Map Projection Distortion

Ask Hjorth Larsen s021864

Supervisor: Steen Markvorsen

DTU - Technical University of Denmark

June 7, 2007

Front cover: Satellite imagery compiled to a world map using the Mercator projection.

Source: Google™Maps, `http://maps.google.com`

Abstract

This text concerns distortions inherent in cartographic projections. It derives measures of the local distortion of shape and area which can be used to obtain a quantitative estimate of the violation of conformality or equiareality of a certain projection. Further the Mercator and Lambert cylindrical projections are derived, and the measures are used to describe their modes of distortion

Table of contents

1 Introduction **1**

2 Terms and basics **1**

3 Geometric properties of the first fundamental form **2**

4 Cylindrical map projections **3**

5 Conclusion **7**

References **8**

1 Introduction

The term *map projection* generally refers to a way of representing features of a curved surface such as the surface of the Earth on a flat surface such as a piece of paper. It is a well-known fact[1, proposition 10.1] that no such representation can be made without distorting distances. For example, the Mercator projection distorts areas grossly near the poles, while approaching perfection close to the equator. Other map projections distort the shapes of land masses in some locations while correctly portraying *areas* on the entire globe.

While distortion of areas versus shapes is a fundamental problem, these are not the only constraints that determine the applicability of a map. For instance some projections have the useful property of mapping geodesics to geodesics (these projections are called *gnomonic*[2] - in our case the geodesics are lines in the plane and great circles on the sphere, though more realistic models of the Earth, notably ellipsoids, may be desirable under certain circumstances), and in some applications there might be a special interest in correctly portraying the distances or directions to a particular point.

This text will attempt to characterize the magnitude of distortion of a map projection by formulating measures of how grossly the map violates properties such as conformality or equiareality across the globe. What complicates this task is that the two mentioned properties, equiareality and conformality, are different in the following sense: a conformal map can be quite precise *locally*, since in arbitrarily small neighbourhoods of any point the scale is largely constant, whereas a non-conformal map will distort directions differently no matter how small a neighbourhood is considered. As we shall see, shape distortion can thus easily be measured *locally* by a number at each point. But distortion of area takes effect over distances, and a measure of such error must therefore take into account the size of the area to be considered, or attempt to obtain information about how rapidly the distortion will increase from the point in question.

The following section will introduce the mathematical foundation to a more formal description of map projections and the presented problems.

2 Terms and basics

To clarify the notions of distortion introduced previously we shall refer to a map projection as a mathematical map $\boldsymbol{r}$ from some subset of the plane U into a subset of the unit sphere[1] $\mathbb{S}$, i.e.

$$\boldsymbol{r} : U \mapsto \mathbb{S}. \tag{1}$$

In other words the map projection is technically a *chart* on the manifold in question. The statement that distances must be distorted follows simply from *Gauss' Theorema Egregium*, which guarantees that no such chart $\boldsymbol{r}$ exists which is *isometric*. However - as we shall see - many such maps can be derived that, while not isometric, instead preserve other geometric properties. The two most prominent examples of such properties are:

- Conformal maps. These maps preserve angles, i.e. if two curves in the plane intersect at some angle, then their images will intersect at the same angle.
- Equiareal maps. These maps preserve areas, i.e. the area of the image of every subset of U is equal to the area of that subset itself.

We shall use the *first fundamental form*[1, pp. 97-98] of a regular parametrized surface $\boldsymbol{r}(u, v)$. The first fundamental form is characterized by the three functions

$$E(u,v) = \left|\left|\frac{\partial \boldsymbol{r}}{\partial u}\right|\right|^2 \tag{2}$$

$$F(u,v) = \left\langle \frac{\partial \boldsymbol{r}}{\partial u}, \frac{\partial \boldsymbol{r}}{\partial v} \right\rangle \tag{3}$$

$$G(u,v) = \left|\left|\frac{\partial \boldsymbol{r}}{\partial v}\right|\right|^2, \tag{4}$$

[1] Some texts equivalently define map projections as mappings *from* the sphere into the plane. Moreover, this text will only concern mappings to the sphere, yet it would mathematically make equal sense to consider any other two-dimensional differentiable manifold of nonzero gaussian curvature.

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $||\cdot||$ its induced norm. It will further prove useful to consider the matrix

$$\mathscr{F}_I = \begin{bmatrix} E & F \\ F & G \end{bmatrix}. \tag{5}$$

The notation $\mathscr{F}_I(\boldsymbol{r})$ or $\mathscr{F}_I(u, v)$ shall be used to stress the dependence on projection and coordinates when appropriate. Now, the projection $\boldsymbol{r}$ is conformal if and only if there exists a function $\alpha : U \mapsto \mathbb{R}$ such that for all $(u, v) \in U$,

$$\mathscr{F}_I(u, v) = \alpha(u, v)\mathbf{I} \tag{6}$$

where $\mathbf{I}$ denotes the unit matrix in $\mathbb{R}^2$. Furthermore the projection is equiareal if and only if

$$\det \mathscr{F}_I = 1 \tag{7}$$

everywhere in U (see [1, theorem 5.2] and [1, theorem 5.3] for proofs of these propositions). Having thus stated the fundamental definitions and most important propositions to be used, it is time to examine the mathematical implications more closely, which will be the subject of the next section.

3 Geometric properties of the first fundamental form

This section will derive actual measures of how badly a map preserves angles and areas by considering the properties of eigenvalues of the matrix $\mathscr{F}_I$.

$\mathscr{F}_I$ is symmetric and can thus be diagonalized by means of an orthogonal substitution[3, theorem 8.33], i.e. there exists an orthogonal matrix $\mathbf{Q}$ containing the eigenvectors of $\mathscr{F}_I$, and a diagonal matrix $\boldsymbol{\Lambda}$ containing the corresponding eigenvalues in consistent order, such that

$$\boldsymbol{\Lambda} = \mathbf{Q}^T \mathscr{F}_I \mathbf{Q}. \tag{8}$$

This guarantees that two linearly independent eigenvectors exist, which will be implicitly assumed from now on. The eigenvalues of $\mathscr{F}_I$ can be found from

$$\begin{aligned} 0 = \det(\mathscr{F}_I - \lambda \mathbf{I}) = \lambda^2 - \lambda(E + G) + EG - F^2 \Leftrightarrow \\ \lambda = \frac{E+G}{2} \pm \frac{1}{2}\sqrt{(E-G)^2 + (2F)^2} \end{aligned} \tag{9}$$

It is easily shown that both of these eigenvalues are positive: the inequality

$$0 \leqslant E + G - \sqrt{(E-G)^2 + (2F)^2} \tag{10}$$

reduces to

$$F^2 \leqslant EG \tag{11}$$

which is true by the well-known Cauchy-Schwarz inequality

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leqslant ||\boldsymbol{x}||\,||\boldsymbol{y}||\,. \tag{12}$$

Thus $\mathscr{F}_I$ is positive-definite.

The eigenvalues of $\mathscr{F}_I$ are of particular interest, as will be shown by the following investigation. Consider a regular curve $\gamma : \mathbb{R} \mapsto U \subseteq \mathbb{R}^2$ and any smooth surface (patch) $\sigma : U \mapsto \mathbb{R}^3$ with first fundamental form characterized by $\mathscr{F}_I$ as usual.

It is known[1, pp. 97-98] that the curve length of some segment of $\sigma \circ \gamma$ starting in $t = t_0$, given $\dot{\gamma} = [\dot{u}, \dot{v}]^T$, is determined by the integral

$$\mathscr{L}(t) = \int_{\text{curve}} \mathrm{d}\ell = \int_{t_0}^{t} \sqrt{\Omega(s)}\mathrm{d}s, \tag{13}$$

where

$$\Omega(t) = \begin{bmatrix} \dot{u} & \dot{v} \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} \dot{u} \\ \dot{v} \end{bmatrix}. \tag{14}$$

Now, if $\left[\begin{array}{cc}\dot{u} & \dot{v}\end{array}\right]$ is some eigenvector of $\mathscr{F}_I$ corresponding to the eigenvalue λ, then the integrand simplifies to

$$\Omega(t) = \left[\begin{array}{cc}\dot{u} & \dot{v}\end{array}\right] \lambda \left[\begin{array}{c}\dot{u} \\ \dot{v}\end{array}\right] = \lambda \, ||\dot{\gamma}||^2 \tag{15}$$

This reveals the useful geometric information that *the square root of the eigenvalue is the amount by which the infinitesimal curve element parallel to the eigenvector is scaled*. Consider now some curve γ with tangent $\dot{\gamma}$ which is not an eigenvector; we may expand $\dot{\gamma}$ in the eigenvector basis into two components (which will be scaled differently by the projection according to the eigenvalues),

$$\dot{\gamma} = \alpha \boldsymbol{v}_1 + \beta \boldsymbol{v}_2, \tag{16}$$

where $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are unit eigenvectors of $\mathscr{F}_I$ with eigenvalues λ_1 and λ_2, respectively. We have

$$\begin{aligned} \Omega(t) &= \dot{\boldsymbol{\gamma}}^T \mathscr{F}_I \dot{\boldsymbol{\gamma}} \\ &= (\alpha \boldsymbol{v}_1 + \beta \boldsymbol{v}_2)^T \mathscr{F}_I (\alpha \boldsymbol{v}_1 + \beta \boldsymbol{v}_2) \\ &= (\alpha \boldsymbol{v}_1 + \beta \boldsymbol{v}_2)^T (\alpha \lambda_1 \boldsymbol{v}_1 + \beta \lambda_2 \boldsymbol{v}_2) \\ &= \alpha^2 \lambda_1 + \beta^2 \lambda_2 \end{aligned} \tag{17}$$

This shows that the scaling of *any* curve through the point is determined directly by the eigenvalues of the matrix $\mathscr{F}_I$, and the difference in scale along different directions is a direct expression for shape distortion. The eigenvalue expressions therefore prompt a natural measure of the degree of shape distortion (i.e. lack of conformality), namely the difference of the eigenvalues divided by their sum

$$\delta = \frac{\sqrt{(E-G)^2 + (2F)^2}}{E+G} \tag{18}$$

Thus, δ is the *maximal possible difference in squared scales between curve segments through a point, relative to the square sum of those two scales*. Note the important property that if both eigenvalues are multiplied by the same constant, the measure of distortion of angles should logically be preserved - this is obviously true for δ.

Since the square root of the (always positive) determinant measures directly the local scaling factor of area it is logical to use its relative difference from 1 as a measure of area distortion, i.e.

$$\eta = \left|1 - \sqrt{\det \mathscr{F}_I}\right| = \left|1 - \sqrt{EG - F^2}\right| \tag{19}$$

However this is not the only natural choice. For example, why would taking the absolute value be superior to squaring? Yet another reasonable measure would be $|1 - \frac{1}{\det \mathscr{F}_I}|$. These measures are of course similar in the sense that they have the same properties of monotonicity and will consistently rate projections correctly relative to each other. The problem of which one to use is more important if one wishes to construct a compromise projection which is neither conformal nor equiareal but reasonably close to both. In that case it will matter how the two kinds of distortion are compared to each other.

In conclusion, we have obtained expressions for the local deviation of a map from equiareality or conformality. These expressions are based on the two eigenvalues of the matrix which characterizes the first fundamental of the map, which completely determine *both* the distortion of area and shape. The next section will derive some simple map projections and apply these measures to describe the distortion on those maps.

4 Cylindrical map projections

The *cylindrical* map projections are arguably some of the simplest examples. This section will consider the derivation of conformal and equiareal cylindrical projections, then study the resulting distortions of area and directions, respectively. Every cylindrical projection has a parametrization of the form

$$\boldsymbol{\sigma}(u,v) = \left[\begin{array}{c}\omega(u)\cos(v) \\ \omega(u)\sin(v) \\ \gamma(u)\end{array}\right] \tag{20}$$

where:

- $(u,v) \in U = I \times [-\pi, \pi]$ where I is some interval in $\mathbb{R}$
- for all $u \in I : \omega(u) \geqslant 0$, $\omega^2(u) + \gamma^2(u) = 1$

The first fundamental form of such a projection can be calculated directly using the chain rule:

$$\mathscr{F}_I = \begin{bmatrix} \frac{\dot{\omega}^2}{1-\omega^2} & 0 \\ 0 & \omega^2 \end{bmatrix} \tag{21}$$

where the dot denotes differentiation. Let us first derive all conformal projections of this type. The off-diagonal elements of the matrix are evidently always 0 in cylindrical projections. Applying the conformality condition (6) means simply equating the diagonal elements, which results in the differential equation

$$\dot{\omega}^2 = \omega^2(1-\omega^2). \tag{22}$$

Taking the square root would yield two differential equations with opposite sign on the right hand side. Now, the relationship between the two unknown functions ω and γ dictates that as one increases, the other must decrease (numerically) in order for $\sigma(u,v)$ to land on the sphere. Knowing this we may discard one of these differential equations since this will merely force γ to conform to our choice. We thus decide to consider only the equation

$$\dot{\omega} = -\omega\sqrt{1-\omega^2}. \tag{23}$$

Separating and integrating yields (using [4, p. 257, entry 22])

$$u - c = -\int \frac{\mathrm{d}\omega}{\omega\sqrt{1-\omega^2}} = \ln \frac{1+\sqrt{1-\omega^2}}{\omega} \tag{24}$$

for some arbitrary constant c. By rearranging it is possible to isolate ω:

$$\begin{aligned} \sqrt{1-\omega^2} &= \omega \mathrm{e}^{u-c} - 1 \Rightarrow \\ \omega^2 \left\{1 + \mathrm{e}^{2(u-c)}\right\} &= 2\omega \mathrm{e}^{u-c} \Rightarrow \\ \omega \left\{\frac{\mathrm{e}^{-(u-c)} + \mathrm{e}^{u-c}}{2}\right\} = \omega\cosh(u-c) &= 1 \Rightarrow \\ \omega &= \operatorname{sech}(u-c) \end{aligned} \tag{25}$$

By inserting ω into the differential equation (22) it can be verified that $\operatorname{sech}(u-c)$ is indeed a solution (which was not necessarily the case after the rearrangements).

Knowing ω, γ can be found directly. Specifically, the relation

$$\tanh^2 u + \sinh^2 u = 1 \tag{26}$$

and the constraint that γ must increase as ω decreases for $u > 0$ prompts (the constraint ensures that similar but uninteresting solutions, like an inverted map, are thrown away)

$$\gamma(u) = \tanh(u-c). \tag{27}$$

Thus we have the family of conformal cylindrical map projections, one for each value of c. However c merely incorporates a trivial reparametrization (translation along the u axis, meaning we can freely choose where on the paper the map should be drawn), and we will therefore discard it, setting $c = 0$:

$$\mathcal{M}(u,v) = \begin{bmatrix} \operatorname{sech} u \cos v \\ \operatorname{sech} u \sin v \\ \tanh u \end{bmatrix}. \tag{28}$$

This is exactly the *Mercator* projection[1, p. 83] which has been mentioned earlier. It is evident that in order to span the entire globe, we must have all $u \in \mathbb{R}$, and even so the poles remain unmapped.

In order to investigate distortion of distances, consider

$$\mathscr{F}_I = \begin{bmatrix} \operatorname{sech}^2 u & 0 \\ 0 & \operatorname{sech}^2 u \end{bmatrix}. \tag{29}$$

It is easily checked that the projection is conformal, using (6). The area scaling factor is equal to the square root of the determintant $\sqrt{\det \mathscr{F}_I} = \operatorname{sech}^2 u$. Thus at $u = 0$ the determinant is 1, meaning that area is preserved quite well around $u = 0$ which we shall refer to as the Equator (though the coordinate system could be changed such that $u = 0$ corresponds to any great circle). As u increases, however, the scale increases greatly. Using the measure η defined in equation (19) and Taylor expanding, we get an expression for how quickly the area scaling factor increases away from the great circle corresponding to $u = 0$.

$$\eta(u, v) = \eta(u) = 1 - \operatorname{sech}^2 u = u^2 + u^3 \epsilon(u), \tag{30}$$

where ϵ is some ϵ-function, i.e. a continuous function with the property that $\epsilon(0) = 0$, guaranteed to exist by *Taylor's theorem*[5, p. 77]. The Taylor expansion shows that the distortion of area near the Equator increases quadratically with the distance from the Equator.

Note that as u approaches infinity, the projection will include points arbitrarily close to the North Pole - however the North Pole will never be mapped. See figure 1. Figure 1(a) shows a generic domain of definition, whereas figure 1(b) shows the standard longitude latitude parametrization for reference. Images of horizontal and vertical lines in the plane are plotted. Figure 1(c) shows the Mercator projection, and it is evident that this projection compresses grid elements near the poles, thus enlarging the representation of these regions on the plane by a factor of $\cosh^2 u$. Note that the grid elements all appear to be square, which is the logical consequence of conformality.

Next, let us find all cylindrical equiareal projections. Using again the matrix $\mathscr{F}_I$ from equation (21), and applying the condition that the determinant must be equal to one, we get the following differential equation:

$$\frac{\omega^2 \dot{\omega}^2}{1 - \omega^2} = 1 \tag{31}$$

Separating and integrating like in the previous case after discarding one of the differential equations emerging from taking the square root, we get

$$u - c = -\int \frac{\omega \mathrm{d}\omega}{\sqrt{1 - \omega^2}} = \sqrt{1 - \omega^2}. \tag{32}$$

The condition that ω be positive eliminates one solution from the resulting quadratic equation, and we are left with

$$\omega(u) = \sqrt{1 - (u - c)}. \tag{33}$$

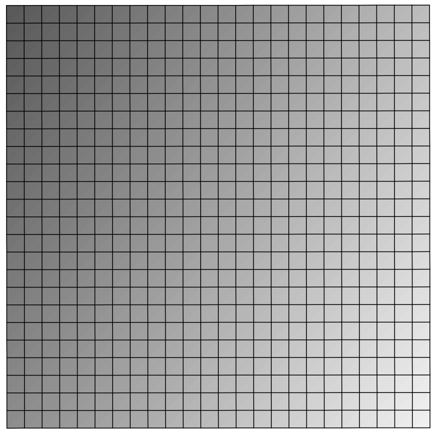
It follows that γ is simply u (or $-u$ which we discard). The resulting projection, where again we discard c since it only determines where on the paper the projection is drawn, is

$$\mathcal{L}(u, v) = \begin{bmatrix} \sqrt{1 - u^2} \cos v \\ \sqrt{1 - u^2} \sin v \\ u \end{bmatrix}, \tag{34}$$

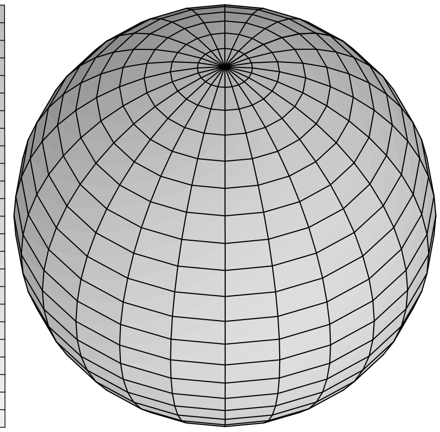
where $-1 \leqslant u \leqslant 1$. This appears to be the *Lambert* projection presented in [6], though from the different notations it is not immediately obvious (but can be shown using the identity $\cos \arcsin x = \sqrt{1 - x^2}$). While there exist many projections that are classified as cylindrical-equal area, they can only differ by scaling the vertical or horizontal axes, the possibility of which we have eliminated using radians such that $-\pi \leqslant v \leqslant \pi$. Depending on stretching, the projection can therefore be made locally conformal along any parallel. We shall let $u = 0$ correspond to the Equator in the following, but like before it could in reality be any great circle. See figure 1(d). The grid elements are all rectangular, but flat near the equator and tall near the poles, meaning that shapes are increasingly distorted near the poles. However all grid elements have the same size, which must obviously be the case given that the projection is equiareal.

Next consider the matrix $\mathscr{F}_I$ for the Lambert projection:

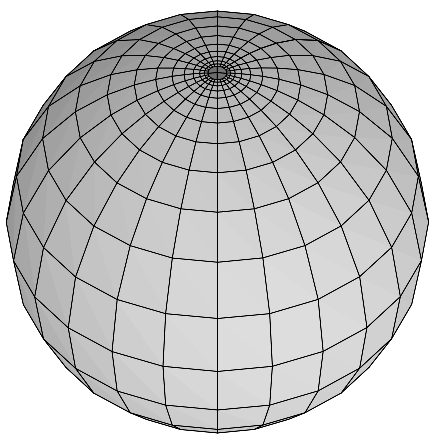
$$\mathscr{F}_I = \begin{bmatrix} \frac{1}{1-u^2} & 0 \\ 0 & 1 - u^2 \end{bmatrix} \tag{35}$$



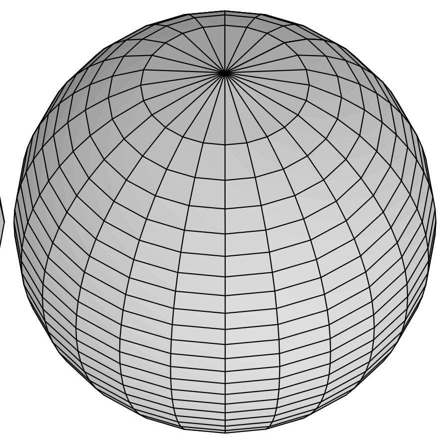
(a) The domain of definition. Varying sizes are used in the following figures.



(b) Standard longitude/latitude parametrization. Neither conformal nor equiareal, the individual grid elements have different area and shape across the globe.



(c) Mercator projection. All grid elements are square except for the very small local distortion. Their areas differ greatly, however, and infinitely much paper is necessary to map the poles. Notice that a small neighbourhood of the North Pole is missing for this reason.



(d) Lambert equal-area cylindrical projection. All grid elements have equal size. However they change shape in order to enforce this, particularly around the poles.

Figure 1: *Different projections. Each grid element represents the same area of paper, so areas with many small elements take up much space on the map.*

The measure of the lack of conformality, δ from equation (18), is

$$\delta(u,v) = \delta(u) = \frac{u^2(2-u^2)}{2-u^2+u^4} \tag{36}$$

This is 0 on the Equator, and a Taylor expansion here reveals the distortion to increase quadratically with distance in the neighbourhood. Thus this particular projection is best near the equator, although, as it has been asserted, it is possible to stretch the axes to achieve local conformality along any parallel.

In conclusion, we have derived two map cylindrical projections with the constraints that one be conformal while the other equiareal. These constraints proved to be sufficient to find a unique projection in each case, discounting trivialities like scaled axes and inverted maps. The conformal one is the Mercator projection, and the equiareal one is the Lambert projection. The derived measure η of area distortion indicated that the Mercator projection is locally exact near the Equator with an error of $\mathcal{O}(u^2)$ where u is the distance from the Equator, but explodes near the poles. The derived measure δ of shape distortion indicated that the Lambert projection is locally exact near the Equator, also with an error of $\mathcal{O}(u^2)$.

While the inclusion of further projections would go beyond the scope of this text, it should be noted that the presented method - namely proposing a projection type (here cylindrical) and a constraint (conformal or equiareal) could likewise be used to derive pseudocylindrical projections, which are similar to cylindrical ones except they allow the interval of definition of v to depend on u:

$$\sigma(u,v) = \begin{bmatrix} \omega(u)\cos(\psi(u)v) \\ \omega(u)\sin(\psi(u)v) \\ \gamma(u) \end{bmatrix} \tag{37}$$

Specifying the width ψ of this interval, which is thus one function of u, would be sufficient to obtain differential equations like the ones seen in this text. For example, setting $\psi(u) = \sec(u)$ such that the domain of definition of v becomes very small around the poles, and setting the determinant of the corresponding matrix $\mathscr{F}_I$ equal to 1 results in the differential equation

$$\omega^2\dot{\omega}^2 = (1-\omega^2)\cos^2 u \tag{38}$$

which clearly resembles the ones encountered above.

5 Conclusion

This text has derived two measures of the distortion of shape and area, respectively, associated with cartographic projections. It is demonstrated that the distortion is determined completely by the two eigenvalues of a certain matrix which characterizes the first fundamental form of a map, and the measures which are based on said eigenvalues have direct geometric interpretations. The measure δ denotes shape distortion and is zero if and only if the projection is locally conformal at the point in question. The measure η denotes area distortion and is zero if and only if the projection is locally equiareal at the point in question. Greater distortion of one of these types consequently results in greater values of the corresponding measure.

No attempt has been made to combine these measures of distortion into a single one since this could be done in different ways depending on specific requirements.

The Mercator and Lambert cylindrical projections, which are conformal and equiareal, respectively, have been derived using only the constraints that they be cylindrical and either conformal or equiareal. Both projections approach isometry near the Equator, but exhibit large distortions near the poles, follows from the derived measures of distortion and 3D plots of the projection aside from common knowledge. The distortion in a short distance u from the Equator is in both cases, according to the respective measures, $\mathcal{O}(u^2)$ which means geometrically - in the case of equiareality - that the greatest difference in scale along different directions through the point of interest increase quadratically, and - in the case of conformality - that the scale of the map increases quadratically.

Finally, in the light of the observations made in this text it would be worthwhile to consider (among others) these questions in a future text:

- Is it possible to define an equiareal or conformal projection with distortion of shapes or areas, respectively, smaller than $\mathcal{O}(u^2)$ where u denotes the distance from some point where the distortion approaches 0?

- Is it possible to define a projection which exhibits both kinds of distortions, but smaller than $\mathcal{O}(u^2)$?

References

[1] Andrew Pressley, *Elementary Differential Geometry* (Springer, 2001. ISBN 1-85233-152-6)

[2] `http://en.wikipedia.org/wiki/Gnomonic_projection`, Wikipedia entry on gnomonic projections

[3] Jens Eising, *Liner Algebra* (Department of Mathematics, DTU, 1999. ISBN 87-88764-39-7)

[4] *Handbook of Chemistry and Physics*, 44th edition (Scientific Publishing, 1963)

[5] P.W. Karlsson and V.L. Hansen, *Matematisk Analyse 2, 4. udgave* (Department of Mathematics, DTU, 1998. ISBN 87-88764-60-5)

[6] Julio Benítez and Néstor Thome, *Applications of Differential Geometry to Cartography* (Int. J. Math. Educ. Sci. Technol. 2004, Vol.35, No. 1, 29-38)