Newtonian gravitational force. In mathematical terms we say that:

$$\vec{F} = -\vec{\nabla}V$$
$$= \frac{GMm}{r^2}\hat{e}_r$$
$$V = -\frac{GMm}{r}$$

This form for the potential energy, V has the energy zero at an infinite distance from the object M, (where r is infinite). The energy gets more negative as the test mass m approaches the mass M. It becomes infinitely large at the origin, but this is inside the mass and the equation does not hold there.

The value of the potential energy per unit mass is related to the nominal gravity on the surface of the earth

$$\begin{aligned}\frac{V(r_e)}{m} &= -\frac{GM_e}{r_e} \\ &= -\frac{GM_e}{r_e^2} r_e \\ &= -g\, r_e\end{aligned}$$

The standard scientific units of the acceleration of gravity are m/s/s. The units of the potential energy per unit mass are therefore m^2/s^2, or velocity squared . (Remember kinetic energy is half the mass times the velocity squared, so the units are consistent with the definition of energy.)

D. Gravity vector and the Gradient Operator

The differential operator $\vec{\nabla}$ operating on a function of position is just a vector that expresses the rate of change of the function in different directions:

$$\begin{aligned}\vec{\nabla}V &= \frac{\partial V}{\partial x}\hat{e}_x + \frac{\partial V}{\partial y}\hat{e}_y + \frac{\partial V}{\partial z}\hat{e}_z \\ &= \left(\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y}, \frac{\partial V}{\partial z}\right)\end{aligned}$$

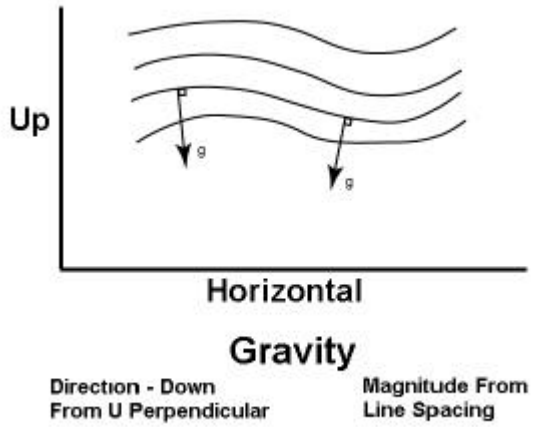
Here the two notations for Cartesian vectors are used. The direction of this vector will be along the greatest change in V, and the magnitude is the rate of change in that direction. To obtain the rate of change in any particular direction, just take the dot product with a unit vector in that direction. (There is an extensive technical note on vectors available. In addition the gradient operator is covered more extensively in a technical note on basic physics.)

There are two important facts about the gradient. If you take the gradient of a potential function V, you get the force. The following is true of this gradient and the force:

1. The direction of the gradient vector is perpendicular to surfaces of constant V,

2. The magnitude of the gradient is inversely proportional to the spacing of surfaces of constant potential. If the surfaces are close together, the gradient is large, if far apart it is small.

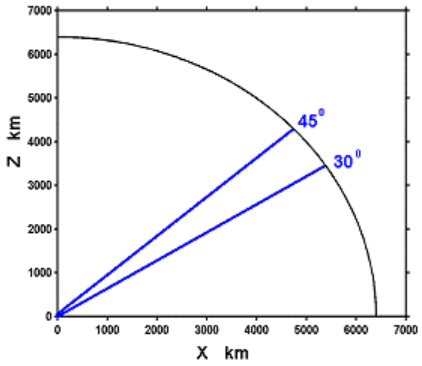
This is shown below where a diagram of a vertical cut of the equal potential surfaces is shown. The lines are spaced at equally differences of potential. The surfaces, of course, come out of the page and are three dimensional.



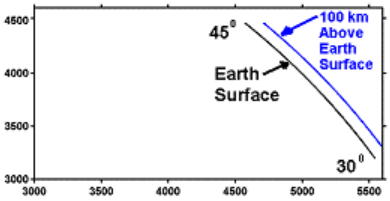
The distortion are greatly magnified for purposes of illustration. The measured direction of gravity, and hence the "normal" way we define down, are perpendicular to these equal potential surfaces. If they have bumps (as they do in the real world) then the up-down line has variations.

E. Spherical Earth Example

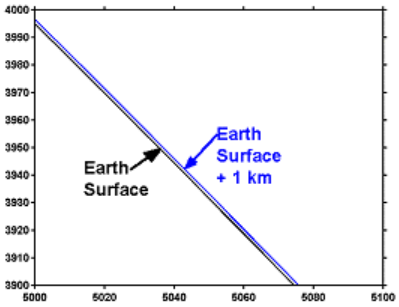
The earth is very large compared to everyday distances we experience. This is why we can assume that the earth is flat and only make small errors in most things we do. Below is a series of diagrams of a segment of the earth, with each step having higher magnification (smaller scale in the mapping terminology.)



Earth Quarter
With 15 Degree Segment



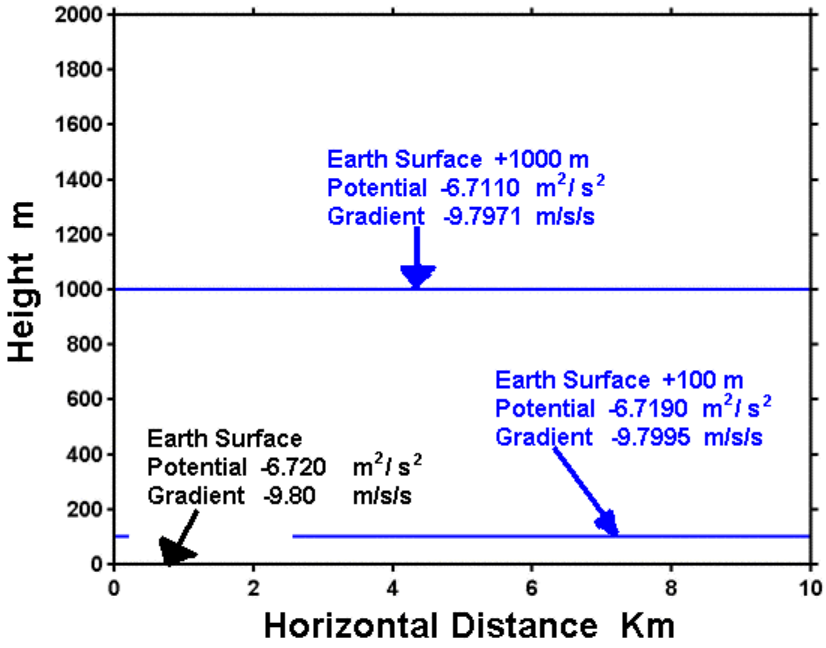
Earth Segment of 15 Degrees



Small Earth of Segment

It is clear that normal heights, such as a 100 m (300 ft) 30 story building, are essentially invisible on true scale plots of any significant segment of the earth.

A plot of some near earth heights and the gravitational values for a spherical earth are shown below. The values are for points very near the earth.



Near Earth Height Lines
With Gravitational Values

The values of the potential energy per unit mass (that is V/m) are shown for a nominal earth size that is spherical. Notice that the potentials get smaller in absolute value very slowly as you increase in height. The difference of these values, divided by the height difference is a good approximation for the derivative in the vertical direction, which is the gradient here. Notice that the gradient, which is the local value of g, also decreases in magnitude as you go up, but also very slowly. We can often treat the acceleration of gravity near the earth surface as a constant. (This is not true for satellite work however.)

F. Potential Differences Near Earth's Surface

In elementary physics texts, we learn that the gravitational potential energy is given by

$$V = mgh$$
$$\Delta V = mg\,\Delta h$$

The second form is more correct, as we are almost always dealing in differences of potential energies. This formula can be derived from the general Newtonian Law of Gravity assuming that the height difference is small compared to the radius of the earth.

Many approximations used for computation are derived using some small, dimensionless parameter. Both being dimensionless and small are important. Here the parameter will be h/r_e, the height above the earth divided by the nominal radius of the earth. The Newtonian potential is given by:

$$V = -\frac{GM_e m}{r}$$

Consider this equation for two locations, one on the surface of the earth and the second at a height h above the earth. It is the difference in potentials that we need. For the second we have

$$V(h) = -\frac{GM_e m}{r_e + h}$$

The only difference between these two is the denominator. We wish to compute the difference of these two potentials

$$\begin{aligned} \Delta V &= V(h) - V(0) \\ &= [-GM_e m]\left[\frac{1}{r_e + h} - \frac{1}{r_e}\right] \end{aligned}$$

The key to finding the most significant term is to let:

$$\begin{aligned} \frac{1}{r_e + h} &= \frac{1}{r_e}\frac{1}{(1 + h/r_e)} \\ &= \frac{1}{r_e}\frac{1}{(1+\alpha)} \end{aligned}$$

where the small parameter h/r_e has been called α. Substituting this form into the equation for ΔV one obtains

$$\begin{aligned} \Delta V &= \frac{GM_e m}{r_e}\left(\frac{\alpha}{1+\alpha}\right) \\ &\approx mgr_e\alpha \\ &= mgh \end{aligned}$$

This is the result desired. In this the value of the acceleration of gravity, g , found above has been used. The approximation $1+\alpha \approx 1$ was also used. Here the height is measured from the surface of the earth.

II. The Real Gravity Field of the Earth - Overview

There are several differences between the real gravity field of the earth and the above simplification. The two major physical differences are:

1. The earth is rotating. This causes an additional effective force to be felt for all objects rotating with the earth. This includes apples, pendulums, and people. It does not include earth satellites. This leads to effects on the order of 1/300 of the Newtonian gravitation values.

2. The earth is not homogeneous. These effects are 1/10,000 or less than the main effects. But they can be measured and have some practical consequences.

The rotational effects are not insignificant. Not only do they modify the value of the acceleration of gravity, they result in a change in the shape of the earth. The earth is approximately an ellipse revolved about the polar axis. This shape change leads to significant changes is the values of latitude. It also indirectly leads to many of the map use problems as different people and organizations use different approximations for the true figure of the earth. This leads to most of the difference in datums. (There is an extensive technical note on datums.)

In addition there are some terminology differences between what Geodesy uses and physics. In the field of Geodesy, there are some terms and concepts that "look and feel" like things studied in elementary physics, but are slightly different. This can cause some confusion, especially if one takes equations from both physics books and geodesy books.

	Physics	**Geodesy**
Gravity	Newtonian Gravitation	Newtonian Gravitation Plus Rotational Effects
Force	Force, F=ma Body Force f= F/m	Force per unit mass, Acceleration, g, As seen on Rotating Earth
Potential	Force = - Gradient (Potential) $\vec{F} = -\vec{\nabla}V$	Acceleration = + Gradient (Potential) $\vec{g} = +\vec{\nabla}V$
Coordinates Polar Angle	Measured from Pole (0 to 180 degrees)	Measured from Equator (ϕ') (-90 to 90 degrees)

The main source of difference is the inclusion of the rotational effects in the terms used in geodesy. This is done to reflect the measurements seen by an observer rotating on the earth's surface, that is measurements made on the real world. The definitions of several items were made before many of the concepts used in today's elementary physics were discovered. Newton or his contemporaries made many definitions. The concepts of potential energy, kinetic energy, and the conservation of energy followed this work by 100 years.

The second difference is the relation between the force or acceleration and a **potential function**. First geodesy deals in accelerations, not forces. This is because the effects of the mass of an object in Newton's second law (F = ma) and his law of gravity ($F = GMm/r^2$) cancel out.

In geodesy the acceleration is taken as the **gradient**[1] of the potential. In physics the force is the negative gradient of the potential energy. In geodesy the negative sign is omitted. In physics terms, the potential energy of something sitting on the earth is negative. You have to add energy to move it off the earth. In geodesy the gravity potential is a positive number. It is the absolute value of the physics value of potential energy per unit mass. The two potentials are related by a minus sign.

We will have three different sets of quantities used in geodesy,

1. Quantities based on a spherical, non-rotating earth,
2. Quantities based on a rotating, ellipsoidal earth without density variations,
3. Quantities based on the real earth.

The first set is used in elementary physics books. They are not used in geodesy. The second set is used as a baseline or reference set in geodesy. Differences from the smooth ellipsoidal model are reported.

Earth Gravity Field Quantities

	Potential	**Gravity Acceleration**
Spherical Earth	V	
Ellipsoidal Earth	$U=V+\Phi$	γ
Real Earth	W=U+T	g

Here,

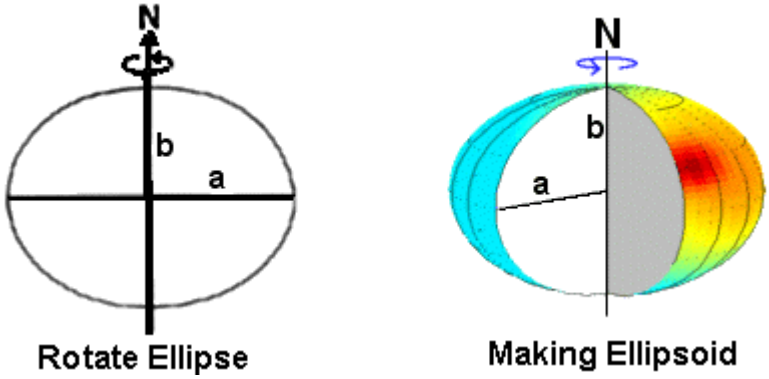
V is the Newtonian Potential for the model earth (spherical or ellipsoidal)
U is the gravity potential for the model earth - with rotational effects included,
Φ is the potential for rotational effects. $\Phi = \omega^2 p^2 / 2$, where p and ω are discussed below,
W is the gravity potential for the real earth,
T is the effect of the inhomogeneities in the earth,
γ is the acceleration due to the model potential, U, and
g is the real gravity acceleration of a body co-rotating with the earth.

[1] The gradient is a measure of how fast a function changes. In this case the changes are with respect to position, a rate of change with distance. The gradient is a derivative. It is also a vector. It points in the direction of maximum positive change. For a spherical earth, the gradient of the gravity potential points to the center of the earth. For the real earth it points in the local down direction. (Inverse of the local vertical).

III. Rotational Effects

A. Ellipsoid Model of Earth

What we call "gravity" in geodesy is the sum of the **Newtonian gravitational attraction** and rotational effects. The rotation effect is called **centrifugal acceleration**. All things co-rotating with the earth feel it. This causes the earth to have a bulge at the equator. For a uniform earth, the shape is an ellipsoid of revolution. (This is also called a spheroid). This is the geodetic model we will discuss in this section. The complexity of the bumps in the real world will be discussed in the later sections.



One of the ways to define the amount the ellipse differs from a circle is called the **flattening**, f. The flattening, f, is given by the equation

$$f = \frac{a-b}{a} = 1-\frac{b}{a},$$

where a is the longer equatorial radius and b is the shorter polar radius. The flattening of the earth is about 1/300. If the earth were a fluid it would be about 1/230. The earth acts almost as a fluid over long time frames. Newton was the first to predict the flattening of the earth. Another common way to describe the shape of the ellipse is with the eccentricity, e. This is defined as:

$$e = \sqrt{1-\frac{a^2}{b^2}}, \quad \text{or}$$

$$e^2 = 1-\frac{a^2}{b^2}.$$

The observed acceleration of gravity is the sum of the Newtonian gravitational attraction and the rotational force. The shape of the earth minimizes the potential energy due to the two forces. This makes

the direction of the observed acceleration of gravity perpendicular to the ellipsoid. This line does not go to the center of the earth except for the poles and points on the equator.

The Newtonian accelerations is given by

$$\vec{g}_n = -\frac{GM}{r^2}\hat{e}_r$$

where

G is the universal gravitational constant, M is the mass of the earth,
r is the radius of the earth (or distance from earth center of mass to observer), and
$\hat{e}_r$ is a unit vector that points outward from the center of the earth to the observation point.

This equation is only valid outside the earth. Inside the earth only the below the observer contributes to the Newtonian acceleration.

The centrifugal acceleration is given by

$$\vec{a}_c = \omega^2 p\hat{e}_p$$

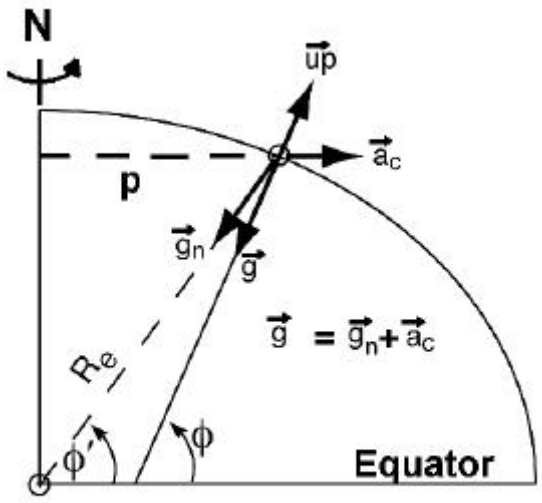
where

$\omega = 2\pi/T$ is the rotation rate of the earth (T is 1 sidereal day),
p is the distance from the axis of rotation to the observer shown in the figure, and
$\hat{e}_p$ is a unit vector along the outward direction of p.

This applies only to objects that rotate with the earth. Satellites in space do not feel this acceleration.

B. Gravity of Perfect Ellipsoid / Geodetic Latitude

The figure below is a summary of the acceleration vectors and the up-down lines for a point on the rotating earth. The up line is defined by the observed gravity acceleration, g, for an object co-rotating with the earth. This is the sum of the rotational and Newtonian gravitation effects. These together cause the shape of the earth to budge at equator and form an ellipsoid in the perfect earth case.



Quarter of Earth Ellipsoid Crossection

The Newtonian gravitational attraction, which we will denote g_n, does point at the center of mass of the earth. But the sum of forces we feel does not. The rotational acceleration, which we will call a_c, acts outward from the axis of rotation. These accelerations are shown in the diagram. Note that the vectors are not drawn to scale.

The line from the observed up-down line generates the usual latitude we use. This is called **Geodetic (ϕ).** It is the kind found on maps. A latitude from the line to the center of the earth also can be defined. This is **geocentric (ϕ') latitude.** It would be the observed latitude on a spherical earth. In the real world, it is used mostly in satellite work.

The observed acceleration, g, is defined by the vector sum of the g_n and a_c . <u>The total is the earth's acceleration of "gravity" as defined in geodesy</u>. The rotational acceleration term is maximum at the equator and zero at the poles because the moment arm p varies with latitude. In the real world, the maximum a_c is smaller than g_n by a factor of 300. Therefore the effects on g, and angles, are small. But the earth is large and even small effects on angles can cause changes of 10's of km.

The nominal Newtonian acceleration is about 980 cm/s^2 and the maximum rotational term about 3 cm/s^2. In geodesy accelerations of gravity are called **gal**'s, short for Galileo. One gal is one cm/s^2. This is a large unit for measuring variations, so the milligal, mgal or 1/1000th gal, commonly occurs.

The gravity value at each geodetic latitude for an ellipsoid is given by:

$$g = \frac{g_e \cos^2\phi + g_p\sqrt{1-e^2}\sin^2\phi}{\sqrt{1-e^2\sin^2\phi}}$$

where g_e is the acceleration at the equator, g_p, is the acceleration at the poles, and e is the eccentricity. This is not as complicated as it looks. It is easy to show that

$$\frac{b}{a} = \sqrt{1-e^2} \quad .$$

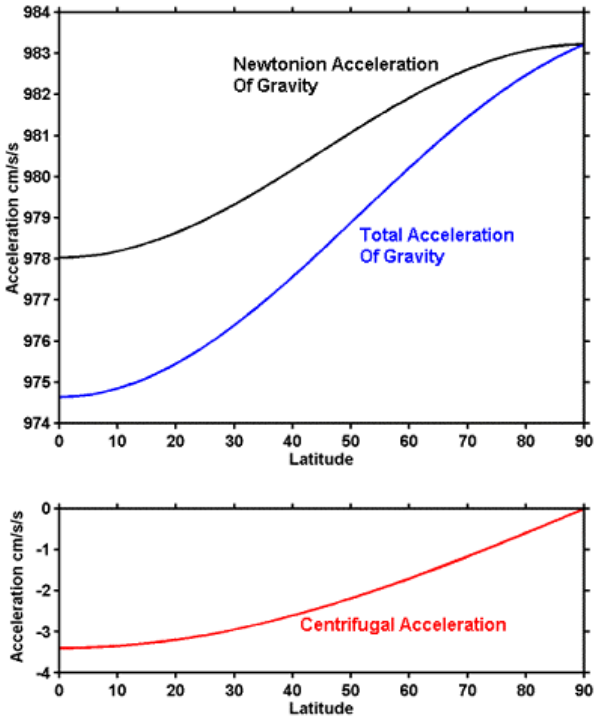
This equation is often written using the symbol gamma, γ, instead of g. Often gamma is used for theoretical gravity from a perfect ellipsoid and g for the real gravity from the lumpy, non-uniform real earth. Another form of this equation is

$$\gamma = g_e \frac{1+k\sin^2\phi}{\sqrt{1+e^2\sin^2\phi}},$$

with the auxiliary value k given by

$$k = \frac{b}{a}\frac{g_p}{g_e} - 1.$$

This value is graphed in the next figure along with the two components.



Components of Gravity Acceleration

Total is vector sum of components

Newtonian Acceleration -- Toward Center of Earth
Centrifugal Acceleration -- Outward From Polar Axis
Total Acceleration - Perpendicular to Ellipsoid

The two components and the total acceleration of gravity on the ellipsoid are a function of latitude as show in the above figure. The total acceleration is obtained by adding the Newtonian term and the centrifugal term as vectors. The Newtonian acceleration is opposed by the rotation terms everywhere except at the poles. It is therefore larger than the total value. The Newtonian term increases at the poles because the surface is closer to the center of mass. The gravity at the poles is about 5 gal (5 cm/s/s) larger than at the equator. About half of this effect is being closer to the center and the rest is the difference in the rotation term, the centrifugal acceleration.

The direction "Up" is also defined from the sum of the forces. It is perpendicular to the ellipsoid surface. It's inverse, "Down", does not point to the center of the earth. We historically defined latitude from observations of stars, with respect to the local up called the **local vertical**. The historical definition of latitude corresponds to the angle made the line perpendicular to the ellipsoid, not the one to the center of the earth. The local vertical is perpendicular to the ellipsoid in the absence of inhomogenates.

The latitude we see on maps, called **geodetic latitude**, comes from this perpendicular to the ellipsoid. This latitude is commonly called geographic latitude, but this term is not well defined in the scientific literature. Officially it is geodetic latitude. Usually when you see a term called geodetic something, it refers to definitions based on the ellipsoid. The coordinates using the vertical sensed by a bubble level, which responds to the real variation in the gravity field, are discussed in section VII.

The angle made by the line to the center of the earth is called **geocentric latitude**. This is the only latitude in the spherical earth model. Today geocentric latitude is used mainly in the field of artificial earth satellites.

IV. Potential Surfaces, Geoid, and Heights

The surfaces of constant W are surfaces of constant potential in the geodesy nomenclature. These are also surfaces of constant potential energy. The minus sign doesn't change the form of a constant surface. Water would assume one of these surfaces in a bowl, lake or ocean. These surfaces are called "**Level Surfaces**"[2]. **Mean sea level** is one particular surface of constant W. This particular surface can be extended over land. This particular level surface is called the **geoid**.

To this point, the earth has been idealized. But the real earth has oceans, mountains and other density variations. These variations are small, but cause the real gravity values to differ from the simple ellipsoid model. In the ellipsoid model, the geoid is an ellipsoid. In the real world the geoid has bumps. Now we will turn to the effects of the inhomogenates of the real world. The largest effects will be on heights.

The form of the surfaces of W must be measured; they cannot be computed from a few constants and a model. They depend on the real world mass variations, which varies on many distance scales. Computing the distance from the geoid to the center of the earth is not possible without large numbers of real world measurements. The ellipsoid is defined mathematically with respect to the center of the earth and its radius is easily computed. Knowing the distance between the ellipsoid and geoid therefore would define the geoid. The form of the geoid over large distances was not well known until satellites were launched.

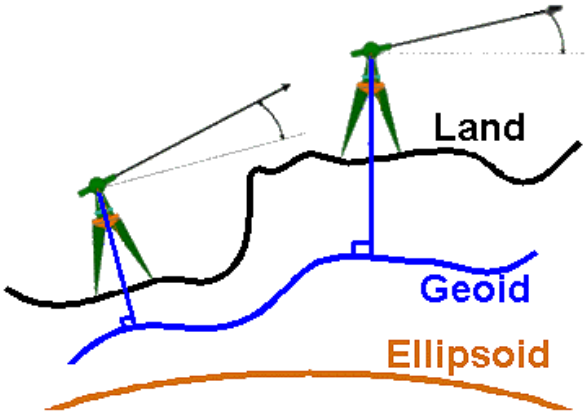
Starting with a "stake in the sand" that defines mean sea level at the shore, the heights of the land can be measured. With classical survey techniques, this is done with a series of measurements made from point to point.

These heights measured by classical methods are measured from the geoid. This happens because of the way heights are measured. Transits (telescopes on a tripod) are used, with angles measured with respect to the **local vertical**. The local up vector measured with a **plumb bob** or **spirit level** is the perpendicular to the real geoid.

[2] Level Surfaces are the three-dimensional analogue of contours in two dimensions. In one dimension if you find the locations where the function has a particular value, you get a set of points. If you have a function of two dimensions, such as height as a function of latitude and longitude, the set of points of a particular value is a line. This is one of the contours. If you have a function of latitude, longitude and height the set of points of a particular value is a surface. Picking one potential defines a surface. One particular value defines the geoid.

1908 Transit Theodolite

As the transit is moved from place to place, the local vertical will follow the hills and valleys of W. Therefore the geoid is the reference surface for classical height measurements. As one chief of the US National Geodetic Survey once said, "we are in the position of knowing heights very accurately with respect to a surface we know poorly".



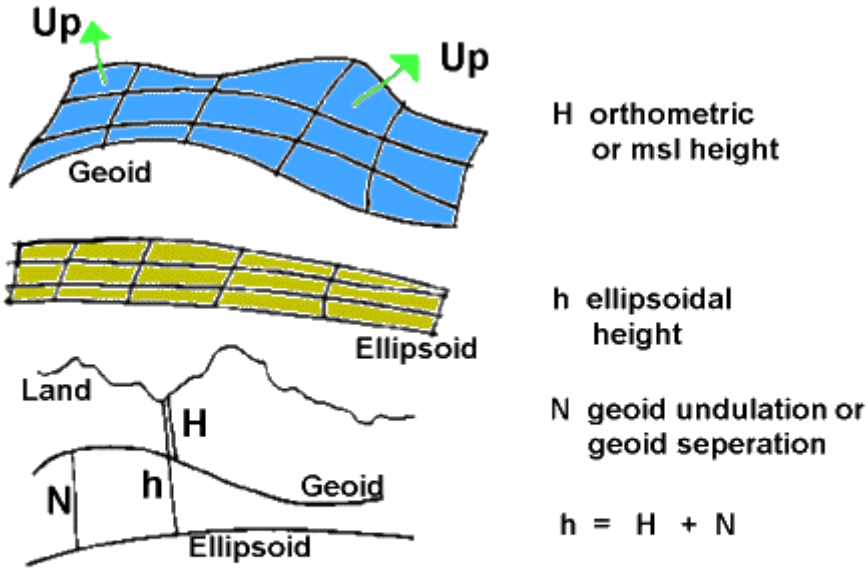
Classical Surveying Follows Geoid

Heights determined in this manor are called **orthometric heights** or **mean sea level (msl)** heights. (Orthometric and msl height are slightly different, but the two terms are commonly used to mean the same thing.)

Heights determined in this manor are called **orthometric heights** or **mean sea level (msl)** heights. Except in geodetic science literature, where you see msl height, it usually means orthometric height.

Technically orthometric and msl height are slightly different. Ocean currents and other effects slightly modify the real surface height of the sea. All the difference between real msl and orthometric heights can be classified as "oceanography".

Heights measured from the ellipsoid are called **ellipsoidal heights** or **geodetic heights**. Satellite systems measure ellipsoidal heights. This is not what we see on maps or in databases. These contain orthometric height. Because satellite based positioning is now very important, there is a need for a good measurement of the geoid location. This is done by measuring the vertical distance between the ellipsoid and the real world geoid as a function of location. This difference is called **undulation of the vertical** or **separation of the geoid**. The symbol N is used for the undulation.



Geoid - Ellisoid Diagram
Two and Three Dimensions

While the symbols h and H are commonly used for the two heights, there is no standard for which means orthometric or ellipsoidal. Here we will use the DoD standard,

h for ellipsoidal, and
H for orthometric height.

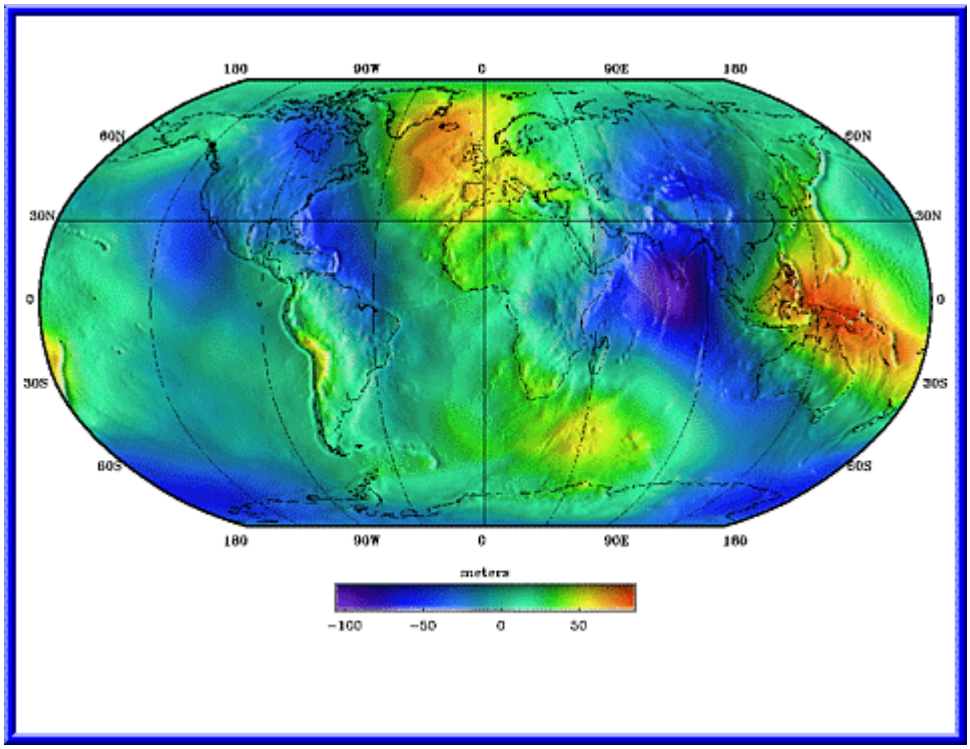
Note that in all cases the undulation, N is the orthometric minus the ellipsoidal heights. It is the distance from the ellipsoid to the geoid.

Thus

$$h = H + N$$

$$N = H - h \text{ ,}$$

where the values are taken at the same point. N is a function of location. Finding N as a function of latitude and longitude is how we determine the geoid.

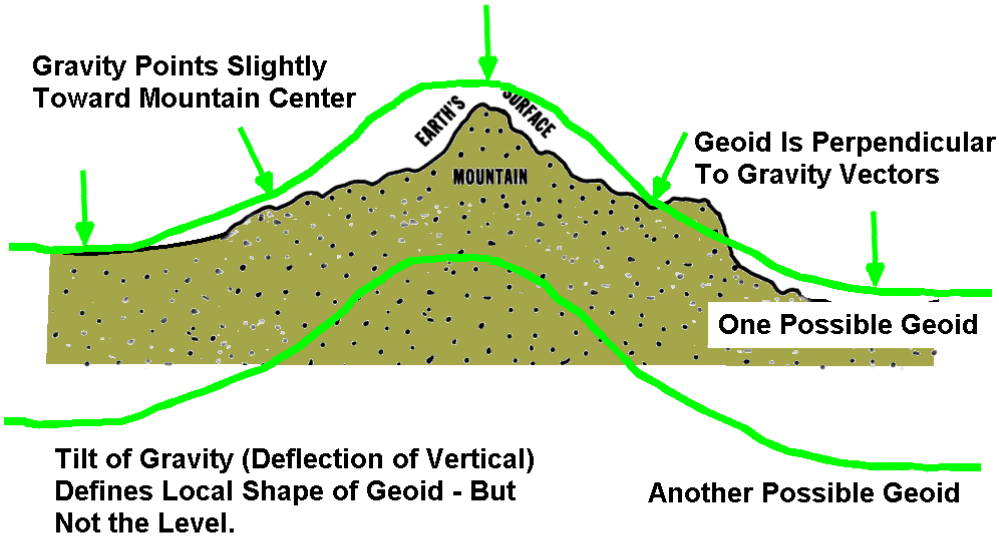


The above figure gives a worldwide view of the geoid surface. The values range from about -100 m to + 100 m. World wide ellipsoids are chosen so that the average undulation is zero over the world. This means that there are as many places with the geoid above the ellipsoid as there are with the geoid below the ellipsoid. There are many small area variations in the geoid that do not show up in this figure.

Because orthometric heights (msl) were the only ones historically measurable, they are the heights shown on maps and in databases. All heights on maps are orthometric. However position measurements made using satellite systems, such as GPS, are inherently done in an earth centered, earth fixed Cartesian X-Y-Z system. These can easily be converted to latitude, longitude, and ellipsoidal height. (Also called geodetic height). The geoid undulation, N, is needed to convert these to map type (orthometric or msl) heights. This value is only know at the meter level in most places.

V. Geoid and Topography

The geoid has variations, or components, at many spatial scales. At long lengths it feels the influences of most of the earth and its variations in density. However on a short scale, small variations are dominated by local effects such as near by mountains etc.



Geoid Follows Mass Bumps

This is illustrated in the above figure. The geoid seems to follow the mountain range up and down. This can be seen from the condition that the geoid is perpendicular to the local gravity vector. The absolute value of the geoid level can vary a lot because it is influenced by mass variations over a very long range, but the small bumps look like local topography - at least on the 100 - 1000 km range.