1995

A General Theory of Geodesics with Applications to Hyperbolic Geometry

Deborah F. Logan
University of North Florida

A GENERAL THEORY OF GEODESICS WITH APPLICATIONS TO HYPERBOLIC GEOMETRY

by

Deborah F. Logan

A thesis submitted to the Department of
Mathematics and Statistics
in partial fulfillment of the requirements for the degree
of

Master of Science in Mathematical Sciences

UNIVERSITY OF NORTH FLORIDA

COLLEGE OF ARTS AND SCIENCES

July, 1995

The thesis of Deborah F. Logan is approved:

Signature Deleted

(Date)

7/20/95

Signature Deleted

7/20/95

Signature Deleted

7/20/95

Committee Chairperson

Accepted for the Department:

Signature Deleted

7/24/95

Chairperson

Accepted for the College of Arts and Sciences:

Signature Deleted

7-31-95

Dean

Accepted for the University:

Signature Deleted

8/2/95

Dean of Graduate Studies

ACKNOWLEDGEMENTS

I wish to thank Dr. Leonard Lipkin for his guidance, encouragement, and his thorough analysis of the contents of this paper. This study of Differential Geometry has helped me see the study of advanced mathematics as one big picture rather than disconnected entities.

I appreciate the time and effort that Dr. Jingcheng Tong and Dr. Faiz Al-Rubaee put into reading my thesis. I wish to thank Dr. Champak Panchal and Dr. Donna Mohr for their planning of my graduate program and their guidance throughout.

Last, but not least, I want to extend my gratitude and love to my husband, Charlie, and my three children, Jonathan, April, and Philip. They have supported and encouraged me during my graduate studies.

TABLE OF CONTENTS

ABSTRACT

A General Theory of Geodesics
With Applications to Hyperbolic Geometry

In this thesis, the geometry of curved surfaces is studied using the methods of differential geometry. The introduction of manifolds assists in the study of classical two-dimensional surfaces. To study the geometry of a surface a metric, or way to measure, is needed. By changing the metric on a surface, a new geometric surface can be obtained. On any surface, curves called geodesics play the role of "straight lines" in Euclidean space. These curves minimize distance locally but not necessarily globally. The curvature of a surface at each point P affects the behavior of geodesics and the construction of geometric objects such as circles and triangles. These fundamental ideas of manifolds, geodesics, and curvature are developed and applied to classical surfaces in Euclidean space as well as models of non-Euclidean geometry, specifically, two-dimensional hyperbolic space.

INTRODUCTION

The geometry of curved surfaces has been studied throughout the history of mathematics. The development of the calculus enabled the theory of curves to flourish. It enabled a deeper study of curved surfaces that led to a better understanding and generalization of the whole subject. It was discovered that some properties of curved surfaces are intrinsic properties, that is, they are geometric properties that belong to the surface itself and not the surrounding space.

To study the geometry of a surface, we need a metric, or a way to measure. Any property of a surface or formula that can be deduced from the metric alone is intrinsic. By changing the metric on a surface, we will obtain a new geometric surface. The two-dimensional plane with the Euclidean metric gives rise to the Euclidean plane. The two-dimensional plane with the Poincaré metric gives rise to a model of non-Euclidean geometry called the Poincaré half-plane. This geometric surface will be studied in section three of chapter two.

The methods of differential geometry are used to study the geometry of curved surfaces. Differential geometry deals with objects such as tangent vectors, tangent fields, tangent spaces, differentiable functions on surfaces, and curves. The introduction of manifolds (a generalization of surfaces) will assist in the study of classical two-dimensional surfaces.

On any manifold, there are special curves called geodesics. These are curves that play the role of "straight lines" in Euclidean space. These curves minimize distance locally but do not necessarily minimize distance globally. From this fact comes the concept of convex neighborhoods, that is, neighborhoods in which pairs of points can be joined to each other by a unique minimizing geodesic.

An important property of a manifold is how the manifold is curving at each point P. Curvature affects the behavior of geodesics, the measure of angles, as well as the sum of the interior angles of geodesic triangles constructed on the manifold.

In chapter one, the fundamental ideas of manifolds, geodesics, and curvature are developed. In this chapter, these fundamental ideas are applied to classical surfaces in Euclidean space. In chapter two, a thorough study is done of models of non-Euclidean geometry, specifically, two-dimensional hyperbolic space. This study includes geodesics and curvature as well as geodesic triangles, hyperbolic circles, congruences, and similarities.

CHAPTER 1

MANIFOLDS AND GEODESIC THEORY

1. Manifolds

1.1 Definition. A geometric surface is an abstract surface M furnished with an inner product, $\cdot$, on each of its tangent planes. This inner product is required to be differentiable in the sense that if $\vec{V}$ and $\vec{W}$ are differentiable vector fields on M, then $\vec{V} \cdot \vec{W}$ is a differentiable real-valued function on M.

Each tangent plane of M at the point P has its own inner product. This inner product is a function which is bilinear, symmetric, and positive definite. An assignment of inner products to tangent planes is called a geometric structure or metric tensor on M. Therefore, the same surface furnished with two different geometric structures gives rise to two different geometric surfaces.

The two-dimensional plane, with the usual dot product on tangent vectors $<\vec{v},\vec{w}>= v_1w_1 + v_2w_2$, is the best-known geometric surface. Its geometry is two-dimensional Euclidean geometry. A simple way to get new geometric structures is to distort old ones. Let $g > 0$ be any differentiable function on the plane.

Define

$$\vec{v} \cdot \vec{w} = \frac{\langle \vec{v}, \vec{w} \rangle}{g^2(P)}$$

for tangent vectors $\vec{v}$ and $\vec{w}$ to the two-dimensional plane at P. This is a new geometric structure on the plane. As long as $g^2(P) \neq 1$, then the resulting geometric surface has properties quite different from the Euclidean plane.

In studying the geometry of a surface, some of the most important geometric properties belong to the surface itself and not the surrounding space. These are called intrinsic properties. In the nineteenth century, Riemann concluded the following: There must exist a geometrical theory of surfaces completely independent of $\mathbb{R}^3$. The properties of a surface M could be discovered by the inhabitants of M unaware of the space outside their surface.

At first, Riemannian geometry was a development of the differential geometry of two-dimensional surfaces in $\mathbb{R}^3$. From this perspective, given a surface $S \subset \mathbb{R}^3$, the inner product $\langle \vec{v}, \vec{w} \rangle$ of two vectors tangent to S at a point P of S is the inner product of these vectors in $\mathbb{R}^3$. This yields the measure of the lengths of vectors tangent to S. To compute the length of a curve, integrate the length of its velocity vector.

1.2 Definition. A regular curve in $\mathbb{R}^3$ is a function $\alpha:(a,b) \to \mathbb{R}^3$ which is of class C^k for some $k \geq 1$ and for which $d\alpha/dt \neq \vec{0}$ for all $t \in (a,b)$. A regular curve segment is a function $\alpha:[a,b] \to \mathbb{R}^3$ together with an open interval (c,d), with $c<a<b<d$, and a regular curve $\gamma:(c,d) \to \mathbb{R}^3$ such that $\alpha(t)=\gamma(t)$ for all $t \in [a,b]$.

1.3 Definition. The length of a regular curve segment $\alpha:[a,b] \to \mathbb{R}^3$ is $\int_a^b |d\alpha/dt|\, dt$.

The definition of inner product allows us to measure area of domains in S, the angle between two curves, and all other "metric" ideas in geometry. Certain special curves on S, called geodesics, will be a major focus in this thesis. These curves play the role of straight lines in Euclidean geometry.

The definition of the inner product at each point $P \in S$ yields a quadratic form I_P, called the first fundamental form of S at P, defined in the tangent plane T_PS by $I_P(\vec{v}) = \langle \vec{v}, \vec{v} \rangle$, $\vec{v} \in T_PS$. In 1827, Gauss defined a notion of curvature for surfaces. Curvature measures the amount that S deviates, at a point $P \in S$, from its tangent plane at P. Curvature, as Gauss defined it, depended only on the manner of measuring in S, which was the first fundamental form of S at P. Curvature will be discussed in section four of this chapter.

During Gauss' time, work was done to show that the fifth postulate of Euclid was independent of the other postulates of geometry. Euclid's fifth (parallel postulate) says [13]: "Given a straight line and a point not on the line then there is a straight line through the point which does not meet the given line." It was also earlier shown that this postulate is equivalent to the fact that the sum of the interior angles of a triangle equals 180°. This led to a new geometry that depended on a fundamental quadratic form that was independent of the surrounding space. In this geometry, straight lines are defined as geodesics and the sum of the interior angles of a triangle depends on the curvature.

In 1854, Riemann continued working on Gauss' ideas and introduced what we call today a differentiable manifold of arbitrary dimension n. Riemann associated to each point of the manifold a fundamental quadratic form and generalized the idea of curvature. Riemann was interested in the relationship between physics and geometry. This relationship motivated the development of non-Euclidean geometries.

<u>1.4</u> <u>Definition.</u> $f:\mathbb{R}\to\mathbb{R}$ is of <u>class</u> $\underline{C}^k$ if all derivatives up through order k exist and are continuous. $f:\mathbb{R}^n\to\mathbb{R}$ is of <u>class</u> $\underline{C}^k$ if all its (mixed) partial derivatives of order k and less exist and are continuous. A function $f:\mathbb{R}^n\to\mathbb{R}^p$ is of <u>class</u> $\underline{C}^k$ if all its components with respect to a given basis are of class C^k.

The concept of a differentiable manifold is necessary for extending the methods of differential calculus to spaces more general than $\mathbb{R}^n$. An example of a manifold is a regular surface in $\mathbb{R}^3$.

1.5 Definition. A C^k coordinate patch (regular surface) is a one-to-one C^k function $\mathbf{x}: U \to \mathbb{R}^3$ for some $k \geq 1$, where U is an open subset of $\mathbb{R}^2$ with coordinates u^1 and u^2 and $(\partial\mathbf{x}/\partial u^1) \times (\partial\mathbf{x}/\partial u^2) \neq \vec{0}$ on U.

The mapping $\mathbf{x}$ is called a parametrization of S at P. A regular surface is intuitively a union of open sets of $\mathbb{R}^2$, organized in such a way that when two such open sets overlap, the change from one to the other can be made in a differentiable manner. The problem with this definition is its dependence on $\mathbb{R}^3$. The definition of a differentiable manifold will be given for an arbitrary dimension n. Differentiable will always mean a class of C^∞.

1.6 Definition. A differentiable manifold of dimension n is a set M and a family of one-to-one mappings $\mathbf{x}_\alpha : U_\alpha \subset \mathbb{R}^n \to M$ of open sets U_α of $\mathbb{R}^n$ into M such that:

(1) $\cup_\alpha \mathbf{x}_\alpha(U_\alpha) = M$.

(2) for any pair α, β with $\mathbf{x}_\alpha(U_\alpha) \cap \mathbf{x}_\beta(U_\beta) = W \neq \emptyset$, the sets $\mathbf{x}_\beta^{-1}(W)$ and $\mathbf{x}_\alpha^{-1}(W)$ are open sets in $\mathbb{R}^n$ and the mappings $\mathbf{x}_\beta^{-1} \circ \mathbf{x}_\alpha$ are differentiable.

(3) The family $\{(U_\alpha, \mathbf{x}_\alpha)\}$ is maximal relative to the conditions (1) and (2), i.e., the family $\{(U_\alpha, \mathbf{x}_\alpha)\}$ contains all possible mappings with these properties.

The pair $(U_\alpha, \mathbf{x}_\alpha)$ with $P \in \mathbf{x}_\alpha(U_\alpha)$ is called a parametrization or system of coordinates of M at P. $\mathbf{x}_\alpha(U_\alpha)$ is then called a coordinate neighborhood at P. A family $\{(U_\alpha, \mathbf{x}_\alpha)\}$ satisfying (1) and (2) is called a differentiable structure on M.

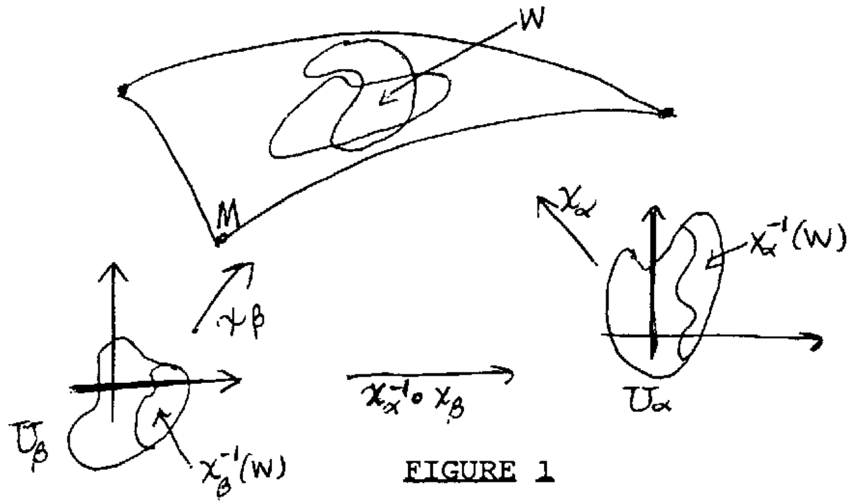


FIGURE 1

A differentiable structure on a set M induces a natural topology on M. Define $A \subset M$ to be an open set in M if and only if $\mathbf{x}_\alpha^{-1}(A \cap \mathbf{x}_\alpha(U_\alpha))$ is an open set in $\mathbb{R}^n$ for all α. The topology is defined such that the sets $\mathbf{x}_\alpha(U_\alpha)$ are open and that the mappings $\mathbf{x}_\alpha$ are continuous. The Euclidean space $\mathbb{R}^n$ is an n-manifold with the family of mappings generated by ($\mathbb{R}^n$, identity). Similarly, any open set in $\mathbb{R}^n$ is an n-manifold.

1.7 Example. Let G=GL(n,R) be the group of all non-singular nxn matrices. We show that G is an n^2-dimensional manifold. G is a metric space with distance function $d(A,B)=\sqrt{\sum(a_{ij}-b_{ij})^2}$ where $A=(a_{ij})$ and $B=(b_{ij})$. If $A=(a_{ij})\in G$, let

$$\phi(A)=(a_{11},a_{12},\ldots,a_{1n},a_{21},\ldots,a_{nn})\in\mathbb{R}^{n^2}.$$

Define a function $\nabla: \mathbf{R}^{n^2} \to \mathbf{R}$ by

$$\nabla(x_{11}, \ldots, x_{nn}) = \sum_{\sigma \in S_n} (-1)^{\sigma}\, x_{1,\sigma(1)}\, x_{2,\sigma(2)} \ldots x_{n,\sigma(n)},$$

where S_n is the group of permutations on n letters.
It can be shown that ∇ is continuous and $\nabla \circ \phi(A) = \det A$.
Therefore $\phi(G) = \nabla^{-1}(\mathbf{R} - \{0\})$ is open. Let M(n) be the set of all nxn matrices. We have identified M(n) and $\mathbf{R}^{n^2}$ by the mapping ϕ.

For surfaces in $\mathbf{R}^3$, a tangent vector at a point P of the surface is defined as the "velocity" in $\mathbf{R}^3$ of a curve in the surface passing through P. For differentiable manifolds, we do not have the support of the surrounding space. In elementary calculus, a vector $\vec{v}$ at a point $P \in \mathbf{R}^n$ may be viewed as a directional derivative. If $\vec{v} = (a^1, a^2, \ldots, a^n)$ and $f: \mathbf{R}^n \to \mathbf{R}$ is differentiable, then the non-normalized directional derivative of f at P in the direction $\vec{v}$ is

$$\vec{v}(f) = \Sigma\ a^i (\partial f/\partial u^i)(P).$$

This concept will be used to define a tangent vector as a real-valued operator on the set of differentiable functions on M which obeys the properties of a derivative.

Let D(M) denote the set of all local smooth (C^∞) functions at the point x in the smooth (C^∞) manifold M. By a local smooth function at the point x of M, we mean a smooth (C^∞) function $f: U \to \mathbf{R}$ defined on an open neighborhood U of x in M. In the set D(M) define scalar multiplication, addition, and multiplication as follows.

For arbitrary $a,b \in \mathbf{R}$ and any two local smooth functions $f:U\to\mathbf{R}$, $g:V\to\mathbf{R}$ in $D(M)$, we have the local smooth functions

$$af + bg: U \cap V \to \mathbf{R}$$

$$fg: U \cap V \to \mathbf{R}$$

defined by

$$(af + bg)(w) = a[f(w)] + b[g(w)],$$

$$(fg)(w) = f(w)g(w)$$

for every $w \in U \cap V$. These operations fail to make $D(M)$ an algebra over $\mathbf{R}$. $D(M)$ is not a linear space over $\mathbf{R}$ because $f + (-f) \neq g + (-g)$ unless $U = V$. To correct this problem, define a relation $\sim$ in the set $D(M)$ as follows. For any two local smooth functions $f:U\to\mathbf{R}$ and $g:V\to\mathbf{R}$ in $D(M)$, $f \sim g$ if and only if there exists an open neighborhood $W \subset U \cap V$ of x in M such that $f(w) = g(w)$ holds for every point w in W. Since this relation in $D(M)$ is reflexive, symmetric, and transitive, it is an equivalence relation. Therefore $\sim$ divides the members of $D(M)$ into disjoint equivalence classes called <u>the germs</u> of local smooth functions at the point x of M.

Let $G(M)=D(M)/\sim$ denote the set of all smooth germs at the point x of M and let $p:D(M) \to G(M)$ denote the natural projection of the set $D(M)$ onto its quotient set $G(M)$. Define scalar multiplication, addition, and multiplication in $G(M)$ for arbitrary smooth germs $w,\theta \in G(M)$, $a,b \in \mathbf{R}$, $f \in W$, and $g \in \theta$ as follows:

$$aw + b\theta = p(af + bg)$$

$$w\theta = p(fg)$$

These operations make G(M) the algebra of smooth germs at the point x in M of C^∞ functions.

1.8 Definition. A tangent vector to M at P is a function $X_P: D(M) \to \mathbf{R}$ whose value at f is denoted by $X_P(f)$, such that for all $f, g \in D(M)$ and $r \in \mathbf{R}$,

(a) $X_P(f + g) = X_P(f) + X_P(g)$;

(b) $X_P(rf) = rX_P(f)$; and

(c) $X_P(fg) = f(P)X_P(g) + g(P)X_P(f)$,

where fg is the ordinary product of functions f and g and $f(P)X_P(g)$ is the product of real numbers f(P) and $X_P(g)$. $X_P(f)$ may be read as the non-normalized directional derivative of f in the direction X_P at P.

Let $\alpha: (-\epsilon, \epsilon) \to M$ be a differentiable curve in M with $\alpha(0) = P$. Let X_P^α be defined by $X_P^\alpha(f) = (d(f\circ\alpha)/dt)(0)$. Let D be the set of functions on M that are differentiable at P. The tangent vector to the curve α at t=0 is a function $X_P^\alpha: D \to \mathbf{R}$ given by

$$X_P^\alpha(f) = d(f\circ\alpha)/dt|_{t=0}\ , \quad f \in D.$$

The tangent vector at P is the tangent vector at t=0 to the curve $\alpha: (-\epsilon, \epsilon) \to M$ with $\alpha(0) = P$. The set of all tangent vectors to M at P will be indicated by T_PM.

Let $\mathbf{x}: U \to M^n$ at $P = \mathbf{x}(0)$ be a parametrization where n indicates the dimension of M. Express the function $f: M^n \to \mathbf{R}$ and the curve α in this parametrization by

$$f\circ\mathbf{x}(q)=f(x_{1,\ldots,}x_n),\quad q=(x_1,\ldots,x_n)\in U,$$

$$\text{and } \mathbf{x}^{-1}\circ\alpha(t)=(x_1(t),\ldots,x_n(t)).$$

Restricting f to α we obtain

$$X_P^\alpha(f) = (d/dt)(f\circ\alpha)\big|_{t=0} = (d/dt)f(x_1(t),\ldots,x_n(t))\big|_{t=0}$$

$$= \Sigma\ (dx_i/dt)\big|_{t=0}(\partial f/\partial x_i) = (\Sigma\ x_i'(0)(\partial/\partial x_i)_0)f.$$

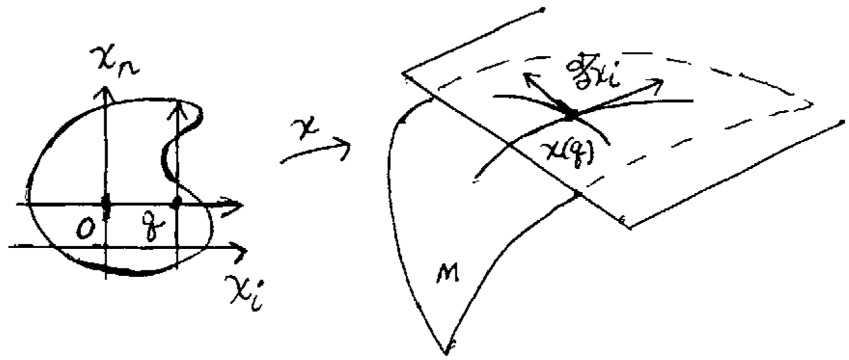


FIGURE 2

Therefore, the vector X_P^α can be expressed in the parametrization $\mathbf{x}$ by

$$X_P^\alpha = \Sigma\ x_i'(0)(\partial/\partial x_i)_0.$$

$(\partial/\partial x_i)_0$ is the tangent vector at P of the "coordinate curve": $x_i \rightarrow \mathbf{x}(0,\ldots,0,x_i,0,\ldots,0)$. Therefore the tangent vector to the curve α at P depends only on the derivative of α in a coordinate system.

1.9 Definition. The tangent space to M at P, T_pM, is the set of all tangent vectors to M at P. The set T_pM, with the usual operations of functions, forms a vector space of dimension n.

The choice of a parametrization $\mathbf{x}:U\rightarrow M$ determines an associated basis $\{(\partial/\partial x_1)_0, \ldots, (\partial/\partial x_n)_0\}$ in $T_P M$. The linear structure in $T_p m$ defined above does not depend on the parametrization $\mathbf{x}$.

1.10 Definition. Let M and N be differentiable manifolds. If $\phi : M \to N$ is differentiable, the differential of ϕ at P is the function

$$(\phi_*)_P : T_P M \to T_{\phi(P)} N$$

defined by

$$(\phi_*)_P (X_p)(f) = X_p(f \circ \phi)$$

where $X_p \in T_p M$, $f \in D(N)$.

1.11 Proposition. Let Φ: $M \to N$ and $P \in M$. Then $(\Phi_*)_p$: $T_p M \to T_{\Phi(p)} N$ is a linear transformation.

Proof: Let $r \in \mathbb{R}$, X_p, $Y_p \in T_p M$. We need to show that $(\Phi_*)_p (rX_p + Y_p) = r(\Phi_*)_p X_p + (\Phi_*)_p Y_p$. For any $f \in D(N)$:

$((\Phi_*)_P (rX_p + Y_p))(f) = (rX_p + Y_p)(f \circ \Phi)$

$= rX_p(f \circ \Phi) + Y_p(f \circ \Phi) = r(\Phi_*)_p (X_p) f + (\Phi_*)_p (Y_p) f$

$= (r(\Phi_*)_p X_p + (\Phi_*)_p Y_p)(f)$ QED

1.12 Definition. Let M^n and N^m be differentiable manifolds of dimensions n and m, respectively. M is an embedded submanifold of N is there is a differentiable function $\Phi : M \to N$ such that Φ is one-to-one and $(\Phi_*)_p$ is one-to-one for each $P \in M$.

If M is a submanifold of N, then $\dim M \le \dim N$. The tangent space to M can be viewed as a subspace of the tangent space to N.

1.13. Definition. A vector field X on a differentiable manifold M is a correspondence that associates to each point $P \in M$ a vector $X_p \in T_PM$.

1.14 Definition. A vector field X on a differentiable manifold M is differentiable in the following sense: if f is a differentiable function on M, then the mapping $P \to X_p f$ is differentiable.

Let X(M) represent the set of all vector fields on M. If $X, Y \in X(M)$, $r \in \mathbb{R}$, and $f \in D(M)$, then:

$(X + Y)_p = X_p + Y_p$, $(rX)_p = r \cdot X_p$, and

$(fX)_p = f(P)X_p$.

1.15 Definition. If $X, Y \in X(M)$, then the Lie Bracket of X and Y, $[X,Y]$, is the vector field defined by

$[X,Y]_p f = X_p(Yf) - Y_p(Xf)$ for $f \in D(M)$ and $P \in M$.

1.1 Lemma. $[X,Y]$ is vector field on M.

(See Millman and Parker,[9])

1.17 Proposition. If $X,Y,Z \in X(M)$ and $r \in \mathbb{R}$, then

(a) $[X,Y] = -[Y,X]$ and $[rX,Y] = r[X,Y]$

(anticommutativity)

$[X,Y] = XY - YX = -[YX - XY] = -[Y,X]$

$[rX,Y] = rXY - YrX = r[XY-YX] = r[X,Y]$.

(b) $[X + Y,Z] = [X,Z] + [Y,Z]$ and $[Z,X + Y] = [Z,X] + [Z,Y]$ (linearity).

(c) $[[X,Y],Z] + [[Y,Z],X] + [[Z,X],Y] = 0$

(Jacobi's Identity)

2. Riemannian Manifolds

Riemannian geometry is a generalization of metric differential geometry of surfaces. Instead of surfaces, one considers n-dimensional Riemannian manifolds. These are obtained from differentiable manifolds by introducing a Riemannian metric. The corresponding geometry is called Riemannian geometry. Surfaces are two-dimensional Riemannian manifolds. These concepts will be discussed in this section.

<u>2.1 Definition.</u> A <u>Riemannian metric</u> (or Riemannian structure) on a differentiable manifold M is a correspondence which associates to each point P of M an inner product $<,>_p$ (that is, a symmetric, bilinear, positive definite function) on the tangent space T_pM which varies differentiably in the following sense. If $\vec{x}:U\subset\mathbf{R}^n \to M$ is a system of coordinates around P, with $\vec{x}(x_1,x_2,\ldots x_n) = q \in \vec{x}(U)$, then

$$<\frac{\partial}{\partial x_i},\frac{\partial}{\partial x_j}>_q = g_{ij}(x_1,\ldots,x_n)$$

is a differentiable function on U.

This definition does not depend on the choice of coordinate system. The function $g_{ij}(x_1,\ldots,x_n)$ is called the local representation of the Riemannian metric in the coordinate system $\vec{x}:U\subset\mathbf{R}^n \to M$. A differentiable manifold with a given Riemannian metric is called a Riemannian manifold.

When are two Riemannian manifolds M and N the same?

<u>2.2 Definition.</u> Let M and N be Riemannian manifolds.

A diffeomorphism $f:M \to N$ (that is, f is a differentiable bijection with a differentiable inverse) is called an <u>isometry</u> if:

$$(*) \quad \langle \vec{u},\vec{v}\rangle_p = \langle (f_*)_p(\vec{u}), (f_*)_p(\vec{v})\rangle_{f(p)},$$

for all $P \in M$, $\vec{u},\vec{v} \in T_pM$.

<u>2.3 Definition.</u> Let M and N be Riemannian manifolds. A differentiable mapping $f:M \to N$ is a <u>local isometry</u> at $P \in M$ if there is a neighborhood $U \subset M$ of P such that $f:U \to f(U)$ is a diffeomorphism satisfying (*). Commonly, it is said that a Riemannian manifold M is <u>locally isometric</u> to a Riemannian manifold N if for every P in M there exists a neighborhood U of P in M and a local isometry $f:U \to f(U) \subset N$.

Some examples of the notion of Riemannian manifold are as follows.

<u>2.4 Example.</u> Let $M = \mathbb{R}^n$ with $\partial/\partial x_i$ identified with $\vec{e}_i=(0,\ldots,1,\ldots,0)$. The metric is given by $\langle\vec{e}_i,\vec{e}_j\rangle = \delta_{ij}$. $\mathbb{R}^n$ is called Euclidean space of dimension n and the Riemannian geometry of this space is metric Euclidean geometry.

<u>2.5 Example.</u> The product metric. Let M_1 and M_2 be Riemannian manifolds and consider the cartesian product M_1 X M_2 with the product structure.

Let $\pi_1:M_1XM_2 \to M_1$ and $\pi_2:M_1XM_2 \to M_2$ be the natural projections. (See Do Carmo,[3]) Introduce on M_1 X M_2 a Riemannian metric as follows:

$$<\vec{u},\vec{v}>_{(pq)} = <(\Pi_1)_*\cdot\vec{u}, (\Pi_1)_*\cdot\vec{v}>_p + <(\Pi_2)_*\cdot\vec{u}, (\Pi_2)_*\cdot\vec{v}>_q ,$$

for all $(p,q) \in M_1 \times M_2$, $\vec{u},\vec{v} \in T_{(pq)}(M_1 \times M_2)$.

The torus $S^1X\ldots XS^1 = T^n$ has a Riemannian structure obtained by choosing the induced Riemannian metric from $\mathbb{R}^2$ on the circle $S^1 \subset \mathbb{R}^2$ and then taking the product metric. The torus T^n with this metric is called the flat torus.

A differentiable mapping $\alpha: I \to M$ of an open interval $I \subset \mathbb{R}$ into a differentiable manifold M is called a (parametrized) curve. A parametrized curve can self-intersect as well as have "corners".

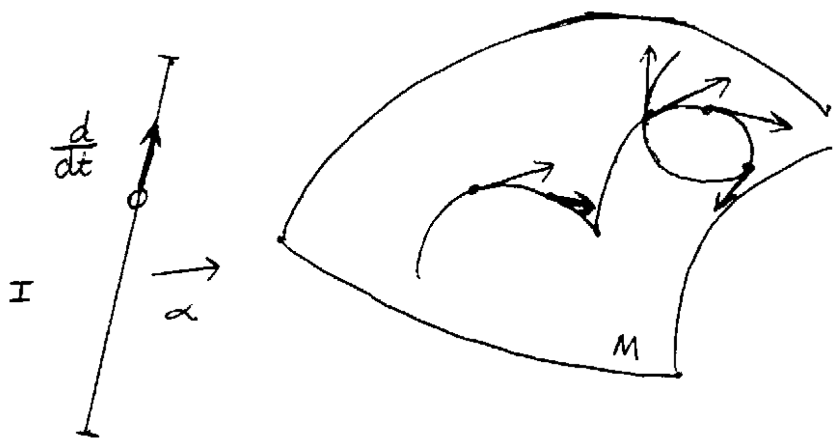


FIGURE 3

<u>2.6 Definition.</u> A <u>vector field V</u> along a curve $\alpha: I \to M$ is a differentiable mapping that associates to every $t \in I$ a tangent vector $V(t) \in T_{\alpha(t)}M$. The vector field $\alpha_*(d/dt)$, denoted by $d\alpha/dt$, is called the <u>velocity field</u> (or tangent vector field) of α.

I is a one-dimensional manifold with tangent vector d/dt as a basis. The restriction of a curve α to a closed interval $[a,b] \subset I$ is called a segment. If M is a

Riemannian manifold, the length of a segment is defined by

$$l_a^b = \int_a^b <\frac{d\alpha}{dt}, \frac{d\alpha}{dt}>^{1/2} dt.$$

Let X(M) be the set of all vector fields of class C^∞ on M and let G(M) be the algebra of germs of C^∞ (smooth) functions. In order to study the change in a vector field with respect to a direction, we introduce the notion of differentiation of vector fields.

<u>2.7 Definition.</u> A <u>linear connection ∇</u> on a differentiable manifold M is a mapping ∇: X(M) **X** X(M) $\to$ X(M) which is denoted by $<X,Y> \overset{\nabla}{\to} \nabla_X Y$ and which satisfies the following properties:

1) $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$ and $\nabla_X rY = r\nabla_X Y$;

2) $\nabla_{x+y} Z = \nabla_x Z + \nabla_y Z$ and $\nabla_{fx} Y = f\nabla_X Y$; and

3) $\nabla_X fY = (Xf)Y + f\nabla_X Y$; $X,Y,Z \in X(M)$ and $f,g \in D(M)$.

$\nabla_X Y$ should be read as the covariant derivative of Y in the direction of X.

Since many developments in the geometry of manifolds are local, we specify a connection by its local coordinates as follows.

<u>2.8 Definition.</u> Let ∇ be a connection on M and let $\bar{x}: U \subset \mathbf{R}^n \to M$ be a system of coordinates about P. The <u>Christoffel symbols</u> of ∇ with respect to the system of coordinates are the functions $\Gamma_{ij}^k \in D(\bar{x}(U))$ defined by

$$\nabla_{X_i} X_j = \nabla_{\partial/\partial x_i}(\partial/\partial x_j) = \sum_k \Gamma_{ij}^k (\partial/\partial x_k) = \sum_k \Gamma_{ij}^k X_k.$$

This shows that there are infinitely many connections on a manifold which can be obtained by prescribing the Christoffel symbols (subject to symmetry conditions).

2.9 Example. (Flat Euclidean space).

Let $M=\mathbb{R}^n$. If $Y \in X(\mathbb{R}^n)$, then $Y = \Sigma f^i \vec{e}_i$ for some $f^i \in D(\mathbb{R}^n)$. Define the flat conncection on $\mathbb{R}^n$ by $\nabla_X Y = \sum_i (Xf^i)\vec{e}_i$. By definition, this defines a linear connection on the manifold $\mathbb{R}^n$.

$\nabla_{X_i} X_j = 0$ for all i and j so that $\Gamma_{ij}{}^k = 0$ for all i,j,k. $\nabla_X Y$ is the usual directional derivative of a vector-valued function.

2.10 Example. Let M be flat Euclidean 2-space, $\alpha(t) = (\cos t, \sin t)$ for $0 < t < 2\pi$ and $Y = y\vec{e}_1 - x\vec{e}_2$. Since $T = -\sin t\vec{e}_1 + \cos t\vec{e}_2$, let $X = -y\vec{e}_1 + x\vec{e}_2$ so that $X_{\alpha(t)} = T_{\alpha(t)}$. If follows that $\nabla_T Y = \nabla_X Y = Xy\vec{e}_1 - Xx\vec{e}_2$ $= x\vec{e}_1 + y\vec{e}_2$ and $\nabla_T Y = \cos t\vec{e}_1 + \sin t\vec{e}_2$. Other connections will be discussed in section three of this chapter.

In the Euclidean plane, two lines are parallel if they have the same slope. Two curves $\alpha,\beta: I \to \mathbb{R}^3$ are parallel if their tangent vectors $\alpha'(s)$ and $\beta'(s)$ are parallel for each s in I, which implies that their tangent vectors have the same slope for each s in I.

2.11 Definition. Let M be a differentiable manifold with a linear connection ∇. Let $\alpha: I \to M$ be a differentiable curve in M. A vector field V along a curve $\alpha: I \to M$ is called parallel when $\nabla_T V = DV/dt = 0$, for all $t \in I$.

2.12 Proposition. Let M be a differentiable manifold with a linear connection ∇. Let $\alpha: I \to M$ be a differentiable curve in M and let V_o be a vector tangent to M at $\alpha(t_o)$, $t_o \in I$. Then there exists a unique parallel vector field $V(t)$ along α, such that $V(t_o) = V_o$. $V(t)$ is called the parallel transport of $V(t_o)$ along α.
(See Do Carmo,[3])

As a consequence of this proposition, if there exists a vector field V in $\mathfrak{X}(U)$ which is parallel along α with $V(t_o) = V_o$, then

$$0 = \frac{DV}{dt} = \sum_j (dv^j/dt)X_j + \sum_{i,j}(dx_i/dt)v^j \nabla_{X_i}X_j$$

where $V = \sum_j v^j X_j$ and $V_o = \sum_j v_o^j X_j$. Setting $\nabla_{X_i}X_j = \sum_k \Gamma_{ij}^k X_k$ and replacing j with k in the first sum, we obtain

$$\nabla_T V = \frac{DV}{dt}$$
$$= \sum_k \{(dv^k/dt) + \sum_{ij} v^j (dx_i/dt)\Gamma_{ij}^k\}X_k = 0.$$

The system of n differential equations in $v^k(t)$,

$$0 = dv^k/dt + \sum_{i,j} \Gamma_{ij}^k v^j (dx_i/dt), \ k=1,\dots,n,$$

possesses a unique solution satisfying the initial conditions $v^k(t_o) = v_o^k$.

2.13 Definition. A linear connection ∇ on a smooth manifold M is said to be symmetric when

$\nabla_X Y - \nabla_Y X = [X,Y]$ for all $X,Y \in \mathfrak{X}(M)$.

Local Riemannian geometry is concerned only with the differential geometric properties of a part of a differentiable manifold which can be covered by one system of coordinates. The fundamental theorem of local

Riemannian geometry states that with a given Riemannian metric there is uniquely associated a symmetric linear connection with the property that parallel transport preserves inner products. This unique linear connection is called the Riemannian connection of the Riemannian manifold.

2.14 Definition. Let V be a vector field along a curve α, and let T be its tangent vector. V is parallel along α if $\nabla_T V = DV/dt = 0$.

2.15 Definition. Let M be a differentiable manifold with a linear connection ∇ and a Riemannian metric $<,>$. A connection is said to be compatible with the metric $<,>$, when for any smooth curve α and any pair of parallel vector fields V and V' along α, then $<V,V'> =$ constant.

2.16 Proposition. Let M be a Riemannian manifold. A connection ∇ on M is compatible with a metric $<,>$ if and only if for any vector fields V and W along the differentiable curve $\alpha : I \to M$ we have

$$\frac{d}{dt}<V,W> = <\nabla_T V, W> + <V, \nabla_T W> = <\frac{DV}{dt}, W> + <V, \frac{DW}{dt}> , \quad t \in I.$$

This implies that if ∇ is compatible with a Riemannian metric $<,>$, then we are able to differentiate the inner product using the "product rule". (See Do Carmo,[3])

2.17 Corollary. A connection ∇ on a Riemannian manifold M is compatible with the metric if and only if

$$X<Y,Z> = <\nabla_X Y, Z> + <Y, \nabla_X Z>, \text{ for all } X,Y,Z \in \mathfrak{X}(M).$$

Proof. If $X\langle Y,Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$ for $X,Y,Z \in \mathcal{X}(M)$, then ∇ is compatible with the metric by Proposition 2.16. Suppose that ∇ is compatible with the metric. Let $P \in M$ and let $\alpha: I \to M$ be a differentiable curve with $\alpha(t_0) = P$, $t_0 \in I$. Let $\frac{d\alpha}{dt}\big|_{t=t_0} = X(P)$. Then

$$X(P)\langle Y,Z\rangle = \frac{d}{dt}\langle Y,Z\rangle\Big|_{t=t_0} = \langle \nabla_{X(p)} Y, Z\rangle_p + \langle Y, \nabla_{X(p)} Z\rangle_p.$$

P is arbitrary. Therefore,

$$X\langle Y,Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle,\quad X,Y,Z \in \mathcal{X}(M). \text{ QED}$$

2.18 Theorem. (The Fundamental Theorem of Local Riemannian geometry.) Given a Riemannian manifold M, there exists a unique linear connection ∇ on M satisfying the conditions:

a) ∇ is symmetric.

b) ∇ is compatible with the Riemannian metric.

This connection is called the Riemannian connection on M. (See Do Carmo[3])

Let $\mathbf{x}: U \subset \mathbf{R}^n \to M$ be a system of coordinates and let $\nabla_{X_i} X_j = \sum_k \Gamma_{ij}^k X_k$ where Γ_{ij}^k are called the coefficients of the unique linear connection on U or the Christoffel symbols of the connection. As a consequence of Theorem 2.18, it follows that

$$\sum_l \Gamma_{ij}^l g_{lk} = 1/2\{(\partial/\partial x_i) g_{jk} + (\partial/\partial x_j) g_{ki} - (\partial/\partial x_k) g_{ij}\}$$

where $g_{ij} = \langle X_i, X_j\rangle$. Since the g_{ij} are the coefficients of a positive definite quadratic form, the matrix (g_{km}) has an inverse (g^{km}).

Therefore,

$$\Gamma_{ij}^{m} = 1/2 \sum_k \{(\partial/\partial x_i) g_{jk} + (\partial/\partial x_j) g_{ki} - (\partial/\partial x_k) g_{ij}\} g^{km}$$

yields the Christoffel symbols of the unique linear connection. This equation is a classical expression for the Christoffel symbols of the Riemannian connection in terms of the g_{ij} given by the metric. In terms of the Christoffel symbols, $\nabla_T V = \frac{DV}{dt}$ has the classical expression

$$\frac{DV}{dt} = \sum_k \{(dv^k/dt) + \sum_{i,j} \Gamma_{ij}^{k} v^j (dx_i/dt)\} X_k.$$

3. Geodesics

"Straight line" and "point" are two of the undefined terms in plane geometry upon which the axioms of plane geometry are built. Straight lines play an important role in the construction and formation of many figures that are studied. What types of curves play the role of "lines" on a Riemannian manifold. These "lines" should be curves whose tangent vectors are all parallel. These "lines" should also be curves of shortest length joining two points on a Riemannian manifold. A geometric surface is considered to be a two-dimensional Riemannian manifold. These "lines" on a surface are referred to as geodesics.

3.1 Definition. A parametrized curve $\alpha: I \to M$ is a geodesic at $t_o \in I$ if $\nabla_{T_\alpha} T_\alpha = \frac{D}{dt}\left(\frac{d\alpha}{dt}\right) = 0$ at the point t_o; if α is a geodesic for all $t \in I$, we say that α is a geodesic. If $[a,b] \subset I$ and $\alpha: I \to M$ is a geodesic, the restriction of α to $[a,b]$ is called a geodesic segment joining α to $\alpha(a)$ to $\alpha(b)$.

If $\alpha: I \to M$ is a geodesic, then

$$\frac{d}{dt}\left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt}\right\rangle = 2\left\langle \frac{D}{dt}\frac{d\alpha}{dt}, \frac{d\alpha}{dt}\right\rangle = 0.$$

This implies that the length of the tangent vector $\frac{d\alpha}{dt}$ is constant and the tangent vectors are all parallel.

Let $\vec{x}: U \subset \mathbb{R}^n \to M$ be a system of coordinates about $\alpha(t_o)$. We need to determine the local equations satisfied by a geodesic α. By Proposition 2.12, in U a curve $\alpha(t) = (x_1(t), \ldots, x_n(t))$ will be a geodesic if and only if

$$0 = \frac{D}{dt}\left(\frac{d\alpha}{dt}\right) = \sum_k \left(\frac{d^2x_k}{dt^2} + \sum_{ij} \Gamma^k_{ij} \frac{dx_i}{dt}\frac{dx_j}{dt} \right) \frac{\partial}{\partial x_k}.$$

Therefore,

$$\frac{d^2x_k}{dt^2} + \sum_{i,j} \Gamma^k_{ij} \frac{dx_i}{dt}\frac{dx_j}{dt} = 0, \quad k=1,\ldots,n \quad \text{(Equation 3A)}$$

and this second order system yields the local equations satisfied by a geodesic α. By the usual existence and uniqueness theorem for Ordinary Differential Equations, we see that for every point P and every tangent vector V_o at P there exists (locally) a unique geodesic through P with tangent vector V_o.

3.2 Example Let $\mathbf{x}:U \subset \mathbb{R}^2 \rightarrow \mathbb{R}^3$ and let $\mathbf{x}(r,s) = (r,s,0)$ represent the two-dimensional plane in $\mathbb{R}^3$. This system of coordinates represents a two-dimensional manifold in $\mathbb{R}^3$ known as a simple surface. The standard classical notation for a Riemannian metric is given by

$ds^2 = Edx_1^2+2Fdx_1dx_2+Gdx_2^2$ where

$E = g_{11} = \langle X_1,X_1\rangle$, $F = g_{12} = \langle X_1,X_2\rangle$,

and $G = g_{22} = \langle X_2,X_2\rangle$, where $X_i = \dfrac{\partial}{\partial x_i}$.

For the two-dimensional plane, $X_1 = (1,0,0)$ and $X_2 = (0,1,0)$ with E=1, F=0, and G=1. Therefore, the Riemannian metric of the two-dimensional plane will be the Euclidean metric given by

$$ds^2 = dx_1^2 + dx_2^2.$$

With $(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $(g^{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, the Riemannian connection is $\Gamma_{ij}{}^k = 0$ for all i,j,k. The differential equations (Equation 3A) reduce to $\frac{d^2x_k}{dt^2}=0, k=1,2$. It follows that

$$\frac{d^2r}{dt^2} = 0, \quad \frac{d^2s}{dt^2} = 0, \text{ with } \frac{dr}{dt}=\vec{a}_1, \quad \frac{ds}{dt}=\vec{a}_2.$$

Therefore, $r = \vec{a}_1 t+\vec{a}_2$ and $s = \vec{a}_3 t+\vec{a}_4$ and the geodesics of the Euclidean plane are straight lines.

<u>3.3 Example.</u> Let $\vec{x}:U\subset\mathbf{R}^2\to\mathbf{R}^3$ and

$$\vec{x}(u,v)=(\cos u \cos v, \cos u \sin v, \sin u),$$

$$(u,v) \in \left(\frac{-\pi}{2},\frac{\pi}{2}\right) \times \mathbf{R}.$$

The image of $\vec{x}$ is the unit sphere S^2 minus the north pole and south pole: $S^2 - \{0,0,\pm1\}$.

$$X_1 = (-\sin u \cos v, -\sin u \sin v, \cos u)$$

$$X_2 = (-\cos u \sin v, \cos u \cos v, 0)$$

With the usual Euclidean metric,

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & \cos^2 u \end{pmatrix} \quad \text{and} \quad (g^{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\cos^2 u} \end{pmatrix}.$$

The unique Riemannian connection is $\Gamma_{12}^{\ 2} = \Gamma_{21}^{\ 2} = \frac{-\sin u}{\cos u}$, $\Gamma_{22}^{\ 1} = \cos u \sin u$, all other $\Gamma_{ij}^{\ k} = 0$. The differential equations (Equation 3A) reduce to

$$(1) \quad \frac{d^2u}{dt^2}+\cos u \sin u\frac{dv}{dt}\frac{dv}{dt}=0$$

$$(2) \quad \frac{d^2v}{dt^2} - 2\frac{\sin u}{\cos u}\frac{du}{dt}\frac{dv}{dt}=0$$

A meridian of the sphere is given by $v(t)$ = constant. It follows that $\frac{dv}{dt}=0$ and $\frac{d^2v}{dt^2}=0$ and equation (2) is satisfied. Along a meridian $u(t) = t$ so that $\frac{du}{dt}=1$ and $\frac{d^2u}{dt^2}=0$. Therefore equation (1) is satisfied. It follows that every meridian of the sphere is a geodesic. The meridian of a sphere is a great circle. The sphere is symmetrical and there is nothing geometrically special about this great circle. Therefore every great circle of the Riemannian manifold S^2 is a geodesic.

3.4 Example. Consider a curve in the (r,z) plane given by $r=r(t) > 0$, $z=z(t)$. If this curve is rotated about the z-axis, we obtain a surface of revolution. Let M be a surface of revolution generated by the unit speed curve $(r(t),z(t))$. M may be parametrized by

$$\vec{x}(t,\theta) = (r(t)\cos\theta,\ r(t)\sin\theta,\ z(t)).$$

t measures position on the curve and θ measures how far the curve has been rotated. The t-curves are called meridians and the θ-curves are called circles of latitude. The z-axis is called the axis of revolution.

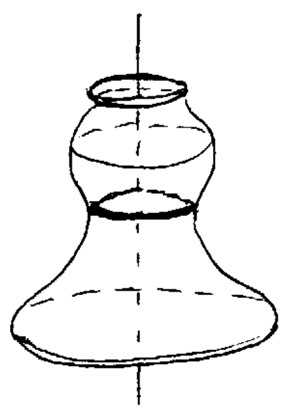


FIGURE 4

With the usual Euclidean metric,

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & r^2(t) \end{pmatrix} \text{ and } (g^{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{r^2(t)} \end{pmatrix}.$$

The Riemannian connection is $\Gamma_{12}^{\ 2} = \Gamma_{21}^{\ 2} = \frac{r'(t)}{r(t)}$, $\Gamma_{22}^{\ 1} = -r(t)r'(t)$, and all other $\Gamma_{ij}^{\ k} = 0$. The differential equations (Equation 3A) reduce to

$$(1) \quad \frac{d^2t}{ds^2} - r(t)\,r'(t)\,\frac{d\theta}{ds}\frac{d\theta}{ds} = 0$$

$$(2) \quad \frac{d^2\theta}{ds^2} + 2\left(\frac{r'(t)}{r(t)}\right)\frac{dt}{ds}\frac{d\theta}{ds} = 0$$

(Note: Comparison of this example with that of the unit sphere S^2 shows clearly that the sphere is a surface of revolution.) A meridian is given by $\theta(s)$ = constant. Then $\frac{d\theta}{ds}$ and $\frac{d^2\theta}{ds^2}$ are zero and equation (2) is satisfied. Along a meridian $t(s) = s$, so that $\frac{dt}{ds}=1$ and $\frac{d^2t}{ds^2}=0$ and equation (1) is satisfied. Therefore every meridian of the surface of revolution is a geodesic.

A circle of latitude is given by $t(s)$ = constant. Then $\frac{dt}{ds}$ and $\frac{d^2t}{ds^2}$ are both zero.
Since $\vec{\gamma}(s)=\vec{x}(t(s),\theta(s))$ has unit speed,

$$1=|\vec{\gamma}'(s)|^2=\left|\frac{\partial\vec{x}}{\partial t}\frac{dt}{ds}+\frac{\partial\vec{x}}{\partial\theta}\frac{d\theta}{ds}\right|^2=g_{22}\left(\frac{d\theta}{ds}\right)^2.$$

This implies that $1 = r^2\left(\frac{d\theta}{ds}\right)^2$ and $0 \neq \frac{d\theta}{ds}=\pm\frac{1}{r}$. It follows that r is constant if t is. Therefore $\frac{d^2\theta}{ds^2}=0$ and a circle of latitude satisfies equation (2). Since $\frac{d\theta}{ds}\neq 0$ and $r>0$, a circle of latitude satisfies equation (1) if and only if $r'(t)=0$. This happens if and only if

$$\frac{\partial \vec{x}}{\partial t} = (r'(t)\cos\theta,\ r'(t)\sin\theta,\ z'(t))$$

is parallel to the axis of rotation (0,0,1). This implies that a circle of latitude is a geodesic if and only if the tangent $\frac{\partial \vec{x}}{\partial t}$ to the meridians is parallel to the axis of revolution at all points on the circle of latitude.

In a later chapter we will study examples in which the metric is not induced by the Euclidean metric.

3.5 Definition. The curvature of a unit speed curve $\alpha(s)$ is given by $K(s) = |T'(s)| = |\alpha''(s)|$.

3.6 Definition. The principal normal vector field of $\alpha(s)$ is a unit vector-field $\vec{N}(s)=\frac{\alpha''(s)}{K(s)}$ which tells the direction in which $\alpha(s)$ is turning at each point.

3.7 Definition. The unit normal to the surface at a point $P=\vec{x}(u,v)=\alpha(s)$ is $\vec{n}=\frac{X_1 x X_2}{|X_1 x X_2|}$ with $X_i=\frac{\partial}{\partial x_i}$.

If $\vec{x}: U\subset \mathbf{R}^2 \to \mathbf{R}^3$ is a two-dimensional manifold (called a simple surface) and $\vec{\gamma}(s)$ is a unit speed curve in the image of $\vec{x}$, then $\vec{S}=\vec{n}X\vec{T}=\vec{n}X\vec{\gamma}'$ is called the intrinsic normal of $\vec{\gamma}$ where $\vec{n}=\frac{X_1 x X_2}{|X_1 x X_2|}$. $\vec{S}$ is well defined on a surface M up to sign.

If $P \in M$, let $N_pM = \{r\vec{n} \mid r \in R\}$. N_pM is the set of all vectors perpendicular to M at P and is called the normal space of M at P. The tangent space of a surface M at $P \in M$ is the set T_pM of all vectors tangent to M at P. $\mathbf{R}^3 = T_pM \oplus N_pM$ and any vector in $\mathbf{R}^3$ can be decomposed uniquely as a sum of a vector tangent to M at P and a vector normal to M at P. If this is done for $\gamma''(s)$, then $\gamma''(s)=\vec{X}(s)+\vec{V}(s)$ where $\vec{X}(s)$ is tangent to M and $\vec{V}(s)$ is normal to M.

Since $\vec{T}(s)=\gamma'(s)$ is tangent to M, $<\vec{V},\vec{T}>=0$. $<\gamma'',\vec{T}>=0$ and therefore $<\vec{X}(s),\vec{T}>=0$. But $<\vec{X}(s),\vec{n}>=0$ and therefore $\vec{X}(s)$ is perpendicular to both $\vec{n}$ and $\vec{T}$ and is thus a multiple of $\vec{S}=\vec{n}X\vec{T}$. Define two functions $k_n(s)$ and $k_g(s)$ by

$$k_n(s)=<\gamma''(s),\vec{n}(\gamma^1(s),\gamma^2(s))> \text{ and}$$

$$k_g(s)=<\gamma''(s),\vec{S}(s)>.$$

Therefore,

$$K(s)\vec{N}(s)=\vec{T}'(s)=\gamma''(s)=k_n(s)\vec{n}(s)+k_g(s)\vec{S}(s).$$

$k_n(s)$ is the normal curvature of a unit speed curve γ and measures how the surface M is curving in $\mathbf{R}^3$. $k_g(s)$ is the geodesic curvature of a unit speed curve γ and measures how γ is curving in M.

<u>3.8 Proposition.</u> (Gauss's formulas) Let $\vec{x}:U\to\mathbf{R}^3$ be a simple surface. Then for any unit speed curve,

$$\alpha(s)=\vec{x}(x_1(s),x_2(s))$$

$$k_n=\sum_{i,j}L_{ij}\frac{dx_i}{dt}\frac{dx_j}{dt} \quad \text{and}$$

$$k_gS=\sum_k\left(\frac{d^2x_k}{dt^2}+\sum_{i,j}\Gamma^k_{ij}\frac{dx_i}{dt}\frac{dx_j}{dt}\right)X_k.$$

(See Millman and Parker,[9])

<u>3.9 Definition.</u> A curve α on a manifold M is a <u>geodesic</u>

(with respect to ∇) if $\nabla_T T_\alpha = 0$.

For a geodesic α,

$$\nabla_{T_\alpha} T_\alpha = \frac{D}{dt}\left(\frac{d\alpha}{dt}\right) = 0$$

$$= \sum_k \left(\frac{d^2 x_k}{dt^2} + \sum_{ij} \Gamma^k_{ij} \frac{dx_i}{dt}\frac{dx_j}{dt}\right) X_k.$$

Therefore if α is a geodesic then $k_g S = 0$ and $k_g = 0$. This implies that a geodesic α on M has geodesic curvature equal to 0 everywhere. Since geodesic curvature measures how α is curving in a surface M, $k_g = 0$ everywhere means that α is not turning, i.e., is the "straight line" of the surface.

<u>3.10 Proposition.</u> A unit speed curve $\alpha(s)$ on a two-dimensional manifold M is a geodesic of M if and only if $\alpha''(s)$ is everywhere normal to the surface (i.e. is a multiple of the normal to M).

Proof: $\alpha''=K\vec{N}=k_g\vec{S}+k_n\vec{n}$. If $k_g=0$, then $\alpha''(s)=K\vec{N}=k_n\vec{n}$ and α'' is everywhere normal to the surface. If α'' is everywhere normal to the surface, then

$\alpha''=K\vec{N}=k_n\vec{n}=0+k_n\vec{n}$ and $k_g = 0$. QED

This proposition implies that along all curved geodesics, the principal normal to the curve coincides with the surface normal. Since α'' is normal to the surface, the inhabitants of M perceive no acceleration at all. For them the geodesic is a "straight line".

<u>3.11 Example.</u> Let $\vec{x}: U\subset\mathbf{R}^2\to\mathbf{R}^3$ and

$\vec{x}(t,v) = (r \cos t, r \sin t, v)$.

The image of $\vec{x}$ is a surface of revolution known as a cylinder of revolution or right circular cylinder whose radius is r and whose axis of revolution is the z-axis.

Let $\alpha:\mathbb{R}\to\mathbb{R}^3$ be given by $\alpha(t)=(r \cos t, r \sin t, bt)$, r,b constants $\neq 0$. This is called the right helix on the cylinder of radius r of pitch $2\pi b$.

The principal normal to the curve is

$$\vec{N}(t)=\frac{\alpha''(t)}{K(t)}=\frac{(-r\cos t,-r\sin t,0)}{r}=(-\cos t,-\sin t,0).$$

A normal to the surface is

$$\vec{n}=\frac{X_1 x X_2}{|X_1 x X_2|}=\frac{(rv'\cos t,rv'\sin t,0)}{rv'}=(\cos t,\sin t,0)$$

where $X_1=(-r\sin t,r\cos t,0)$ and $X_2=(0,0,v')$.

Therefore, $\vec{n}=\pm\vec{N}$ and circular helices on the cylinder are geodesics of the circular cylinder. Since the cylinder is a surface of revolution, other geodesics of the cylinder would be the generators of the cylinder (meridians) and the circles of latitude for which the vector tangent to the meridians is parallel to the axis of revolution at all points on the circle of latitude.

FIGURE 5

3.12 Example. Let $\vec{x}: U \subset \mathbb{R}^2 \to \mathbb{R}^3$ be given by

$\vec{x}(u,v) = ((a+b\cos u)\cos v,\ (a+b\cos u)\sin v,\ b\sin u)$

$0<b<a$, $(u,v) \in R \times R$.

The image of $\vec{x}$ is a surface of revolution known as a torus. $(a+b\cos u)$ represents the distance from the z-axis. $(b\sin u)$ represents the distance along the z-axis. Since the torus is a surface of revolution, the geodesics of the torus include the meridians and the circles of latitude where the tangents to the meridians are parallel to the axis of revolution at all points on the circles of latitude. These circles of latitude would be the outer equator and the inner equator of the torus.

3.13 Definition. A geodesic segment $\vec{\gamma}$ from P to Q locally minimizes arc length from P to Q provided there exists an $\epsilon>0$ such that for any $\vec{\alpha}$ which is sufficiently close (ϵ-close) to $\vec{\gamma}$ then the length of $\vec{\alpha}$ is greater than or equal to the length of $\vec{\gamma}: L(\vec{\alpha}) \geq L(\vec{\gamma})$.

3.14 Theorem. Let $\vec{\gamma}$ be a unit speed curve in a surface M between points $P=\vec{\gamma}(a)$ and $Q = \vec{\gamma}(b)$. If $\vec{\gamma}$ is the shortest curve between P and Q, then $\vec{\gamma}$ is a geodesic.

The proof will proceed along the following lines. Start with a length minimizing curve $\vec{\gamma}$ and assume that the geodesic curvature is not zero. Then "wiggle" the curve to form a family of curves with the same endpoints as $\vec{\gamma}$

with $\alpha_0=\gamma$. Let L(t) be the function that gives the length of α_t and it must have a minimum at t=0 ($\alpha_0=\gamma$). Therefore L'(0)=0. Using this fact and integrating by parts leads to a contradiction.

Proof: Let $a<x_0<b$ and let k_g be the geodesic curvature of γ. To prove that γ is a geodesic, show that $k_g(s_0)=0$.

Suppose that $k_g(s_0) \neq 0$. There exist numbers c and d with $a<c<s_0<d<b$, $k_g \neq 0$ on [c,d]. The image of [c,d] under γ is contained in a coordinate patch $\vec{x}$. The segment of γ from $\gamma(c)$ to $\gamma(d)$ must be the shortest curve joining $\gamma(c)$ to $\gamma(d)$ or there must be a piece-wise regular curve from $\gamma(a)$ to $\gamma(c)$ to $\gamma(d)$ to $\gamma(b)$ that is shorter than γ.

But γ is the shortest curve from $\gamma(a)$ to $\gamma(b)$.

Let $\lambda(s)$ be a C^2 function defined for $c\le s\le d$ such that $\lambda(c)=\lambda(d)=0$, $\lambda(s_0)\neq 0$, and $\lambda(s)k_g(s)\ge 0$ for $c\le s\le d$.

$\vec{S}=\vec{n}X\vec{T}=\vec{n}X\gamma'$ and in the coordinate patch $\vec{x}$ we have $\lambda(s)\vec{S}=\sum v^i(s)\vec{X}_i$ for some $v^i:[c,d]\to\mathbb{R}$. $\lambda(s)$ moves in and out with endpoints fixed.

Let $\gamma(s)$ be given by $\gamma(s)=\vec{x}(\gamma^1(s),\gamma^2(s))$.

Define a family of curves by

$$\alpha_t(s)=\vec{x}(\gamma^1(s)+tv^1(s),\gamma^2(s)+tv^2(s))$$

with $|t|$ small enough (ϵ-close). α_t is a curve from $\gamma(c)$ to $\gamma(d)$ for each choice of t with $\alpha_0=\gamma$ or $\alpha_t(s)=\alpha(s;t)$. The length of $\alpha(s;t)$ is

$$L(t)=\int_c^d<\frac{\partial\alpha}{\partial s},\frac{\partial\alpha}{\partial s}>^{\frac{1}{2}}ds \;.$$

L(t) has a minimum for t=0. $\alpha(s;0)=\gamma(s)$ yields the shortest path.

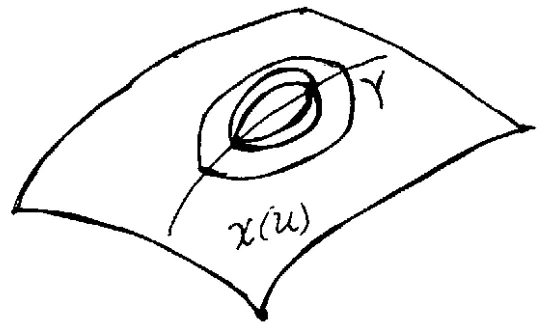


FIGURE 6

$$L'(t) = \frac{d}{dt}\int_c^d<\frac{\partial\alpha}{\partial s},\frac{\partial\alpha}{\partial s}>^{\frac{1}{2}}ds$$

$$= \int_c^d\frac{\partial}{\partial t}<\frac{\partial\alpha}{\partial s},\frac{\partial\alpha}{\partial s}>^{\frac{1}{2}}ds$$

$$= \int_c^d\frac{1}{2}\frac{2<\frac{\partial^2\alpha}{\partial t\partial s},\frac{\partial\alpha}{\partial s}>}{<\frac{\partial\alpha}{\partial s},\frac{\partial\alpha}{\partial s}>^{\frac{1}{2}}}\,ds \qquad \text{(Chain Rule)}$$

$$= \int_c^d\frac{<\frac{\partial^2\alpha}{\partial s\partial t},\frac{\partial\alpha}{\partial s}>}{<\frac{\partial\alpha}{\partial s},\frac{\partial\alpha}{\partial s}>^{\frac{1}{2}}}\,ds.$$

At t=0 ($\alpha_o=\gamma$)

$$<\frac{\partial\alpha}{\partial s},\frac{\partial\alpha}{\partial s}>=<\frac{d\gamma}{ds},\frac{d\gamma}{ds}>=1 \quad (\gamma \text{ is unit speed.})$$

Therefore,

$$L'(0) = \int_c^d<\frac{\partial^2\alpha}{\partial s\partial t},\frac{\partial\alpha}{\partial s}>|_{t=0}ds$$

$$= \int_c^d \left[\frac{d}{ds}<\frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial s}>|_{t=0} - <\frac{\partial \alpha}{\partial t}, \frac{\partial^2 \alpha}{\partial s^2}>|_{t=0}\right] ds$$

$$= <\frac{\partial \alpha}{\partial t}, \frac{\partial \alpha}{\partial s}>|_{t=0}\Big|_c^d - \int_c^d <\frac{\partial \alpha}{\partial t}, \frac{\partial^2 \alpha}{\partial s^2}>|_{t=0}\, ds$$

$\frac{\partial \alpha}{\partial t}|_{t=0} = \sum v^i(s)\vec{X}_i = \lambda(s)\vec{S}$. λ was constructed so that $\lambda(c) = \lambda(d) = 0$.

Therefore,

$$0 = L'(0) = 0 - \int_c^d <\lambda(s)\vec{S}, k_g(s)\vec{S} + k_n(s)\vec{n}> ds$$

$$= - \int_c^d \lambda(s) k_g(s)\, ds < 0 . \qquad (\gamma'' = k_g\vec{S} + k_n\vec{n})$$

This contradiction implies that the geodesic curvature is everywhere 0 and γ is a geodesic. QED

The converse of the previous theorem is false: If γ is a geodesic, then γ is the shortest curve between P and Q. A geodesic need not minimize distance. Let P and Q be two points on the unit sphere S^2 with $P \neq \pm Q$. There are two geodesics of different lengths joining P to Q. They correspond to the two arcs of the great circle through P and Q. The longer geodesic does not minimize distance. However, geodesics locally minimize length.

4. Curvature.

An (n - 1) submanifold of an n-manifold is called a hypersurface. Let M be a hypersurface of $\mathbb{R}^n$, and let ∇ be the natural connection on $\mathbb{R}^n$, and assume N is a unit normal vector field that is C^∞ on M. Thus $\langle N_pM, N_pM\rangle = 1$ and $\langle N_pM, X\rangle = 0$ for all P in M and X in T_PM. Such an N always exists locally.

For any P in M and any vector X in T_PM, define the linear map $L: T_PM \to T_PM$ by $L(X) = \nabla_X N$. The vector L(X) lies in T_PM. L is called the Weingarten map and in the case of $\mathbb{R}^n$, it has the geometric interpretation of the Jacobian of the sphere map (Gauss map).

Let $N=(a_1,\ldots,a_n)$, so the a_i are real-valued C^∞ functions on M and $\sum(a_i)^2 = 1$. Then the mapping of $g:M \to S^{n-1}$ in $\mathbb{R}^n$ is a C^∞ map of M into the unit (n-1)-sphere S^{n-1}. g is called the sphere map (or Gauss map). If X is in T_PM and $\sigma(t)$ is a curve fitting X with $\sigma(0)=P$ and $T_\sigma(0)=X$, then

$$g^*(X)\ \ T_{g\circ\sigma}(0)\ \ (Xa_1,\ldots,Xa_n)\ \ \nabla_X N\ L(X).$$

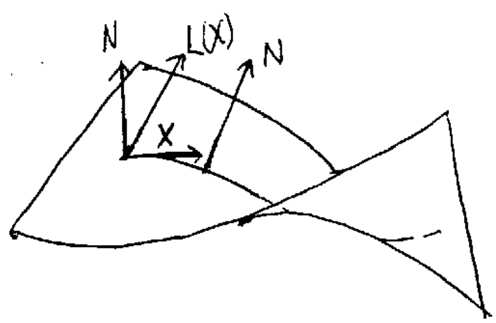


FIGURE 7

The map L is C^∞ on M in the sense that if X is C^∞ on the subset A of M, then $L(X) = (Xa_1, \ldots, Xa_n)$ is also C^∞ on A since each a_i is C^∞ on M. The Weingarten Map is the derivative of the normal and therefore gives the change in the normal.

In order to study how a two-dimensional manifold M (simple surface) is curving at a point P, without reference to a direction, find the eigenvalues of the matrix $(L_k^{\,l}) = L$ (the Weingarten map). These eigenvalues at a point P will tell us how M curves at that point.

4.1 Definition. Let $\vec{n}$ be a unit normal vector on U.

The coefficients of the second fundamental form of a simple surface $\vec{X}: U \subset \mathbf{R}^2 \to \mathbf{R}^3$ are the functions L_{ij} defined on U by $L_{ij} = \langle X_{ij}, \vec{n} \rangle$.

Since $X_{ij} = X_{ji}$, then $L_{ij} = L_{ji}$. The L_{ij} are called the coefficients of the second fundamental form because the assignment $II(X,Y) = \sum_{i,j} L_{ij}\, x^i y^j$ is a symmetric bilinear form on T_PM, as is the first fundamental form $I(X,Y) = \langle X,Y \rangle = \sum x^i y^j \langle X_i, X_j \rangle = \sum x^i y^j g_{ij}$.

4.2 Theorem. Let M be a surface. Then L is a linear transformation from T_PM to T_PM.

(See Millman and Parker,[9])

4.3 Theorem. Let M be a surface. If $L(X_k) = \sum L_k^{\,l} X_l$; $L_k^{\,l} = \sum L_{ik}\, g^{il}$ where $(L_k^{\,l})$ is the matrix representing L with respect to the basis $\{X_1, X_2\}$.

Proof: Since X_i is tangent to M, $\langle \vec{n}, X_i \rangle = 0$. This implies that

$$0 = \frac{\partial<\vec{n},X_i>}{\partial x_k} = <\frac{\partial\vec{n}}{\partial x_k},X_i> <\vec{n},X_{ik}>$$

$$= -<L(X_k),X_i> + L_{ik} = L_{ik} - <\sum L_k^j X_j,X_i>$$

$$= L_{ik} - \sum L_k^j <X_j,X_i> = L_{ik} - \sum L_k^j g_{ji}.$$

Therefore $L_{ik} = \sum L_k^j g_{ji}$ and

$\sum L_{ik} g^{il} = \sum L_k^j g_{ji} g^{il} = \sum L_k^j \delta_j^l = L_k^l$. QED

The normal curvature k_n of α at P depends only on the unit tangent of α at P. If we know all the possible values that k_n takes on at P, we would know how M curves at that point. One way to find this information would be to find the maximum and minimum values that k_n obtains called k_1 and k_2, respectively. The following results are from elementary linear algebra. (See Ortega,[11]) To find these values, we determine the maximum and minimum of II(X,X) as X runs over all unit vectors in T_PM. This means we are maximizing and minimizing II(X,X) subject to the constraints <X,X> = 1.

Find the critical values of

$$f(X,\lambda) = II(X,X) - \lambda(<X,X> - 1)$$

$$= <L(X),X> - \lambda<X,X> + \lambda = <L(X)-\lambda X,X> + \lambda \text{ at P.}$$

The problem has a solution since II(X,X) does have a maximum and minimum: the set of unit vectors in T_PM is closed and bounded, i.e., compact. The eigenvalues are the roots of

$$0 = \det(L - \lambda I) = \lambda^2 - (\text{trace } L)\lambda + \det L.$$

Denotes these roots by k_1 and k_2, with $k_1 \geq k_2$.

4.4 Proposition. At each point of a surface M there are two orthogonal directions such that the normal curvature takes its maximum value in one direction and its minimum value along the other. (See Ortega, [11])

4.5 Definition. The principal curvatures of a surface M at a point P are the eigenvalues of L (k_1 and k_2) at the point P. Corresponding unit eigenvectors are called principal directions at P.

4.6 Definition. The Gaussian curvature of M at P is $K = k_1 k_2 = \det L$. The mean curvature of M at P is $H = 1/2(k_1 + k_2) = 1/2\ \mathrm{trace}(L)$.

In the previous examples of two-dimensional manifolds, the geodesics of these surfaces were discussed. In each case, the Weingarten map L and the curvature of these surfaces can be computed.

4.7 Example. For the Euclidean plane $\vec{x}(r,s) = (r,s,0)$, the metric coefficients were found to be

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ with inverse } (g^{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Therefore $L_{ij} = \langle X_{ij}, \vec{n} \rangle = \langle \vec{0}, \vec{n} \rangle\ 0$ for all i and j and

$$L = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

The det L = 0 and 1/2 trace(L) = 0. This implies that the Gaussian and mean curvatures of the Euclidean plane at any point P is 0.

4.8 Example. The unit sphere $S^2 - \{0,0,\pm 1\}$ was given by $\vec{x}(u,v) = (\cos u \cos v, \cos u \sin v, \sin u)$, $(u,v) \in (-\frac{\pi}{2}, \frac{\pi}{2}) \times \mathbf{R}$.

The metric coefficients were given by

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & \cos^2 u \end{pmatrix} \text{ with inverse } (g^{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & \dfrac{1}{\cos^2 u} \end{pmatrix}.$$

$\vec{n} = -(\cos u \cos v, \cos u \sin v, \sin u)$,

$L_{11} = 1$, $L_{12} = L_{21} = 0$, and $L_{22} = \cos^2 u$.

Therefore

$$L = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ with } \det L = 1 \text{ and } 1/2 \operatorname{trace}(L) = 1.$$

The Gaussian and mean curvatures of the unit sphere S^2 at any point P is 1.

4.9 Example.

Let $\vec{x}(u,v) = (r \cos v \cos u, r \cos v \sin u, r \sin v)$ be the sphere of radius r.

$X_1 = (-r \cos v \sin u, r \cos v \cos u, 0)$

$X_2 = (-r \sin v \cos u, -r \sin v \sin u, r \cos v)$

$$g_{ij} = \langle X_i, X_j \rangle \text{ with } (g_{ij}) = \begin{pmatrix} r^2\cos^2 v & 0 \\ 0 & r^2 \end{pmatrix}$$

$$\text{and } (g^{ij}) = \begin{pmatrix} \dfrac{1}{r^2\cos^2 u} & 0 \\ 0 & \dfrac{1}{r^2} \end{pmatrix}.$$

$L_{ij} = \langle X_{ij}, \vec{n} \rangle$

$X_{11} = (-r \cos v \cos u, -r \cos v \sin u, 0)$

$X_{12} = (r \sin v \sin u, -r \sin v \cos u, 0)$

$X_{22} = (-r \cos v \cos u, -r \cos v \sin u, -r \sin v)$

$$\vec{n}\ \frac{X_1 x X_2}{|X_1 x X_2|} = (\cos v \cos u, \cos v \sin u, \sin v)$$

$L_{11} = -r(\cos^2 v)$, $L_{12} = L_{21} = 0$, $L_{22} = -r$

$$L = (L_k^{\,l}) = \sum L_{ik} g^{il} \text{ and } L = \begin{pmatrix} \frac{1}{r} & 0 \\ 0 & \frac{1}{r} \end{pmatrix}.$$

Therefore, $\det L = \frac{1}{r^2}$ and $1/2 \operatorname{trace}(L) = \frac{1}{r}$ which yield the Gaussian and mean curvatures, respectively, of the sphere at a given point P.

4.10 Example. For the circular cylinder $\vec{x}(t,v) = (r \cos t, r \sin t, v)$ the metric coefficients are

$$(g_{ij}) = \begin{pmatrix} r^2 & 0 \\ 0 & v'^2 \end{pmatrix} \text{ with inverse } (g^{ij}) = \begin{pmatrix} 1/r^2 & 0 \\ 0 & 1/v'^2 \end{pmatrix}.$$

Let $\vec{n} = (\cos t, \sin t, 0)$ with $L_{11} = -r$, $L_{12} = L_{21} = 0$, and $L_{22} = 0$. Therefore,

$$L = \begin{pmatrix} 1/r & 0 \\ 0 & 0 \end{pmatrix} \text{ with } \det L = 0 \text{ and } 1/2 \operatorname{trace}(L) = \frac{1}{2r}.$$

It follows that the Gaussian curvature of the cylinder is 0, while the mean curvature is $\frac{1}{2r}$.

The above examples all had constant curvature ≥ 0. Now we turn to an example with variable curvature: positive, negative, and 0.

4.11 Example. In studying the torus, we find that the Gaussian curvature of a point P depends upon the point's position on the surface. Let $\vec{x}(u,v) = ((a + b \cos u)\cos v, (a + b \cos u)\sin v, b \sin u)$ $0<b<a$, $(u,v) \in \mathbb{R} \times \mathbb{R}$. The metric coefficients are

$$(g_{ij}) = \begin{pmatrix} b^2 & 0 \\ 0 & (a\ b\cos u)^2 \end{pmatrix}$$

with inverse

$$(g^{ij}) = \begin{pmatrix} \frac{1}{b^2} & 0 \\ 0 & \frac{1}{(a\ b\cos u)^2} \end{pmatrix}.$$

$\vec{n} = -(\cos u \cos v, \cos u \sin v, \sin u)$ and

$L_{11} = b,\ L_{12} = L_{21} = 0,\ L_{22} = (a + b\cos u)\cos u.$

Therefore

$$L = \begin{pmatrix} \frac{1}{b} & 0 \\ 0 & \frac{\cos u}{(a\ b\cos u)} \end{pmatrix} \text{ with}$$

$$\text{Det } L = \frac{\cos u}{b(a\ b\cos u)} \text{ which is}$$

a) > 0 for $-\frac{\pi}{2}<u<\frac{\pi}{2}$

b) $= 0$ for $u = \pm\frac{\pi}{2}$

c) < 0 for $\frac{\pi}{2}<u<\frac{3\pi}{2}$.

These three cases represent the outside, the top and bottom circle, and the inside respectively.

One of Gauss's deepest and most surprising observations in his investigation of curved surfaces is that the curvature of a surface can be expressed in terms of its metric and the derivatives of its component functions. This metric is expressed in the form

$Edx^2 + 2Fdxdy + Gdy^2$ where

$E = \langle\partial/\partial x,\partial/\partial x\rangle,\ F = \langle\partial/\partial x,\partial/\partial y\rangle,\ G = \langle\partial/\partial y,\partial/\partial y\rangle.$

Gauss called this theorem "egregium" because it is so remarkable. K is defined very extrinsically, in terms of

$\vec{n}$ or L, none of which are intrinsic. Yet K is intrinsic.

4.12 Theorem. Gauss's Theorema Egregium

The Gaussian curvature K of a surface is intrinsic.

(See McCleary,[8])

Stahl,[13], gives a version of the formula for finding K as follows, where $E_1 = \partial E/\partial x$, $E_2 = \partial E/\partial y$, $E_{11} = \partial^2 E/\partial x^2$, etc.

$$4(EG - F^2)^2\ K$$

$$= E[E_2G_2\ \ 2F_1G_2\ \ (G_1)^2] + F[E_1G_2\ \ E_2G_1\ \ 2E_2F_2\ \ 4F_1F_2\ \ 2F_1G_1]$$

$$+ G[E_1G_1\ \ 2E_1F_2\ \ (E_2)^2]\ \ 2(EG\ \ F^2)[E_{22}\ \ 2F_{12}\ \ G_{11}].$$

This formula will be used to compute the Gaussian curvature for some of the models of non-Euclidean geometry discussed in chapter two.

5. Cut Locus

Take two points P and Q of a connected Riemannian manifold M and join them by a continuous piecewise differentiable curve. We can measure the arc length of this curve using the Riemannian metric. All possible piecewise differentiable curves joining P and Q will be considered. Define the distance $d(P,Q)$ between P and Q as the infimum of their arc lengths. The distance function d satisfies the usual three axioms of a metric space:

(1) $d(P,Q) = d(Q,P)$.

(2) For all points P and Q in S, $d(P,Q) \geq 0$ and $d(P,Q) = 0$ if and only if $P = Q$.

(3) (The triangle inequality). For all P, Q, and R in the metric space $d(P,Q) + d(Q,R) \geq d(P,R)$.

This allows us to talk about Cauchy sequences of points of M and also the completeness of M. A metric space is complete if every Cauchy sequence of points in the space converges to a point in the space.

Assume that M is a complete Riemannian manifold. Since a compact metric space is complete, a compact Riemannian manifold is always complete. A geodesic $g(t)$ can be parametrized by arclength. A geodesic $g(t)$ defined on the interval $a \leq t \leq b$, is said to be infinitely extendable if it can be extended to a geodesic $g(t)$ defined for the whole interval, $-\infty < t < \infty$.

<u>5.1 Definition</u>. A surface S is said to be <u>geodesically</u>

complete if every geodesic $g:[a,b]\to S$ can be extended to a geodesic $g:\mathbb{R}\to S$.

Take the Euclidean plane and delete the origin. This yields an incomplete Riemannian manifold because the positive x-axis is not infinitely extendable. The following important theorem relates metric completeness and geodesic completeness.

5.2 Theorem. (Hopf-Rinow) On a complete Riemannian manifold, every geodesic is infinitely extendable and any two points can be joined by a minimizing geodesic. (see McCleary,[8])

At each point P of a complete Riemannian manifold M, define a mapping of the tangent space T_PM at P onto M as follows. If X is a tangent vector at P, draw a geodesic g(t) starting at P in the direction of X parametrized by arclength. Parametrize the geodesic in such a way that g(0)=P. If X has length a, then we map X onto the point g(a) of the geodesic. Denote this mapping by $\exp_p:T_PM\to M$.

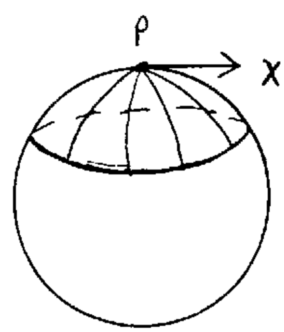


FIGURE 8

This mapping is called the exponential map at P. Exp_p maps a line in the tangent space T_pM through its origin onto the geodesic of M through P in the direction of the line. Since every point Q of M can be joined by a geodesic to P, $\exp_p$ maps T_pM onto M.

5.3 Definition. Fix a point P of a complete Riemannian manifold M and a geodesic g(t) starting at P. A cut point of P along g(t) is the first point Q on g(t) such that, for any point R on g(t) beyond Q, there is a geodesic from P to R shorter than g(t). This implies that Q is the first point where g(t) ceases to minimize distance (or arclength).

Let A be the set of positive real numbers s such that the geodesic g(t), $0 \le t \le s$, is minimizing where $s=d(P,g(s))$. Either $A=(0,\infty)$ or $A=(0,r)$ where r is some positive number. If $A=(0,r)$, then g(r) is the cut point of P along the geodesic g(t). If $A=(0,\infty)$, then we say that P has no cut point along the geodesic g(t). Therefore, if Q is a point on g(t) which comes after the cut point $P'=g(r)$, such that $Q=g(s)$ with $s>r$, then we can find a geodesic from P to Q which is shorter than g(t). If Q is a point which comes before the cut point P', then we cannot find a shorter geodesic from P to Q and there is not another geodesic from P to Q of the same length.

5.4 Theorem. If Q comes before P', then g(t) is the unique minimizing geodesic joining P and Q.

Proof: Let h(t) be another minimizing geodesic from P to Q. By moving from P to Q along h(t) and continuing from Q to P' along $g(t)$, we obtain a nongeodesic curve α from P to P' with an arclength equal to the distance $d(P,P')$.

Choose a point M on $h(t)$ before Q and also a point W on $g(t)$ after Q. Taking M and W sufficiently close to Q, replace the portion of α from M to W by the minimizing geodesic from M to W. We obtain a curve from P to P' with an arclength less than the distance $d(P,P')$ which is impossible. QED

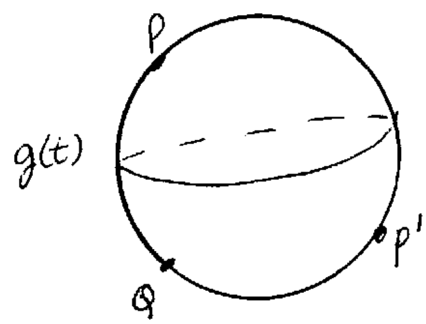


FIGURE 9

5.5 Definition. The set of cut points of P is called the cut locus of P and is denoted by $C(P)$.

If M is a compact Riemannian manifold, then on each geodesic starting from a point P, there is a cut point of P. Let M be an n-dimensional unit sphere. As in the unit sphere S^2, the geodesics are the great circles. If P is the north pole, then the cut locus $C(P)$ reduces to the south pole. The cut locus of a point P of a cylinder in $\mathbf{R}^3$ is the opposite generator to that which passes through P. However, there is a geodesic, namely, the generator through P, that extends infinitely far.

6. Geodesic Circles, Normal Neighborhoods, and Conjugate Points

6.1 Definition. A region D is convex if any two points of it can be joined by a geodesic arc lying wholly in D. A convex region is called simple if there is not more than one such geodesic arc. In the Euclidean plane, every convex region is simple, but this is not so for a surface in general. The surface of a sphere is convex but not simple.

6.2 Theorem. (J.H.C. Whitehead (1932))
Every point P of a surface has a neighborhood which is convex and simple and every point can be joined uniquely to every other point.

A particular form of Whitehead's theorem is concerned with a geodesic circle of given center P and radius r. This geodesic circle (or geodesic disk) is defined as the set of points Q such that there is a geodesic arc PQ of length not greater than r. This geodesic circle and normal neighborhood will be discussed in this section.

In this section, it will be shown that short enough geodesic segments behave as well in an arbitrary geometric surface as they do in $\mathbb{R}^3$. In the Euclidean plane, if we are interested in the distance to the origin, we use polar coordinates as a convenience. The distance from the origin to the point $\vec{x}(u,v) = (u \cos v, u \sin v)$ is simply u. In $\mathbb{R}^2$, the u-parameter curves are geodesics radiating out from some fixed point P of M.

Such geodesics may be described as follows: If W is a unit tangent vector at P, let α_W be the unique geodesic which starts at P with initial velocity W. Assembling all these geodesics into a single mapping yields the following.

<u>6.3 Definition.</u> Let $\vec{i},\vec{k}$ be orthogonal unit vectors tangent to M at P. Then $\vec{x}(u,v) = \alpha_{\cos v i + \sin v k}(u)$ is the <u>geodesic polar mapping</u> of M with pole P.

The domain of $\vec{x}$ is the largest region of $\mathbf{R}^2$ on which the formula makes sense. A choice of v fixes a unit tangent vector $W = \cos v\vec{i} + \sin v\vec{k}$ at P. Then the u-parameter curve $u \to \vec{x}(u,v) = \alpha_W(u)$ is the radial geodesic with initial velocity W. Since $\|W\| = 1$, this geodesic has unit speed, so that the length of α_W from $P=\alpha_W(0)$ to $\alpha_W(u)$ is just u. At the origin of $\mathbf{R}^2$, the geodesic polar mapping becomes

$$\vec{x}(u,v) = \alpha_{\cos v i + \sin v k}(u) = \vec{0} + u(\cos v\vec{i} + \sin v\vec{k})$$
$$= (u \cos v, u \sin v).$$

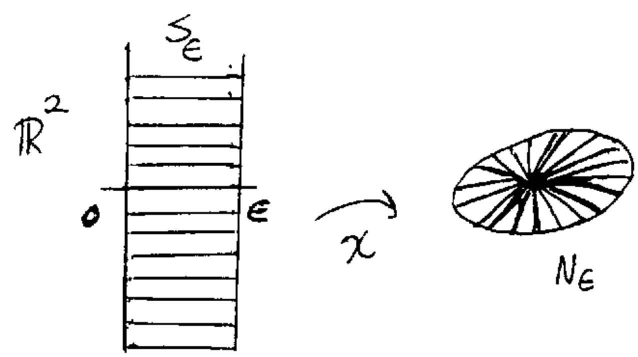
Therefore $\vec{x}$ is a generalization of polar coordinates in the plane.

The pole P is a trouble spot for a geodesic polar mapping. To clarify the situation near P, define a new mapping

$$\vec{y}(u,v) = \alpha_{ui + vk}(1).$$

$\vec{y}$ is differentiable and regular at the origin. By the inverse function theorem, $\vec{y}$ is a diffeomorphism of some disc D_ε: $u^2 + v^2 < \varepsilon^2$ onto a neighborhood N_ε of P. N_ε is called a normal neighborhood of P. In the special case M = $\mathbb{R}^2$, $\vec{y}$ is just the identity map $\vec{y}(u,v) = (u,v)$. Therefore for arbitrary M, $\vec{y}$ is a generalization of the rectangular coordinates of $\mathbb{R}^2$.

<u>6.4 Lemma.</u> For a sufficiently small number $\epsilon > 0$, let S_ϵ be the strip $0<u<\epsilon$ in $\mathbb{R}^2$. Then a geodesic polar mapping $\vec{x}:S_\epsilon \to M$ with pole P parametrizes a normal neighborhood N_ϵ of P - omitting P itself.



<u>FIGURE 10</u>

Proof: Note that $\vec{x}$ bears to $\vec{y}$ the usual relationship of polar coordinates to rectangular coordinates. This implies that

$$\vec{x}(u,v) = \alpha_{\cos v i + \sin v k}(u) = \alpha_{u\cos v i + u\sin v k}(1)$$
$$= \vec{y}(u \cos v, u \sin v).$$

This formula expresses $\vec{x}$ as the composition of two regular mappings:

(1) The Euclidean polar mapping

$(u,v) \to (u \cos v, u \sin v)$

which wraps the strip S_ϵ around the disc D_ϵ, and

(2) The one-to-one mapping $\vec{y}$ of D_ϵ onto N_ϵ.

Therefore $\vec{x}$ is regular and carries S_ϵ in usual polar-coordinate fashion onto the neighborhood N_ϵ - omitting only the pole. QED

A fundamental consequence of the previous lemma is that if $Q = \vec{x}(u_o, v_o)$ is any point in a normal neighborhood N_ϵ of P, then there is only one unit speed geodesic from P to Q which lies entirely in N, namely, the radial geodesic $\alpha(u) = \vec{x}(u, v_o)$, $0 \le u \le u_o$.

6.5 Example. Given the unit sphere S^2, let P be the north pole (0,0,1). The geographical parametrization

$\vec{x}(u,v) = (\sin u \cos v, \sin u \sin v, \cos u)$

yields the geodesics radiating out from P.

Each u-parameter curve is a unit-speed parametrization of a great circle and is therefore a geodesic.

$X_1 = (\cos u \cos v, \cos u \sin v, -\sin u)$ and for $u = 0$,

$X_1(0,v) = (\cos v, \sin v, 0) = \cos v\vec{i} + \sin v\vec{k}$

with $\vec{i} = (1,0,0)P$ and $\vec{k} = (0,1,0)P$.

By the uniqueness of geodesics,

$\vec{x}(u,v) = \alpha_{\cos v i + \sin v k}(u)$ which shows that $\vec{x}$ as defined above is the geodesic polar mapping of S^2 with pole P. Therefore the largest possible normal neighborhood N_ϵ of P occurs when $\epsilon = \pi$, for on the strip S_π, $\vec{x}$ is a polar parametrization of all the sphere except the north and south poles.

6.6 Theorem. For each point Q in a normal neighborhood

N_ϵ of P the radial geodesic segment in N_ϵ from P to Q uniquely minimizes arclength. (See O'Neill,[10])

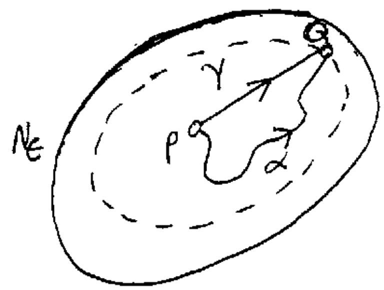


FIGURE 11

As a result of this theorem, if points P and Q are close enough together, then as in Euclidean space, there is a unique geodesic segment from P to Q which is shorter than any other curve from P to Q. Unlike the Euclidean case, there may be many other nonshortest geodesics from P to Q. If $\vec{x}$ is a geodesic polar parametrization at P, we shall call the v-parameter curve $u=\epsilon$, the geodesic circle of radius ϵ whose center is P. C_ϵ consists of all points at a distance ϵ from P.

6.7 Example. On a sphere of radius r the geographical parametrization would be

$\vec{x}(u,v)$ = (rsinu cosv, rsinu sinv, rcosu) with

P=(0,0,1).

X_1 = (rcosu cosv, rcosu sinv, -rsinu) and

$X_1(0,v)$ = (rcosv, rsin v, 0) = cosrv$\vec{i}$ + sinrv$\vec{k}$.

Therefore $\vec{x}(u,v)=\alpha_{\cos rvi + \sin rvk}(u)$ and each point P of the sphere of radius r has a normal neighborhood N_ϵ when $\epsilon = \pi r$. This is all of the sphere except the point, -P, antipodal to the pole P. Therefore, if two points P and Q are not antipodal ($Q \neq -P$) then there is a unique shortest curve α from P to Q. Intrinsic distance on the sphere is given by the formula $d(P,Q) = r\theta$ where θ $(0 \le \theta \le \pi)$ is the angle from P to Q in $\mathbb{R}^3$. If P and Q are not antipodal, then $d(P,Q) = L(\alpha) = r\theta$. As Q moves toward the antipodal point -P of P, by continuity $d(P,-P) = r\pi$. Therefore no geodesic segment α of length $L(\alpha) > r\pi$ can minimize arc length between its endpoints.

The Gaussian curvature K=(det L) of a geometric surface M affects the geodesics of the surface. (See section four for Gaussian curvature.)

<u>6.8 Definition.</u> A geodesic segment α from P to Q <u>locally minimizes arclength</u> from P to Q provided that for any curve segment β from P to Q which is sufficiently near (ϵ-close) to α, then $L(\beta) \ge L(\alpha)$ where $L(\beta) = d(P,Q)$. This local minimization is strict (or unique) provided we get strictly inequality $L(\beta) > L(\alpha)$ unless β is a reparametrization of α.

Think of α as an elastic string or rubber band which is constrained to lie in M, is under tension, and has its endpoints pinned down at P and Q. Because α is a geodesic, it is in equilibrium. If it were not a geodesic, its tension would pull it into a new shorter position. If α is pulled aside slightly to a new curve β and released, will it return to its original position α? If β is longer than α, then its tension will pull it back to α.

The study of local minimization on two-dimensional manifolds depends on the notion of conjugate points. If α is a unit-speed geodesic starting at P, then α is a u-parameter curve, $v=v_o$, of a geodesic polar mapping $\mathbf{x}$ with pole P.

$\mathbf{x}(u,v) = \alpha_{\cos vi + \sin vk}(u) = (u \cos v, u \sin v)$.

$G=\langle X_2, X_2\rangle=u^2$ where $X_2 = (-u \sin v, u \cos v)$

Therefore at u=0, G is zero but is nonzero immediately thereafter.

<u>6.9 Definition.</u> A point $\alpha(s) = \mathbf{x}(s,v_o)=(s \cos v_o, s \sin v_o)$ with $s > 0$ is a conjugate point of $\alpha(0) = P$ on α provided $G(s,v_o) = 0$ where $G=\langle X_2,X_2\rangle=s^2$. (Such points may or may not exist.)

The geometric meaning of conjugacy rests on the interpretion of $\sqrt{G} = \|X_2\|$ as the rate at which the radial geodesic u-parameter curves are spreading apart. For fixed $\epsilon > 0$, if $\sqrt{G}$ is large, then the distance from $\mathbf{x}(u,v)$ to $\mathbf{x}(u,v+\epsilon)$ is large. This means that the radial geodesics are spreading rapidly. When $\sqrt{G}$ is small, then the distance from $\mathbf{x}(u,v)$ to $\mathbf{x}(u,v+\epsilon)$ is small. Therefore the radial geodesics are pulling back together again. It follows that when G vanishes at a conjugate point $\alpha(s_1) = \mathbf{x}(s_1,v_o)$, for v near v_o, the u-parameter curves have all reached this same point after traveling at unit speed the same distance s_1. However, this meeting may not occur.

6.10 Example. The Euclidean plane gives the standard rate at which radial geodesics spread apart. For

$\vec{x}(u,v) = (u \cos v,\ u \sin v)$ with $X_2=(-u \sin v,\ u \cos v)$,

then $G=\langle X_2, X_2\rangle= u^2$ and $\sqrt{G} = u$. Therefore G does not vanish and there are no conjugate points in the Euclidean plane.

6.11 Example. The unit sphere S^2 with P=(0,0,1), the north pole, has parametrization

$\vec{x}(u,v) = (\sin u \cos v,\ \sin u \sin v,\ \cos u)$.

Therefore, $X_2 = (-\sin u \sin v,\ \sin u \cos v,\ 0)$ and $\sqrt{G} = \sin u$. Since $\sin u < u$ for $u > 0$, the radial geodesics starting at the north pole P of S^2 spread less rapidly than in $\mathbf{R}^2$. Since $\sqrt{G(\pi, r)}=\sin\pi=0$, the radial geodesics all have their first conjugate point after traveling a distance of π.

6.12 Example. Let $\vec{x}$ be a geodesic polar mapping defined on a region where $G > 0$. Then $\sqrt{G}=\|X_2\|$ satisfies the Jacobi differential equation

$(\sqrt{G})_{11} + K \sqrt{G} = 0$ subject to the initial conditions

$\sqrt{G}(0,v) = 0$, $(\sqrt{G})_1(0,v) = 1$ for all v, where

$(\sqrt{G})_1 = (\partial\sqrt{G}/\partial u)$ and $(\sqrt{G}\ )_{11} = (\partial^2\sqrt{G}/\partial u^2)$ and

K = Gaussian curvature. The restriction $G > 0$ is needed to ensure that $\sqrt{G}$ is differentiable. $\sqrt{G}(u,v)$ is well-defined for u=0 since $\sqrt{G}(0,v) = \|X_2(0,v)\| = 0$. $\sqrt{G}$ need not be differentiable at u = 0, so interpret

$(\sqrt{G})_1(0,v)$ and $(\sqrt{G})_{11}(0,v)$ as limits such as

$$(\sqrt{G})_1(0,v) = \lim_{u\to 0} (\sqrt{G})_1(u,v).$$

For the Euclidean plane $\sqrt{G} = u$, $\sqrt{G}(0,v) = 0$ and $(\sqrt{G})_1(0,v) = 1$ for all v. Therefore the initial conditions show that as the radial geodesics leave the pole P in any geometric surface, they are spreading at the same rate as in the Euclidean plane. But the Jacobi equation shows that immediately thereafter the rate of spreading depends on the Gaussian curvature of the surface. For $K < 0$, radial geodesics spread apart faster than in $\mathbb{R}^2$. For $K > 0$, the rate of spreading is less than in $\mathbb{R}^2$.

In the Euclidean plane, we found the Gaussian curvature to be zero. By measuring a short distance ϵ in all directions from P, we obtain the polar geodesic circle C_ϵ of radius ϵ. The circumference of C_ϵ is $L(C_\epsilon) = 2\pi\epsilon$. For $K > 0$, the radial geodesics from P are not spreading as rapidly as in the Euclidean plane, so C_ϵ will be shorter than $2\pi\epsilon$. This implies that geodesic circles on a surface of positive curvature are always "too small". For $K < 0$ the radial geodesics from P are spreading more rapidly than in the Euclidean plane, and C_ϵ will be longer than $2\pi\epsilon$. Therefore geodesic circles on a surface of negative curvature are always "too large".

As a consequence of the theorem, we can find $\sqrt{G}$ on a geodesic α by solving the Jacobi equation on α, subject to the given initial conditions. Let α be a unit-speed geodesic starting at the point P in M. Let $g(u) = \sqrt{G}(u,v_o)$ be the unique solution of the Jacobi equation on α, $g'' + K(\alpha)g = 0$ such that $g(0) = 0$, $g'(0) = 1$. Then the first conjugate point of $\alpha(0) = P$ on α (if it exists) is $\alpha(s_1)$, where s_1 is the smallest positive number such that $g(s_1) = 0$.

6.13. Example. Let α be a unit speed geodesic starting at any point P of the sphere of radius r. The Gaussian curvature along this geodesic (great circle) is $1/r^2$. The Jacobi equation for α is given

$g'' + g/r^2 = 0$ which has the general solution

$g(s) = A \sin(s/r) + B \cos(s/r)$.

The initial conditions $g(0) = 0$, $g'(0) = 1$, yield the equation $g(s) = r \sin(s/r)$. The first zero of this function with $s_1 > 0$ occurs at $s_1 = \pi r$. Therefore the first conjugate point of $\alpha(0) = P$ on α is at the antipodal point of P.

6.14 Example. Let α be a unit-speed parametrization of the outer equator of a torus of revolution T. On α the Gaussian curvature is

$$\frac{\cos u}{b(a+b\cos u)} \quad \text{with } u = 0.$$

It follows that α has constant positive Gaussian curvature $\frac{1}{b(a+b)}$. The Jacobi equation for α is

$$g'' + \frac{g}{b(a+b)} = 0 \quad \text{which has the general solution}$$

$$g(s) = A \sin \frac{s}{\sqrt{b(a+b)}} + B \cos \frac{s}{\sqrt{b(a+b)}} .$$

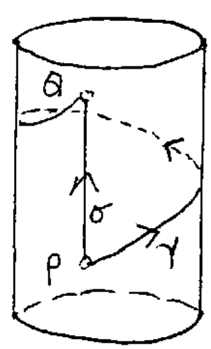
Therefore the first conjugate point $\alpha(s_1)$ of $\alpha(0) = P$ on α will occur at exactly the same distance s_1 along α as if α were on a sphere with this curvature. The initial conditions $g(0)=0$, $g'(0)=1$, yield the equation

$$g(s) = \sqrt{b(a+b)} \sin\left(\frac{s}{\sqrt{b(a+b)}}\right).$$

The first zero of this function with $s_1 > 0$ occurs at $s_1 = \pi \sqrt{b(a+b)}$.

<u>6.15 Corollary.</u> There are no conjugate points on any geodesic in a surface with curvature $K \le 0$. Hence every geodesic segment on such a surface is locally minimizing.

Proof: Let α be a geodesic in M. Since $g(0)=0$ and $g'(0)=1$, we have $g(s) \ge 0$ for $s \ge 0$ at least up to the first conjugate point (if it exists). But $K \le 0$ implies that $g'' = -Kg \ge 0$, so g' is an increasing function with $g' \ge 1$. Therefore $g(s) \ge s$ up to the first conjugate point which can never occur. QED



<u>FIGURE 12</u>

If α is a geodesic segment from P to Q such that there are no conjugate points of $P = \alpha(0)$ on α, then α locally minimizes arc length from P to Q. On the circular cylinder with $K=0$, the helical geodesic α from P to Q, with Q directly above P, is locally minimizing, but it is certainly not minimizing. The straight line segment σ provides a much shorter way to get from P to Q.

In Ordinary Differential Equations, we say that $y(x)$ oscillates on the interval $[c,\infty)$ if y is nonconstant and has infinitely many zeros on $[c,\infty)$. Equations of the form $y'' + p(x)y = 0$ are oscillatory. The Jacobi equation is of this form. Therefore all solutions of the Jacobi equation are oscillatory. (See Derrick and Grossman,[2]) This implies that if $K \geq \epsilon > 0$, i.e., if K is positive and bounded away from 0, then there must be conjugate points.

7. Gauss-Bonnet Theorem.

A triangle in the plane is determined by three line segments and the region they enclose. A triangle on a surface is determined by three geodesic segments that enclose a region and is called a geodesic triangle. Gauss deduced some of the basic properties of geodesic triangles including a general relation between area and angle sum.

If a region D is simply connected in a surface S, then any closed curve in D can be contracted to a point without leaving the region. From H. Hopf we know that the tangent along a closed piecewise differentiable curve enclosing a simply connected region turns through 2π. $\sum \int \alpha_j k_g(s)ds$ represents the total geodesic curvature along α, the boundary of the region D. $\iint D\ K\ dA$ represents the total Gaussian curvature of D with dA the element of area of D.

7.1 Theorem. (Gauss-Bonnet) If D is a simply connected region in a regular surface S bounded by a piecewise differentiable curve α making exterior angles $\epsilon_1,\ldots,\epsilon_n$ at the vertices of α, then

$$\sum \int\alpha_j\ k_g(s)ds + \iint D\ K\ dA = 2\pi - \sum \epsilon_j.$$

(See McCleary,[8]) This formula contains geodesic curvature, Gaussian curvature, and exterior angles, mixing up curves, angles,and areas into a remarkable relation.

Gauss's version of the Gauss-Bonnet theorem is concerned with geodesic triangles. Let R = ΔABC be such a triangle with sides given by geodesic segments. The interior angles are given by $\angle A = \pi - \epsilon_A$, $\angle B = \pi - \epsilon_B$, and $\angle C = \pi - \epsilon_C$. Therefore $2\pi - \epsilon_A - \epsilon_B - \epsilon_C$ $= 2\pi + \angle A - \pi + \angle B - \pi + \angle C - \pi = \angle A + \angle B + \angle C - \pi$. Since geodesic curvature vanishes on geodesics, then:

<u>7.2 Corollary</u>. (Gauss) For a geodesic triangle ΔABC on a surface, $\iint_{\Delta ABC} K\,dA = \angle A + \angle B + \angle C - \pi$. (See McCleary, [8])

In the Euclidean plane, the Gaussian curvature (K) is zero. Therefore $\angle A + \angle B + \angle C = \pi$, and the interior angle sum of a triangle in the Euclidean plane is π.

On a surface of constant positive curvature, such as the sphere, let $c = \iint_{\Delta ABC} K\,dA = K \cdot \text{area}(\Delta ABC) > 0$. Therefore $c = \angle A + \angle B + \angle C - \pi$, and $\angle A + \angle B + \angle C = \pi + c$ with $c > 0$, and the interior angle sum of a triangle on a surface of constant positive curvature is greater than π.

On a surface of constant negative curvature, let $d = \iint_{\Delta ABC} K\,dA = K \cdot \text{area}(\Delta ABC) < 0$. Then $d = \angle A + \angle B + \angle C - \pi$, $\angle A + \angle B + \angle C = \pi + d$ with $d < 0$, and the interior angle sum of a triangle on a surface of constant negative curvature is less than π.

There are many generalizations of the Gauss-Bonnet Theorem. In its simplest form it states that positive curvature everywhere implies positive Euler characteristic. The Euler characteristic of a surface M is the integer $\chi = F - E + V$. Let M be a compact surface in $\mathbb{R}^3$. Suppose

M can be broken into regions bounded by polygons, each region contained in a simply connected geodesic coordinate patch, with the segments of the polygon being geodesics. V equals the number of vertices, E the number of edges, and F the number of faces or number of polygonally bounded regions. This version of the Gauss-Bonnet Theorem relates two seemingly unrelated quantities, curvature (a differential geometric quantity) with the Euler characteristic χ (a topological or combinatorial quantity).

7.3 Theorem. (Gauss-Bonnet) If M is compact then $\iint_M K\, dA = 2\pi\chi$. (See Millman and Parker,[9])

CHAPTER 2

NON-EUCLIDEAN PLANE

1. BELTRAMI DISK

The notion of an abstract surface frees us to seek models of non-Euclidean geometry without the restriction of finding a subset of a Euclidean space. A set with a system of coordinates and a Riemannian metric determines a geometric surface. The models of non-Euclidean geometry in this chapter are due to E. Beltrami (1835-1906) and J. Henri Poincaré (1854-1912).

1.1 Definition A function $f:S_1 \to S_2$ between surfaces S_1 and S_2 is a *diffeomorphism* if it is differentiable, one-to-one, and onto, and has a differentiable inverse function. Two surfaces S_1 and S_2 are said to be diffeomorphic if there is a diffeomorphism $f:S_1 \to S_2$.

1.2a Definition. An *isometry* $f:M \to N$ of surfaces is a diffeomorphism such that

$$<\vec{v},\vec{w}> = <f_*(\vec{v}),f_*(\vec{w})>$$

for any pair of tangent vectors $\vec{v},\vec{w}$ to M.

1.2b Definition. A mapping of surfaces $f:M \to N$ is *conformal* provided there exists a real-valued function $g>0$ on M such that

$$<\vec{u},\vec{v}> = g(P)<f_*(\vec{u}), f_*(\vec{v})>$$

for all tangent vectors $\vec{u},\vec{v}$ to M. The function g is called the scale factor of f.

If f is a conformal mapping for which g has constant value 1, then f is a local isometry. A local isometry preserves lengths of tangent vectors. Therefore distances and angles are preserved. Otherwise a conformal mapping is a generalized isometry in which angles are preserved but in which lengths of tangent vectors need not be preserved. At each point P of M the tangent vectors at P all have their lengths stretched by the same factor.

Beltrami studied and sought local conditions on a pair of surfaces, S_1 and S_2, that guarantee that there is a local diffeomorphism of S_1-S_2 such that geodesics on S_1 are taken to geodesics on S_2. Beltrami solved the problem when one of the surfaces is the Euclidean plane. He found conditions for the existence of a mapping taking geodesics on a surface S to straight lines in the plane. Such a mapping is called a geodesic mapping.

1.3 Theorem (Beltrami 1865). If there is a geodesic mapping from a surface S to the Euclidean plane, then the Gaussian curvature of the surface S is constant. (See McCleary,[8])

In the case of the sphere, central projection takes great circles to straight lines. To construct the projection, fix the tangent plane at a point R on the sphere and join a point P in the adjacent hemisphere to the center of the sphere. Extend this segment to the tangent plane and this is the image of the point P. Central projection is defined on the open hemisphere with R as the center.

If R is taken to be the south pole, then

$$\vec{Y}(\lambda,\Phi) = (-\cos\lambda \cot\Phi, -\sin\lambda \cot\Phi, -1).$$

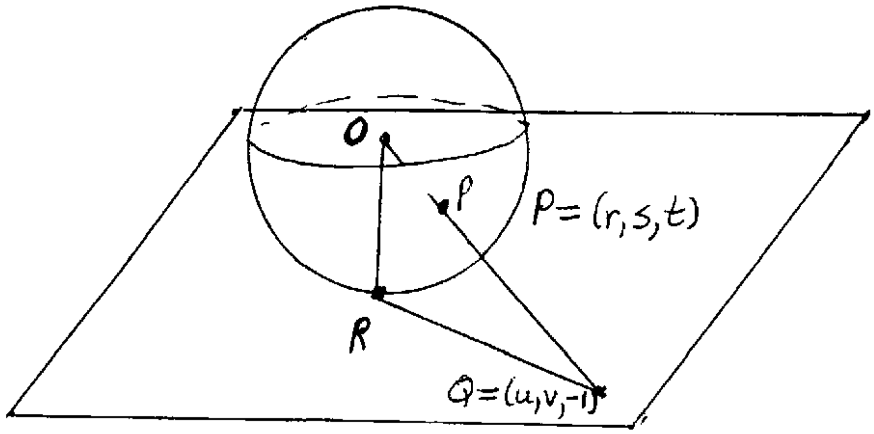


FIGURE 13

A great circle is determined by the intersection of the sphere and a plane through the center of the sphere. The image of the great circle under central projection is the intersection of this plane with the tangent plane, and therefore determines a line. This projection will be important for models of non-Euclidean geometry, so consider the inverse of central projection.

1.4 Proposition The inverse of central projection of the lower hemisphere of a sphere of radius R centered at the origin to the plane tangent to the south pole (0,0,-R) has the form

$$\vec{X}:\mathbb{R}^2\to S^2, \quad \vec{X}(u,v)=\frac{R}{\sqrt{R^2+u^2+v^2}}(u,v,-R).$$

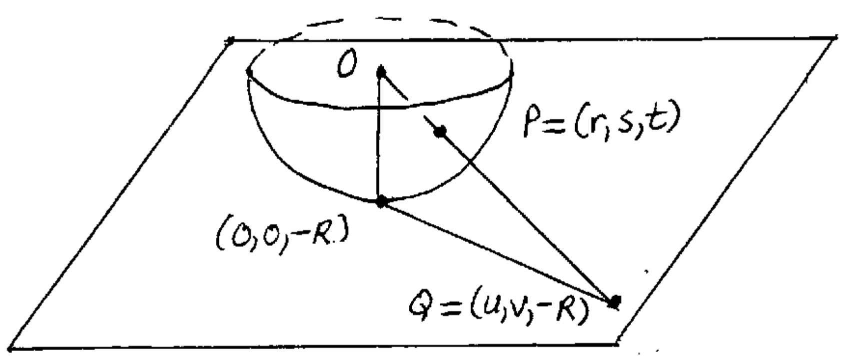


FIGURE 14

Proof: Let Q=(u,v,-R) denote a point in the plane tangent to the south pole and consider the line segment in $\mathbb{R}^3$ joining Q to the origin. This line segment passes through a point P on the sphere. Write the coordinates of the point $\vec{x}(u,v)$ = P = (r,s,t). The linear dependence of OP and OQ implies that

$$(0,0,0) = OP \times OQ = (-Rs-tv,\ Rr+tu,\ rv-su).$$

It follows that $r = \frac{tu}{-R}$ and $s = \frac{tv}{-R}$.

The condition $r^2+s^2+t^2=R^2$ implies that $t=\frac{-R^2}{\sqrt{R^2+u^2+v^2}}$.

It follows that $\vec{x}(u,v)=(r,s,t)=\frac{R}{\sqrt{R^2+u^2+v^2}}(u,v,-R)$. QED

From the inverse of the central projection we can endow the plane with the geometry of the sphere by inducing a Riemannian metric on $\mathbb{R}^2$ via the mapping $\vec{x}:\mathbb{R}^2\to S^2$. Since the sphere is a surface in $\mathbb{R}^3$, compute directly $\vec{X}_1=\vec{X}_u$ and $\vec{X}_2=\vec{X}_v$.

$$\vec{X}_1=\frac{R}{(R^2+u^2+v^2)^{\frac{3}{2}}}(R^2+v^2,-uv,Ru) \text{ and}$$

$$\vec{X}_2=\frac{R}{(R^2+u^2+v^2)^{\frac{3}{2}}}(-uv,R^2+u^2,Rv).$$

As in Example 3.2 of Chapter 1, the standard classical notation for the Riemannian metric of a Riemannian manifold is given by:

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2$$

where $E=<X_1,X_1>$, $F=<X_1,X_2>=<X_2,X_1>$, and $G=<X_2,X_2>$. We call ds the element of arc length or the line element on S. The coordinate vectors $\vec{X}_1, \vec{X}_2$ determine the following line element on the sphere and thus on the plane in the induced metric:

$$ds^2 = R^2 \frac{(R^2+v^2)\,du^2-2uvdudv+(R^2+u^2)\,dv^2}{(R^2+u^2+v^2)^2}.$$

Computing the curvature associated with the metric induced by the central projection yields $K(P) = \frac{1}{R^2}$ for all P. This was computed by using Gauss's Theorema Egregium, Stahl's formula for curvature, and many tedious calculations.

Beltrami observed that the above calculation depends on R^2 and not on R. He therefore replaced R with $\sqrt{-1}R$ to develop a model of non-Euclidean geometry. The abstract surface is the interior of the disk of radius R in $\mathbb{R}^2$ centered at (0,0) and given by

$$D=\{(u,v)\in\mathbb{R}^2 \mid u^2+v^2<R^2\}.$$

The line element, or element of arc length, becomes

$$ds^2 = -R^2 \frac{(v^2-R^2)\,du^2-2uvdudv+(u^2-R^2)\,dv^2}{(-R^2+u^2+v^2)^2}$$

$$= R^2 \ \frac{(R^2-v^2)\,du^2+2uvdudv+(R^2-u^2)\,dv^2}{(R^2-u^2-v^2)^2}. \quad \text{(Formula 1)}$$

This formula determines a Riemannian metric on the given abstract surface. Since $R=\sqrt{-1}R$, the curvature is constant and equal to $\frac{-1}{R^2}$. The local differential equations satisfied by the geodesic mapping in Theorem 1.1 will continue to hold since we have only changed the constant R^2. It follows that the geodesics on this abstract surface are Euclidean line segments.

Now fix the value of R as 1. Our abstract surface is the interior of the unit disk, denoted by

$$D_B = \{(u,v) \in \mathbb{R}^2 \,|\, u^2+v^2<1\}.$$

This disk is called the Beltrami disk. Let R = 1 in Formula 1. The metric for the Beltrami disk is then given by

$$ds_B{}^2 = \frac{(1-v^2)\,du^2+2uvdudv+(1-u^2)\,dv^2}{(1-u^2-v^2)^2}$$

where B represents Beltrami. The curvature which is constant becomes $\frac{-1}{R^2}=\frac{-1}{1}=-1$.

Choose a convenient point in the Beltrami disk. When u=0 or v=0 the middle term of $ds_B{}^2$ vanishes and so the geodesics $u \to (u,v_0)$ and $v \to (u_0,v)$ are perpendicular to the v-axis or u-axis, respectively. At the center of the disk the axes themselves are perpendicular geodesics. Our convenient point is the origin (0,0). To discuss distance from (0,0) and Euclidean circles centered at (0,0) polar coordinates on the Beltrami disk are convenient. Let

$$u=r\cos\theta,\ v=r\sin\theta,\ du=\cos\theta dr-r\sin\theta d\theta,\ dv=\sin\theta dr+r\cos\theta d\theta.$$

Rewriting ds_B^2 as

$$ds_B^2 = \frac{(1-u^2-v^2)(du^2+dv^2)+(udu+vdv)^2}{(1-u^2-v^2)^2}$$

and substituting yields

$$= \frac{(1-r^2)(dr^2+r^2d\theta^2)+r^2dr^2}{(1-r^2)^2} = \frac{dr^2}{(1-r^2)^2}+\frac{r^2d\theta^2}{1-r^2}.$$

A Euclidean line segment through the origin in the Beltrami model, which is a geodesic, has polar equation $\theta=\theta_\circ$, a constant. If $O=(0,0)$ and $P=(r\cos\theta_\circ, r\sin\theta_\circ)$, then

$$ds_B^2 = \frac{dr^2}{(1-r^2)^2}$$

and the distance in the Beltrami model, denoted by $d_B(O,P)$, is given by

$$d_B(O,P) = \int_0^r \frac{dt}{1-t^2} = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right).$$

As r approaches 1 the distance goes to infinity. Therefore Euclidean lines through the origin have infinite length.

2. Poincaré Disk

One problem with the Beltrami disk is the representation of angles - the rays may be Euclidean line segments, but the angles can be far from their Euclidean appearance in measure. Fix a Euclidean angle, say $\frac{\pi}{2}$, between two line segments and place this Euclidean figure at different points in the Beltrami disk. The angle measure depends on the position of the vertex. For example, away from the u- and v-axes, the line segments $u \to (u, v_\circ)$ and $v \to (u_\circ, v)$ cross at $(u_\circ, v_\circ)$ in an angle determined by the function $F(u_\circ, v_\circ)$ where $F = \frac{\langle X_1, X_2 \rangle}{(1-u^2-v^2)^2}$.

This next model of non-Euclidean geometry again has the interior of a Euclidean disk as the underlying abstract surface. In this model, angle measurement agrees with its Euclidean measure. An orthographic projection is a mapping for which all lines of projection are orthogonal to the plane of projection.

2.1 Proposition. Orthographic projection of the Beltrami model to the lower hemisphere of the sphere of radius one centered at (0,0,1) is a conformal mapping.

Proof: Orthographic projection is given by mapping the Beltrami disk into $S^2 + (0,0,1)$,

$$(u,v) \mapsto (u, v, 1-\sqrt{1-u^2-v^2}).$$

$$\vec{X}_1 = (1, 0, \frac{u}{\sqrt{1-u^2-v^2}}) \text{ and } \vec{X}_2 = (0, 1, \frac{v}{\sqrt{1-u^2-v^2}})$$

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2$$

$$= \frac{(1-v^2)\,du^2+2uv\,du\,dv+(1-u^2)\,dv^2}{1-u^2-v^2}.$$

Comparing this equation with the line element on the Beltrami disk yields $ds^2 = (1-u^2-v^2)\,ds_B^2$ and the mapping is conformal. QED

Orthographic projection takes the geodesics in the Beltrami disk, Euclidean line segments, to semicircles on the sphere that meet the equator at right angles.

Follow this mapping by stereographic projection from the north pole to the plane that is tangent to the south pole, which is a conformal mapping of a sphere. The image point P* of a point P on the sphere is the point of intersection of the straight line through the north pole and P and the tangent plane at the south pole. In this way, the whole sphere, except the north pole, is mapped onto the open plane in a one-to-one fashion. The images of the circles parallel to the equator are concentric circles with their common center at the south pole. The images of the meridians are straight lines through the south pole. The projection of the north pole is the point at infinity. The mapping is represented by

$$P = (p_1, p_2, p_3) \mapsto \left(\frac{Rp_1}{r}, \frac{Rp_2}{r}\right) = P^* \quad \text{where}$$

r and R are the distances from P and P*, respectively, to the z-axis. By using the similar triangles in Figure 15,

$$\frac{R}{2} = \frac{r}{(2-p_3)} \quad \text{and} \quad R = \frac{2r}{(2-p_3)}.$$

Therefore,

$$P = (p_1, p_2, p_3) \mapsto \left(\frac{2p_1}{2-p_3}, \frac{2p_2}{2-p_3}\right) = P^*.$$

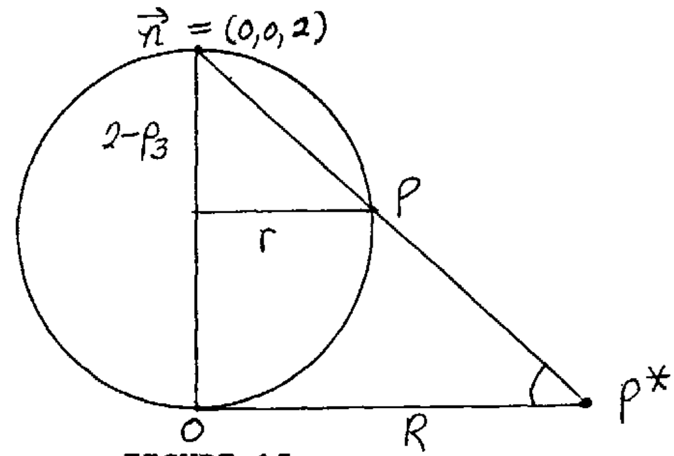


FIGURE 15

As a result of the two mappings, the lower hemisphere of the unit sphere maps to the disk of radius 2, centered at the origin,

$$D_2 = \{(x,y) \mid x^2 + y^2 < 4\}.$$

A Riemannian metric is induced on this disk D_2, called the Poincaré disk, by transfering the Beltrami metric from the Beltrami disk via the diffeomorphism given by the composite of orthographic and stereographic projection:

$$(u,v) \mapsto (u, v, 1-\sqrt{1-u^2-v^2}) \mapsto (x, y, 0)$$

$$\text{where } x = \frac{2u}{1+\sqrt{1-u^2-v^2}} \text{ and } y = \frac{2v}{1+\sqrt{1-u^2-v^2}}.$$

Let $w = -\sqrt{1-u^2-v^2}$, then

$$dw = \frac{udu+vdv}{\sqrt{1-u^2-v^2}}, \quad \frac{dw^2}{w^2} = \frac{(udu+vdv)^2}{(1-u^2-v^2)^2}, \quad x = \frac{2u}{1-w},$$

$$y = \frac{2v}{1-w}, \quad dx = \frac{2(1-w)\,du+2u\,dw}{(1-w)^2}, \quad dy = \frac{2(1-w)\,dv+2v\,dw}{(1-w)^2}.$$

This allows us to write

$$ds_B^2=\frac{(1-u^2-v^2)\,(du^2+dv^2)+(udu+vdv)^2}{(1-u^2-v^2)^2} = \frac{1}{w^2}(du^2+dv^2+dw^2).$$

By expanding $dx^2 + dy^2$ and letting $w^2 = 1-u^2-v^2$, $1-w^2=u^2+v^2$, and $-wdw=udu+vdv$, we obtain

$$dx^2+dy^2=\frac{4}{(1-w)^2}(du^2+dv^2+dw^2)=\frac{4w^2}{(1-w)^2}ds_B^2.$$

Let $ds_R^2=\frac{(1-w)^2}{4w^2}(dx^2+dy^2)$, where R refers to Riemann.

Since $\frac{x^2+y^2}{4}-1=\frac{u^2+v^2}{(1-w)^2}-1=\frac{2w(1-w)}{(1-w)^2}=\frac{2w}{1-w}$, the metric induced by the mapping on D_2 takes the form

$$ds_R^2=\frac{dx^2+dy^2}{(1-\frac{x^2+y^2}{4})^2}.$$

It can be shown that the Poincaré disk has constant Gaussian curvature K=-1 as does the Beltrami disk. (See McCleary,[8])

As a point (u,v) approaches the rim of the Poincaré disk, that is, the circle $u^2 + v^2 = 4$, then $(1 - \frac{x^2+y^2}{4})$ approaches zero. Therefore rulers must shrink as they approach the rim, so that the disk is bigger than our Euclidean intuition may suggest. Let σ be a constant polar angle and compute the arc-length function s(t) of the Euclidean line segment $\alpha(t) = (t\cos\sigma, t\sin\sigma)$, $0 \le t < 2$, which runs from the origin almost to the rim. $\alpha'(t)=(\cos\sigma,\sin\sigma)$ and with the given Riemannian metric,

$<\alpha'(t),\alpha'(t)>_{D_2} = \dfrac{1}{(1-\frac{t^2}{4})^2}$. Therefore, $\alpha(t)$ has hyperbolic speed

$$\sqrt{<\alpha'(t),\alpha'(t)>_{D_2}} = \|\alpha'(t)\| = \frac{1}{(1-\frac{t^2}{4})}.$$

It follows that

$$s(t) = \int_0^t \frac{dt}{1-\frac{t^2}{4}} = 2\tanh^{-1}\frac{t}{2} = \log\frac{2+t}{2-t}.$$

As t approaches 2, arclength s(t) from the origin $\alpha(0)$ to $\alpha(t)$ approaches infinity. This "short" segment α actually has infinite hyperbolic length.

When working in the Poincaré disk, polar coordinates are a natural choice. This allows the metric to depend only on the distance to the origin. Let

$\vec{x}(u,v)=(u\cos v, u\sin v)$, $0<u<2$, with $\dfrac{du^2+dv^2}{(1-\frac{u^2}{4})^2}$,

$\vec{x}_1=(\cos v,\sin v)$ and $\vec{x}_2=(-u\sin v, u\cos v)$.

It follows that

$$E = \frac{1}{(1-\frac{u^2}{4})^2},\ F = 0,\ G = \frac{u^2}{(1-\frac{u^2}{4})^2}.$$

2.2 <u>Definition.</u> A <u>Clairaut parametrization</u> $\vec{x}:U\subset R^2\to M$ is a parametrization in which $F=<X_1,X_2>=0$ and $E=<X_1,X_1>$ and $G=<X_2,X_2>$ depend only on x_1 (the u-parameter curves).

2.3 <u>Lemma.</u> If $\vec{x}$ is a Clairaut parametrization, then all the x_1=u-parameter curves of $\vec{x}$ are geodesics.

Proof: A regular curve α in M is a geodesic if and only if α has constant speed and geodesic curvature

$k_g=0$. α has geodesic curvature 0 if and only if α' and α'' are always collinear. For a Clairaut parametrization we need to show that $\vec{X}_1$ and $\vec{X}_{11}=\frac{\partial^2\vec{X}}{\partial u^2}$ are collinear. Since $\vec{X}_1$ and $\vec{X}_2$ are orthogonal, this is equivalent to $\vec{X}_2\cdot\vec{X}_{11}=0$.

$$0=E_2=(\vec{X}_1\cdot\vec{X}_1)=2\vec{X}_1\cdot\vec{X}_{12} \text{ and } 0=F_1=(\vec{X}_1\cdot\vec{X}_2)_1=\vec{X}_{11}\cdot\vec{X}_2+\vec{X}_1\cdot\vec{X}_{21}.$$

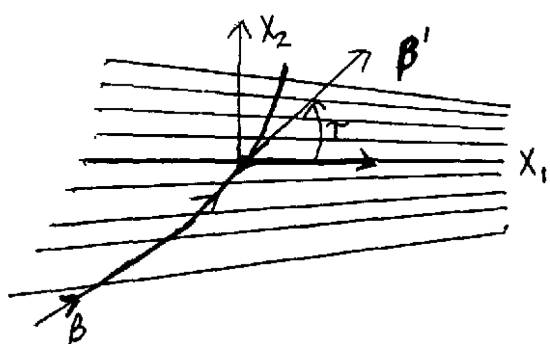
Therefore, $\vec{X}_2\cdot\vec{X}_{11}=0$ and all the u-parameter curves of $\vec{X}$ are geodesics. QED

The polar coordinate parametrization of the Poincaré disk, $\vec{X}(u,v)=(u\cos v, u\sin v)$, $0<u<2$, is a Clairaut parametrization , by Definition 2.2. Therefore the u-parameter curves, Euclidean lines through the origin, are geodesics of the disk.

<u>2.4 Theorem.</u> A curve $\beta(u)=\vec{X}(u,v(u))$, where $\vec{X}$ is a Clairaut parametrization, is a geodesic if and only if

$$\frac{dv}{du} = \frac{\pm c\sqrt{E}}{\sqrt{G}\sqrt{G-c^2}} .$$

The constant c is then the slant of β, i.e., in combination with G it determines the angle τ at which the geodesic β is cutting across the u-parameter streamlines of $\vec{X}$. (See O'Neill,[10])



<u>FIGURE 16</u>

Since $\vec{x}$ is a Clairaut parametrization, E and G depend only on u. Therefore the formula for $\frac{dv}{du}$ depends only on u. By the fundamental theorem of calculus, it can be written in integral form as

$$v(u)=v(u_0) \pm \int_{u_0}^{u} \frac{c\sqrt{E}du}{\sqrt{G}\sqrt{G-c^2}}.$$

Therefore, in our given parametrization of the Poincaré disk, $\beta(u) = \vec{x}(u, v(u))$ is a geodesic provided

$$\frac{dv}{du} = \frac{\pm c\sqrt{E}}{\sqrt{G}\sqrt{G-c^2}} = \pm \frac{(cg/u^2)}{\sqrt{1-(cg/u)^2}} \text{ where } g = 1- \frac{u^2}{4}.$$

To carry out the required the integration, set

$$w = \frac{a}{u}\left(1 + \frac{u^2}{4}\right), \text{ where } a = \frac{c}{\sqrt{1+c^2}} .$$

It follows that $\frac{dv}{du} = \pm \frac{dw/du}{\sqrt{1-w^2}}$. Therefore $v-v_0=\pm\cos^{-1} w$, or

$\cos(v-v_0) = w = \frac{a}{u}(1+\frac{u^2}{4})$. Also

$$(*) \quad u^2 + 4 - \frac{4u}{a}\cos(v-v_0) = 0.$$

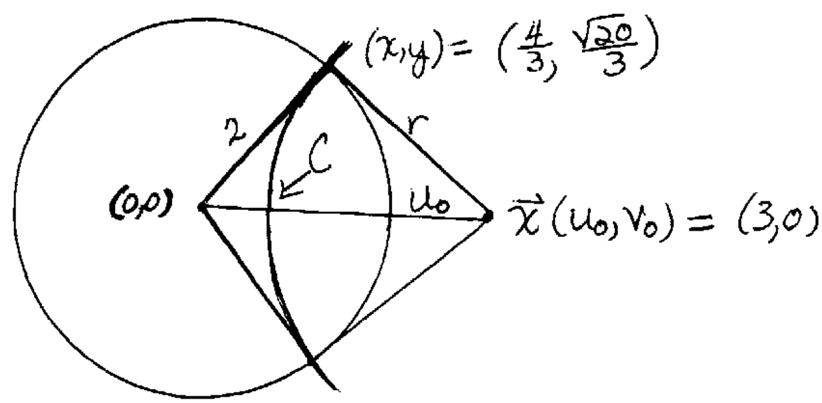


FIGURE 17

Using the law of cosines in the diagram yields the polar equation of a circle of radius r, centered at $\vec{x}(u_0,v_0)$:

$$u^2+u_o^2-2u_o u\cos(v-v_o)=r^2.$$

Comparing this equation with (*) above shows that β is a Euclidean circle with $u_o^2-r^2=4$. Since $u_o > 2$, the center of the circle lies outside the Poincaré disk: $x^2+y^2<4$. Orthographic projection is a conformal mapping that takes the geodesics in the Beltrami disk, Euclidean line segments, to semicircles on the sphere that meet the equator at right angles. Then stereographic projection, which is a conformal mapping, takes the equator to the rim of the Poincaré disk. It follows that the semicircles on the sphere that meet the equator at right angles are taken to arcs on the disk that are orthogonal to the rim of the Poincaré disk. Therefore the Euclidean circle C is orthogonal to the rim of the disk.

If follows that the geodesics of the Poincaré disk $x^2 + y^2 < 4$ are all Euclidean lines through the origin, and all Euclidean circles orthogonal to the rim of the disk.

The geodesics of the Poincaré disk should be compared to those of the Euclidean plane. Around 300 B.C. Euclid established a set of axioms for the straight lines of his plane. The goal was to derive its geometry from axioms so reasonable as to be "self-evident". The most famous of these is equivalent to the parallel postulate:

> If P is a point not on a line α, then there is a unique line β through P which does not meet α.

Over time, this postulate began to seem less self-evident than the others. The axiom that two points determine a unique straight line can be checked by laying down a long but finite straight edge touching both points. But to check the parallel postulate, also called Euclid's V, one would have to travel the whole infinite length of ß to be sure it never touches α.

Much effort was given to trying to deduce the parallel postulate from the other axioms. The Poincaré disk offers convincing proof that this cannot be done.
If we replace "line" by "route of geodesic", then every Euclidean axiom holds in the Poincaré disk except the parallel postulate.

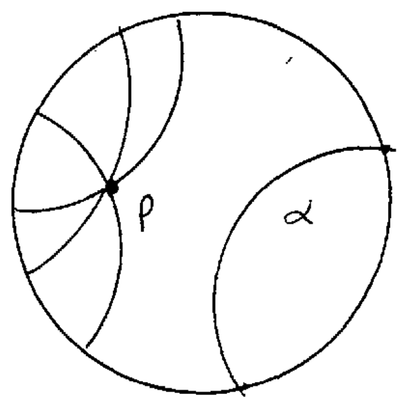


FIGURE 18

Given two points one and only one geodesic route runs through them. But, as Figure 18 shows, in the Poincaré disk there are always an infinite number of geodesic routes through P that do not meet α. When the implications of this discovery were worked out, the hope of deducing the parallel postulate was destroyed. The whole idea that $\mathbf{R}^2$ is, in some sense, an Absolute whose properties are "self-evident" was also destroyed. $\mathbb{R}^2$ became but one geometric surface among infinitely many others discovered by Riemann.

3. Poincaré Half-Plane

A. Geodesics and Curvature

In 1882, Poincaré used the methods of stereographic and orthographic projection to provide another conformal model of the non-Euclidean plane, the Poincaré half-plane. The Poincaré half-plane can be constructed by projecting the Beltrami disk orthographically to the lower hemisphere of the sphere of radius one centered at (0,0,1). This orthographic projection is a mapping $D \subset \mathbb{R}^2$ to the lower hemisphere of S^2 for which all lines of projection are perpendicular to the plane containing D_B (Beltrami disk). This is accomplished by the mapping

$$(u,v) \mapsto (u,v,1-\sqrt{1-u^2-v^2}).$$

The sphere is then rotated around the axis through the center parallel to the x-axis through a right angle to move the lower hemisphere to the half-space y>0. The lower hemisphere is now the "right hemisphere." This is represented by the mapping

$$(u,v,1-\sqrt{1-u^2-v^2}) \mapsto (u,\sqrt{1-u^2-v^2},v+1).$$

Now stereographically project from the north pole. This takes the new "right hemisphere" of the sphere to the upper half-plane in $\mathbb{R}^2$ given by $H = \{(u,v) | v>0\}$. This stereographic projection is represented by the mapping

$$(u,\sqrt{1-u^2-v^2},v+1) \mapsto \left(\frac{2u}{2-(v+1)}, \frac{2\sqrt{1-u^2-v^2}}{2-(v+1)}\right)$$

$$= \left(\frac{2u}{1-v}, \frac{2\sqrt{1-u^2-v^2}}{1-v}\right).$$

The composite mapping of the Beltrami disk into the Poincaré half-plane is

$$(u,v) \mapsto \left(\frac{2u}{1-v}, \frac{2\sqrt{1-u^2-v^2}}{1-v} \right).$$

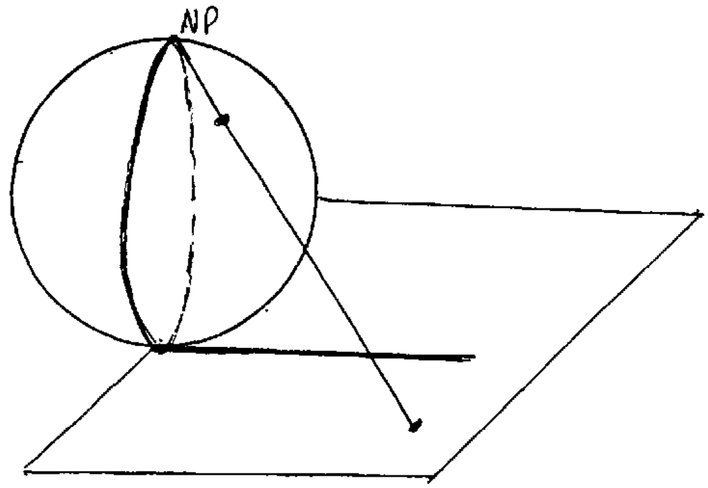


FIGURE 19

To find the metric on the half-plane induced by this mapping let

$$w=\sqrt{1-u^2-v^2},\quad dw = \frac{1}{2}\frac{-2udu-2vdv}{w},\quad -wdw=udu+vdv,$$

$$x = \frac{2u}{1-v},\quad y = \frac{2w}{1-v},\quad dx = \frac{2(1-v)du+2udv}{(1-v)^2},\ \text{and}$$

$$dy = \frac{2(1-v)dw+2wdv}{(1-v)^2}.$$

The Beltrami metric can again be expressed as

$$ds_B^2 = \frac{1}{w^2}(du^2+dv^2+dw^2)\ \text{and}$$

$$dx^2 + dy^2 = \frac{4}{(1-v)^2}(du^2+dv^2+dw^2) = \frac{4w^2}{(1-v)^2}ds^2=y^2ds^2.$$

This equation yields the induced metric (called the Poincaré metric)

$$ds_p^2 = \frac{dx^2+dy^2}{y^2}.$$

The Poincaré half-plane represents the abstract surface with the given induced Riemannian metric.

3.1 Proposition Let h be a Euclidean circle with center C(c,0) and radius r. If P and Q are points of h such that the radii CP and CQ make angles α and β ($\alpha<\beta$), respectively, with the positive x-axis, then the hyperbolic length of arc PQ = $\ln \dfrac{\csc\beta-\cot\beta}{\csc\alpha-\cot\alpha}$.

Proof: Let t be the angle from the positive x-axis to the radius through an arbitrary point (x,y) on h, then

$x=c+r\cos t$, $y=r\sin t$, $dx=-r\sin t\,dt$, and $dy=r\cos t\,dt$.

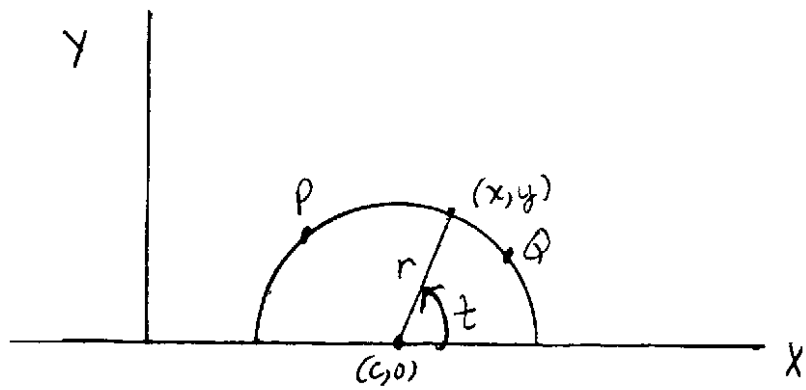


FIGURE 20

Hyperbolic length of arc PQ =

$$\int\frac{\sqrt{dx^2+dy^2}}{y} = \int\frac{\sqrt{(-r\sin t\,dt)^2+(r\cos t\,dt)^2}}{r\sin t}$$

$$= \int\frac{r\,dt}{r\sin t} = \int_\alpha^\beta \csc t\,dt=\ln\frac{\csc\beta-\cot\beta}{\csc\alpha-\cot\alpha}.\ \text{QED}$$

3.2 Proposition. The hyperbolic length of the Euclidean line segment joining the points $P(a,y_1)$ and $Q(a,y_2)$, $0<y_1\le y_2$, is $\ln \dfrac{y_2}{y_1}$.

Proof: x=a constant and dx=0. The hyperbolic length of PQ =

$$\int \frac{\sqrt{dx^2+dy^2}}{y} = \int_{y_1}^{y_2} \frac{dy}{y} = \ln \frac{y_2}{y_1} \text{ . QED}$$

A geodesic segment in the Poincaré half-plane is a curve whose hyperbolic length is the shortest among all the curves that join a given pair of points in the hyperbolic plane.

<u>3.3 Theorem</u> The geodesic segments of the Poincaré half-plane are either

a) segments of Euclidean straight lines that are perpendicular to the x-axis or

b) arcs of Euclidean semicircles that are centered on the x-axis.

Proof: Let $P(x_1,y_1)$ and $Q(x_2,y_2)$ be any two points of the Poincaré half-plane, and let h be a curve joining them.

a) Case I: $x_1 = x_2$

In this case, the Euclidean line segment PQ is perpendicular to the x-axis. Let the geodesic segment h have equation $x=f(y)$ and dx/dy is given by f' so that $dx=f'dy$. The hyperbolic length of h is

$$\int \frac{\sqrt{(f')^2dy^2+dy^2}}{y} = \int_{y_1}^{y_2} \frac{\sqrt{(f')^2+1}}{y}\,dy$$

$$\geq \int_{y_1}^{y_2} \frac{dy}{y} = \ln \frac{y_2}{y_1}.$$

In Proposition 3.2, $\ln \frac{y_2}{y_1}$ was shown to be the hyperbolic length of the Euclidean line segment PQ joining points $P(a,y_1)$ and $Q(a,y_2)$. This implies that segments of Euclidean straight lines perpendicular to the x-axis are geodesics of the Poincaré half-plane.

b) Case II: $x_1 \neq x_2$

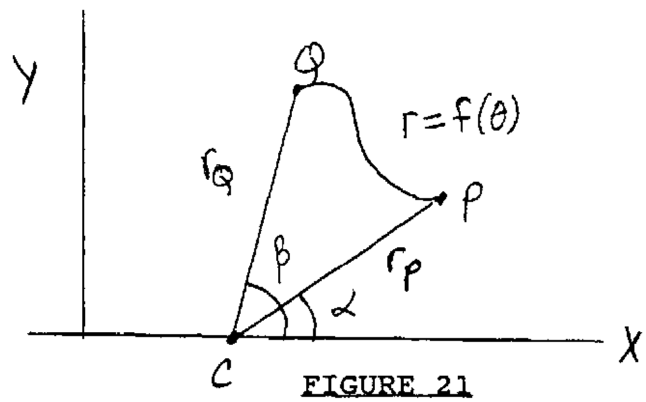


FIGURE 21

In this case, the Euclidean line segment PQ is not perpendicular to the x-axis. Let C(c,0) be the x-intercept of the perpendicular bisector to PQ. Place a polar coordinate system so that its origin coincides with C and its initial ray points in the same direction as the positive x-axis. Let the geodesic segment h be a part of the curve whose equation is $r=f(\theta)$. Let the coordinates of P and Q be (r_p,α) and (r_Q,β), respectively. The hyperbolic length of h is

$$\int_h = \frac{\sqrt{dx^2+dy^2}}{y} .$$

Relate these polar coordinates to the defining Cartesian coordinates by the equations:

$$x = c + r\cos\theta, \quad y = r\sin\theta.$$

It follows that

$$\frac{dx}{d\theta}=\frac{dr}{d\theta}\cos\theta+r\frac{d\cos\theta}{d\theta}=r'\cos\theta-r\sin\theta \quad \text{and}$$

$$\frac{dy}{d\theta}=\frac{dr}{d\theta}\sin\theta+r\frac{d\sin\theta}{d\theta}=r'\sin\theta+r\cos\theta \ .$$

$dx^2 + dy^2 =$

$$(r'\cos\theta - r\sin\theta)^2 \, d\theta^2 + (r'\sin\theta + r\cos\theta)^2 \, d\theta^2$$

$$= (r'^2 + r^2) \ d\theta^2.$$

It follows that

$$\int_h \frac{\sqrt{dx^2+dy^2}}{y} = \int_h \frac{\sqrt{(r')^2+r^2}}{r\sin\theta} \ d\theta$$

$$\geq \int_h \frac{\sqrt{r^2}}{r\sin\theta} \ d\theta = \int_\alpha^\beta \csc\theta \ d\theta = \ln \frac{\csc\beta-\cot\beta}{\csc\alpha-\cot\alpha} \ .$$

In Proposition 3.1 this expression was shown to be the hyperbolic length of the arc of the Euclidean circle with center C(c,0) and radii CP=CQ. It follows that this arc is the geodesic segment joining P and Q and that arcs of Euclidean semicircles with centers on the x-axis are geodesic segments of the Poincaré half-plane. QED

3.4 Example . Find the hyperbolic length of the geodesic segment joining the points A(8,4) and B(0,8).

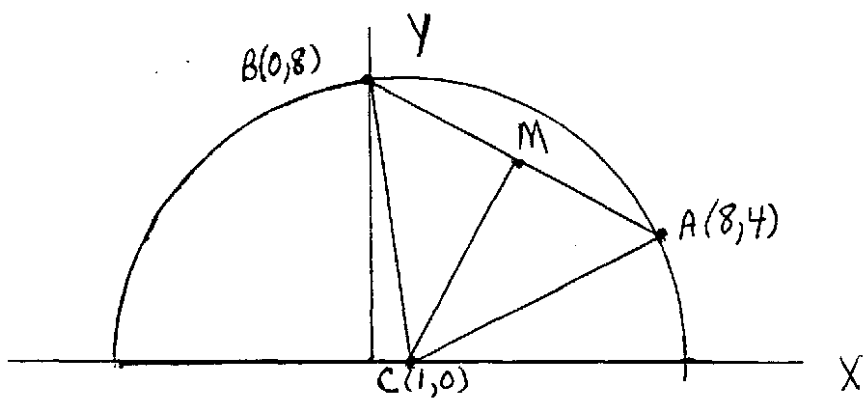


FIGURE 22

Find the intersection of the perpendicular bisector to the Euclidean line segment AB with the x-axis. This gives point C which is the center of the geodesic joining A and B. The Euclidean straight line joining A and B has slope $\frac{4-8}{8-0}=\frac{-1}{2}$ and the midpoint M of the Euclidean line segment AB has coordinates $(\frac{0+8}{2},\frac{8+4}{2})=(4,6)$. Therefore the perpendicular bisector has equation $y-6=2(x-4)$ and the x-intercept is $(1,0)$. $CA=CB=\sqrt{(8-1)^2+(4-0)^2}=\sqrt{65}$. It follows that the hyperbolic length of the geodesic segment from A to B is

$$\ln\frac{\csc\beta-\cot\beta}{\csc\alpha-\cot\alpha} = \ln\frac{\frac{\sqrt{65}}{8}-(\frac{-1}{8})}{\frac{\sqrt{65}}{4}-\frac{7}{4}} \approx 1.45.$$

Consider the Poincaré metric $\frac{dx^2+dy^2}{y^2}$. Using Stahl's formula for Gaussian curvature yields the following:

$E=\frac{1}{y^2}$, $F=0$, $G=\frac{1}{y^2}$, $E_1=0$, $F_1=0$, $G_1=0$, $E_2=\frac{-2}{y^3}$,

$F_2=0$, $G_2=\frac{-2}{y^3}$, $E_{22}=\frac{6}{y^4}$, $F_{12}=0$, $G_{11}=0$;

$$4(\frac{1}{y^4}-0)^2K = \frac{1}{y^2}(\frac{4}{y^6})+\frac{1}{y^2}(\frac{4}{y^6})-2(\frac{1}{y^4})(\frac{6}{y^4}));$$

$\frac{4K}{y^8}=\frac{-4}{y^8}$; $4K=-4$; and $K=-1$.

Therefore the Poincaré half-plane with the given Poincaré metric has constant curvature $K=-1$. By Corollary 6.15, there are no conjugate points an any geodesic in a surface with Gaussian curvature $K\leq 0$. Hence every geodesic segment on such a surface is locally minimizing and the circumference of the polar geodesic circle will be longer than $2\pi\epsilon$ for a given $\epsilon>0$, ϵ small.

B. Relationship Between Euclidean Circles and Hyperbolic Circles

Given any point C, a positive real number r, and any ray (half-geodesic) h emanating from C, there is a point P_h on h that is at a hyperbolic distance of r from C. The locus of all such points P_h is the hyperbolic circle with center C and hyperbolic radius r.

<u>3.5 Proposition</u> If a Euclidean circle has Euclidean center (h,k) and a Euclidean radius r, then it has the hyperbolic center (H,K), and the hyperbolic radius R, with

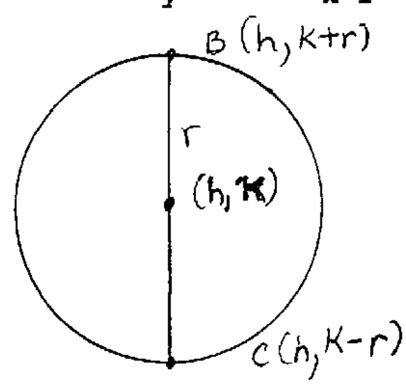
$$H=h,\ K=\sqrt{k^2-r^2},\ R=\frac{1}{2}\ln\frac{k+r}{k-r}\ \text{, and}$$

$$h=H,\ k=K\cosh(R),\ r=K\sinh(R).$$

Proof: Let B and C be, respectively, the points of the circle that lie directly above and below (h,k). Their coordinates are B(h,k+r) and C(h,k-r).

The hyperbolic length of BC (dx=0)

$$= \int \frac{\sqrt{dx^2+dy^2}}{y} = \int_{k-r}^{k+r}\frac{dy}{y} = \ln\frac{k+r}{k-r}\ .$$



<u>FIGURE 23</u>

This is the hyperbolic diameter of the circle and the hyperbolic radius R = $\frac{1}{2}\ln\frac{k+r}{k-r}$. H=h since x=h on the Euclidean line segment perpendicular to the x-axis at h. (H,K) should be the hyperbolic midpoint of segment BC. By Proposition 3.2 of chapter 2, the hyperbolic length of the Euclidean line segment joining points (h,k+r) and (h,K)=(H,K) is $\ln\frac{k+r}{K}$ and the hyperbolic length of the Euclidean line segment joining the points (h,K)=(H,K) and (h,k-r) is $\ln\frac{K}{k-r}$. For (H,K)=(h,K) to be the hyperbolic midpoint of segment BC, then

$$\ln\frac{k+r}{K}=\ln\frac{K}{k-r} \quad \text{and} \quad K^2=k^2-r^2, K=\sqrt{k^2-r^2} \ .$$

When R= $\frac{1}{2}\ln\frac{k+r}{k-r}$ is inverted it yields

$$\frac{k}{r}=\frac{e^{2R}+1}{e^{2R}-1}=\coth(R).$$

Let $K^2=k^2-r^2$ so that $k^2=K^2+r^2$ and solve the above equations simultaneously, then

$$r^2=\frac{K^2}{\coth^2(R)-1}=K^2\sinh^2(R) \quad \text{and} \quad r=K\sinh(R).$$

$$k^2=K^2+r^2=K^2+K^2\sinh^2(R)=K^2(1+\sinh^2(R))=K^2\cosh^2(R)$$

and k=Kcosh(R). QED

<u>3.6 Corollary</u> Every Euclidean circle in the upper half-plane is a hyperbolic circle. Every hyperbolic circle is also a Euclidean circle.

C. Rigid Motions

<u>3.7 Definition</u> A mapping $f:\mathbf{R}^n\to\mathbf{R}^n$ is an <u>isometry</u> if, for all $\vec{v},\vec{w},\in\mathbf{R}^n$, $\|f(\vec{v})-f(\vec{w})\|=\|\vec{v}-\vec{w}\|$.

The notion of isometry makes precise the ideas of rigid motions and congruence. We define a congruence to be a self-isometry of a surface, $\phi: S \to S$. Two figures, that is subsets of S, are congruent if there is an isometry with $\phi(\text{figure}_1) = \text{figure}_2$. A figure made up of segments of curves on a surface may be thought of as rods in a configuration and the term rigid motion is synonymous with congruence. Therefore, a rigid motion is considered to be a transformation of a surface that does not distort its configurations.

In Euclidean plane geometry, two triangles are congruent if there is a rigid motion of the plane which carries one triangle exactly onto the other. For these congruent triangles, corresponding angles and corresponding sides are congruent and the areas enclosed are equal. Any geometric property of a given triangle is shared by every congruent triangle.

<u>3.8 Example</u> A translation of $\mathbf{R}^3$ is an isometry. Fix a point A in $\mathbf{R}^3$ and let T be the mapping that adds A to every point of $\mathbf{R}^3$. Therefore, T(P)=P+A for all points P. T is called translation by A. T is an isometry since

$$\|T(P)-T(Q)\|=\|(P+A)-(Q+A)\|=\|P-Q\|.$$

<u>3.9 Example</u> In $\mathbf{R}^2$ the rotations given by

$$\tau_\theta(\vec{v}) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

in the standard basis are isometries.

$$\|\tau_\theta\vec{v}\|^2=\|(\cos\theta v_1-\sin\theta v_2,\sin\theta v_1+\cos\theta v_2)\|^2$$
$$=\cos^2\theta v_1^2-2\cos\theta\sin\theta v_1v_2+\sin^2\theta v_2^2+\sin^2\theta v_1^2+2\cos\theta\sin\theta v_1v_2+\cos^2\theta v_2^2$$
$$= v_1^2+v_2^2=\|\vec{v}\|^2.$$

3.10 Theorem If F and G are isometries of $\mathbb{R}^n$, then the composite mapping GF is also an isometry of $\mathbb{R}^n$.

Proof: Since G is an isometry, $\|G(F(P))-G(F(Q))\| = \|F(P)-F(Q)\|$. Since F is an isometry $\|F(P)-F(Q)\| = \|P-Q\|$. Therefore GF is also an isometry. QED

The rigid motions (isometries) of Euclidean and hyperbolic space have formulations in terms of complex numbers. The Poincaré metric brought to light the role played by complex numbers in this and other geometries. Many mathematicians have come to think of the points of the Poincaré half-plane as complex numbers $z=x+yi$ with a positive imaginary part.

If c is any fixed complex number, then the function $f(z)=z+c$ is a translation of the Euclidean plane and every translation of the plane is expressible in this manner. For any angle α and any complex number c, the function

$$f(z)=e^{i\alpha}(z-c)+c = e^{i\alpha}z + (1-e^{i\alpha})c$$

is the rotation about c, $R_{c,\alpha}$ where $e^{i\alpha} = \cos\alpha + i\sin\alpha$. The 90° counterclockwise rotation of the plane about the point (0,1) has the expression

$$R_{i,\pi/2} = e^{i(\pi/2)}(z-i)+i = i(z-i)+i = iz+1+i$$

in terms of complex numbers. If m is any line with inclination α to the positive x-axis, and c is a point on m, then the function $f(z)=e^{2i\alpha}\,\overline{z}-\overline{c}+c$ is the reflection in the line m where $\overline{z}=x-yi$ is the conjugate of $z=x+yi$.

3.11 Theorem The rigid motions of the Euclidean plane all have the form

$$f(z)=e^{i\alpha}z+c \text{ or } f(z)=e^{i\alpha}\overline{z}+c$$

where α is an arbitrary real number and c is an arbitrary complex number. Conversely, every function of either of these forms is a rigid motion of the Euclidean plane. (See Stahl,[13])

This leads us to the complex description of the rigid motions of the hyperbolic half-plane. Horizontal translations and reflections in vertical lines, which are geodesics, are also Euclidean rigid motions. They can be expressed as $f(z)=z+r$ or $f(z)=-\overline{z}+r$, respectively, where r is an arbitrary real number. A hyperbolic reflection is either a Euclidean reflection in vertical lines, which are geodesics, or an inversion centered at some point on the x-axis.

3.12 Theorem The following transformations of the hyperbolic plane preserve both hyperbolic lengths and measures of angles:

a) inversions $I_{c,k}$ where C is the center of a circle on the x-axis and k is the radius;

b) reflections σ_m where m is perpendicular to the x-axis;

c) translations τ_{AB} where AB is a Euclidean segment parallel to the x-axis.

Proof: Inversions of the form $I_{c,k}$ are conformal transformations of the plane. Therefore Euclidean measures of angles are preserved. The hyperbolic measure of an angle coincides with its Euclidean measure, and inversions of the form $I_{c,k}$ preserve hyperbolic measures of angles. Place a polar coordinate system with its origin at C and its initial ray along the positive x-axis. Let h denote the curve $r=f(\theta)$, $\alpha\le\theta\le\beta$. The inversion $I_{c,k}$ maps h to h' given by $r=F(\theta)=\dfrac{k^2}{f(\theta)}$, $\alpha\le\theta\le\beta$.

As in the proof of Theorem 3.3, the hyperbolic length of h' is

$$\int_{h'}\frac{\sqrt{(r')^2+r^2}}{r\sin\theta}\,d\theta = \int_\alpha^\beta\frac{\sqrt{(F')^2+F^2}}{F\sin\theta}\,d\theta =$$

$$\int_\alpha^\beta\frac{\sqrt{\left(\frac{-k^2f'}{f^2}\right)^2+\left(\frac{k^2}{f}\right)^2}}{k^2\sin\frac{\theta}{f}}\,d\theta = \int_\alpha^\beta\frac{\sqrt{(f')^2+f^2}}{f\sin\theta}\,d\theta$$

$$=\int_h\frac{\sqrt{(r')^2+r^2}}{r\sin\theta}\,d\theta = \text{hyperbolic length of h.}$$

It follows that the given inversion does preserve hyperbolic lengths and hyperbolic measures of angles. For the proofs of parts b & c, see Stahl,[13]. QED

$I_{c,k}$ represents an inversion whose fixed points are exactly those that makeup the circle centered at C with radius k. Euclidean reflections are defined so that their axes are perpendicular bisectors of the line segment joining any point to its image.

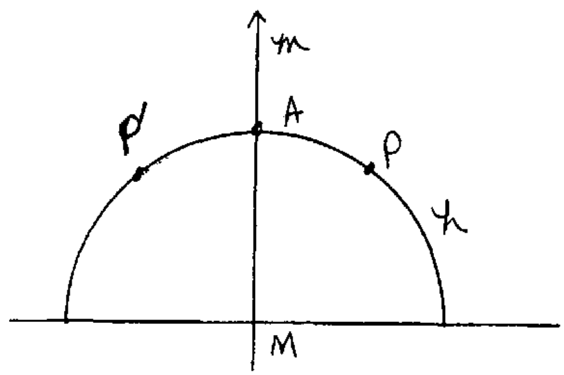


FIGURE 24

Consider the reflection σ_m whose axis is the vertical line m, a geodesic, which is above the point M on the x-axis. If P is any point of the upper half-plane, let h be the geodesic through P that is an arc of a semicircle centered at M. Since h and m are orthogonal at their intersection A, it follows that $\sigma_m(h)=h$ and so $P'=\sigma_m(P)$ is also on h. σ_m is a hyperbolic rigid motion (as well as a Euclidean rigid motion), and so the geodesic segments PA and P'A have equal hyperbolic lengths. Therefore, m is the hyperbolic perpendicular bisector of the geodesic segment h joining P to its image $P'=\sigma_m(P)$.

Next look at the hyperbolic reflection that consists of the inversion $I_{c,k}$ where C is some point on the x-axis. This inversion fixes every point on the geodesic n through B which is an arc of a semicircle centered at C with radius k. Therefore think of the geodesic n as the axis of $I_{c,k}$.

For any point Q between the x-axis and the geodesic n, let $Q'=I_{c,k}(Q)$ and let w be the geodesic that contains Q and is orthogonal to n at their intersection B.

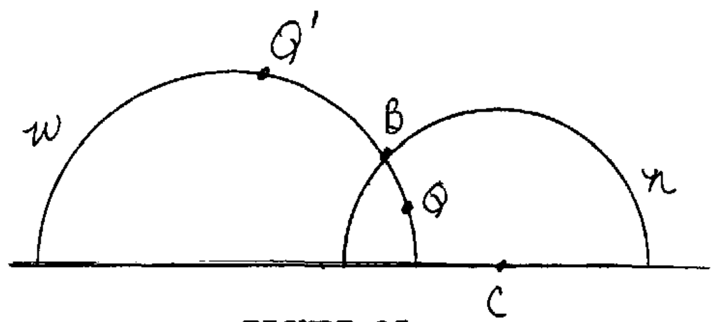


FIGURE 25

It follows that $I_{c,k}(w)=w$ and so Q' too is on w. $I_{c,k}$ is a hyperbolic rigid motion and so the geodesic segments QB and Q'B have equal hyperbolic lengths. Thus n is the hyperbolic perpendicular bisector of the geodesic segment joining Q to its image $Q'=I_{c,k}(Q)$. Therefore hyperbolic reflections satisfy the definition of reflections for Euclidean motions.

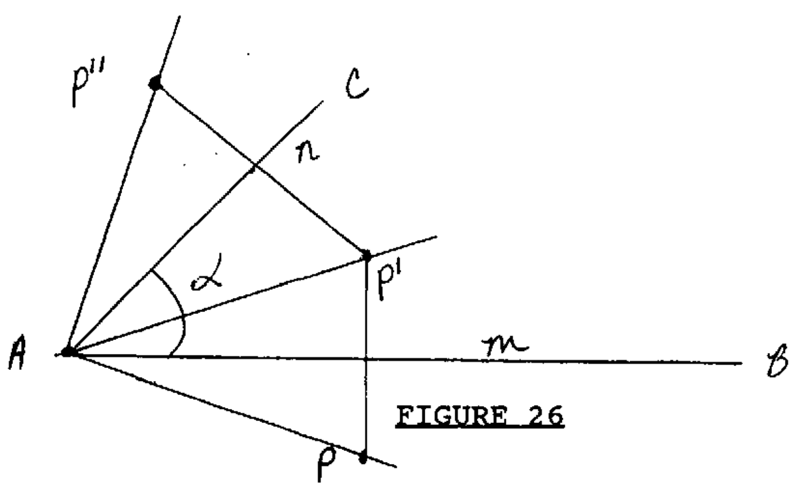


FIGURE 26

3.13 Proposition Let m and n be two Euclidean lines (geodesics) that intersect at a point A, and let α be the counterclockwise angle from m to n at A. Then $\sigma_n \circ \sigma_m = R_{A,2\alpha}$. (See Stahl,[13], and Figure 26)

Since hyperbolic reflections satisfy the definition of reflections for Euclidean motions, this proposition implies that the composition of two hyperbolic reflections with intersecting axes is a hyperbolic rotation. The center of this rotation is the intersection of the axes and the angle of the rotation is twice the angle between the axes.

3.14 Theorem Every Euclidean rigid motion is the composition of at most three reflections. (See Stahl,[13])

Since hyperbolic reflections satisfy the definition of reflections for Euclidean motions, the following theorem is implied.

3.15 Theorem Every hyperbolic rigid motion is the composition of at most three hyperbolic reflections.

3.16 Example In the geodesic which is the arc of a semicircle with center A=(a,0) on the x-axis and radius k, the inversion $I_{A,k(z)}$ is given by

$$I_{A,k(z)} = \frac{k^2}{\overline{z}-a}+a \quad .$$

If A=(0,0), then $I_{0,k(z)}= \frac{k^2}{\overline{z}}$. The inversion $I_{0,2}$ has the expression $\frac{2^2}{\overline{z}}$ and so it maps the point 1+i to the point $\frac{2^2}{\overline{1+i}}=\frac{4}{1-i}=$ 2+2i. (See Figure 27)

If A=(3,0), then $I_{A,4}$ has the expression $\frac{4^2}{\overline{z}-3}+3 = \frac{3\overline{z}+7}{\overline{z}-3}$ and so it maps 1+i to the point $\frac{-17+16i}{5}$. (See Figure

27) The inversions given are hyperbolic reflections whose axes intersect. Therefore the composition $R=I_{A,4} \circ I_{0,2}$ is a hyperbolic rotation.

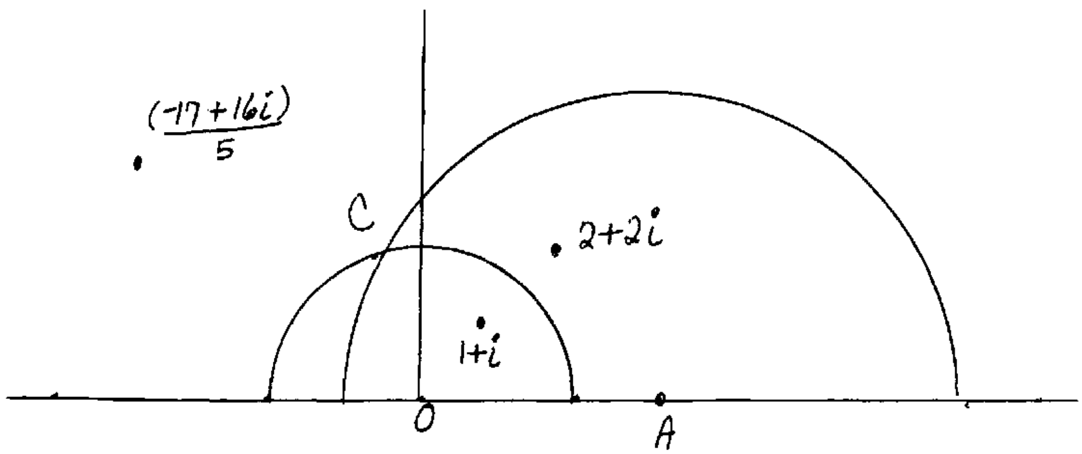


FIGURE 27

3.17 Theorem The rigid motions of the hyperbolic plane coincide with the complex functions that have the following forms:

$$\text{i) } f(z) = \frac{\alpha z+\beta}{\gamma z+\delta} \quad \text{or ii) } f(z) = \frac{\alpha(-\bar{z})+\beta}{\gamma(-\bar{z})+\delta}$$

where α, β, γ, δ are real numbers and $\alpha\delta-\beta\gamma>0$.

Proof: Horizontal translations have the form $\frac{1z+r}{0z+1}$, and the reflections in vertical lines (geodesics), have the form $\frac{1(-\bar{z})+r}{0(-\bar{z})+1}$. Reflections in geodesic arcs have the form

$$\frac{k^2}{\bar{z}-a}+a=\frac{a\bar{z}+k^2-a^2}{\bar{z}-a}=\frac{-a(-\bar{z})+(k^2-a^2)}{-(-\bar{z})-a}$$

with $\alpha\delta-\beta\gamma>0$. If f and g are two functions that have this format, then so does their composition. Since every hyperbolic rigid motion is the composition of some hyperbolic reflections, it follows that all hyperbolic rigid motions have either form i) or form ii). QED

Transformations of the type $f(z)=\frac{az+b}{cz+d}$ where a,b,c,d are allowed to be any complex numbers as long as $ad-bc\neq 0$, are called Moebius transformations. If $z_0=a+bi$ is any point of the half-plane, then the Moebius transformation $f(z)=\frac{z-a}{b}$ is a hyperbolic translation that carries z_0 to i, and its inverse $f^{-1}(z)=bz+a$ carries i to z_0. Therefore, given any other point $z_1=c+di$ of the half-plane, the composition

$$d\left(\frac{z-a}{b}\right)+c = \frac{dz+(bc-ad)}{0z+b}$$

is a hyperbolic rigid motion that carries z_0 onto z_1. If b=d, then this is a horizontal Euclidean translation. If $z_0=2-3i$ and $z_1=3+4i$, then $\frac{4z+(-3\cdot 3-2\cdot 4)}{-3}=\frac{4z+17}{-3}$ is a hyperbolic rigid motion that carries z_0 onto z_1. Hyperbolic rotations are characterized by a single fixed point in the half-plane. To find the rotation, find the center and locate the image of one other point.

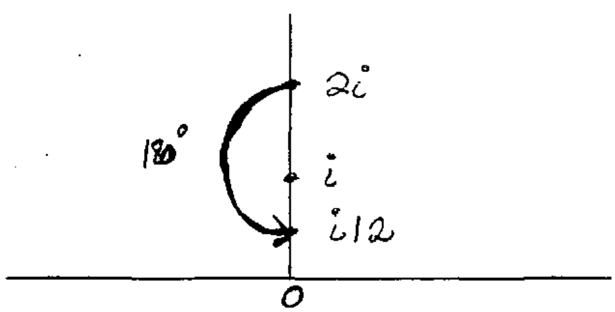


FIGURE 28

3.18 Example Let $f(z)=\frac{-1}{z}=\frac{0z-1}{1z+0}$. Let z=i. $f(i)=\frac{-1}{i}=i$ and this rotation is pivoted at i. $f(2i)=\frac{-1}{2i}=\frac{i}{2}$ and it follows that the angle of rotation is 180°. Therefore f(z) is a Moebius transformation that is the hyperbolic rotation by 180° about the point i. (See Figure 28)

D. Geodesic Triangles

3.19 Definition A hyperbolic triangle ABC consists of three points A,B,C (vertices) that do not lie on a single geodesic and the three geodesic segments (sides) that join each pair of vertices. A hyperbolic triangle is said to be in standard position if the vertices A,B,C have coordinates (0,k), (s,t), (0,1), respectively, where k>1 and s>0.

3.20 Proposition Every hyperbolic triangle can be brought into standard position by a hyperbolic rigid motion.

Proof: Let the hyperbolic triangle ABC have its vertices C at (0,1) and B at (s,t) with s>0, but A is at (0,k) with k<1. Let z=ci with c<1. The reflection $I_{0,1}$- $\frac{k^2}{\bar{z}}=\frac{1}{\overline{ci}}=\frac{1}{-ci}=\frac{i}{c}>1$ will transform triangle ABC into triangle A'B'C' that is in standard position. (See Figure 29)

If the hyperbolic triangle ABC has both of its vertices A(0,a) and C(0,c) on the y-axis, then by reflecting this triangle in the y-axis, if necessary, then B(s,t) has s>0. The reflection $I_{0,\sqrt{c}}$ transforms triangle ABC into a triangle in standard position where $I_{0,\sqrt{c}}=\frac{c}{\overline{ci}}=\frac{c}{-ci}=\frac{ci}{c}=i$. (See Figure 29)

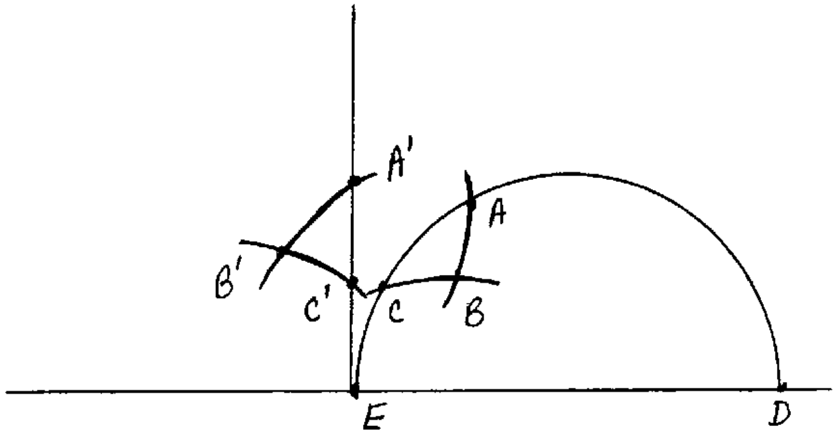


FIGURE 29

If the hyperbolic triangle ABC is in an arbitrary position, assume that its side AC is a segment of a geodesic arc g that joins two points D and E on the x-axis. Horizontal translations are hyperbolic rigid motions. This allows us to assume that E is at the origin. If D=(d,0) then the inversion $I_{D,d}$ transforms g onto the y-axis where

$$I_{D,d} = \frac{k^2}{\overline{z}-D}+D = \frac{d^2}{\overline{z}-D}+D \ .$$

The given triangle is transformed into triangle A'B'C' which has one of its sides on the y-axis. By the previous steps in this proof, triangle A'B'C' can be brought into standard position. (See Figure 29) QED

3.21 Proposition At any given point P, let g be a vertical geodesic ray and let g_1, g_2, g_3 be geodesic arcs centered at C_1, C_2, C_3 respectively. Then

$\angle(g,g_1)=\angle DC_1P$, $\angle(g_1,g_2)=\angle C_1PC_2$, and $\angle(g_3,g_1)=\Pi-C_1PC_3$.

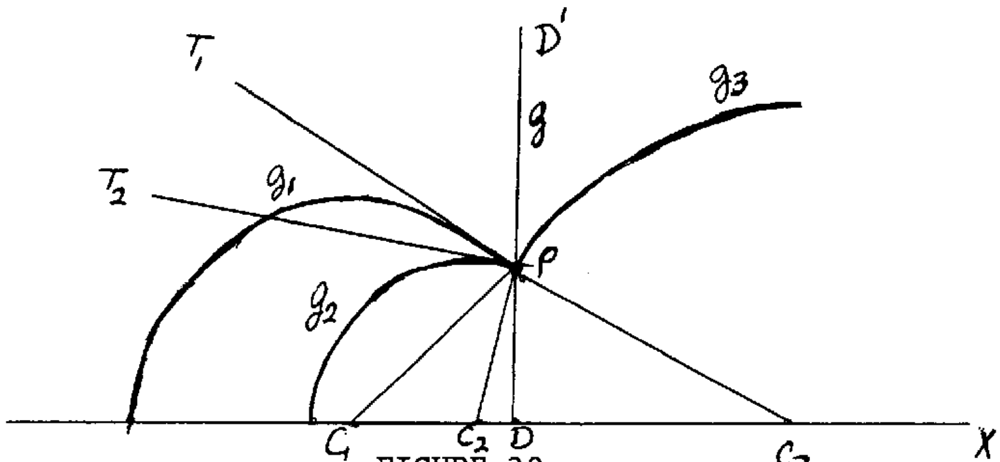


FIGURE 30

Proof: Let D' be a point of g such that P is between D and D'. Let PT_1 and PT_2 be the Euclidean tangent lines to g_1 and g_2, respectively, at P. Since the tangent to a circle is perpendicular to the radius at the point of contact, then

$$\beta = \angle(g_1,g_2) = \angle T_1PT_2 = \angle T_1PC_1 - \angle T_2PC_1$$

$$= \frac{\pi}{2} - \angle T_2PC_1 = \angle T_2PC_2 - \angle T_2PC_1 = \angle C_1PC_2;$$

$$\alpha = \angle(g,g_1) = \angle D'PT_1 = \pi - \angle T_1PC_1 - \angle C_1PD$$

$$= \frac{\pi}{2} - \angle C_1PD = \angle DC_1P;$$

$$\gamma = \angle(g_3,g_1) = \angle(g_3,g) + \angle(g,g_1) = \angle PC_3D + \angle DC_1P$$

$$= \pi - \angle C_1PC_3. \text{ QED}$$

The Poincaré half-plane has constant curvature $K=-1$. By the Gauss-Bonnet Theorem and its corollaries, the interior angle sum of a triangle in the Poincaré half-plane is less than π. This is the only constraint on the angles of the hyperbolic triangle. The next theorem's proof contains a blueprint for the construction of hyperbolic triangles with specified angles such that $\alpha+\beta+\gamma<\pi$.

3.22 Theorem Given any three angles whose sum is less than π, they are indeed the angles of some hyperbolic triangle.

Proof: Let α, β, γ be three arbitrary positive angles such that $\alpha+\beta+\gamma<\pi$. Let triangle ABC be a hyperbolic triangle in standard position such that $\angle CAB=\alpha$, $\angle ABC=\beta$, and $\angle BCA=\gamma$.

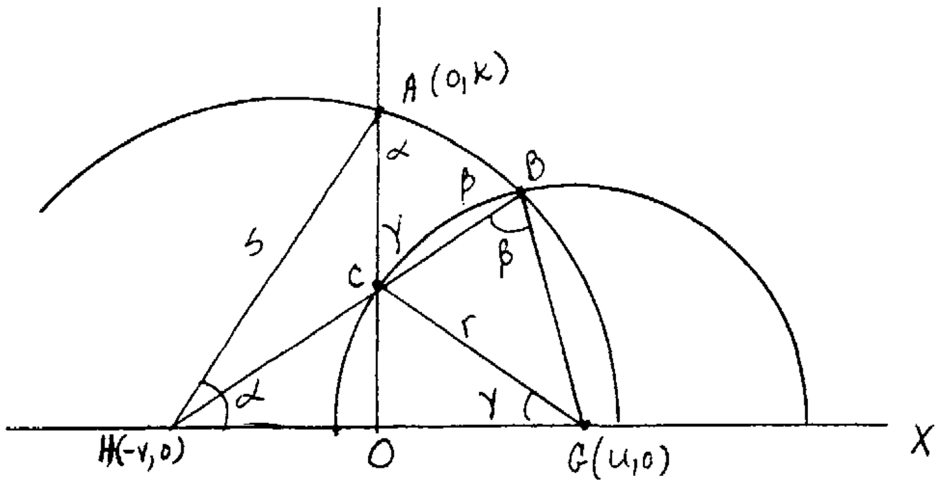


FIGURE 31

Let G(u,0) and H(-v,0) be the Euclidean centers of the geodesic arcs BC and AB, respectively. Let r and s be their respective Euclidean radii. $\angle CGO=\gamma$, $\angle OHA=\alpha$, $\angle HBG=\beta$ by proposition 3.21. The trigonometry of the Euclidean triangles GCO, AHO, and BGH yields the constraints:

(1) $u = r\cos\gamma$

(2) $v = s\cos\alpha$

(3) $(u+v)^2 = r^2 + s^2 - 2rs\cos\beta$.

By the Pythagorean Theorem $r^2=u^2+1$. Along with (1) it follows that

$$r = \csc\gamma \text{ and } u = \cot\gamma.$$

Substitute u, r, and v in (3) to get

$$(\cot\gamma + s\cos\alpha)^2 = \csc^2\gamma + s^2 - 2s\cos\beta\csc\gamma$$

which simplifies to the quadratic equation

$$(4)\quad s^2\sin^2\alpha - 2s(\cos\alpha\cot\gamma + \cos\beta\csc\gamma) + 1 = 0.$$

It follows that if triangle ABC is a hyperbolic triangle in standard position that satisfies these constraints, then its angles are indeed α,β,γ. The quadratic equation (4) has discriminant

$$4(\cos\alpha\cot\gamma + \cos\beta\csc\gamma)^2 - 4\sin^2\alpha.$$

If this discriminant is necessarily positive, this guarantees that the quadratic equation (4) does have solution in s. Since $\alpha+\beta+\gamma<\pi$, then $\alpha+\gamma<\pi-\beta$. The cosine function is monotone decreasing in the first two quadrants. Therefore we have the following sequence of equivalent statements:

$$\cos(\alpha+\gamma) > \cos(\pi-\beta) = -\cos\beta,$$

$$\cos\alpha\cos\gamma - \sin\alpha\sin\gamma > -\cos\beta,$$

$$\cos\alpha\cos\gamma + \cos\beta > \sin\alpha\sin\gamma > 0,$$

$$\cos\alpha\cot\gamma + \cos\beta\csc\gamma > \sin\alpha > 0,$$

$$(\cos\alpha\cot\gamma + \cos\beta\csc\gamma)^2 > \sin^2\alpha.$$

The last inequality establishes the positivity of the discriminant of (4). It follows that this quadratic has two real solutions for any given positive angles α,β,γ such that $\alpha+\beta+\gamma<\pi$. If

$v=s\cos\alpha$, then s, v, $u=\cot\gamma$, and $r=\csc\gamma$ satisfy equations (1), (2), and (3) and so the given hyperbolic triangle does exist. QED

3.23 Example

1) Construct a hyperbolic triangle in standard position with given angles: $\alpha=30°$, $\beta=50°$, $\gamma=60°$.

$r=\csc\gamma=1.1547$, $u=r\cos\gamma=0.57735$;

$s^2\sin^2\alpha-2s(\cos\alpha\cot\gamma+\cos\beta\csc\gamma)+1=0$ yields

$s^2(.25)-2s(1.2422)+1=0$ and $s=\dfrac{2.4844\pm\sqrt{5.1702}}{0.5}$.

$s\approx 9.52$ or $s\approx 0.421$. Let $s\approx 9.52$, then

$v=s\cos\alpha=8.244$ and $k=s\sin\alpha=4.76$.

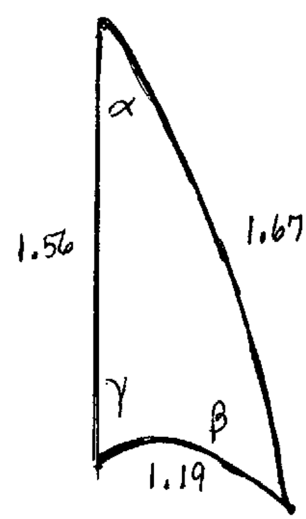


FIGURE 32

2) Construct a hyperbolic triangle in standard position with the given angles: $\alpha=60°$, $\beta=50°$, $\gamma=30°$.

$u=r\cos\gamma=1.73$, $r=\csc\gamma=2$, $s^2(.75)-2s(2.152)+1=0$,

and $s=\dfrac{4.304\pm\sqrt{15.52}}{1.5}$. $s\approx 5.5$ or $s\approx 0.24$.

Let $s\approx 5.5$, then $v=s\cos\alpha=2.75$ and $k=s\sin\alpha=4.76$.

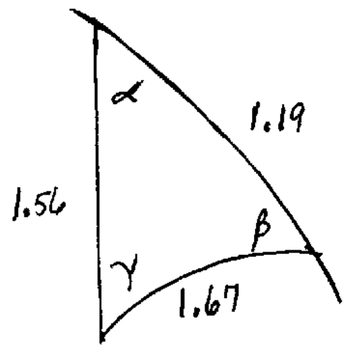


FIGURE 33

3) Construct a hyperbolic triangle in standard position with the given angles: $\alpha=\beta=\gamma=50°$.

$r=\csc\gamma=1.305$, $u=r\cos\gamma=0.839$, $s^2(0.587)-2s(1.378)+1=0$,

and $s=\dfrac{2.756\pm\sqrt{5.25}}{1.174}$. $s\approx4.3$ or $s\approx0.396$.

Let $s\approx4.3$, then $v=s\cos\alpha=2.76$ and $k=s\sin\alpha=3.29$

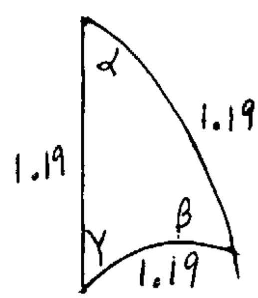


FIGURE 34

In each of the three constructions, the hyperbolic lengths of the sides of the triangles were found using Propositon 3.1, Proposition 3.2, and Example 3.4 of this chapter as well as the hyperbolic trigonometric formula:

$$\frac{\sin\alpha}{\sinh a}=\frac{\sin\beta}{\sinh b}=\frac{\sin\gamma}{\sinh c}.$$

Let triangle ABC be any hyperbolic triangle with angles α, β, and γ with $\angle CAB=\alpha$, $\angle ABC=\beta$, and $\angle BCA=\gamma$. Place this triangle in standard positon, and let r and s be as in the previous constructions. C and A have coordinates (0,1) and (0,s $\sin\alpha$) respectively. By Proposition 3.2 of this chapter, the hyperbolic length of the side AC is $|\ln(s \sin\alpha)|$. s was found by solving the quadratic equation as in the previous constructions. It would seem possible that there would be two triangles in standard position determined by these angles. But there is only one. Let s_1 and s_2 denote the two possible values of s. Since these are the roots of the quadratic equation (4) in Theorem 3.22, it follows that

$$s_1 s_2=\frac{1}{\sin^2\alpha} \text{ and therefore } s_2\sin\alpha=\frac{1}{s_1\sin\alpha}.$$

However, $|\ln\frac{1}{x}| = |\ln x|$ and so the length of side AC as given by $|\ln (s \sin\alpha)|$ is completely determined by the angles α, β, and γ. This argument could have been applied to any side of the hyperbolic triangle ABC. It follows that every hyperbolic triangle is completely determined by its angles.

<u>3.24. Theorem.</u> If two hyperbolic triangles have their respective angles equal, then they are hyperbolically congruent.

Generally speaking, a Euclidean triangle is completely determined by any of its three sides and three angles. An exception to this rule is the fact that a Euclidean triangle is determined only up to similarity when only its three angles are given. In the hyperbolic plane, this exception does not occur and a triangle is completely determined by its angles. In construction (3), $\alpha=\beta=\gamma=50°$. The hyperbolic length of each side of the hyperbolic triangle was found to be approximately 1.19. This will be true for every hyperbolic triangle with $\alpha=\beta=\gamma=50°$. Therefore, in the hyperbolic plane, similar triangles are congruent! So if you lived in a 30°-30°-30° triangular house in hyperbolic space, you could buy carpet and furniture custom made to fit.

BIBLIOGRAPHY

[1] Chern, S.S., MAA Studies in Mathematics- Studies in Global Geometry and Analysis, New Jersey: Prentice-Hall, 1967.

[2] Derrick, William R. and Grossman, Stanley I., Introduction to Differential Equations with Boundary Value Problems, St. Paul, Minnesota: West Publishing Company, 1987.

[3] Do Carmo, Manfredo Perdigao, Riemannian Geometry, Boston, Massachusetts: Birkhauser, 1992.

[4] Hicks, Noel J., Notes on Differential Geometry, New Jersey: D. Van Nostrand Co., Inc., 1965.

[5] Hsiung, Chuan-Chih, A First Course in Differential Geometry, New York: John Wiley & Sons, 1981.

[6] Klingenberg, Wilhelm, A Course in Differential Geometry, New York: Springer-Verlag, 1978.

[7] Kreyszig, Erwin, Introduction to Differential Geometry and Riemannian Geometry, Great Britain: University of Toronto Press, 1968.

[8] McCleary, John, Geometry From a Differentiable Viewpoint, New York: Cambridge University Press, 1994.

[9] Millman, Richard S. and Parker, George D., Elements of Differential Geometry, New Jersey: Prentice-Hall, 1977.

[10] O'Neill, Barrett, Elementary Differential Geometry, New York: Academic Press, 1966.

[11] Ortega, James M., Matrix Theory, New York: Plenum Press, 1987.

[12] Spivak, Michael, A Comprehensive Introduction to Differential Geometry, Volume Two, Massachusetts: Publish or Perish, Inc., 1970.

[13] Stahl, Saul, The Poincaré Half-Plane: A Gateway to Modern Geometry, London: Jones and Bartlett Publishers, 1993.

[14] Struik, Dirk J., Lectures on Classical Differential Geometry, Massachusetts: Addison-Wesley Publishing Company, Inc., 1950.

[15] Willmore, T.J., An Introduction to Differential Geometry, Great Britain: Oxford University Press, 1959.

VITA

Name of Author: Deborah F. Logan

Place of Birth:

Date of Birth:

EDUCATION

M.S. in Mathematical Sciences, 1995
University of North Florida,
Jacksonville, Florida

B.S. in Secondary Education, 1975
Mathematics Proficiency
Tennessee Temple University,
Chattanooga, Tennessee

TEACHING EXPERIENCE

1990-Present
Adjunct Professor of Mathematics
Florida Community College at Jacksonville
South Campus and Kent Campus

Mathematics Instructor, 1976-1990
Various high schools in
Spartanburg, South Carolina
Riverview, Florida
Jacksonville, Florida