Analysis of Universal Transverse Mercator projection and Coordinate System

Adviser: Dr. Mahmoud Zolfaghari

Students:
Erfan Amini
Hamid Mozaffari

December 2012
Acknowledgement

Many thanks go to our adviser, Dr. Mahmoud Zolfaghari who learned us how to think like a researcher.
Abstract

The familiar Mercator projection used on so many world maps is a cylindrical projection, meaning the globe is encircled by an imaginary cylinder touching at the equator, and the earth is projected onto the cylinder. Contrary to a lot of geography texts, the projection is not what would result from placing a light at the center of the earth and projecting the surface onto the cylinder. (That projection called the cylindrical gnomonic, results in extreme distortion in polar areas and has virtually no practical uses.) The term "map projection" actually means any mathematical transformation of the globe onto some other surface, including many that can't be physically realized by any actual optical projection system.

The Mercator projection is a conformal projection, meaning that angles and small shapes on the globe project as the same angles or shapes on the map. The price paid by all conformal projections is great variation in scale away from the central portions of the map. Greenland on a Mercator map looks as big as South America, though it has actually only 1/8 the area. However, a small portion of the Greenland coast (or any small region, in fact) has the same shape on the map as it does on the ground.

Because the Transverse Mercator projection is very accurate in narrow zones, it has become the basis for a global coordinate system called the Universal Transverse Mercator System or UTM System. The globe is subdivided into narrow longitude zones, which are projected onto a Transverse Mercator projection. A grid is constructed on the projection, and used to locate points. The upside of the grid system is that, since the grid is rectangular and decimal, it is far easier to use than latitude and longitude. The downside is that, unlike latitude and longitude, there is no way to determine grid locations independently.

The scale along the central meridian of each zone is 0.9996 of true scale. True scale occurs about 180 kilometers east and west of the central meridian.
TABLE OF CONTENTS

Introduction..................................................................................................................1
1- Basic Concepts of Universal Transverse Mercator (UTM)..............................2
2- UTM Geographic Coordinate System............................................................7
  2-1 Definitions.....................................................................................................9
  2-2 Simplified formulas......................................................................................10
3- Spherical transverse Mercator.................................................................12
  3-1 Normal and transverse spherical projections...........................................13
  3.2 Formulae for the spherical transverse Mercator...............................15
  3-2-1 Spherical normal Mercator revisited.................................................15
  3-2-2 Normal and transverse graticules.........................................................16
4- Ellipsoidal transverse Mercator...............................................................19
  4-1 Implementations of the Gauss–Krüger projection...........................21
  4-1-1 Exact Gauss-Krüger and accuracy of the truncated series........23
5- State Plane Coordinate System.................................................................24
6- NATO System...............................................................................................25
  6-1 Determining Grid Coordinates..............................................................27
7- Military Grid Reference System (MGRS)................................................28
Introduction

The Universal Transverse Mercator (UTM) geographic coordinate system uses a 2-dimensional Cartesian coordinate system to give locations on the surface of the Earth. It is a horizontal position representation, i.e. it is used to identify locations on the Earth independently of vertical position, but differs from the traditional method of latitude and longitude in several respects.

The UTM system is not a single map projection. The system instead divides the Earth into sixty zones, each a six-degree band of longitude, and uses a secant transverse Mercator projection in each zone.

The Universal Transverse Mercator coordinate system was developed by the United States Army Corps of Engineers in the 1940s.[1] The system was based on an ellipsoidal model of Earth. For areas within the contiguous United States the Clarke 1866 ellipsoid[2] was used. For the remaining areas of Earth, including Hawaii, the International Ellipsoid[3] was used. The WGS84 ellipsoid is now used to model the Earth in the UTM coordinate system, which means current UTM northing at a given point can be 200+ meters different from the old.

Prior to the development of the Universal Transverse Mercator coordinate system, several European nations demonstrated the utility of grid-based conformal maps by mapping their territory during the interwar period. Calculating the distance between two points on these maps could be performed more easily in the field (using the Pythagorean Theorem) than was possible using the trigonometric formulas required under the graticule-based system of latitude and longitude. In the post-war years, these concepts were extended into the Universal Transverse Mercator / Universal Polar Stereographic (UTM/UPS) coordinate system, which is a global (or universal) system of grid-based maps.

The transverse Mercator projection is a variant of the Mercator projection, which was originally developed by the Flemish geographer and cartographer Gerardus Mercator, in 1570. This projection is conformal, so it preserves angles and approximates shape but distorts distance and area. UTM involves non-linear scaling in both Easting and Northing ensures the projected map of the ellipsoid is conformal.

[1]
1- Basic Concepts of Universal Transverse Mercator (UTM)

The UTM system applies the Transverse Mercator projection to mapping the world, using 60 pre-defined standard zones to supply parameters. UTM zones are six degrees wide. Each zone exists in a North and South variant.

Europeans can best understand UTM by thinking of it as a world-wide version of the Gauss Kruger system, which is also based on a regular system of Transverse Mercator projections that each map a zone six degrees wide.

| Universal Transverse Mercator | The Northern Hemisphere projections for the infamous UTM system consisting of 120 zones (60 different zones with North and South variants of each). Originally developed for military use and now widely misused in civil mapping. |
| Universal Transverse Mercator (South) | The Southern Hemisphere projections for UTM. These are mainly distinguished by each having a Northing parameter of 10 million so that no coordinates need involve negative numbers. |

Limitations

The accuracy of any Transverse Mercator projection quickly decreases from the central meridian. Therefore, it is strongly recommended to restrict the longitudinal extent of the projected region when using Universal Transverse Mercator projections to +/- 6 degrees from the central meridian. This requirement is met within all State Plane zones that use Transverse Mercator projections.
Each UTM Zone is a Different Projection

The Mercator projection maps the world onto a cylinder where the central ring of tangency is the Earth's Equator.

Near the Equator, the Mercator projection provides low distortion. Away from the Equator distortion becomes very high. This limits the utility of the Mercator projection to regions near the Equator. That is a big limitation because most places that people live (and thus, most of the regions that people most frequently map) are located not along the Equator but along North-South directions, such as from North America to South America.

Turning the Mercator projection's cylinder so that it is tangent to the Earth along a meridian (longitude line) instead of the Equator results in what is called a Transverse Mercator projection. If we created a Transverse Mercator projection that had a meridian as the central ring of the cylinder we could make local maps anywhere along the North-South line of tangency. If the maps are limited to the thin, vertical region near the meridian of tangency they will be relatively free of distortion.

The problem is that any Transverse Mercator projection created by choosing any one meridian as a line of tangency is useful only near that meridian. If we pick a North-South line running through Athens we can make maps all the way from Scandinavia down the length of Africa, but any maps using this projection in North and South America would be hopelessly distorted.

The Universal Transverse Mercator system of projections deals with this by defining 60 different standard projections, each one of which is a different Transverse Mercator projection that is slightly rotated to use a different meridian as the central line of tangency. Each different centerline defines a UTM Zone. The "UTM Zone" is a shorthand way of naming a specific, different projection that consists of a Transverse Mercator projection using a different meridian as the centerline. By rotating the cylinder in 60
steps (six degrees per step) UTM assures that all spots on the Earth will be within 3 degrees of the centerline of one of the 60 cylindrical projections.

To map any spot on Earth, one picks the UTM Zone centerline that is closest to it and then makes a map using that "UTM Zone" cylindrical projection.

UTM Zones should not be Combined

Novice UTM users usually do not realize that each UTM Zone is in fact a different projection using a different system of coordinates. New users of UTM therefore will frequently attempt to "combine" different maps created in different UTM zones into one map with the expectation that the combined map will show all objects with low distortion as did the original maps. The motivating factor is often a desire to create a map centered on a region of interest that spans several UTM zones or which is centered between two zones. Such plans fail to take into account that UTM is an intrinsically inflexible system. In effect, the UTM system assumes objects from different zones will never be seen together in the same map.

Combining objects from different UTM zones into a map that is projected using only one of those UTM zones will result in distortion in the locations and shapes of the objects that originated in a different zone map. Geographic shapes that look good in a transverse Mercator projection centered upon a given UTM zone line will be much distorted when illustrated in a UTM projection centered upon a different zone line.
The illustration above shows part of Europe projected into UTM Zone 2 in the yellow map. Overlaid on the yellow map is an Orthographic projection centered on the same map center shown in blue color. The numbers are positioned at the center of UTM Zones 1, 2 and 3.

The Orthographic map is essentially accurate over the entire illustration. In contrast, the UTM map is highly inaccurate only one half zones away from the "home" zone. Note that it distorts the coast of France so much that it has France (in the yellow, UTM projection color) crossing the Channel.

If we need to combine objects from several different UTM zones, the correct solution is to choose a different projection (such as a conic or azimuthally projection) for the combined map that provides low distortion over the entire region of interest. The illustration above shows a Lambert Conformal Conic projection in black outline and
darker blue color overlaid over the Orthographic projection. Note that both projections are so close to each other it is difficult to pick out places where they differ. For example, in the region of France where the UTM projection had the continental landmass crossing the channel there is a very slight North/South offset but otherwise the two projections are virtually the same.

Remember, although no projection is perfect for all uses some projections are better than others in the uses for which they were designed. UTM was designed to map objects within one zone at a time. It is a very bad choice if objects from several zones must be shown together on the same map.

The transverse version is widely used in national and international mapping systems around the world, including the UTM. When paired with a suitable geodetic datum, the transverse Mercator delivers high accuracy in zones less than a few degrees in east-west extent.

The standard (or Normal) Mercator and the transverse Mercator are two different aspects of the same mathematical construction. Because of the common foundation, the transverse Mercator inherits many traits from the normal Mercator:

- Both projections are cylindrical: for the Normal Mercator, the axis of the cylinder coincides with the polar axis and the line of tangency with the equator. For the transverse Mercator, the axis of the cylinder lies in the equatorial plane, and the line of tangency is any chosen meridian, thereby designated the central meridian.
- Both projections may be modified to secant forms, which mean the scale has been reduced so that the cylinder slices through the model globe.
- Both exist in spherical and ellipsoidal versions.
- Both projections are conformal, so that the point scale is independent of direction and local shapes are well preserved;
- Both projections have constant scale the line of tangency (the equator for the normal Mercator and the central meridian for the transverse).

Since the central meridian of the transverse Mercator can be chosen at will, it may be used to construct highly accurate maps (of narrow width) anywhere on the globe. The secant, ellipsoidal form of the transverse Mercator is the most widely applied of all projections for accurate large scale maps.
2- UTM Geographic Coordinate System

The idea of the transverse Mercator projection has its roots in the 18th century, but it did not come into common usage until after World War II. It has become the most used because it allows precise measurements in meters to within 1 meter.

A Mercator projection is a ‘pseudo cylindrical’ conformal projection (it preserves shape). What you often see on poster-size maps of the world is an equatorial Mercator projection that has relatively little distortion along the equator, but quite a bit of distortion toward the poles.

What a transverse Mercator projection does, in effect, is orient the ‘equator’ north-south (through the poles), thus providing a north-south oriented swath of little distortion. By changing slightly the orientation of the cylinder onto which the map is projected, successive swaths of relatively undistorted regions can be created.

This is exactly what the UTM system does. Each of these swaths is called a UTM zone and is six degrees of longitude wide. The first zone begins at the International Date Line (180°, using the geographic coordinate system). The zones are numbered from west to east, so zone 2 begins at 174°W and extends to 168°W. The last zone (zone 60) begins at 174°E and extends to the International Date Line.
The zones are then further subdivided into an eastern and western half by drawing a line, representing a transverse Mercator projection, down the middle of the zone. This line is known as the ‘central meridian’ and is the only line within the zone that can be drawn between the poles and is perpendicular to the equator (in other words, it is the new ‘equator’ for the projection and suffers the least amount of distortion). For this reason, vertical grid lines in the UTM system are oriented parallel to the central meridian. The central meridian is also used in setting up the origin for the grid system.

Any point can then be described by its distance east of the origin (its ‘easting’ value). By definition the Central Meridian is assigned a false easting of 500,000 meters. Any easting value greater than 500,000 meters indicates a point east of the central meridian. Any easting value less than 500,000 meters indicates a point west of the central meridian. Distances (and locations) in the UTM system are measured in meters, and each UTM zone has its own origin for east-west measurements.

To eliminate the necessity for using negative numbers to describe a location, the east-west origin is placed 500,000 meters west of the central meridian. This is referred to as the zone’s ‘false origin’. The zone doesn’t extend all the way to the false origin. The origin for north-south values depends on whether you are in the northern or southern hemisphere. In the northern hemisphere, the origin is the equator and all distances north (or ‘nothings’) are measured from the equator. In the southern
hemisphere the origin is the South Pole and all nothings are measured from there. Once again, having separate origins for the northern and southern hemispheres eliminates the need for any negative values. The average circumference of the earth is 40,030,173 meters, meaning that there are 10,007,543 meters of northing in each hemisphere.

UTM coordinates are typically given with the zone first, then the easting, then the northing. So, in UTM coordinates, Red Hill is located in zone twelve at 328204 E (easting), 4746040 N (northing). Based on this, you know that you are west of the central meridian in zone twelve and just under halfway between the equator and the North Pole. The UTM system may seem a bit confusing at first, mostly because many people have never heard of it, let alone used it. Once you’ve used it for a little while, however, it becomes an extremely fast and efficient means of finding exact locations and approximating locations on a map.

Many topographic maps published in recent years use the UTM coordinate system as the primary grids on the map. On older topographic maps published in the United States, UTM grids are shown along the edges of the map as small blue ticks.

2.1 Definitions

The UTM system divides the Earth between 80°S and 84°N latitude into 60 zones, each 6° of longitude in width. Zone 1 covers longitude 180° to 174° W; zone numbering increases eastward to zone 60 that cover longitude 174 to 180 east.

Each of the 60 zones uses a transverse Mercator projection that can map a region of large north-south extent with low distortion. By using narrow zones of 6° of longitude (up to 800 km) in width, and reducing the scale factor along the central meridian to
0.9996 (a reduction of 1:2500), the amount of distortion is held below 1 part in 1,000 inside each zone. Distortion of scale increases to 1.0010 at the zone boundaries along the equator.

In each zone the scale factor of the central meridian reduces the diameter of the transverse cylinder to produce a secant projection with two standard lines, or lines of true scale, about 180 km on each side of, and about parallel to, the central meridian (Arc cos 0.9996 = 1.62° at the Equator). The scale is less than 1 inside the standard lines and greater than 1 outside them, but the overall distortion is minimized.

**Overlapping grids**

Distortion of scale increases in each UTM zone as the boundaries between the UTM zones are approached. However, it is often convenient or necessary to measure a series of locations on a single grid when some are located in two adjacent zones. Around the boundaries of large scale maps (1:100,000 or larger) coordinates for both adjoining UTM zones are usually printed within a minimum distance of 40 km on either side of a zone boundary. Ideally, the coordinates of each position should be measured on the grid for the zone in which they are located, but because the scale factor is still relatively small near zone boundaries, it is possible to overlap measurements into an adjoining zone for some distance when necessary.

**Latitude bands**

Latitude bands are not a part of UTM, but rather a part of MGRS. They are however sometimes used.

**2.2 Simplified formulas**

These formulas are truncated version of Transverse Mercator: flattening series, which were originally derived by Johann Heinrich Louis Krüger in 1912. They are accurate to around a millimeter within 3,000 km of the central meridian. Concise commentaries for their derivation have also been given.

The WGS 84 spatial reference system describes Earth as an oblate spheroid along North-South axis with an equatorial radius of \( a = 6378.137 \) km and an inverse flattening of \( 1/f = 298.257,223,563 \). Let's take a point of latitude \( \varphi \) and of longitude \( \lambda \) and compute its UTM coordinates as well as point scale factor \( k \) and meridian convergence \( \gamma \) using a reference meridian of longitude \( \lambda_0 \). By convention, in the northern hemisphere \( N_0 = 0 \) km and in the southern hemisphere \( N_0 = 10000 \) km. By convention also \( k_0 = 0.9996 \) and \( E_0 = 500 \) km.
In the following formulas, the distances are in kilometers. In advance let's compute some preliminary values:

\[
\begin{align*}
    n &= \frac{f}{2 - f}, \quad A = \frac{a}{1 + n} \left(1 + \frac{n^2}{4} + \frac{n^4}{64} + \cdots \right), \\
    \alpha_1 &= \frac{1}{2} \left(-\frac{2}{3} n^2 + \frac{5}{16} n^3 \right), \quad \alpha_2 = \frac{13}{48} n^2 - \frac{3}{5} n^3, \quad \alpha_3 = \frac{61}{240} n^3, \\
    \delta_1 &= 2n - \frac{2}{3} n^2 - 2n^3, \quad \delta_2 = \frac{7}{3} n^2 - \frac{8}{5} n^3, \quad \delta_3 = \frac{56}{15} n^3.
\end{align*}
\]

From latitude, longitude \((\phi, \lambda)\) to UTM coordinates \((E, N)\)

First let's compute some intermediate values:

\[
\begin{align*}
    t &= \sinh \left(\tanh^{-1} \sin \phi - \frac{2\sqrt{n}}{1 + n} \tanh^{-1} \left(\frac{2\sqrt{n}}{1 + n} \sin \phi \right)\right), \\
    \xi' &= \tan^{-1} \left(\frac{t}{\cos(\lambda - \lambda_0)}\right), \quad \eta' = \tanh^{-1} \left(\frac{\sin(\lambda - \lambda_0)}{\sqrt{1 + t^2}}\right), \\
    \sigma &= 1 + \sum_{j=1}^{3} 2j\alpha_j \cos 2j\xi' \cosh 2j\eta', \quad \tau = \sum_{j=1}^{3} 2j\alpha_j \sin 2j\xi' \sinh 2j\eta'.
\end{align*}
\]

The final formulas are:

\[
\begin{align*}
    E &= E_0 + k_0 A \left(\eta' + \sum_{j=1}^{3} \alpha_j \cos 2j\xi' \sinh 2j\eta'\right), \\
    N &= N_0 + k_0 A \left(\xi' + \sum_{j=1}^{3} \alpha_j \sin 2j\xi' \cosh 2j\eta'\right), \\
    k &= \frac{k_0 A}{a} \sqrt{\left\{1 + \left(\frac{1 - n}{1 + n} \tan \phi \right)^2\right\}} \frac{\sigma^2 + \tau^2}{t^2 + \cos^2(\lambda - \lambda_0)}, \\
    \gamma &= \tan^{-1} \left(\frac{\tau \sqrt{1 + t^2} + \sigma t \tan(\lambda - \lambda_0)}{\sigma \sqrt{1 + t^2} - \tau t \tan(\lambda - \lambda_0)}\right).
\end{align*}
\]
From UTM coordinates \((E, N)\) to latitude, longitude \((\phi, \lambda)\)

First let's compute some intermediate values:

\[
\xi = \frac{N - N_0}{k_0 A_3}, \quad \eta = \frac{E - E_0}{k_0 A},
\]

\[
\xi' = \xi - \sum_{j=1}^{3} \beta_j \sin 2j \xi \cosh 2j \eta, \quad \eta' = \eta - \sum_{j=1}^{3} \beta_j \cos 2j \xi \sinh 2j \eta,
\]

\[
\sigma' = 1 - \sum_{j=1}^{3} 2j \beta_j \cos 2j \xi \cosh 2j \eta, \quad \tau' = \sum_{j=1}^{3} 2j \beta_j \sin 2j \xi \sinh 2j \eta,
\]

\[
\chi = \sin^{-1} \left( \frac{\sin \xi'}{\cosh \eta'} \right).
\]

The final formulas are:

\[
\varphi = \chi + \sum_{j=1}^{3} \delta_j \sin 2j \chi,
\]

\[
\lambda = \lambda_0 + \tan^{-1} \left( \frac{\sinh \eta'}{\cos \xi'} \right),
\]

\[
k = \frac{k_0 A}{a} \sqrt{\left\{ 1 + \left( \frac{1 - n}{1 + n} \tan \varphi \right)^2 \right\} \frac{\cos^2 \xi' + \sinh^2 \eta'}{\sigma'^2 + \tau'^2}},
\]

\[
\gamma = \tan^{-1} \left( \frac{\tau' + \sigma' \tan \xi' \tanh \eta'}{\sigma' - \tau' \tan \xi' \tanh \eta'} \right).
\]

3- Spherical transverse Mercator

In constructing a map on any projection, a sphere is normally chosen to model the earth when the extent of the mapped region exceeds a few hundred kilometers in length in both dimensions. For maps of smaller regions, an ellipsoidal model must be chosen if greater accuracy is required; see next section. The spherical form of the transverse Mercator projection was one of the seven 'new' projections presented, in 1772, by Johann Heinrich Lambert\(^1\)\(^2\) (also available in a modern English translation\(^3\) Lambert did not name his projections; the name transverse Mercator dates from the second half of the nineteenth century.\(^4\) The principal properties of the transverse projection are here presented in comparison with the properties of the normal projection.
3.1 Normal and transverse spherical projections

**Normal Mercator**

- The central meridian projects to the straight line $x = 0$. Other meridians project to straight lines with $x$ constant.

- The equator projects to the straight line $y = 0$ and parallel circles project to straight lines of constant $y$.

**Transverse Mercator**

- The central meridian projects to the straight line $x = 0$. Meridians 90 degrees east and west of the central meridian project to lines of constant $y$ through the projected poles. All other meridians project to complicated curves.

- The equator projects to the straight line $y = 0$ but all other parallels are complicated closed curves.

Spherical Normal (equatorial) Mercator (truncated at $y = \pm \pi$, corresponding to approximately 85 degrees).

Spherical transverse Mercator (truncated at $x = \pm \pi$ in units of Earth radius.)
- Projected meridians and parallels intersect at right angles.

- The projection is unbounded in the \( y \) direction. The poles lie at infinity.

- The projection is conformal. The shapes of small elements are well preserved.

- Distortion increases with \( y \). The projection is not suited for world maps. Distortion is small near the equator and the projection (particularly in its ellipsoidal form) is suitable for accurate mapping of equatorial regions.

- Greenland is almost as large as Africa; the actual area is about one thirteenth that of Africa.

- The point scale factor is independent of direction. It is a function of \( y \) on the projection. (On the sphere it depends on latitude only.) The scale is true on the equator.

- The projection is reasonably accurate near the equator. Scale at an angular distance of 5° (in latitude) away from the equator is less than 0.4% greater than scale at the equator, and is about 1.54%

- Projected meridians and parallels intersect at right angles.

- The projection is unbounded in the \( x \) direction. The points on the equator at ninety degrees from the central meridian are projected to infinity.

- The projection is conformal. The shapes of small elements are well preserved.

- Distortion increases with \( x \). The projection is not suited for world maps. Distortion is small near the central meridian and the projection (particularly in its ellipsoidal form) is suitable for accurate mapping of narrow regions.

- Greenland and Africa are both near to the central meridian; their shapes are good and the ratio of the areas is a good approximation to actual values.

- The point scale factor is independent of direction. It is a function of \( x \) on the projection. (On the sphere it depends on both latitude and longitude.) The scale is true on the central meridian.

- The projection is reasonably accurate near the central meridian. Scale at an angular distance of 5° (in longitude) away from the central meridian is less than 0.4% greater than scale at the central meridian, and is about 1.54% at an
greater at an angular distance of 10°.

- In the secant version the scale is reduced on the equator and it is true on two lines parallel to the projected equator (and corresponding to two parallel circles on the sphere).

- Convergence (the angle between projected meridians and grid lines with \( x \) constant) is identically zero. Grid north and true north coincides.

- Rhumb lines (of constant compass bearing on the sphere) project to straight lines.

<table>
<thead>
<tr>
<th>greater at an angular distance of 10°.</th>
<th>angular distance of 10°.</th>
</tr>
</thead>
<tbody>
<tr>
<td>In the secant version the scale is reduced on the central meridian and it is true on two lines parallel to the projected central meridian.</td>
<td>Convergence is zero on the equator and non-zero everywhere else. It increases as the poles are approached. Grid north and true north does not coincide.</td>
</tr>
<tr>
<td>Convergence is zero on the equator and non-zero everywhere else. It increases as the poles are approached. Grid north and true north does not coincide.</td>
<td>Rhumb lines project to complex curves.</td>
</tr>
</tbody>
</table>

### 3.2 Formulae for the spherical transverse Mercator

#### 3.2.1 Spherical normal Mercator revisited

The normal aspect of a tangent cylindrical projection of the sphere

The normal cylindrical projections are described in relation to a cylinder tangential at the equator with axis along the polar axis of the sphere. The cylindrical projections are constructed so that all points on a meridian are projected to points with \( x = a\lambda \) and \( y_a \) prescribed function of \( \phi \). For a tangent Normal Mercator projection the (unique) formulae which guarantee co formality are \([25]\).
\[ x = a\lambda, \quad y = a \ln \left[ \tan \left( \frac{\pi}{4} + \frac{\phi}{2} \right) \right] = \frac{a}{2} \ln \left[ \frac{1 + \sin \phi}{1 - \sin \phi} \right]. \]

Co formality implies that the point scale of the projection is independent of direction: it is a function of latitude only:

\[ k(\phi) = \sec \phi. \]

For the secant version of the projection there is a factor of \( k_0 \) on the right hand side of all these equations: this ensures that the scale is equal to \( k_0 \) on the equator.

### 3.2.2 Normal and transverse graticules

Transverse Mercator graticules

The figure on the left shows how a transverse cylinder is related to the conventional graticule on the sphere. It is tangential to some arbitrarily chosen meridian and its axis is perpendicular to that of the sphere. The \( x \) and \( y \) axes defined on the figure are related to the equator and central meridian exactly as they are for the normal projection. In the figure on the right a rotated graticule is related to the transverse cylinder in the same way that the normal cylinder is related to the standard graticule. The 'equator', 'poles' (E and W) and 'meridians' of the rotated graticule are identified with the chosen central meridian, points on the equator 90 degrees east and west of the central meridian, and great circles through those points.

Transverse Mercator geometry
The position of an arbitrary point \((\phi, \lambda)\) on the standard graticule can also be identified in terms of angles on the rotated graticule: \((\phi', \lambda')\) and \(-\lambda'\) (angle M'CO) becomes an effective longitude. (The minus sign is necessary so that \((\phi, \lambda)\) are related to the rotated graticule in the same way that \((\phi', \lambda')\) are related to the standard graticule). The Cartesian \((x', y')\) axes are related to the rotated graticule in the same way that the axes \((x, y)\) axes are related to the standard graticule.

The tangent transverse Mercator projection defines the coordinates \((x', y')\) in terms of \(-\lambda'\) and \(\phi'\) by the transformation formulae of the tangent Normal Mercator projection:

\[
x' = -a\lambda' \quad y' = \frac{a}{2} \ln \left[ \frac{1 + \sin \phi'}{1 - \sin \phi'} \right].
\]

This transformation projects the central meridian to a straight line of finite length and at the same time projects the great circles through E and W (which include the equator) to infinite straight lines perpendicular to the central meridian. The true parallels and meridians (other than equator and central meridian) have no simple relation to the rotated graticule and they project to complicated curves.

**The relation between the graticules**

The angles of the two graticules are related by using spherical trigonometry on the spherical triangle NM'P defined by the true meridian through the origin, OM'N, the true meridian through an arbitrary point, MPN, and the great circle WM'PE. The results are [25].

\[
\sin \phi' = \sin \lambda \cos \phi, \\
\tan \lambda' = \sec \lambda \tan \phi.
\]

**Direct transformation formulae**

The direct formulae giving the Cartesian coordinates \((x, y)\) follow immediately from the above. Setting \(x = y'\) and \(y = -x'\) (and restoring factors of \(k_0\) to accommodate secant versions)
The above expressions are given in Lambert\textsuperscript{11} and also (without derivations) in Snyder,\textsuperscript{13} Maling\textsuperscript{26} and online\textsuperscript{25} (with full details).

**Inverse transformation formulae**

Inverting the above equations gives

\[
x(\lambda, \phi) = \frac{1}{2} k_0 a \ln \left[ \frac{1 + \sin \lambda \cos \phi}{1 - \sin \lambda \cos \phi} \right],
\]
\[
y(\lambda, \phi) = k_0 a \arctan \left[ \sec \lambda \tan \phi \right],
\]

Point scale

In terms of the coordinates with respect to the rotated graticule the point scale factor is given by \( k = \sec \phi' \): this may be expressed either in terms of the geographical coordinates or in terms of the projection coordinates:

\[
k'(\lambda, \phi) = \frac{k_0}{(1 - \sin^2 \lambda \cos^2 \phi)^{1/2}},
\]
\[
k(x, y) = k_0 \cosh \left( \frac{x}{k_0 a} \right).
\]

The second expression shows that the scale factor is simply a function of the distance from the central meridian of the projection. A typical value of the scale factor is \( k_0 = 0.9996 \) so that \( k = 1 \) when \( x \) is approximately 180 km. When \( x \) is approximately 255 km and \( k_0 = 1.0004 \): the scale factor is within 0.04\% of unity over a strip of about 510 km wide.
Convergence

The angle of convergence is defined by the angle measured from the projected meridian, which defines true north, to a grid line of constant x, defining grid north. Therefore \( \gamma \) is positive in the quadrant north of the equator and east of the central meridian and also in the quadrant south of the equator and west of the central meridian. The convergence must be added to a grid bearing to obtain a compass bearing from true north. For the secant transverse Mercator the convergence may be expressed either in terms of the geographical coordinates or in terms of the projection coordinates:

\[
\gamma(\lambda, \phi) = \arctan(\tan \lambda \sin \phi),
\]

\[
\gamma(x, y) = \arctan \left( \tanh \frac{x}{k_0 a} \tan \frac{y}{k_0 a} \right).
\]

4. Ellipsoidal transverse Mercator

The ellipsoidal form of the transverse Mercator projection was developed by Carl Friedrich Gauss in 1825 and further analyzed by Johann Heinrich Louis Krüger in 1912. The projection is known by several names: Gauss Conformal or Gauss-Krüger in Europe; the transverse Mercator in the US; or Gauss-Krüger transverse Mercator generally. The projection is conformal with a constant scale on the central meridian. (There are other conformal generalizations of the transverse Mercator from the sphere to the ellipsoid but only Gauss-Krüger has a constant scale on the central meridian.) Throughout the twentieth century the Gauss-Krüger transverse Mercator was adopted, in one form or another, by many nations (and international bodies); in addition it provides the basis for the Universal Transverse Mercator series of projections. The
Gauss-Krüger projection is now the most widely used projection in accurate large scale mapping.

The projection, as developed by Gauss and Krüger, was expressed in terms of low order power series which were assumed to diverge in the east-west direction, exactly as in the spherical version. This was proved to be untrue by British cartographer E.H. Thompson, whose unpublished exact (closed form) version of the projection, reported by L.P. Lee in 1976, [8] showed that the ellipsoidal projection is finite (below). This is the most striking difference between the spherical and ellipsoidal versions of the transverse Mercator projection: Gauss-Krüger gives a reasonable projection of the whole ellipsoid to the plane, although its principal application is to accurate large scale mapping "close" to the central meridian.

![Ellipsoidal Transverse Mercator: a finite projection.](image)

**Features**

- Near the central meridian (Greenwich in the above example) the projection has low distortion and the shapes of Africa, Western Europe, Britain, Greenland, and Antarctica compare favourably with a globe.
- The central regions of the transverse projections on sphere and ellipsoid are indistinguishable on the small scale projections shown here.
- The meridians at 90° east and west of the chosen central meridian project to horizontal lines through the poles. The more distant hemisphere is projected above the North Pole and below the South Pole.
• The equator bisects Africa, crosses South America and then continues onto the complete outer boundary of the projection; the top and bottom edges and the right and left edges must be identified (i.e. they represent the same lines on the globe). (Indonesia is bisected).

• Distortion increases towards the right and left boundaries of the projection but it does not increase to infinity. Note the Galapagos Islands where the 90° west meridian meets the equator at bottom left.

• The map is conformal. Lines intersecting at any specified angle on the ellipsoid project into lines intersecting at the same angle on the projection. In particular parallels and meridians intersect at 90°.

• The point scale factor is independent of direction at any point so that the shape of a small region is reasonably well preserved. The necessary condition is that the magnitude of scale factor must not vary too much over the region concerned. Note that while South America is distorted greatly the island of Ceylon is small enough to be reasonably shaped although it is far from the central meridian.

• The choice of central meridian greatly affects the appearance of the projection. If 90°W is chosen then the whole of the Americas is reasonable. If 145°E is chosen the Far East is good and Australia is oriented with north up.

In most applications the Gauss–Krüger is applied to a narrow strip near the central meridians where the differences between the spherical and ellipsoidal versions are small, but nevertheless important in accurate mapping. Direct series for scale, convergence and distortion are functions of eccentricity and both latitude and longitude on the ellipsoid: inverse series are functions of eccentricity and both x and y on the projection. In the secant version the lines of true scale on the projection are no longer parallel to central meridian; they curve slightly. The convergence angle between projected meridians and the x constant grid lines is no longer zero (except on the equator) so that a grid bearing must be corrected to obtain a true compass bearing. The difference is small, but not negligible, particularly at high latitudes.

### 4.1 Implementations of the Gauss–Krüger projection

In his 1912[6] paper Krüger presented two distinct solutions, distinguished here by the expansion parameter:

• **Krüger–n** (paragraphs 5 to 8). Formulae for the direct projection, giving the coordinates \(x\) and \(y\), are fourth order expansions in terms of the third flattening,
n (the ratio of the difference and sum of the major and minor axes of the ellipsoid). The coefficients are expressed in terms of latitude (φ), longitude (λ), major axis (a) and eccentricity (e). The inverse formulae for φ and λ are also fourth order expansions in n but with coefficients expressed in terms of x, y, a and e. (See Transverse Mercator: flattening series)

- **Krüger–λ** (paragraphs 13 and 14). Formulae giving the projection coordinates x and y are expansions (of orders 5 and 4 respectively) in terms of the longitude λ, expressed in radians: the coefficients are expressed in terms of φ, a and e. The inverse projection for φ and λ are sixth order expansions in terms of the ratio x/a, with coefficients expressed in terms of y, a and e. (See Transverse Mercator: Redfearn series)

The Krüger–λ series were the first to be implemented, possibly because they were much easier to evaluate on the hand calculators of the mid twentieth century.

- **Lee–Redfearn–OSGB.** In 1946 L.P. Lee[^9] confirmed the λ expansions of Krüger and proposed their adoption by the OSGB[^10] but Redfearn (1948)[^11] pointed out that they were not accurate because of (a) the relatively high latitudes of Great Britain and (b) the great width of the area mapped, over 10 degrees of longitude. Redfearn extended the series to eighth order and examined which terms were necessary to attain an accuracy of 1mm (ground measurement). The Redfearn series are still the basis of the OSGB map projections.[^10]

- **Thomas–UTM** The λ expansions of Krüger were also confirmed by Paul Thomas in 1952:[^12] they are readily available in Snyder.[^13] His projection formulae, completely equivalent to those presented by Redfearn, were adopted by the United States Defense Mapping Agency as the basis for the UTM.[^14] They are also incorporated into the Geotrans[^15] coordinate converter made available by the United States National Geospatial-Intelligence Agency.[^5]

- **Other countries.** The Redfearn series are the basis for geodetic mapping in many countries: Australia, Germany, Canada, South Africa to name but a few. (A list is given in Appendix A.1 of Stuifbergen 2009.)[^16]

- Many variants of the Redfearn series have been proposed but only those adopted by national cartographic agencies are of importance. For an example of modifications which do not have this status see Transverse Mercator: Bowring series). All such modifications have been eclipsed by the power of modern computers and the development of high order n-series outlined below. The precise Redfearn series, although of low order, cannot be disregarded as they are still enshrined in the quasi-legal definitions of OSGB and UTM etc.
The Krüger–n series are described on the page Transverse Mercator: series in "n" (third flattening). They have been implemented (to fourth order in n) by the following nations.

- **France**[^17]
- **Finland**[^18]
- **Sweden**[^19]

Higher order versions of the Krüger–n series have been implemented to seventh order by Ensager and Poder[^20] and to tenth order by Kawase.[^21] Apart from a series expansion for the transformation between latitude and conformal latitude, Karney has implemented the series to thirtieth order.[^22]

### 4.1.1 Exact Gauss-Krüger and accuracy of the truncated series

The exact solution of E. H. Thompson, described by L.P. Lee,[^8] is summarized on the page Transverse Mercator: exact solution. It is constructed in terms of elliptic functions (defined in chapters 19 and 22 of the NIST[^23] handbook) which can be calculated to arbitrary accuracy using algebraic computing systems such as Maxima.[^24] Such an implementation of the exact solution is described by Karney (2011).[^22]

The exact solution is a valuable tool in assessing the accuracy of the truncated n and λ series. For example, the original 1912 Krüger–n series compares very favourably with the exact values: they differ by less than 0.31 μm within 1000 km of the central meridian and by less than 1 mm out to 6000 km. On the other hand the difference of the Redfearn series used by Geotrans and the exact solution is less than 1 mm out to a longitude difference of 3 degrees, corresponding to a distance of 334 km from the central meridian at the equator but a mere 35 km at the northern limit of an UTM zone. Thus the Krüger–n series are very much better than the Redfearn λ series.

The Redfearn series become much worse as the zone widens. Karney discusses Greenland as an instructive example. The long thin landmass is centered on 42W and, at its broadest point, is no more than 750 km from that meridian while the span in longitude reaches almost 50 degrees. Krüger–n is accurate to within 1mm but the Redfearn version of the Krüger–λ series has a maximum error of 1 kilometer.

Karney's own 8th order (in n) series is accurate to 5 nm within 3900 km of the central meridian.

5. State Plane Coordinate System

The State Plane Coordinate System (SPCS) was developed in the 1930s by the U.S. Coast and Geodetic Survey to provide a common reference system for surveyors and mappers. The goal was to design a conformal mapping system for the country with a maximum scale distortion of one part in 10,000, which at the time was considered the limit of surveying accuracy. The State Plane Coordinate System (SPCS) is used for local surveying and engineering applications, but isn’t used if crossing state lines.

The State Plane grid system is very similar to that used with the UTM system, with the exception of where the origin for the grids is located. The easting origin for each zone is always placed an arbitrary number of feet west of the western boundary of the zone, eliminating the need for negative easting values. The northing origin, however, is not at the equator as in UTM, but rather it is placed at an arbitrary number of feet south of the state border.

To maintain the accuracy of one part in 10,000 and minimize distortion, large states were divided into zones, and depending on their orientation different projections were chosen. The three conformal projections used are listed below.

- Lambert Conformal Conic... for states that are longer east–west, such as Tennessee and Kentucky.
- Transverse Mercator projection... for states that are longer north–south, such as Illinois and Vermont.
- The Oblique Mercator projection... for the panhandle of Alaska, because it lies at an angle.
The number of zones in a state is usually determined by the area the state covers and ranges from one to as many as ten in Alaska. Each zone has a unique central meridian.

6. NATO System

Like so many things in cartography, the UTM system was inspired by the military. Some scientists object to using it for that reason. My philosophy is that if you refuse to use a useful tool because of its origins, you probably have other problems that will keep you from being an effective scientist.

Maps in the U.S., Canada, and many western European countries use a grid system common to NATO, the North Atlantic Treaty Organization.

Grid Zones

The map above shows how the NATO UTM system divides the earth into 60 longitude zones, each six degrees wide. The numbering begins at Zone 1 at 180 degrees west and proceeds eastward. To find the grid zone for any longitude:
- Treat west longitude as negative and east as positive.
- Add 180 degrees; this converts the longitude to a number between zero and 360 degrees.
- Divide by 6 and round up to the next higher number.

For example, Green Bay, Wisconsin is at 88 degrees west, or -88 degrees. Adding 180, we get 92 degrees (Green Bay is 92 degrees east of the International Date Line). 92/6 = 15.33, which we round up to 16. So Green Bay is in Zone 16. This information will usually be somewhere on just about every topographic map.

Latitude is also divided into zones, but less regularly. Zones are lettered from A at the South Pole to Z at the North. The circle south of 80 degrees is divided into two zones, A and B. Thereafter zones are 8 degrees wide. Zone M is just south of the equator and N is north. Zone T, between 40 and 48 degrees north, includes Green Bay. Zone X, from 72 to 84 degrees north, is 12 degrees wide and zones Y and Z cover the North Polar Region north of 84 degrees. I and O are not used because they can be too easily confused with numbers.

In the unlikely event you ever have to use the UTM grid on a map of the Polar Regions, those areas are covered by a different conformal projection called the Polar Stereographic. Since compass directions have little meaning at the poles, one direction on the grid is arbitrarily designated "north-south" and the other "east-west" regardless of the actual compass direction. The UTM coordinates are called "false northing" and "false easting."

**Grid North**

Since meridians converge, the UTM grid coincides exactly with compass directions only along the central meridian of each zone. For Green Bay, in zone 16, that meridian is 87 degrees west. Elsewhere, the grid makes an angle to the meridians, and at the edges of the zone it's pretty noticeable. It's important to note whether azimuths are being plotted with respect to true north (meridians) or *grid north* (the north-south lines of the grid). This information, along with the direction of magnetic north, is plotted on most topographic maps. Military people, who work with UTM grids all the time, tend to rely on grid north. Scientists in the field, who are more likely to use the UTM grid only for recording locations, might opt to use whatever is most convenient.
100-kilometer Digraphs

NATO maps subdivide grid zones into 100-kilometer squares, which are labeled with two-letter designations called *digraphs*. The actual lettering scheme is complex, but is designed so that the same digraph is not repeated within 15 degrees in any direction.

The longitude zones in the UTM system are extremely important because the whole map projection changes at the boundary. The latitude zones are comparatively unimportant because the projection cylinder used in any zone circles the entire globe. They are used principally to give a rough idea of latitude. So digraph letters change at the boundaries of grid zones, but not at latitude zones.

### 6.1 Determining Grid Coordinates

On NATO-style maps, the fundamental unit is the 100-km square. Imagine that Point X lays .72417 of the way across the square from west to east, and .43762 of the way up from south to north. The digraph identifier for this square is DQ.

On many maps, one-kilometer grid squares are printed. The one-kilometer grid square containing Point X is shown at lower right. It's .417 of the way across the square from west to east and .762 of the way from south to north.

We could give the locations of point X accurate to the nearest meter by saying it is at .72417 easts, and .43762 norths, just like reading Cartesian coordinates. The decimal point is cumbersome and unnecessary, so we omit it. We give the coordinate as 72417 43762, understanding that the first five digits are the fraction of the way across and the second is the fraction of the way up.
To specify the location accurate to only ten meters, we would say 7241 4376, understanding that the point is somewhere between .7241 and .7242 of the way across, and somewhere between .4376 and .4377 of the way up. A location accurate to 100 meters would be 724 437, and a location accurate to a kilometer would be 72 43.

Can we round? Depends on the application. If you want the best estimate for a given level of accuracy and you don't expect to have to refine it, then probably. If your project supervisor says don't round, then don't (not rounding is the norm in the military). If you want to specify something as being within a box of a certain size, then no. Especially with one-kilometer squares, don't round. One kilometer squares are printed on many maps. Point X above is between 72 and 73 kilometers from the west edge of the grid square and 43 and 44 kilometers from the bottom. It is in grid square 72 43. Rounding the value 43762 to 44 would put you in the next grid square north.

- Grid coordinates always have an even number of digits. The first half is the easting or fraction of the way east across the 100-km square, the second is the northing or fraction of the distance north across the square.
- The number of digits indicates the level of accuracy. Take the number of digits in each portion of the coordinate. Raise 10 to that power and divide 100 km (100,000 meters) by the result to find the level of accuracy. Example: 3 digits. $10^3$ =1000. 100,000/1000 = 100. Hence 3 digits for each portion (6 digits overall) means 100-meter accuracy.
- To remove ambiguity, always include the digraph. Point X, to 100-meter accuracy, is at DQ 724 437.
- In cases where information is being reported to someone far away, it may be necessary to include the zone designators as well.
- Often coordinates are run together, like DQ724437. No problem; simply count the digits and split them in half.
- Never round 1-kilometer grid locations because the identifier refers to a specific grid box on the map.

7. **Military Grid Reference System (MGRS)**

The Military Grid Reference System (MGRS) is an extension of the UTM system. UTM zone number and zone character are used to identify area 6 degrees in east-west extent and 8 degrees in north-south extent. UTM zone number and designator are followed by 100 km square easting and northing identifiers. The system uses a set of alphabetic characters for the 100 km grid squares. Starting at the 180 degree meridian the characters A to Z (omitting me and O) are used for 18 degrees before starting over. From the equator north the characters A to V (omitting I and O) are used for 100 km
squares, repeating every 2,000 km. Northing designators normally begin with 'A' at the equator for odd numbered UTM easting zones.

For even numbered easting zones the northing designators are offset by five characters, starting at the equator with 'F'. South of the equator, the characters continue the pattern set north of the equator. Complicating the system, ellipsoid junctions (spheroid junctions in the terminology of MGRS) require a shift of 10 characters in the northing 100 km grid square designators. Different geodetic datums using different reference ellipsoids use different starting row offset numbers to accomplish this.

If 10 numeric characters are used, a precision of 1 meter is assumed. 2 characters imply a precision of 10 km. From 2 to 10 numeric characters the precision changes from 10 km, 1 km, 100 m 10 m, to 1 m.

**MGRS 100,000-meter square identification**

- The 100,000-meter columns, including partial columns along zone, datum, and ellipsoid junctions, are lettered alphabetically, A through Z (with I and O omitted), north and south of the Equator, starting at the 180° meridian and proceeding easterly for 18°. The alphabetical sequence repeats at 18° intervals.
- To prevent ambiguity of identifications along ellipsoid junctions changes in the order of the row letters are necessary. The row alphabet (second letter) is shifted ten letters. This decreased the maximum distance in which the 100,000-meter square identification is repeated.
- The 100,000-meter row lettering is based on a 20-letter alphabetical sequence (A through V with I and O omitted). This alphabetical sequence is read from south to north, and repeated at 2,000,000-meter intervals from the Equator.
- The row letters in each odd numbered 6° grid zone are read in an A through V sequence from south to north.
- In each even-numbered 6° grid zone, the same lettering sequence is advanced five letters to F, continued sequentially through V and followed by A through V.
- The advancement or staggering of row letters for the even-numbered zones lengthens the distance between 100,000-meter squares of the same identification.
- Deviations from the preceding rules were made in the past. These deviations were an attempt to provide unique grid references within a complicated and disparate world-wide mapping system.
- Determination of 100,000-meter grid square identification is further complicated by the use of different ellipsoids.
The military grid reference

The MGRS coordinate for a position consists of a group of letters and numbers which include the following elements:

- The Grid Zone Designation.
- The 100,000-meter square letter identification.
- The grid coordinates (also referred to as rectangular coordinates); the numerical portion of the reference expressed to a desired refinement.
- A reference is written as an entity without spaces, parentheses, dashes, or decimal points.
Conclusion

While The UTM is the better known regular cylindrical projection, there is no "best" projection for either a map or a digital cartographic data set. Projection decisions are always tradeoffs between desired map characteristics, which in turn depend on the desired use of the data. Projection decisions made many years ago for published paper maps are not necessarily the right decisions for corresponding digital data sets.

The Mercator was designed for sea navigation because it shows Rhumb lines as straight lines. Unfortunately, it is often and inappropriately used for world maps in atlases and wall charts. It presents a misleading view of the world because of excessive area distortion. The Mercator shows meridians and parallels as straight lines. Regular cylindrical projections are the only commonly used projections that show both meridians and parallels as straight lines.

Having digital data does not remove the necessity to make projection decisions. It is possible to store digital data in unprojected form, but this only defers projection decisions until run time. Displaying or plotting a digital data set requires projecting the data to a flat surface. Because runtime projection is computationally intensive, the USGS (U.S. Geological Survey) has chosen not to distribute unprojected data for most digital data products.

The USGS projects most of its digital products on the UTM, in the belief that this will serve more user needs, more often, than any other single projection.
References

[1]- Lambert, Johann Heinrich. 1772. *Ammerkungen und Zusatze zurder Land und Himmelscharten Entwerfung*. In Beyträge zum Gebrauche der Mathematik und deren Anwendung, part 3, section 6)

[2]- Albert Wangerin (Editor), 1894. *Ostwald’s Klassiker der exacten Wissenschaften* (54). Published by Wilhelm Engelmann. This is Lambert’s paper with additional comments by the editor. Available at the University of Michigan Historical Math Library.


time (2010) it is necessary to purchase several units in order to obtain the relevant pages: pp 1–14, 92–101 and 107–114.


[10]- A guide to coordinate systems in Great Britain. This is available as a pdf document at [3]


[21]- C. F. F. Karney (2011), Transverse Mercator with an accuracy of a few nanometers, J. Geodesy 85(8), 475-485 (2011); preprint of paper and C++ implementation of algorithms are available at tm.html.


[24]- The Mercator Projections Detailed derivations of all formulae quoted in this article