LOVELY
PROFESSIONAL
UNIVERSITY

DIFFERENTIAL GEOMETRY

Edited By
Dr.Sachin Kaushal

Printed by
EXCEL BOOKS PRIVATE LIMITED
A-45, Naraina, Phase-I,
New Delhi-110028
for
Lovely Professional University
Phagwara

SYLLABUS

Complex Analysis and Differential Geometry

Objectives:

- To emphasize the role of the theory of functions of a complex variable, their geometric properties and indicating some applications. Introduction covers complex numbers; complex functions; sequences and continuity; and differentiation of complex functions. Representation formulas cover integration of complex functions; Cauchy's theorem and Cauchy's integral formula; Taylor series; and Laurent series. Calculus of residues covers residue calculus; winding number and the location of zeros of complex functions; analytic continuation.
- To understand classical concepts in the local theory of curves and surfaces including normal, principal, mean, and Gaussian curvature, parallel transports and geodesics, Gauss's theorem Egregium and Gauss-Bonnet theorem and Joachimsthal's theorem, Tissot's theorem.

Sr. No.	Content
1	Differentiation of Cartesians tensors, metric tensor, contra-variant, Covariant and mixed tensors, Christoffel symbols, Transformation of christoffel symbols and covariant differentiation of a tensor,
2	Theory of space curves: - Tangent, principal normal, binormal, curvature and torsion, Serret-Frenet formulae Contact between curves and surfaces, Locus of Centre of curvature, spherical curvature
3	Helices, Spherical indicatrix, Bertrand curves, surfaces, envelopes, edge of regression, Developable surfaces, Two fundamental forms
4	Curves on a surface, Conjugate direction, Principal directions, Lines of Curvature, Principal Curvatures, Asymptotic Lines, Theorem of Beltrami and Enneper, Mainardi-Codazzi equations,
5	Geodesics, Differential Equation of Geodesic. Torsion of Geodesic, Geodesic Curvature, Geodesic Mapping, Clairaut.s theorem, Gauss- Bonnet theorem, Joachimsthal.s theorem, Tissot.s theorem

CONTENT

Unit 1: Notation and Summation Convention

CONTENTS

Objectives

After studying this unit, you will be able to:

- Discuss the concept rectangular Cartesian coordinate systems
- Describe the suffix and symbolic notation
- Discuss the orthogonal transformation

Introduction

In this unit, we will discuss an elementary introduction to Cartesian tensor analysis in a three-dimensional Euclidean point space or a two-dimensional subspace. A Euclidean point space is the space of position vectors of points. The term vector is used in the sense of classical vector analysis, and scalars and polar vectors are zeroth- and first-order tensors, respectively. The distinction between polar and axial vectors is discussed later in this chapter. A scalar is a single quantity that possesses magnitude and does not depend on any particular coordinate system, and a vector is a quantity that possesses both magnitude and direction and has components, with respect to a particular coordinate system, which transform in a definite manner under change of coordinate system. Also vectors obey the parallelogram law of addition. There are quantities that possess both magnitude and direction but are not vectors, for example, the angle of finite rotation of a rigid body about a fixed axis.

A second-order tensor can be defined as a linear operator that operates on a vector to give another vector. That is, when a second-order tensor operates on a vector, another vector, in the same Euclidean space, is generated, and this operation can be illustrated by matrix multiplication. The components of a vector and a second-order tensor, referred to the same rectangular Cartesian coordinate system, in a three-dimensional Euclidean space, can be expressed as a (3×1) matrix and a (3×3) matrix, respectively. When a second-order tensor operates on a vector, the components of the resulting vector are given by the matrix product of the (3×3) matrix of components of the

Notes

second-order tensor and the matrix of the (3 × 1) components of the original vector. These components are with respect to a rectangular Cartesian coordinate system, hence, the term Cartesian tensor analysis. Examples from classical mechanics and stress analysis are as follows. The angular momentum vector, h, of a rigid body about its mass center is given by h = Jω, where J is the inertia tensor of the body about its mass center and ω is the angular velocity vector. In this equation the components of the vectors, *h* and ω can be represented by (3 × 1) matrices and the tensor J by a (3 × 3) matrix with matrix multiplication implied. A further example is the relation t = σn, between the stress vector t acting on a material area element and the unit normal n to the element, where *s* is the Cauchy stress tensor. The relations $h = J\omega$ and $t = \sigma n$ are examples of coordinate-free symbolic notation, and the corresponding matrix relations refer to a particular coordinate system.

We will meet further examples of the operator properties of second order tensors in the study of continuum mechanics and thermodynamics. Tensors of order greater than two can be regarded as operators operating on lower-order tensors. Components of tensors of order greater than two cannot be expressed in matrix form.

It is very important to note that physical laws are independent of any particular coordinate system. Consequently, equations describing physical laws, when referred to a particular coordinate system, must transform in definite manner under transformation of coordinate systems. This leads to the concept of a tensor, that is, a quantity that does not depend on the choice of coordinate system. The simplest tensor is a scalar, a zeroth-order tensor. A scalar is represented by a single component that is invariant under coordinate transformation. Examples of scalars are the density of a material and temperature.

Higher-order tensors have components relative to various coordinate systems, and these components transform in a definite way under transformation of coordinate systems. The velocity *v* of a particle is an example of a first-order tensor; henceforth we denote vectors, in symbolic notation, by lowercase bold letters. We can express v by its components relative to any convenient coordinate system, but since *v* has no preferential relationship to any particular coordinate system, there must be a definite relationship between components of *v* in different coordinate systems. Intuitively, a vector may be regarded as a directed line segment, in a three-dimensional Euclidean point space E_3, and the set of directed line segments in E_3, of classical vectors, is a vector space V_3. That is, a classical vector is the difference of two points in E_3. A vector, according to this concept, is a first-order tensor.

There are many physical laws for which a second-order tensor is an operator associating one vector with another. Remember that physical laws must be independent of a coordinate system; it is precisely this independence that motivates us to study tensors.

1.1 Rectangular Cartesian Coordinate Systems

The simplest type of coordinate system is a rectangular Cartesian system, and this system is particularly useful for developing most of the theory to be presented in this text.

A rectangular Cartesian coordinate system consists of an orthonormal basis of unit vectors (e_1, e_2, e_3) and a point 0 which is the origin. Right-handed Cartesian coordinate systems are considered, and the axes in the (e_1, e_2, e_3) directions are denoted by $0x_1$, $0x_2$, and $0x_3$, respectively, rather than the more usual 0x, 0y, and 0z. A right-handed system is such that a 90° right-handed screw rotation along the $0x_1$ direction rotates $0x_2$ to $0x_3$, similarly a right-handed rotation about $0x_2$ rotates $0x_3$ to $0x_1$, and a right-handed rotation about $0x_3$ rotates $0x_1$ to $0x_2$.

Notes

A right-handed system is shown in Figure 14.1. A point, $x \in E_3$, is given in terms of its coordinates (x_1, x_2, x_3) with respect to the coordinate system $0x_1x_2x_3$ by

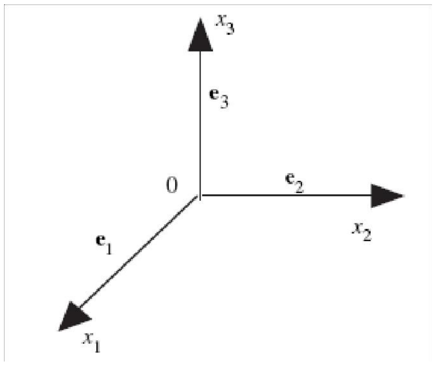
$$x = x_1e_1 + x_2e_2 + x_3e_3,$$

which is a bound vector or position vector.

If points $x, y \in E_3$, $u = x - y$ is a vector, that is, $u \in V_3$. The vector u is given in terms of its components (u_1, u_2, u_3), with respect to the rectangular coordinate system, $0x_1x_2x_3$ by

$$u = u_1e_1 + u_2e_2 + u_3e_3.$$

Figure 1 .1: Right-handed Rectangular Cartesian Coordinate System



Henceforth in this unit, when the term coordinate system is used, a rectangular Cartesian system is understood. When the components of vectors and higher-order tensors are given with respect to a rectangular Cartesian coordinate system, the theory is known as Cartesian tensor analysis.

1 .2 Suffix and Symbolic Notation

Suffixes are used to denote components of tensors, of order greater than zero, referred to a particular rectangular Cartesian coordinate system. Tensor equations can be expressed in terms of these components; this is known as suffix notation. Since a tensor is independent of any coordinate system but can be represented by its components referred to a particular coordinate system, components of a tensor must transform in a definite manner under transformation of coordinate systems. This is easily seen for a vector. In tensor analysis, involving oblique Cartesian or curvilinear coordinate systems, there is a distinction between what are called contra-variant and covariant components of tensors but this distinction disappears when rectangular Cartesian coordinates are considered exclusively.

Bold lower- and uppercase letters are used for the symbolic representation of vectors and second-order tensors, respectively. Suffix notation is used to specify the components of tensors, and the convention that a lowercase letter suffix takes the values 1, 2, and 3 for three-dimensional and 1 and 2 for two-dimensional Euclidean spaces, unless otherwise indicated, is adopted. The number of distinct suffixes required is equal to the order of the tensor. An example is the suffix representation of a vector u, with components (u_1, u_2, u_3) or u_i, $i \in \{1, 2, 3\}$. The vector is then given by

$$u = \sum_{i=1}^{3} u_i e_i. \qquad ...(1)$$

Notes

It is convenient to use a summation convention for repeated letter suffixes. According to this convention, if a letter suffix occurs twice in the same term, a summation over the repeated suffix from 1 to 3 is implied without a summation sign, unless otherwise indicated. For example, equation (1) can be written as

$$u = u_i e_i = u_1 e_1 + u_2 e_2 + u_3 e_3 \qquad ...(2)$$

without the summation sign. The sum of two vectors is commutative and is given by

$$u + v = v + u = (u_i + v_i)e_i,$$

which is consistent with the parallelogram rule. A further example of the summation convention is the scalar or inner product of two vectors,

$$u \,.\, v = u_i v_i = u_1 v_1 + u_2 v_2 + u_3 v_3 \qquad ...(3)$$

Repeated suffixes are often called dummy suffixes since any letter that does not appear elsewhere in the expression may be used, for example,

$$u_i v_i = u_j v_j.$$

Equation (3) indicates that the scalar product obeys the commutative law of algebra, that is,

$$u \,.\, v = v \,.\, u.$$

The magnitude $|u|$ of a vector u is given by

$$|u| = \sqrt{u.u} = \sqrt{u_i u_i}.$$

Other examples of the use of suffix notation and the summation convention are

$$C_{ii} = C_{11} + C_{22} + C_{33}$$

$$C_{ij} b_j = C_{i1} b_1 + C_{i2} b_2 + C_{i3} b_3.$$

A suffix that appears once in a term is known as a free suffix and is understood to take in turn the values 1, 2, 3 unless otherwise indicated. If a free suffix appears in any term of an equation or expression, it must appear in all the terms.

1 .3 Orthogonal Transformations

The scalar products of orthogonal unit base vectors are given by

$$e_i \,.\, e_j = \delta_{ij}, \qquad ...(1)$$

where δ_{ij} is known as the Kronecker delta and is defined as

$$\delta_{ij} = \begin{cases} 1 \text{ for } i = j \\ 0 \text{ for } i \neq j \end{cases}. \qquad ...(2)$$

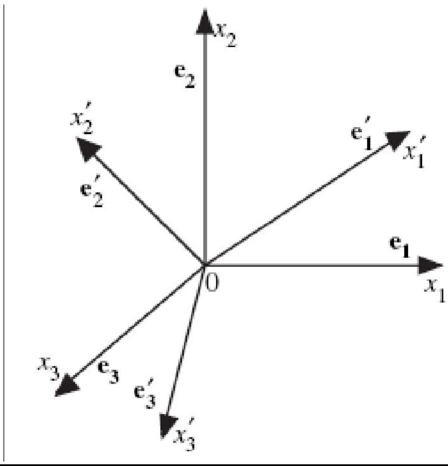
The base vectors ei are orthonormal, that is, of unit magnitude and mutually perpendicular to each other. The Kronecker delta is sometimes called the substitution operator because

$$u_j \delta_{ij} + u_1 \delta_{i1} + u_2 \delta_{i2} + u_3 \delta_{i3} = u_i. \qquad ...(3)$$

Consider a right-handed rectangular Cartesian coordinate system $0x'_i$ with the same origin as $0x_i$ as indicated in Figure 14.2. Henceforth, primed quantities are referred to coordinate system $0x'_i$.

Notes

Figure 1 .2: Change of Axes



The coordinates of a point P are x_i with respect to $0x_i$ and x_i' with respect to $0x_i$. Consequently,

$$x_i e_i = x_i' e_j', \qquad ...(4)$$

where the e_i' are the unit base vectors for the system $0x_i$. Forming the inner product of each side of equation (4) with e'_k and using equation (1) and the substitution operator property equation (3) gives

$$x_k' = a_{ki} x_i, \qquad ...(5)$$

where

$$a_{ki} = e_k'.e_i = \cos(x_k' 0x_i). \qquad ...(6)$$

Similarly

$$x_i = a_{ki} x_k'. \qquad ...(7)$$

It is evident that the direction of each axis $0x_k'$ can be specified by giving its direction cosines $a_{ki} = e_k'.e_i = \cos(x_k' 0x_i)$ referred to the original axes $0x_i$. The direction cosines, $a_{ki} = e_k'.e_i$, defining this change of axes are tabulated in Table 1 .1.

The matrix [a] with elements a_{ij} is known as the transformation matrix; it is not a tensor.

Table 1 .1: Direction cosines for Rotation of Axes

	e_1'	e_2'	e_3'
e_1	a_{11}	a_{21}	a_{31}
e_2	a_{12}	a_{22}	a_{32}
e_3	a_{13}	a_{23}	a_{33}

Notes

It follows from equations (5) and (7) that

$$a_{ki} = \frac{\partial x'_k}{\partial x_i} = \frac{\partial x_i}{\partial x'_k}, \qquad ...(8)$$

and from equation (4) that

$$e_i \frac{\partial x_i}{\partial x'_k} = e'_j \frac{\partial x'_j}{dx'_k} = e'_k, \qquad ...(9)$$

since $\partial x_j/\partial x_k = \partial_{jk}$, and from equations (8) and (9) that

$$e'_k = a_{ki} e_i, \qquad ...(10)$$

and

$$e_i = a_{ki} e'_k. \qquad ...(11)$$

Equations (10) and (11) are the transformation rules for base vectors. The nine elements of a_{ij} are not all independent, and in general,

$$a_{ki} \neq a_{ik}.$$

A relation similar to equations (5) and (7),

$$u'_k = a_{ki} u_i, \text{ and } u_i = a_{ki} u'_k \qquad ...(12)$$

is obtained for a vector u since $u_i e_i = u'_k e'_k$, which is similar to equation (4) except that the ui are the components of a vector and the x_i are coordinates of a point.

The magnitude $|u| = (u_i u_i)^{1/2}$ of the vector u is independent of the orientation of the coordinate system, that is, it is a scalar invariant; consequently,

$$u_i u_i = u'_k = a_{ki} a_{ji} u'_k. \qquad ...(13)$$

Eliminating u_i from equation (12) gives

$$u'_k = a_{ki} a_{ji} u'_j,$$

and since $u'_k = \delta_{kj} u'_j$,

$$a_{ki} a_{ji} = \delta_{kj}. \qquad ...(14)$$

Similarly, eliminating u_k from equation (12) gives

$$a_{ik} a_{jk} = \delta_{ij}. \qquad ...(15)$$

It follows from equation (14) or (15) that

$$\{\det[a_{ij}]\}^2 = 1, \qquad ...(16)$$

where det $[a_{ij}]$ denotes the determinant of a_{ij}. The negative root of equation (16) is not considered unless the transformation of axes involves a change of orientation since, for the identity transformation $x_i = x'_i$, $a_{ik} = \delta_{ik}$ and $\det[\delta_{ik}] = 1$. Consequently, $\det[a_{ik}] = 1$, provided the transformations involve only right-handed systems (or left-handed systems).

The transformations (5), (7), and (12) subject to equation (14) or (15) are known as orthogonal transformations. Three quantities u_i are the components of a vector if, under orthogonal transformation, they transform according to equation (12). This may be taken as a definition of a vector. According to this definition, equations (5) and (7) imply that the representation x of a point is a bound vector since its origin coincides with the origin of the coordinate system.

If the transformation rule (7) holds for coordinate transformations from right-handed systems to left-handed systems (or vice versa), the vector is known as a polar vector. There are scalars and vectors known as pseudo scalars and pseudo or axial vectors; there have transformation rules that involve a change in sign when the coordinate transformation is from a right-handed system to a left-handed system (or vice versa), that is, when $\det[a_{ij}] = -1$. The transformation rule for a pseudo scalar is

$$\phi' = \det[a_{ij}]\phi, \qquad \text{...(17)}$$

and for a pseudo vector

$$u_i' = \det[a_{ij}]a_{ij}u_j. \qquad \text{...(18)}$$

A pseudo scalar is not a true scalar if a scalar is defined as a single quantity invariant under all coordinate transformations. An example of a pseudo vector is the vector product u × v of two polar vectors u and v. The moment of a force about a point and the angular momentum of a particle about a point are pseudo vectors. The scalar product of a polar vector and a pseudo vector is a pseudo scalar; an example is the moment of a force about a line. The distinction between pseudo vectors and scalars and polar vectors and true scalars disappears when only right- (or left-) handed coordinate systems are considered. For the development of continuum mechanics presented in this book, only right-handed systems are used.

Example: Show that a rotation through angle π about an axis in the direction of the unit vector n has the transformation matrix

$$a_{ij} = -\delta_{ij} + 2n_in_j, \det[aij] = 1.$$

Solution. The position vector of point A has components x_i and point B has position vector with components x_i'.

1 .4 Summary

- A rectangular Cartesian coordinate system consists of an orthonormal basis of unit vectors (e_1, e_2, e_3) and a point 0 which is the origin. Right-handed Cartesian coordinate systems are considered, and the axes in the (e_1, e_2, e_3) directions are denoted by $0x_1$, $0x_2$, and $0x_3$, respectively, rather than the more usual 0x, 0y, and 0z. A right-handed system is such that a 90º right-handed screw rotation along the $0x_1$ direction rotates $0x_2$ to $0x_3$, similarly a right-handed rotation about $0x_2$ rotates $0x_3$ to $0x_1$, and a right-handed rotation about $0x_3$ rotates $0x_1$ to $0x_2$.
- Suffixes are used to denote components of tensors, of order greater than zero, referred to a particular rectangular Cartesian coordinate system. Tensor equations can be expressed in terms of these components; this is known as suffix notation. Since a tensor is independent of any coordinate system but can be represented by its components referred to a particular coordinate system, components of a tensor must transform in a definite manner under transformation of coordinate systems. This is easily seen for a vector. In tensor analysis, involving oblique Cartesian or curvilinear coordinate systems, there is a distinction between what are called contra-variant and covariant components of tensors but this distinction disappears when rectangular Cartesian coordinates are considered exclusively.

Bold lower- and uppercase letters are used for the symbolic representation of vectors and second-order tensors, respectively. Suffix notation is used to specify the components of tensors, and the convention that a lowercase letter suffix takes the values 1, 2, and 3 for three-dimensional and 1 and 2 for two-dimensional Euclidean spaces, unless otherwise indicated, is adopted. The number of distinct suffixes required is equal to the order of the tensor. An example is the suffix representation of a vector u, with components (u_1, u_2, u_3) or ui, $i \in \{1, 2, 3\}$. The vector is then given by

$$u = \sum_{i=1}^{3} u_i e_i .$$

1 .5 Keywords

Rectangular Cartesian coordinate: A rectangular Cartesian coordinate system consists of an orthonormal basis of unit vectors (e_1, e_2, e_3) and a point 0 which is the origin.

Right-handed Cartesian coordinate systems are considered, and the axes in the (e_1, e_2, e_3) directions are denoted by $0x_1$, $0x_2$, and $0x_3$, respectively.

Suffixes are used to denote components of tensors, of order greater than zero, referred to a particular rectangular Cartesian coordinate system.

Tensor equations can be expressed in terms of these components; this is known as suffix notation.

1 .6 Self Assessment

1. A .... system consists of an orthonormal basis of unit vectors (e_1, e_2, e_3) and a point 0 which is the origin.
2. .... coordinate systems are considered, and the axes in the (e_1, e_2, e_3) directions are denoted by $0x_1$, $0x_2$, and $0x_3$, respectively.
3. .... are used to denote components of tensors, of order greater than zero, referred to a particular rectangular Cartesian coordinate system.
4. Tensor equations can be expressed in terms of these components; this is known as ....
5. Suffix notation is used to specify the components of tensors, and the convention that a lowercase letter suffix takes the values 1, 2, and 3 for three-dimensional and 1 and 2 for two-dimensional ...., unless otherwise indicated, is adopted.

1 .7 Review Questions

1. Discuss the concept rectangular Cartesian coordinate systems.
2. Describe the suffix and symbolic notation.
3. Discuss the orthogonal transformation.

Answers: Self Assessment

1.	rectangular Cartesian coordinate	2.	Right-handed Cartesian
3.	Suffixes	4.	suffix notation
5.	Euclidean spaces		

1 .8 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 2 : Tensors in Cartesian Coordinates

Notes

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define convectors
- Discuss the scalar products of vector and convectors
- Describe bilinear and quadratic forms
- Explain the dot product and metric tensor

Introduction

In the last unit, you have studied about notation and summation of convention. A tensor written in component form is an indexed array. The order of a tensor is the number of indices required. (The rank of tensor used to mean the order, but now it means something different). The rank of the tensor is the minimal number of rank-one tensor that you need to sum up to obtain this higher-rank tensor. Rank-one tensors are given the generalization of outer product to m-vectors, where m is the order of the tensor.

2.1 Covectors

In previous sections, we learned the following important fact about vectors: a vector is a physical object in each basis of our three-dimensional Euclidean space E represented by three numbers such that these numbers obey certain transformation rules when we change the basis.

Now suppose that we have some other physical object that is represented by three numbers in each basis, and suppose that these numbers obey some certain transformation rules when we change the basis. Is it possible? One can try to find such an object in nature. However, in mathematics we have another option. We can construct such an object mentally, then study its properties, and finally look if it is represented somehow in nature.

Let's denote our hypothetical object by a, and denote by a_1, a_2, a_3 that three numbers which represent this object in the basis e_1, e_2, e_3. By analogy with vectors we shall call them coordinates. But in contrast to vectors, we intentionally used lower indices when denoting them by a_1, a_2, a_3. Let's prescribe the following transformation rules to a_1, a_2, a_3 when we change e_1, e_2, e_3 to $\tilde{e}_1, \tilde{e}_2, \tilde{e}_3$:

$$\tilde{a}_j = \sum_{i=1}^{3} S^i_j a_i, \qquad ...(1)$$

$$a_j = \sum_{i=1}^{3} T^i_j a_i, \qquad ...(2)$$

Note that (1) is sufficient, formula (2) is derived from (1).

Example: Using the concept of the inverse matrix $T = S^{-1}$ derive formula (2) from formula (1).

Definition: A geometric object a in each basis represented by a triple of coordinates a_1, a_2, a_3 and such that its coordinates obey transformation rules (1) and (2) under a change of basis is called a covector.

Looking at the above considerations one can think that we arbitrarily chose the transformation formula (1). However, this is not so. The choice of the transformation formula should be self-consistent in the following sense. Let e_1, e_2, e_3 and $\tilde{e}_1, \tilde{e}_2, \tilde{e}_3$ be two bases and let $\tilde{\tilde{e}}_1, \tilde{\tilde{e}}_2, \tilde{\tilde{e}}_3$ be the third basis in the space. Let's call them basis one, basis two and basis three for short. We can pass from basis one to basis three directly. Or we can use basis two as an intermediate basis. In both cases, the ultimate result for the coordinates of a covector in basis three should be the same: this is the self-consistence requirement. It means that coordinates of a geometric object should depend on the basis, but not on the way that they were calculated.

Exercise 2 .1: Replace S by T in (1) and T by S in (2). Show that the resulting formulas are not self-consistent.

What about the physical reality of covectors? Later on we shall see that covectors do exist in nature. They are the nearest relatives of vectors. And moreover, we shall see that some well-known physical objects we thought to be vectors are of covectorial nature rather than vectorial.

2.2 Scalar Product of Vector and Covector

Suppose we have a vector x and a covector a. Upon choosing some basis e_1, e_2, e_3, both of them have three coordinates: x^1, x^2, x^3 for vector x, and a_1, a_2, a_3 for covector a. Let's denote by $\langle a, x\rangle$ the following sum:

$$\langle a, x\rangle = \sum_{i=1}^{3} a_i x^i. \qquad ...(1)$$

The sum (1) is written in agreement with Einstein's tensorial notation. It is a number depending on the vector x and on the covector a. This number is called the scalar product of the vector x and the covector a. We use angular brackets for this scalar product in order to distinguish it from the scalar product of two vectors in E, which is also known as the dot product.

Defining the scalar product $\langle a, x\rangle$ by means of sum (1) we used the coordinates of vector x and of covector a, which are basis-dependent. However, the value of sum (1) does not depend on any basis. Such numeric quantities that do not depend on the choice of basis are called scalars or true scalars.

Exercise 2 .2: Consider two bases e_1, e_2, e_3 and $\tilde{e}_1, \tilde{e}_2, \tilde{e}_3$, and consider the coordinates of vector x and covector a in both of them. Prove the equality

$$\sum_{i=1}^{3} a_i x^i = \sum_{i=1}^{3} \tilde{a}_i \tilde{x}^i. \qquad ...(2)$$

Thus, you are proving the self-consistence of formula (1) and showing that the scalar product $\langle a, x\rangle$ given by this formula is a true scalar quantity.

Exercise 2 .3: Let α be a real number, let a and b be two covectors, and let x and y be two vectors. Prove the following properties of the scalar product:

(1) $\langle a + b; x\rangle = \langle a, x\rangle + \langle b, x\rangle$;

(2) $\langle a, x + y\rangle = \langle a, x\rangle + \langle a, y\rangle$;

(3) $\langle \alpha a, x\rangle = \alpha\langle a, x\rangle$;

(4) $\langle a, \alpha x\rangle = \alpha\langle a, x\rangle$.

Exercise 2 .4: Explain why the scalar product $\langle a, x\rangle$ is sometimes called the bilinear function of vectorial argument x and covectorial argument a. In this capacity, it can be denoted as f(a, x). Remember our discussion about functions with non-numeric arguments.

Important note. The scalar product $\langle a, x\rangle$ is not symmetric. Moreover, the formula

$$\langle a, x\rangle = \langle x, a\rangle$$

is incorrect in its right hand side since the first argument of scalar product by definition should be a covector. In a similar way, the second argument should be a vector. Therefore, we never can swap them.

2 .3 Linear Operators

In this section we consider more complicated geometric objects. For the sake of certainty, let's denote one of such objects by F. In each basis e_1, e_2, e_3, it is represented by a square 3 × 3 matrix F^i_j of real numbers. Components of this matrix play the same role as coordinates in the case of vectors or covectors. Let's prescribe the following transformation rules to F^i_j :

$$\tilde{F}^i_j = \sum_{p=1}^{3}\sum_{q=1}^{3} T^i_p S^q_j F^p_q, \qquad \text{...(1)}$$

$$F^i_j = \sum_{p=1}^{3}\sum_{q=1}^{3} S^i_p T^q_j \tilde{F}^p_q. \qquad \text{...(2)}$$

Exercise 2 .5: Using the concept of the inverse matrix $T = S^{-1}$ prove that formula (2) is derived from formula (1).

If we write matrices F^i_j and $\tilde{F}^p_q$, then (1) and (2) can be written as two matrix equalities:

$$\tilde{F} = T\,F\,S, \qquad\qquad F = S\,\tilde{F}\,T. \qquad \text{...(3)}$$

Definition: A geometric object F in each basis represented by some square matrix F^i_j and such that components of its matrix F^i_j obey transformation rules (1) and (2) under a change of basis is called a linear operator.

Let's take a linear operator F represented by matrix F^i_j in some basis e_1, e_2, e_3 and take some vector x with coordinates x^1, x^2, x^3 in the same basis. Using F^i_j and x^j we can construct the following sum:

$$y^i = \sum_{j=1}^{3} F^i_j x^j. \qquad \text{...(4)}$$

Index i in the sum (4) is a free index; it can deliberately take any one of three values: i = 1, i = 2, or i = 3. For each specific value of i we get some specific value of the sum (4). They are denoted by y^1, y^2, y^3 according to (4). Now suppose that we pass to another basis $\tilde{e}_1, \tilde{e}_2, \tilde{e}_3$ and do the same things. As a result we get other three values $\tilde{y}^1, \tilde{y}^2, \tilde{y}^3$ given by formula

$$\tilde{y}^p = \sum_{q=1}^{3} \tilde{F}^p_q \tilde{x}^q. \qquad \text{...(5)}$$

Relying upon (1) and (2) prove that the three numbers y^1, y^2, y^3 and the other three numbers $\tilde{y}^1, \tilde{y}^2, \tilde{y}^3$ are related as follows:

$$\tilde{y}^j = \sum_{i=1}^{3} T^j_i y^i, \qquad\qquad y^j = \sum_{i=1}^{3} S^j_i\, \tilde{y}^i. \qquad \text{...(6)}$$

Notes

Thus, formula (4) defines the vectorial object y. As a result, we have vector y determined by a linear operator F and by vector x. Therefore, we write

$$y = F(x) \quad ...(7)$$

and say that y is obtained by applying linear operator F to vector x. Some people like to write (7) without parentheses:

$$y = Fx. \quad ....(8)$$

Formula (8) is a more algebraic form of formula (7). Here the action of operator F upon vector x is designated like a kind of multiplication. There is also a matrix representation of formula (8), in which x and y are represented as columns:

$$\begin{Vmatrix} y^1 \\ y^2 \\ y^3 \end{Vmatrix} = \begin{Vmatrix} F^1_1 & F^1_2 & F^1_3 \\ F^2_1 & F^2_2 & F^2_3 \\ F^3_1 & F^3_2 & F^3_3 \end{Vmatrix} \begin{Vmatrix} x^1 \\ x^2 \\ x^3 \end{Vmatrix}. \quad ...(9)$$

Exercise 2 .6: Derive (9) from (4).

Exercise 2 .7: Let α be some real number and let x and y be two vectors. Prove the following properties of a linear operator (7):

(1) $F(x + y) = F(x) + F(y)$,

(2) $F(\alpha\, x) = \alpha F(x)$.

Write these equalities in the more algebraic style introduced by (8). Are they really similar to the properties of multiplication?

Exercise 2 .8: Remember that for the product of two matrices

$$\det(AB) = \det A \det B. \quad ...(10)$$

Also remember the formula for $\det(A^{-1})$. Apply these two formulas to (3) and derive

$$\det F = \det \tilde{F}. \quad ...(11)$$

Formula (10) means that despite the fact that in various bases linear operator F is represented by various matrices, the determinants of all these matrices are equal to each other. Then we can define the determinant of linear operator F as the number equal to the determinant of its matrix in any one arbitrarily chosen basis e_1, e_2, e_3:

$$\det F = \det F. \quad ...(12)$$

Exercise 2 .9 (for deep thinking). Square matrices have various attributes: eigenvalues, eigenvectors, a characteristic polynomial, a rank (maybe you remember some others). If we study these attributes for the matrix of a linear operator, which of them can be raised one level up and considered as basis-independent attributes of the linear operator itself? Determinant (12) is an example of such attribute.

Exercise 2 .10: Substitute the unit matrix for F^i_j into (1) and verify that $\tilde{F}^i_j$ is also a unit matrix in this case. Interpret this fact.

Exercise 2 .11: Let $x = e_i$ for some basis e_1, e_2, e_3 in the space. Substitute this vector x into (7) and by means of (4) derive the following formula:

$$F(e_i) = \sum_{j=1}^{3} F^j_i e_j. \quad ...(13)$$

The fact is that in some books the linear operator is determined first, then its matrix is introduced by formula (13). Explain why if we know three vectors $F(e_1)$, $F(e_2)$, and $F(e_3)$, then we can reconstruct the whole matrix of operator F by means of formula (13).

Suppose we have two linear operators F and H. We can apply H to vector x and then we can apply F to vector H(x). As a result we get

$$F \circ H(x) = F(H(x)). \quad ...(14)$$

Here $F \circ H$ is new linear operator introduced by formula (14). It is called a composite operator, and the small circle sign denotes composition.

Exercise 2 .12: Find the matrix of composite operator $F \circ H$ if the matrices for F and H in the basis e_1, e_2, e_3 are known.

Exercise 2 .13: Remember the definition of the identity map in mathematics (see on-line Math. Encyclopedia) and define the identity operator id. Find the matrix of this operator.

Exercise 2 .14: Remember the definition of the inverse map in mathematics and define inverse operator F^{-1} for linear operator F. Find the matrix of this operator if the matrix of F is known.

2 .4 Bilinear and Quadratic Forms

Vectors, covectors, and linear operators are all examples of tensors (though we have no definition of tensors yet). Now we consider another one class of tensorial objects. For the sake of clarity, let's denote by a one of such objects. In each basis e_1, e_2, e_3 this object is represented by some square 3 × 3 matrix a_{ij} of real numbers. Under a change of basis these numbers are transformed as follows:

$$\tilde{a}_{ij} = \sum_{p=1}^{3}\sum_{q=1}^{3} S_i^p S_j^q a_{pq}, \quad ...(1)$$

$$a_{ij} = \sum_{p=1}^{3}\sum_{q=1}^{3} T_i^p T_j^q \tilde{a}_{pq}. \quad ...(2)$$

Transformation rules (1) and (2) can be written in matrix form:

$$\tilde{a} = S^T aS, \qquad a = T^T \tilde{a} T. \quad ...(3)$$

Here by S^T and T^T we denote the transposed matrices for S and T respectively.

Exercise 2 .15: Derive (2) from (1), then (3) from (1) and (2).

Definition: A geometric object a in each basis represented by some square matrix aij and such that components of its matrix a_{ij} obey transformation rules (1) and (2) under a change of basis is called a bilinear form.

Let's consider two arbitrary vectors x and y. We use their coordinates and the components of bilinear form a in order to write the following sum:

$$a(x, y) = \sum_{i=1}^{3}\sum_{j=1}^{3} a_{ij} x^i y^j. \quad ...(4)$$

Exercise 2 .16: Prove that the sum in the right hand side of formula (4) does not depend on the basis, i.e. prove the equality

$$\sum_{i=1}^{3}\sum_{j=1}^{3} a_{ij} x^i y^j = \sum_{p=1}^{3}\sum_{q=1}^{3} \tilde{a}_{pq} \tilde{x}^p \tilde{y}^q.$$

This equality means that a(x, y) is a number determined by vectors x and y irrespective of the choice of basis. Hence we can treat (4) as a scalar function of two vectorial arguments.

Exercise 2 .17: Let α be some real number, and let x, y, and z be three vectors. Prove the following properties of function (4):

(1) a(x + y, z) = a(x, z) + a(y, z);

(2) a(α x, y) = α a(x, y);

(3) a(x, y + z) = a(x, y) + a(x, z);

(4) a(x, αy) = α a(x, y).

Due to these properties function (4) is called a bilinear function or a bilinear form. It is linear with respect to each of its two arguments.

Note that scalar product is also a bilinear function of its arguments. The arguments of scalar product are of a different nature: the first argument is a covector, the second is a vector. Therefore, we cannot swap them. In bilinear form (4) we can swap arguments. As a result we get another bilinear function

$$b(x, y) = a(y, x). \qquad \text{...(5)}$$

The matrices of a and b are related to each other as follows:

$$b_{ij} = a_{ji},\ b = a^T. \qquad \text{...(6)}$$

Definition: A bilinear form is called symmetric if a(x, y) = a(y, x).

Exercise 2 .18: Prove the following identity for a symmetric bilinear form:

$$a(x, y) = \frac{a(x + y, x + y) - a(x, x) - a(y, y)}{2}. \qquad \text{...(7)}$$

Definition: A quadratic form is a scalar function of one vectorial argument f(x) produced from some bilinear function a(x; y) by substituting y = x:

$$f(x) = a(x, x). \qquad \text{...(8)}$$

Without a loss of generality a bilinear function a in (8) can be assumed symmetric. Indeed, if a is not symmetric, we can produce symmetric bilinear function

$$c(x, y) = \frac{a(x, y) + a(y, x)}{2}, \qquad \text{...(9)}$$

and then from (8) due to (9) we derive

$$f(x) = a(x, x) = \frac{a(x, x) + a(x, x)}{2} = c(x, x).$$

This equality is the same as (8) with a replaced by c. Thus, each quadratic function f is produced by some symmetric bilinear function a. And conversely, comparing (8) and (7) we get that a is produced by f:

$$a(x, y) = \frac{f(x + y) - f(x) - f(y)}{2}. \qquad \text{...(10)}$$

Formula (10) is called the recovery formula. It recovers bilinear function a from quadratic function f produced in (8). Due to this formula, in referring to a quadratic form we always imply some symmetric bilinear form like the geometric tensorial object introduced by definition.

2 .5 General Definition of Tensors

Vectors, covectors, linear operators, and bilinear forms are examples of tensors. They are geometric objects that are represented numerically when some basis in the space is chosen. This numeric representation is specific to each of them: vectors and covectors are represented by one-dimensional arrays, linear operators and quadratic forms are represented by two-dimensional arrays. Apart from the number of indices, their position does matter. The coordinates of a vector are numerated by one upper index, which is called the contra-variant index. The coordinates of a covector are numerated by one lower index, which is called the covariant index. In a matrix of bilinear form we use two lower indices; therefore bilinear forms are called twice-covariant tensors. Linear operators are tensors of mixed type; their components are numerated by one upper and one lower index. The number of indices and their positions determine the transformation rules, i.e. the way the components of each particular tensor behave under a change of basis. In the general case, any tensor is represented by a multidimensional array with a definite number of upper indices and a definite number of lower indices.

Let's denote these numbers by r and s. Then we have a tensor of the type (r; s), or sometimes the term valency is used. A tensor of type (r; s), or of valency (r; s) is called an r-times contravariant and an s-times covariant tensor. This is terminology; now let's proceed to the exact definition. It is based on the following general transformation formulas:

$$X^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \sum_{\substack{h_1, \ldots, h_r \\ k_1, \ldots, k_s}}^{3} \ldots \sum^{3} S^{i_1}_{h_1} \ldots S^{i_r}_{h_r} T^{k_1}_{j_1} \ldots T^{k_s}_{j_s} \tilde{X}^{h_1 \ldots h_r}_{k_1 \ldots k_s}, \quad \ldots(1)$$

$$\tilde{X}^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \sum_{\substack{h_1, \ldots, h_r \\ k_1, \ldots, k_s}}^{3} \ldots \sum^{3} T^{i_1}_{h_1} \ldots T^{i_r}_{h_r} S^{k_1}_{j_1} \ldots S^{k_s}_{j_s} X^{h_1 \ldots h_r}_{k_1 \ldots k_s}, \quad \ldots(2)$$

Definition: A geometric object X in each basis represented by (r + s) dimensional array $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ of real numbers and such that the components of this array obey the transformation rules (1) and (2) under a change of basis is called tensor of type (r, s), or of valency (r, s).

Formula (2) is derived from (1), so it is sufficient to remember only one of them. Let it be the formula (1). Though huge, formula (1) is easy to remember.

Indices $i_1, \ldots i_r$ and $j_1, \ldots, j_s$ are free indices. In right hand side of the equality (1) they are distributed in S-s and T-s, each having only one entry and each keeping its position, i.e. upper indices $i_1, \ldots; i_r$ remain upper and lower indices $j_1, \ldots, j_s$ remain lower in right hand side of the equality (1).

Other indices $h_1, \ldots h_r$ and $k_1, \ldots, k_s$ are summation indices; they enter the right hand side of (1) pairwise: once as an upper index and once as a lower index, once in S-s or T-s and once in components of array $\tilde{X}^{h_1 \ldots h_r}_{k_1 \ldots k_s}$.

When expressing $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ through $\tilde{X}^{h_1 \ldots h_r}_{k_1 \ldots k_s}$ each upper index is served by direct transition matrix S and produces one summation in (1):

Notes

$$X^{\dots i_\alpha \dots}_{\dots\dots\dots} = \sum \dots \sum_{h_\alpha=1}^{3} \dots \sum \dots S^{i_\alpha}_{h_\alpha} \dots \tilde{X}^{\dots h_\alpha \dots}_{\dots\dots\dots} \qquad \dots(3)$$

In a similar way, each lower index is served by inverse transition matrix T and also produces one summation in formula (1):

$$X^{\dots\dots\dots}_{\dots j\alpha \dots} = \sum \dots \sum_{k_\alpha=1}^{3} \dots \sum \dots T^{k_\alpha}_{j_\alpha} \dots \tilde{X}^{\dots\dots\dots}_{\dots k_\alpha \dots} \qquad \dots(4)$$

Formulas (3) and (4) are the same as (1) and used to highlight how (1) is written. So tensors are defined. Further we shall consider more examples showing that many well-known objects undergo the definition.

Task What are the valencies of vectors, covectors, linear operators, and bilinear forms when they are considered as tensors.

Exercise 2 .19: Let a_{ij} be the matrix of some bilinear form a. Let's denote by b^{ij} components of inverse matrix for a_{ij}. Prove that matrix b^{ij} under a change of basis transforms like matrix of twice-contravariant tensor. Hence, it determines tensor b of valency (2, 0). Tensor b is called a dual bilinear form for a.

2 .6 Dot Product and Metric Tensor

The covectors, linear operators, and bilinear forms that we considered above were artificially constructed tensors. However there are some tensors of natural origin. Let's remember that we live in a space with measure. We can measure distance between points (hence, we can measure length of vectors) and we can measure angles between two directions in our space. Therefore, for any two vectors x and y we can define their natural scalar product (or dot product):

$$(x, y) = |x| \; |y| \cos(\varphi) \qquad \dots(1)$$

where φ is the angle between vectors x and y.

Task Remember the following properties of the scalar product (1):

(1) $(x + y, z) = (x, z) + (y, z)$;

(2) $(a\,x, y) = a\,(x, y)$;

(3) $(x, y + z) = (x, y) + (x, z)$;

(4) $(x, a\,y) = a\,(x, y)$;

(5) $(x, y) = (y, x)$;

(6) $(x, x) \geq 0$ and $(x, x) = 0$ implies $x = 0$.

These properties are usually considered in courses on analytic geometry or vector algebra, see Vector Lessons on the Web.

Notes The first four properties of the scalar product (1) are quite similar to those or quadratic forms. This is not an occasional coincidence.

Exercise 2 .20: Let's consider two arbitrary vectors x and y expanded in some basis e_1, e_2, e_3. This means that we have the following expressions for them:

$$x = \sum_{i=1}^{3} x^i e_i, \qquad y = \sum_{j=1}^{3} x^j e_j. \qquad ...(2)$$

Substitute (2) into (1) and using properties (1)-(4) listed in exercise 2 .17 derive the following formula for the scalar product of x and y:

$$(x, y) = \sum_{i=1}^{3}\sum_{j=1}^{3} (e_i, e_j) x^i y^j. \qquad ...(3)$$

Exercise 2 .21: Denote $g_{ij} = (e_i, e_j)$ and rewrite formula (3) as

$$(x, y) = \sum_{i=1}^{3}\sum_{j=1}^{3} g_{ij} x^i y^j. \qquad ...(4)$$

Consider some other basis $\tilde{e}_1, \tilde{e}_2, \tilde{e}_3$, denote $\tilde{g}_{pq} = (\tilde{e}_p, \tilde{e}_q)$ and prove that matrices g_{ij} and $\tilde{g}_{pq}$ are components of a geometric object under a change of base. Thus you prove that the Gram matrix

$$g_{ij} = (e_i, e_j) \qquad ...(5)$$

defines tensor of type (0; 2). This is very important tensor; it is called the metric tensor. It describes not only the scalar product in form of (4), but the whole geometry of our space. Evidences for this fact are below.

Matrix (5) is symmetric due to property (5) in task on previous page. Now, keeping in mind the tensorial nature of matrix (5), we conclude that the scalar product is a symmetric bilinear form:

$$(x, y) = g(x, y) \qquad ...(6)$$

The quadratic form corresponding to (6) is very simple: $f(x) = g(x, x) = |x|^2$. The inverse matrix for (5) is denoted by the same symbol g but with upper indices: g^{ij}. It determines a tensor of type (2, 0), this tensor is called dual metric tensor.

2 .7 Multiplication by Numbers and Addition

Tensor operations are used to produce new tensors from those we already have. The most simple of them are multiplication by number and addition. If we have some tensor X of type (r, s) and a real number α, then in some base e_1, e_2, e_3 we have the array of components of tensor X; let's denote it $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$. Then by multiplying all the components of this array by α we get another array

$$Y^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \alpha X^{i_1 \ldots i_r}_{j_1 \ldots j_s}. \qquad ...(1)$$

Notes

Choosing another base $\tilde{e}_1, \tilde{e}_2, \tilde{e}_3$, and repeating this operation we get

$$\tilde{Y}^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \alpha \tilde{X}^{i_1 \ldots i_r}_{j_1 \ldots j_s}. \quad ...(2)$$

Exercise 2 .22: Prove that arrays $\tilde{Y}^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ and $Y^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ are related to each other in the same way as arrays $\tilde{X}^{i_1 \ldots i_r}_{j_1 \ldots j_s} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$, i.e. according to transformation formulas. In doing this you prove that formula (1) applied in all bases produces new tensor Y = αX from initial tensor X.

Formula (1) defines the multiplication of tensors by numbers. In exercise 2 .22 you prove its consistence. The next formula defines the addition of tensors:

$$X^{i_1 \ldots i_r}_{j_1 \ldots j_s} + Y^{i_1 \ldots i_r}_{j_1 \ldots j_s} = Z^{i_1 \ldots i_r}_{j_1 \ldots j_s}. \quad ...(3)$$

Having two tensors X and Y both of type (r, s) we produce a third tensor Z of the same type (r, s) by means of formula (3). It's natural to denote Z = X + Y.

Exercise 2.1: By analogy with exercise 2 .22 prove the consistence of formula (3).

Exercise 3.1: What happens if we multiply tensor X by the number zero and by the number minus one? What would you call the resulting tensors?

2 .8 Tensor Product

The tensor product is defined by a more tricky formula. Suppose we have tensor X of type (r, s) and tensor Y of type (p, q), then we can write:

$$Z^{i_1 \ldots i_{r+p}}_{j_1 \ldots j_{s+p}} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s} Y^{i_{r+1} \ldots i_{r+p}}_{j_{s+1} \ldots j_{s+p}}. \quad ...(1)$$

Formula (1) produces new tensor Z of the type (r + p, s + q). It is called the tensor product of X and Y and denoted Z = X ⊗ Y. Don't mix the tensor product and the cross product. They are different.

Exercise 2 .23: By analogy prove the consistence of formula (1).

Exercise 2 .24: Give an example of two tensors such that X ⊗ Y ≠ Y ⊗ X.

2 .9 Contraction

As we have seen above, the tensor product increases the number of indices. Usually the tensor Z = X ⊗ Y has more indices than X and Y. Contraction is an operation that decreases the number of indices. Suppose we have tensor X of the type (r + 1, s + 1). Then we can produce tensor Z of type (r, s) by means of the following formula:

$$Z^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \sum_{\rho=1}^{n} X^{i_1 \ldots i_{m-1}\, \rho\, i_m \ldots i_r}_{j_1 \ldots j_{k-1}\, \rho\, j_k \ldots j_s}. \quad ...(1)$$

What we do ? Tensor X has at least one upper index and at least one lower index. We choose the m-th upper index and replace it by the summation index ρ. In the same way, we replace the *k*-th lower index by ρ. Other *r* upper indices and *s* lower indices are free. They are numerated in some convenient way, say as in formula (1). Then we perform summation with respect to index ρ. The contraction is over. This operation is called a contraction with respect to *m*-th upper and k-th lower indices. Thus, if we have many upper an many lower indices in tensor X, we can perform various types of contractions to this tensor.

Task Prove the consistence of formula (1).

Interpret this formula as the contraction of the tensor product ax.

2 .10 Raising and Lowering Indices

Suppose that X is some tensor of type (r, s). Let's choose its α-th lower index: $X^{\dots\dots}_{\dots k\dots}$. The symbols used for the other indices are of no importance. Therefore, we denoted them by dots. Then let's consider the tensor product Y = g ⊗ X:

$$Y^{\dots pq\dots}_{\dots k\dots} = g^{pq} X^{\dots\dots}_{\dots k\dots}. \qquad \dots(1)$$

Here g is the dual metric tensor with the components g^{pq}. In the next step, let's contract (1) with respect to the pair of indices k and q. For this purpose we replace them both by s and perform the summation:

$$X^{\dots p\dots}_{\dots\dots} = \sum_{s=1}^{3} g^{ps} X^{\dots\dots}_{\dots s\dots}. \qquad \dots(2)$$

This operation (2) is called the index raising procedure. It is invertible. The inverse operation is called the index lowering procedure:

$$X^{\dots\dots}_{\dots p\dots} = \sum_{s=1}^{3} g_{ps} X^{\dots s\dots}_{\dots\dots}. \qquad \dots(3)$$

Like (2), the index lowering procedure (3) comprises two tensorial operations: the tensor product and contraction.

2 .11 Some Special Tensors and some useful Formulas

Kronecker symbol is a well known object. This is a two-dimensional array representing the unit matrix. It is determined as follows:

$$\delta^i_j = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j \end{cases} \qquad \dots(1)$$

We can determine two other versions of Kronecker symbol:

$$\delta_{ij} = d_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j \end{cases} \qquad \dots(2)$$

Exercise 2 .25: Prove that definition (1) is invariant under a change of basis, if we interpret the Kronecker symbol as a tensor. Show that both definitions in (2) are not invariant under a change of basis.

Exercise 2 .26: Lower index i of tensor (1). What tensorial object do you get as a result of this operation?

Notes

Exercise 2 .27: Likewise, raise index J in (1).

Another well known object is the Levi-Civita symbol. This is a three dimensional array determined by the following formula:

$$\in_{jkq} = \in^{jkq} = \begin{cases} 0, & \text{if among } j,k,q, \text{there at least two equal numbers;} \\ 1, & \text{if (j k q) is even permutation of numbers (1 2 3);} \\ -1, & \text{if (jkq) is odd permutation of numbers (1 2 3).} \end{cases} \quad \text{...(3)}$$

The Levi-Civita symbol (3) is not a tensor. However, we can produce two tensors by means of Levi-Civita symbol. The first of them

$$\omega_{ijk} = \sqrt{\det(g_{ij})} \in_{ijk} \quad \text{...(4)}$$

is known as the volume tensor. Another one is the dual volume tensor:

$$\omega^{ijk} = \sqrt{\det(g^{ij})} \in^{ijk} . \quad \text{...(5)}$$

Let's take two vectors x and y. Then using (4) we can produce covector a:

$$a_i = \sum_{j=1}^{3}\sum_{k=1}^{3} \omega_{ijk} x^j y^k . \quad \text{...(6)}$$

Then we can apply index raising procedure and produce vector a:

$$a^r = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} g^{ri} \omega_{ijk} x^j y^k . \quad \text{...(7)}$$

Formula (7) is known as formula for the vectorial product (cross product) in skew-angular basis.

Exercise 2 .28: Prove that the vector a with components (7) coincides with cross product of vectors x and y, i.e. a = [x, y].

2 .12 Summary

- Suppose we have a vector x and a covector a. Upon choosing some basis e_1, e_2, e_3, both of them have three coordinates: x^1, x^2, x^3 for vector x, and a_1, a_2, a_3 for covector a. Let's denote by $\langle a, x \rangle$ the following sum:

  $$\langle a, x \rangle = \sum_{i=1}^{3} a_i x^i .$$

  The sum is written in agreement with Einstein's tensorial notation. It is a number depending on the vector x and on the covector a. This number is called the scalar product of the vector x and the covector a. We use angular brackets for this scalar product in order to distinguish it from the scalar product of two vectors in E, which is also known as the dot product.

  Defining the scalar product $\langle a, x \rangle$ by means of sum we used the coordinates of vector x and of covector a, which are basis-dependent.

- A geometric object F in each basis represented by some square matrix F^i_j and such that components of its matrix F^i_j obey transformation rules $\tilde{F}^i_j = \sum_{p=1}^{3}\sum_{q=1}^{3} T^i_p S^q_j F^p_q$, and $F^i_j = \sum_{p=1}^{3}\sum_{q=1}^{3} S^i_p T^q_j \tilde{F}^p_q$. under a change of basis is called a linear operator.
- Vectors, covectors, and linear operators are all examples of tensors (though we have no definition of tensors yet). Now we consider another one class of tensorial objects. For the sake of clarity, let's denote by a one of such objects. In each basis e_1, e_2, e_3 this object is represented by some square 3 × 3 matrix a_{ij} of real numbers. Under a change of basis these numbers are transformed as follows $\tilde{a}_{ij} = \sum_{p=1}^{3}\sum_{q=1}^{3} S^p_i S^q_j a_{pq}$.
- The covectors, linear operators, and bilinear forms that we considered above were artificially constructed tensors. However there are some tensors of natural origin. Let's remember that we live in a space with measure. We can measure distance between points (hence we can measure length of vectors) and we can measure angles between two directions in our space. Therefore for any two vectors x and y we can define their natural scalar product (or dot product):

  $$(x, y) = |x|\ |y| \cos(\varphi)$$

  where φ is the angle between vectors x and y.
- The tensor product is defined by a more tricky formula. Suppose we have tensor X of type (r, s) and tensor Y of type (p, q), then we can write:

  $$Z^{i_1 \ldots i_{r+p}}_{j_1 \ldots j_{s+p}} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s} Y^{i_{r+1} \ldots i_{r+p}}_{j_{s+1} \ldots j_{s+p}}.$$

  The above formula produces new tensor Z of the type (r + p, s + q). It is called the tensor product of X and Y and denoted Z = X ⊗ Y. Don't mix the tensor product and the cross product. They are different.

2 .13 Keywords

Linear operator: A geometric object F in each basis represented by some square matrix F^i_j and such that components of its matrix F^i_j obey transformation rules $\tilde{F}^i_j = \sum_{p=1}^{3}\sum_{q=1}^{3} T^i_p S^q_j F^p_q$, and $F^i_j = \sum_{p=1}^{3}\sum_{q=1}^{3} S^i_p T^q_j \tilde{F}^p_q$. under a change of basis is called a linear operator.

Bilinear form: A geometric object a in each basis represented by some square matrix aij and such that components of its matrix aij obey transformation rules $\tilde{a}_{ij} = \sum_{p=1}^{3}\sum_{q=1}^{3} S^p_i S^q_j a_{pq}$ and aij = $\sum_{p=1}^{3}\sum_{q=1}^{3} T^p_i T^q_j \tilde{a}_{pq}$ under a change of basis is called a bilinear form.

Notes

2 .14 Self Assessment

1. Defining the scalar product $\langle a, x\rangle$ by means of sum we used the coordinates of vector x and of covector a, which are ....
2. A geometric object a in each basis represented by some square matrix aij and such that components of its matrix aij obey transformation rules $\tilde{a}_{ij} = \sum_{p=1}^{3}\sum_{q=1}^{3} S_i^p S_j^q a_{pq}$ and $a_{ij} = \sum_{p=1}^{3}\sum_{q=1}^{3} T_i^p T_j^q \tilde{a}_{pq}$ under a change of basis is called a ....
3. The coordinates of a vector are numerated by one upper index, which is called the ....
4. The number of indices and their positions determine the ...., i.e. the way the components of each particular tensor behave under a change of basis.

2 .15 Review Questions

1. Explain why the scalar product $\langle a, x\rangle$ is sometimes called the bilinear function of vectorial argument x and covectorial argument a. In this capacity, it can be denoted as f(a, x). Remember our discussion about functions with non-numeric arguments.
2. Derive $\begin{Vmatrix} y^1 \\ y^2 \\ y^3 \end{Vmatrix} = \begin{Vmatrix} F_1^1 & F_2^1 & F_3^1 \\ F_1^2 & F_2^2 & F_3^2 \\ F_1^3 & F_2^3 & F_3^3 \end{Vmatrix} \begin{Vmatrix} x^1 \\ x^2 \\ x^3 \end{Vmatrix}$ from $y^i = \sum_{j=1}^{3} F_j^i x^j$.
3. Let α be some real number and let x and y be two vectors. Prove the following properties of a linear operator:

   (1) $F(x + y) = F(x) + F(y)$,

   (2) $F(\alpha x) = \alpha F(x)$.
4. Find the matrix of composite operator $F \circ H$ if the matrices for F and H in the basis e_1, e_2, e_3 are known.
5. Remember the definition of the identity map in mathematics (see on-line Math. Encyclopedia) and define the identity operator id. Find the matrix of this operator.
6. Remember the definition of the inverse map in mathematics and define inverse operator F^{-1} for linear operator F. Find the matrix of this operator if the matrix of F is known.
7. Let a_{ij} be the matrix of some bilinear form a. Let's denote by b^{ij} components of inverse matrix for a_{ij}. Prove that matrix b^{ij} under a change of basis transforms like matrix of twice-contravariant tensor. Hence it determines tensor b of valency (2, 0). Tensor b is called a dual bilinear form for a.
8. By analogy with exercise $Y^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \alpha X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ prove the consistence of formula $Z^{i_1 \ldots i_{r+p}}_{j_1 \ldots j_{s+p}} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s} Y^{i_{r+1} \ldots i_{r+p}}_{j_{s+1} \ldots j_{s+p}}$.
9. Give an example of two tensors such that $X \otimes Y \neq Y \otimes X$.

10. Prove that definition $\delta^i_j = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j \end{cases}$ is invariant under a change of basis, if we interpret the Kronecker symbol as a tensor. Show that both definitions in $\delta_{ij} = d_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j \end{cases}$ are not invariant under a change of basis.

11. Lower index i of tensor $\delta^i_j = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j \end{cases}$ by means of $X^{\cdots\cdots}_{\cdots p \cdots} = \sum_{s=1}^{3} g_{ps} X^{\cdots s \cdots}_{\cdots\cdots}$. What tensorial object do you get as a result of this operation? Likewise, raise index J in $\delta^i_j = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \neq j \end{cases}$.

12. Prove that the vector a with components $a^r = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} g^{ri}\omega_{ijk} x^j y^k$ coincides with cross product of vectors x and y, i.e. a = [x, y].

Answers: Self Assessment

1. basis-dependent
2. bilinear form.
3. contravariant index.
4. transformation rules

2 .16 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 3 : Tensor Fields Differentiation of Tensors

CONTENTS

Objectives

After studying this unit, you will be able to:

- Discuss the tensor fields in Cartesian coordinates
- Describe the change of Cartesian coordinates system
- Explain the differentiation of tensors fields
- Discuss the gradient, divergency and rotor

Introduction

Cartesian tensors are widely used in various branches of continuum mechanics, such as fluid mechanics and elasticity. In classical continuum mechanics, the space of interest is usually 3-dimensional Euclidean space, as is the tangent space at each point. If we restrict the local coordinates to be Cartesian coordinates with the same scale centered at the point of interest, the metric tensor is the Kronecker delta. This means that there is no need to distinguish covariant and contra variant components, and furthermore there is no need to distinguish tensors and tensor densities. All Cartesian-tensor indices are written as subscripts. Cartesian tensors achieve considerable computational simplification at the cost of generality and of some theoretical insight.

3 .1 Tensor Fields in Cartesian Coordinates

The tensors that we defined in the previous unit are free tensors. Indeed, their components are arrays related to bases, while any basis is a triple of free vectors (not bound to any point). Hence, the tensors previously considered are also not bound to any point.

Notes

Now suppose we want to bind our tensor to some point in space, then another tensor to another point and so on. Doing so we can fill our space with tensors, one per each point. In this case, we say that we have a tensor field. In order to mark a point P to which our particular tensor is bound we shall write P as an argument:

$$X = X(P) \qquad ...(1)$$

Usually the valencies of all tensors composing the tensor field are the same. Let them all be of type (r, s). Then if we choose some basis e_1, e_2, e_3, we can represent any tensor of our tensor field as an array $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ with r + s indices:

$$X^{i_1 \ldots i_r}_{j_1 \ldots j_s} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(P). \qquad ...(2)$$

Thus, the tensor field (1) is a tensor-valued function with argument P being a point of three-dimensional Euclidean space E, and (2) is the basis representation for (1). For each fixed set of numeric values of indices $i_1, \ldots, i_r, j_1, \ldots, j_s$ in (2), we have a numeric function with a point-valued argument. Dealing with point-valued arguments is not so convenient, for example, if we want to calculate derivatives. Therefore, we need to replace P by something numeric. Remember that we have already chosen a basis. If, in addition, we fix some point O as an origin, then we get Cartesian coordinate system in space and hence can represent P by its radius-vector $r_p = \overline{OP}$ and by its coordinates x^1, x^2, x^3:

$$X^{i_1 \ldots i_r}_{j_1 \ldots j_s} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x^1, x^2, x^3). \qquad ...(3)$$

Conclusion. In contrast to free tensors, tensor fields are related not to bases, but to whole coordinate systems (including the origin). In each coordinate system they are represented by functional arrays, i.e. by arrays of functions (see (3)).

A functional array (3) is a coordinate representation of a tensor field (1). What happens when we change the coordinate system ? Dealing with (2), we need only to recalculate the components of the array $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$:

$$\tilde{X}^{i_1 \ldots i_r}_{j_1 \ldots j_s}(P) = \sum_{h_1,}^{3} \ldots \sum_{h_r}^{3} T^{i_1}_{h_1} \ldots T^{i_r}_{h_r} S^{k_1}_{j_1} \ldots S^{k_s}_{j_s} X^{h_1 \ldots h_r}_{k_1 \ldots k_s}(P). \qquad ...(4)$$

In the case of (3), we need to recalculate the components of the array $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ in the new basis

$$\tilde{X}^{i_1 \ldots i_r}_{j_1 \ldots j_s}(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3) = \sum_{h_1,}^{3} \ldots \sum_{h_r}^{3} T^{i_1}_{h_1} \ldots T^{i_r}_{h_r} S^{k_1}_{j_1} \ldots S^{k_s}_{j_s} X^{h_1 \ldots h_r}_{k_1 \ldots k_s}(x^1, x^2, x^3), \qquad ...(5)$$

We also need to express the old coordinates x^1, x^2, x^3 of the point P in right hand side of (5) through new coordinates of the same point:

$$\begin{cases} x^1 = x^1(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3), \\ x^2 = x^2(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3), \\ x^3 = x^3(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3). \end{cases} \qquad ...(6)$$

Formula (5) can be inverted :

$$X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x_1, x_2, x_3) = \sum_{h_1,}^{3} \ldots \sum_{h_r}^{3} S^{i_1}_{h_1} \ldots T^{i_r}_{h_r} S^{k_1}_{j_1} \ldots T^{k_s}_{j_s} \tilde{X}^{h_1 \ldots h_r}_{k_1 \ldots k_s}(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3). \qquad ...(7)$$

Notes

But now, apart from (7), we should have inverse formulas for (6) as well:

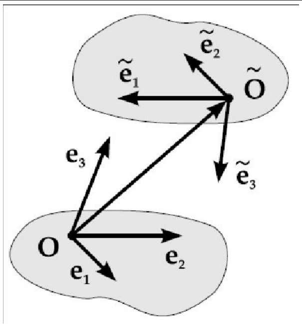
$$\begin{cases} \tilde{x}_1 = x^1(x^1, x^2, x^3); \\ \tilde{x}^2 = x^2(x^1, x^2, x^3); \\ \tilde{x}^3 = x^3(x^1, x^2, x^3). \end{cases} \qquad ...(8)$$

The couple of formulas (5) and (6), and another couple of formulas (7) and (8), in the case of tensor fields play the same role as transformation formulas in the case of free tensors.

3 .2 Change of Cartesian Coordinate System

Note that formulas (6) and (8) are written in abstract form. They only indicate the functional dependence of new coordinates of the point P from old ones and vice versa. Now we shall specify them for the case when one Cartesian coordinate system is changed to another Cartesian coordinate system. Remember that each Cartesian coordinate system is determined by some basis and some fixed point (the origin). We consider two Cartesian coordinate systems. Let the origins of the first and second systems be at the points O and $\tilde{O}$, respectively. Denote by e_1, e_2, e_3 the basis of the first coordinate system, and by $\tilde{e}_1, \tilde{e}_2, \tilde{e}_3$ the basis of the second coordinate system (see Fig. 16.1 below).

Figure 3 .1



Let P be some point in the space for whose coordinates we are going to derive the specializations of formulas (6) and (8). Denote by rP and $\tilde{r}_P$ the radius-vectors of this point in our two coordinate systems. Then rP = $\overrightarrow{OP}$ and $\tilde{r}_p = \overrightarrow{\tilde{O}P}$. Hence,

$$r_P = \overrightarrow{O\tilde{O}} + \tilde{r}_P. \qquad ...(1)$$

Vector $\overrightarrow{\tilde{O}O}$ determines the origin shift from the old to the new coordinate system. We expand this vector in the basis e_1, e_2, e_3:

$$a = \overrightarrow{O\tilde{O}} = \sum_{i=1}^{3} a^i e_i. \qquad ...(2)$$

Radius-vectors r_p and $\tilde{r}_P$ are expanded in the bases of their own coordinate systems:

$$r_P = \sum_{i=1}^{3} x^i e_i,$$

$$\tilde{r}_P = \sum_{i=1}^{3} \tilde{x}^i \tilde{e}_i, \qquad ...(3)$$

Exercise 1.1: Using (1), (2) and (3) derive the following formula relating the coordinates of the point P in the two coordinate systems in Fig. 16.1 :

$$x^i = a^i + \sum_{j=1}^{3} S^i_j \tilde{x}^j. \qquad ...(4)$$

Exercise 2.1: Derive the following inverse formula for (4):

$$\tilde{x}^i = \tilde{a}^i + \sum_{j=1}^{3} T^i_j x^j. \qquad ...(5)$$

Prove that ai in (4) and $\tilde{a}^i$ in (5) are related to each other as follows:

$$\tilde{a}^i = -\sum_{j=1}^{3} T^i_j x^j. \qquad a^i = -\sum_{j=1}^{3} S^i_j \tilde{a}^j. \qquad ...(6)$$

Explain the minus signs in these formulas. Formula (4) can be written in the following expanded form:

$$\begin{cases} x^1 = S^1_1\tilde{x}^1 + S^1_2\tilde{x}^2 + S^1_3\tilde{x}^3 + a^1, \\ x^2 = S^2_1\tilde{x}1 + S^2_2\tilde{x}^2 + S^2_3\tilde{x}^3 + a^2, \\ x^3 = S^3_1\tilde{x}^1 + S^3_2\tilde{x}^2 + S^3_3\tilde{x}^3 + a^3. \end{cases} \qquad ...(7)$$

This is the required specialization. In a similar way we can expand (5) :

$$\begin{cases} \tilde{x}^1 = T^1_1 x^1 + T^1_2 x^2 + T^1_3 x^3 + \tilde{a}^1, \\ \tilde{x}^2 = T^2_1 x^1 + T^2_2 x^2 + T^2_3 x^3 + \tilde{a}^2, \\ \tilde{x}^3 = T^3_1 x^1 + T^3_2 x^2 + T^3_3 x^3 + \tilde{a}^3. \end{cases} \qquad ...(8)$$

This is the required specialization. Formulas (7) and (8) are used to accompany the main transformation formulas.

3 .3 Differentiation of Tensor Fields

In this section we consider two different types of derivatives that are usually applied to tensor fields: differentiation with respect to spacial variables x^1, x^2, x^3 and differentiation with respect to external parameters other than x^1, x^2, x^3, if they are present. The second type of derivatives are simpler to understand. Let's consider them to start. Suppose we have tensor field X of type (r, s) and depending on the additional parameter t (for instance, this could be a time variable). Then, upon choosing some Cartesian coordinate system, we can write

$$\frac{\partial X^{i1\ldots ir}_{j1\ldots js}}{\partial t} = \lim_{h\to 0} \frac{X^{i_1\ldots i_r}_{j_1\ldots j_s}(t+h, x^1, x^2, x^3) - X^{i_1\ldots i_r}_{j_1\ldots j_s}(t, x^1, x^2, x^3)}{h}. \qquad ...(1)$$

The left hand side of (1) is a tensor since the fraction in right hand side is constructed by means of tensorial operations. Passing to the limit $h \to 0$ does not destroy the tensorial nature of this fraction since the transition matrices S and T are all time-independent.

Notes

Differentiation with respect to external parameters (like t in (1)) is a tensorial operation producing new tensors from existing ones.

Exercise 1.1: Give a more detailed explanation of why the time derivative (1) represents a tensor of type (r, s).

Now let's consider the spacial derivative of tensor field X, i.e. its derivative with respect to a spacial variable, e.g. with respect to x^1. Here we also can write

$$\frac{\partial X^{i_1 \ldots i_r}_{j_1 \ldots j_s}}{\partial t} = \lim_{h \to 0} \frac{X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x^1 + h, x^2, x^3) - X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x^1, x^2, x^3)}{h}, \qquad \ldots(2)$$

but in numerator of the fraction in the right hand side of (2) we get the difference of two tensors bound to different points of space: to the point P with coordinates x^1, x^2, x^3 and to the point P′ with coordinates $x^1 + h$, x^2, x^3. To which point should be attributed the difference of two such tensors ? This is not clear. Therefore, we should treat partial derivatives like (2) in a different way.

Let's choose some additional symbol, say it can be q, and consider the partial derivative of $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ with respect to the spacial variable x^q:

$$Y^{i_1 \ldots i_r}_{q j_1 \ldots j_s} = \frac{\partial X^{i_1 \ldots i_r}_{j_1 \ldots j_s}}{\partial x^q}. \qquad \ldots(3)$$

Partial derivatives (2), taken as a whole, form an (r + s + 1)-dimensional array with one extra dimension due to index q. We write it as a lower index in $Y^{i_1 \ldots i_r}_{q j_1 \ldots j_s}$ due to the following theorem concerning (3).

Theorem 1.1: For any tensor field X of type (r, s) partial derivatives (3) with respect to spacial variables x^1, x^2, x^3 in any Cartesian coordinate system represent another tensor field Y of the type (r, s + 1).

Thus differentiation with respect to x^1, x^2, x^3 produces new tensors from already existing ones. For the sake of beauty and convenience this operation is denoted by the nabla sign: $Y = \nabla X$. In index form this looks like

$$Y^{i_1 \ldots i_r}_{q j_1 \ldots j_s} = \nabla_q X^{i_1 \ldots i_r}_{j_1 \ldots j_s}. \qquad \ldots(4)$$

Simplifying the notations we also write:

$$\nabla_q = \frac{\partial}{\partial x^q}. \qquad \ldots(5)$$

Warning: Theorem 1.1 and the equality (5) are valid only for Cartesian coordinate systems. In curvilinear coordinates things are different.

Exercise 2.1: Prove theorem 1.1. For this purpose consider another Cartesian coordinate system $\tilde{x}^1, \tilde{x}^2, \tilde{x}^3$ related to x^1, x^2, x^3. Then in the new coordinate system consider the partial derivatives

$$\tilde{Y}^{i_1 \ldots i_r}_{q j_1 \ldots j_s} = \frac{\partial \tilde{X}^{i_1 \ldots i_r}_{j_1 \ldots j_s}}{\partial \tilde{x}^q} \qquad \ldots(6)$$

and derive relationships binding (6) and (3).

Notes

3 .4 Gradient, Divergency, and Rotor

The tensorial nature of partial derivatives established by theorem 1.1 is a very useful feature. We can apply it to extend the scope of classical operations of vector analysis. Let's consider the gradient, grad = ∇. Usually the gradient operator is applied to scalar fields, i.e. to functions $\varphi = \varphi(P)$ or $\varphi = \varphi(x^1, x^2, x^3)$ in coordinate form:

$$a_q = \nabla_q \varphi = \frac{\partial \varphi}{\partial x^q}. \qquad \text{...(1)}$$

Note that in (1) we used a lower index q for a_q. This means that a = grad φ is a covector. Indeed, according to theorem 1.1, the nabla operator applied to a scalar field, which is tensor field of type (0, 0), produces a tensor field of type (0, 1). In order to get the vector form of the gradient one should raise index q:

$$a^q = \sum_{i=1}^{3} g^{qi} a_i = \sum_{i=1}^{3} g^{qi} \nabla_i \varphi. \qquad \text{...(2)}$$

Let's write (2) in the form of a differential operator (without applying to φ):

$$\Delta^q = \sum_{i=1}^{3} g^{qi} \Delta_i. \qquad \text{...(3)}$$

In this form the gradient operator (3) can be applied not only to scalar fields, but also to vector fields, covector fields and to any other tensor fields.

Usually in physics we do not distinguish between the vectorial gradient ∇^q and the covectorial gradient ∇_q because we use orthonormal coordinates with ONB as a basis. In this case, dual metric tensor is given by unit matrix ($g^{ij} = \delta^{ij}$) and components of Δ^q and Δ_q coincide by value.

Divergency is the second differential operation of vector analysis. Usually it is applied to a vector field and is given by formula:

$$\text{div } X = \sum_{i=1}^{3} \Delta_i X^i. \qquad \text{...(4)}$$

As we see, (4) is produced by contraction (see section 16) from tensor $\nabla q X^i$. Therefore we can generalize formula (4) and apply divergency operator to arbitrary tensor field with at least one upper index:

$$(\text{div } X)^{\ldots\ldots}_{\ldots\ldots} = \sum_{s=1}^{3} \Delta_s X^{\ldots s \ldots}_{\ldots\ldots}. \qquad \text{...(5)}$$

The Laplace operator is defined as the divergency applied to a vectorial gradient of something, it is denoted by the triangle sign: Δ = div grad. From (3) and (5) for Laplace operator Δ we derive the following formula:

$$\Delta = \sum_{i=1}^{3} \sum_{j=1}^{3} g^{ij} \nabla_i \nabla_j. \qquad \text{...(6)}$$

Denote by $\square$ the following differential operator:

$$\square = \frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \Delta. \qquad \text{...(7)}$$

Notes

Operator (7) is called the d'Alambert operator or wave operator. In general relativity upon introducing the additional coordinate x^0 = ct one usually rewrites the d'Alambert operator in a form quite similar to (6).

And finally, let's consider the rotor operator or curl operator (the term "rotor" is derived from "rotation" so that "rotor" and "curl" have approximately the same meaning). The rotor operator is usually applied to a vector field and produces another vector field: Y = rotX. Here is the formula for the r-th coordinate of rot X:

$$(\text{rot } X)r = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} g^{ri}\omega_{ijk}\nabla^{j}X^{k}. \qquad ...(8)$$

Exercise 1.1: Formula (8) can be generalized for the case when X is an arbitrary tensor field with at least one upper index. By analogy with (5) suggest your version of such a generalization.

Note that formulas (6) and (8) for the Laplace operator and for the rotor are different from those that are commonly used. Here are standard formulas:

$$\Delta = \left(\frac{\partial}{\partial x^1}\right)^2 + \left(\frac{\partial}{\partial x^2}\right)^2 + \left(\frac{\partial}{\partial x^3}\right)^2, \qquad ...(9)$$

$$\text{rot } X = \det\begin{Vmatrix} e_1 & e_2 & e_3 \\ \dfrac{\partial}{\partial x^1} & \dfrac{\partial}{\partial x^2} & \dfrac{\partial}{\partial x^3} \\ X^1 & X^2 & X^3 \end{Vmatrix}. \qquad ...(10)$$

The truth is that formulas (6) and (8) are written for a general skew-angular coordinate system with a SAB as a basis. The standard formulas (10) are valid only for orthonormal coordinates with ONB as a basis.

Exercise 2.1: Show that in case of orthonormal coordinates, when $g^{ij} = \delta^{ij}$, formula (6) for the Laplace operator 4 reduces to the standard formula (9).

The coordinates of the vector rot X in a skew-angular coordinate system are given by formula (8). Then for vector rot X itself we have the expansion:

$$\text{rot } X = \sum_{r=1}^{3} (\text{rot } X)^r e_r. \qquad ...(11)$$

Exercise 3.1: Substitute (8) into (11) and show that in the case of a orthonormal coordinate system the resulting formula (11) reduces to (10).

3 .5 Summary

- Indeed, their components are arrays related to bases, while any basis is a triple of free vectors (not bound to any point). Hence, the tensors previously considered are also not bound to any point.

  Now suppose we want to bind our tensor to some point in space, then another tensor to another point and so on. Doing so we can fill our space with tensors, one per each point. In this case we say that we have a tensor field. In order to mark a point P to which our particular tensor is bound we shall write P as an argument:

  $$X = X(P)$$

Usually the valencies of all tensors composing the tensor field are the same. Let them all be of type (r, s). Then if we choose some basis e_1, e_2, e_3, we can represent any tensor of our tensor field as an array $X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$ with r + s indices:

$$X^{i_1 \ldots i_r}_{j1 \ldots j_s} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(P).$$

The left hand side is a tensor since the fraction in right hand side is constructed by means of tensorial operations. Passing to the limit $h \to 0$ does not destroy the tensorial nature of this fraction since the transition matrices S and T are all time-independent.

Differentiation with respect to external parameters is a tensorial operation producing new tensors from existing ones.

- The tensorial nature of partial derivatives established by theorem is a very useful feature. We can apply it to extend the scope of classical operations of vector analysis. Let's consider the gradient, grad = ∇. Usually the gradient operator is applied to scalar fields, i.e. to functions $\varphi = \varphi(P)$ or $\varphi = \varphi(x^1, x^2, x^3)$ in coordinate form:

$$a_q = \nabla_q \varphi = \frac{\partial \varphi}{\partial x^q}.$$

- The Laplace operator is defined as the divergency applied to a vectorial gradient of something, it is denoted by the triangle sign: Δ = div grad. From Laplace operator Δ, we derive the following formula:

$$\Delta = \sum_{i=1}^{3}\sum_{j=1}^{3} g^{ij}\nabla_i \nabla_j.$$

Denote by □ the following differential operator:

$$\square = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \Delta.$$

Above operator is called the d'Alambert operator or wave operator. In general relativity upon introducing the additional coordinate x^0 = ct one usually rewrites the d'Alambert operator in a form quite similar.

And finally, let's consider the rotor operator or curl operator (the term "rotor" is derived from "rotation" so that "rotor" and "curl" have approximately the same meaning). The rotor operator is usually applied to a vector field and produces another vector field: Y = rot X. Here is the formula for the r-th coordinate of rot X:

$$(\text{rot } X)r = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} g^{ri}\omega_{ijk}\nabla^j X^k.$$

3 .6 Keywords

Cartesian coordinate system in space and hence can represent P by its radius-vector $r_p = \overrightarrow{OP}$ and by its coordinates x^1, x^2, x^3.

Partial derivatives $\frac{\partial X^{i_1 \ldots i_r}_{j_1 \ldots j_s}}{\partial t} = \lim_{h \to 0} \frac{X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x^1 + h, x^2, x^3) - X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x^1, x^2, x^3)}{h}$, taken as a whole, form an

Notes

(r + s + 1)-dimensional array with one extra dimension due to index q. We write it as a lower index in $Y^{i_1 \ldots i_r}_{qj_1 \ldots j_s}$ due to the following theorem concerning $Y^{i_1 \ldots i_r}_{qj_1 \ldots j_s} = \frac{\partial X^{i_1 \ldots i_r}_{j_1 \ldots j_s}}{\partial x^q}$.

Divergency is the second differential operation of vector analysis. Usually it is applied to a vector field and is given by formula:

$$\operatorname{div} X = \sum_{i=1}^{3} \Delta_i X^i.$$

3 .7 Self Assessment

1. .... in space and hence can represent P by its radius-vector $r_p = \overrightarrow{OP}$ and by its coordinates x^1, x^2, x^3:

$$X^{i_1 \ldots i_r}_{j_1 \ldots j_s} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x^1, x^2, x^3).$$

2. A .... $X^{i_1 \ldots i_r}_{j_1 \ldots j_s} = X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x^1, x^2, x^3)$ is a coordinate representation of a tensor field $X = X(P)$.

3. .... $\frac{\partial X^{i_1 \ldots i_r}_{j_1 \ldots j_s}}{\partial t} = \lim_{h \to 0} \frac{X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x^1 + h, x^2, x^3) - X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(x^1, x^2, x^3)}{h}$, taken as a whole, form an (r + s + 1)-dimensional array with one extra dimension due to index q. We write it as a lower index in $Y^{i_1 \ldots i_r}_{qj_1 \ldots j_s}$ due to the following theorem concerning $Y^{i_1 \ldots i_r}_{qj_1 \ldots j_s} = \frac{\partial X^{i_1 \ldots i_r}_{j_1 \ldots j_s}}{\partial x^q}$.

3 .8 Review Questions

1. Using $r_P = \overrightarrow{O\tilde{O}} + \tilde{r}_P$, $a = \overrightarrow{O\tilde{O}} = \sum_{i=1}^{3} a^i e_i$, $\tilde{r}_P = \sum_{i=1}^{3} \tilde{x}^i \tilde{e}_i$, and $\tilde{e}_i = \sum_{j=1}^{3} S^j_i e_j$ derive the following formula relating the coordinates of the point P in the two coordinate systems.

$$x^i = a^i + \sum_{j=1}^{3} S^i_j \tilde{x}^j.$$

Compare $x^i = a^i + \sum_{j=1}^{3} S^i_j \tilde{x}^j$ with $x^j = \sum_{i=1}^{3} S^j_i \tilde{x}^i$. Explain the differences in these two formulas.

2. Give a more detailed explanation of why the time derivative $\frac{\partial X^{i1 \ldots ir}_{j1 \ldots js}}{\partial t} = \lim_{h \to 0} \frac{X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(t + h, x^1, x^2, x^3) - X^{i_1 \ldots i_r}_{j_1 \ldots j_s}(t, x^1, x^2, x^3)}{h}$ represents a tensor of type (r, s)

Notes

3. Prove theorem $\dfrac{\partial X^{i1\ldots ir}_{j1\ldots js}}{\partial t} = \lim_{h\to 0} \dfrac{X^{i_1\ldots i_r}_{j_1\ldots j_s}(t+h, x^1, x^2, x^3) - X^{i_1\ldots i_r}_{j_1\ldots j_s}(t, x^1, x^2, x^3)}{h}$. For this purpose consider another Cartesian coordinate system $\tilde{x}^1, \tilde{x}^2, \tilde{x}^3$ related to x^1, x^2, x^3 via $\begin{cases} x^1 = S^1_1\tilde{x}^1 + S^1_2\tilde{x}^2 + S^1_3\tilde{x}^3 + a^1, \\ x^2 = S^2_1\tilde{x}1 + S^2_2\tilde{x}^2 + S^2_3\tilde{x}^3 + a^2, \\ x^3 = S^3_1\tilde{x}^1 + S^3_2\tilde{x}^2 + S^3_3\tilde{x}^3 + a^3. \end{cases}$ and $\begin{cases} \tilde{x}^1 = T^1_1 x^1 + T^1_2 x^2 + T^1_3 x^3 + \tilde{a}^1, \\ \tilde{x}^2 = T^2_1 x^1 + T^2_2 x^2 + T^2_3 x^3 + \tilde{a}^2, \\ \tilde{x}^3 = T^3_1 x^1 + T^3_2 x^2 + T^3_3 x^3 + \tilde{a}^3. \end{cases}$. Then in the new coordinate system consider the partial derivatives

$$\tilde{Y}^{i_1\ldots i_r}_{qj_1\ldots j_s} = \frac{\partial \tilde{X}^{i_1\ldots i_r}_{j_1\ldots j_s}}{\partial \tilde{x}^q}$$

and derive relationships binding $\tilde{Y}^{i_1\ldots i_r}_{qj_1\ldots j_s} = \dfrac{\partial \tilde{X}^{i_1\ldots i_r}_{j_1\ldots j_s}}{\partial \tilde{x}^q}$ and $Y^{i_1\ldots i_r}_{qj_1\ldots j_s} = \dfrac{\partial X^{i_1\ldots i_r}_{j_1\ldots j_s}}{\partial x^q}$.

4. Formula $(\text{rot } X)r = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} g^{ri}\omega_{ijk}\nabla^j X^k$ can be generalized for the case when X is an arbitrary tensor field with at least one upper index. By analogy with $(\text{div } X)^{\ldots\ldots}_{\ldots\ldots} = \sum_{s=1}^{3} \Delta_s X^{\ldots\ldots s\ldots\ldots}_{\ldots\ldots\ldots\ldots}$. suggest your version of such a generalization.

Note that formulas $\Delta = \sum_{i=1}^{3}\sum_{j=1}^{3} g^{ij}\nabla_i\nabla_j$ and $(\text{rot } X)r = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} g^{ri}\omega_{ijk}\nabla^j X^k$ for the Laplace operator and for the rotor are different from those that are commonly used. Here are standard formulas:

$$\Delta = \left(\frac{\partial}{\partial x^1}\right)^2 + \left(\frac{\partial}{\partial x^2}\right)^2 + \left(\frac{\partial}{\partial x^3}\right)^2,$$

$$\text{rot } X = \det\begin{Vmatrix} e_1 & e_2 & e_3 \\ \dfrac{\partial}{\partial x^1} & \dfrac{\partial}{\partial x^2} & \dfrac{\partial}{\partial x^3} \\ X^1 & X^2 & X^3 \end{Vmatrix}.$$

The truth is that formulas $\Delta = \sum_{i=1}^{3}\sum_{j=1}^{3} g^{ij}\nabla_i\nabla_j$ and $(\text{rot } X)r = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} g^{ri}\omega_{ijk}\nabla^j X^k$ are written for a general skew-angular coordinate system with a SAB as a basis. The standard formulas $\text{rot } X = \det\begin{Vmatrix} e_1 & e_2 & e_3 \\ \dfrac{\partial}{\partial x^1} & \dfrac{\partial}{\partial x^2} & \dfrac{\partial}{\partial x^3} \\ X^1 & X^2 & X^3 \end{Vmatrix}$ are valid only for orthonormal coordinates with ONB as a basis.

Notes

5. Show that in case of orthonormal coordinates, when $g^{ij} = d^{ij}$, formula $\Delta = \sum_{i=1}^{3}\sum_{j=1}^{3} g^{ij}\nabla_i\nabla_j$ for the Laplace operator 4 reduces to the standard formula $\Delta = \left(\frac{\partial}{\partial x^1}\right)^2 + \left(\frac{\partial}{\partial x^2}\right)^2 + \left(\frac{\partial}{\partial x^3}\right)^2$.

   The coordinates of the vector rot X in a skew-angular coordinate system are given by formula $(\text{rot } X)r = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} g^{ri}\omega_{ijk}\nabla^j X^k$. Then for vector rot X itself we have the expansion:

$$\text{rot } X = \sum_{r=1}^{3} (\text{rot } X)^r e_r.$$

6. Substitute $(\text{rot } X)r = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3} g^{ri}\omega_{ijk}\nabla^j X^k$ into $\text{rot } X = \sum_{r=1}^{3} (\text{rot } X)^r e_r$ and show that in the case of a orthonormal coordinate system the resulting formula $\text{rot } X = \sum_{r=1}^{3} (\text{rot } X)^r e_r$ reduces to

$$\text{rot } X = \det \begin{Vmatrix} e_1 & e_2 & e_3 \\ \frac{\partial}{\partial x^1} & \frac{\partial}{\partial x^2} & \frac{\partial}{\partial x^3} \\ X^1 & X^2 & X^3 \end{Vmatrix}.$$

Answers: Self Assessment

1. Cartesian coordinate system
2. functional array
3. Partial derivatives

3 .9 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 4 : Tensor Fields in Curvilinear Coordinates

CONTENTS

Objectives

After studying this unit, you will be able to:

- Discuss the general idea of curvilinear coordinates
- Describe the auxiliary Cartesian coordinate system
- Explain the coordinate lines and coordinate grid.
- Discuss the moving frame of curvilinear coordinates
- Explain the formula for Christoffel symbols

Introduction

In the last unit, you have studied about tensor fields differentiation of tensors and tensor fields in Cartesian coordinates. Curvilinear coordinates are a coordinate system for Euclidean space in which the coordinate lines may be curved. These coordinates may be derived from a set of Cartesian coordinates by using a transformation that is locally invertible (a one-to-one map) at each point. This means that one can convert a point given in a Cartesian coordinate system to its curvilinear coordinates and back. The name curvilinear coordinates, coined by the French mathematician Lamé, derives from the fact that the coordinate surfaces of the curvilinear systems are curved.

Notes

4 .1 General idea of Curvilinear Coordinates

What are coordinates, if we forget for a moment about radius-vectors, bases and axes? What is the pure idea of coordinates? The pure idea is in representing points of space by triples of numbers. This means that we should have one to one map $P \rightleftarrows (y^1, y^2, y^3)$ in the whole space or at least in some domain, where we are going to use our coordinates y^1, y^2, y^3. In Cartesian coordinates this map $P \rightleftarrows (y^1, y^2, y^3)$ is constructed by means of vectors and bases. Arranging other coordinate systems one can use other methods. For example, in spherical coordinates $y^1 = r$ is a distance from the point P to the center of sphere, $y^2 = q$ and $y^3 = \varphi$ are two angles. By the way, spherical coordinates are one of the simplest examples of curvilinear coordinates. Furthermore, let's keep in mind spherical coordinates when thinking about more general and hence more abstract curvilinear coordinate systems.

4 .2 Auxiliary Cartesian Coordinate System

Now we know almost everything about Cartesian coordinates and almost nothing about the abstract curvilinear coordinate system y^1, y^2, y^3 that we are going to study. Therefore, the best idea is to represent each point P by its radius vector r_P in some auxiliary Cartesian coordinate system and then consider a map $r_P \rightleftarrows (y_1, y_2, y_3)$. The radius-vector itself is represented by three coordinates in the basis e_1, e_2, e_3 of the auxiliary Cartesian coordinate system:

$$r_P = \sum_{i=1}^{3} x^i e_i. \qquad ...(1)$$

Therefore, we have a one-to-one map $(x^1, x^2, x^3) \rightleftarrows (y^1, y^2, y^3)$. Hurrah! This is a numeric map. We can treat it numerically. In the left direction it is represented by three functions of three variables:

$$\begin{cases} x^1 = x^1(y^1, y^2, y^3), \\ x^2 = x^2(y^1, y^2, y^3), \\ x^3 = x^3(y^1, y^2, y^3). \end{cases} \qquad ...(2)$$

In the right direction we again have three functions of three variables:

$$\begin{cases} y^1 = y^1(x^1, x^2, x^3), \\ y^2 = y^2(x^1, x^2, x^3), \\ y^3 = y^3(x^1, x^2, x^3). \end{cases} \qquad ...(3)$$

Further we shall assume all functions in (2) and (3) to be differentiable and consider their partial derivatives. Let's denote

$$S^i_j = \frac{\partial x^i}{\partial y^j}, \qquad T^i_j = \frac{\partial y^i}{\partial x^j}. \qquad ...(4)$$

Partial derivatives (4) can be arranged into two square matrices S and T respectively. In mathematics such matrices are called Jacobi matrices. The components of matrix S in that form, as they are defined in (4), are functions of y^1, y^2, y^3. The components of T are functions of x^1, x^2, x^3:

$$S^i_j = S^i_j(y^1, y^2, y^3), \qquad T^i_j(x^1, x^2, x^3). \qquad ...(5)$$

Notes

However, by substituting (3) into the arguments of S^i_j, or by substituting (2) into the arguments of T^i_j, we can make them have a common set of arguments:

$$S^i_j, = S^i_j(x^1, x^2, x^3), \qquad T^i_j = T^i_j(x^1, x^2, x^3), \qquad ...(6)$$

$$S^i_j, = S^i_j(y^1, y^2, y^3), \qquad T^i_j = T^i_j(y^1, y^2, y^3), \qquad ...(7)$$

When brought to the form (6), or when brought to the form (7) (but not in form of (5)), matrices S and T are inverse of each other:

$$T = S^{-1}. \qquad ...(8)$$

This relationship (8) is due to the fact that numeric maps (2) and (3) are inverse of each other.

Exercise 1.1: You certainly know the following formula:

$$\frac{df(x^1(y), x^2(y), x^3(y))}{dy} = \sum_{i=1}^{3} f_i^{'}(x^1(y), x^2(y), x^3(y))\frac{dx^i(y)}{dy}, \text{ where } f_i^{'} = \frac{\partial f}{\partial x^i}.$$

It's for the differentiation of composite function. Apply this formula to functions (2) and derive the relationship (8).

4 .3 Coordinate Lines and the Coordinate Grid

Let's substitute (2) into (1) and take into account that (2) now assumed to contain differentiable functions. Then the vector-function

$$R(y^1, y^2, y^3) = r_P = \sum_{i=1}^{3} x^i(y^1, y^2, y^3)e_i \qquad ...(1)$$

is a differentiable function of three variables y^1, y^2, y^3. The vector-function $R(y^1, y^2, y^3)$ determined by (1) is called a basic vector-function of a curvilinear coordinate system. Let P_0 be some fixed point of space given by its curvilinear coordinates y_0^1, y_0^2, y_0^3. Here zero is not the tensorial index, we use it in order to emphasize that P_0 is fixed point, and that its coordinates y_0^1, y_0^2, y_0^3 are three fixed numbers. In the next step let's undo one of them, say first one, by setting

$$y^1 = y_0^1 + t, \qquad y^2 = y_0^2, \qquad y^3 = y_0^3. \qquad ...(2)$$

Substituting (2) into (1) we get a vector-function of one variable t:

$$R_1(t) = R(y_0^1 + t, y_0^2, y_0^3). \qquad ...(3)$$

If we treat t as time variable (though it may have a unit other than time), then (3) describes a curve (the trajectory of a particle). At time instant t = 0 this curve passes through the fixed point P_0. Same is true for curves given by two other vector-functions similar to (4):

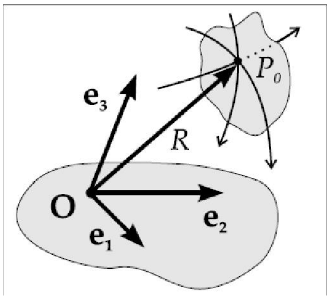
$$R_2(t) = R(y_0^1, y_0^2 + t, y_0^3), \qquad ...(4)$$

$$R_3(t) = R(y_0^1, y_0^2, y_0^3 + t). \qquad ...(5)$$

Notes

This means that all three curves given by vector-functions (3), (4), and (5) are intersected at the point P_0 as shown on Fig. 4 .1.

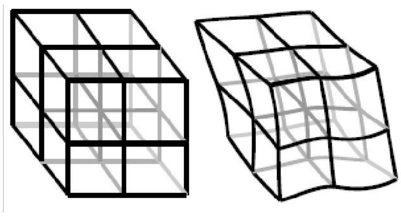
Figure 4 .1



Arrowheads on these lines indicate the directions in which parameter t increases. Curves (3), (4), and (5) are called coordinate lines. They are subdivided into three families. Curves within one family do not intersect each other. Curves from different families intersect so that any regular point of space is an intersection of exactly three coordinate curves (one per family).

Coordinate lines taken in whole form a coordinate grid. This is an infinitely dense grid. But usually, when drawing, it is represented as a grid with finite density. On Fig. 4 .2 the coordinate grid of curvilinear coordinates is compared to that of the Cartesian coordinate system.

Figure 4 .2



Indeed, meridians and parallels are coordinate lines of a spherical coordinate system. The parallels do not intersect, but the meridians forming one family of coordinate lines do intersect at the North and at South Poles. This means that North and South Poles are singular points for spherical coordinates.

Exercise 1.1: Remember the exact definition of spherical coordinates and find all singular points for them.

4 .4 Moving Frame of Curvilinear Coordinates

Let's consider the three coordinate lines shown on Fig. 2 .1 again. And let's find tangent vectors to them at the point P_0. For this purpose, we should differentiate vector-functions (3), (4), and (5) with respect to the time variable t and then substitute t = 0 into the derivatives:

$$E_i = \left.\frac{dR_i}{dt}\right|_{t=0} = \left.\frac{\partial R}{\partial y_i}\right|_{\text{at the point } P_0.} \qquad \text{...(1)}$$

Now let's substitute the expansion (1) into (1) and remember (4):

$$E_i = \frac{\partial R}{\partial y^i} = \sum_{j=1}^{3} \frac{\partial x^j}{\partial y^i} e_j = \sum_{j=1}^{3} S^j_i e_j. \qquad \text{...(2)}$$

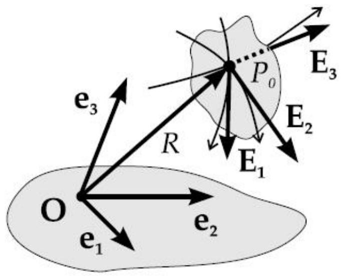
Notes

All calculations in (2) are still in reference to the point P_0. Though P_0 is a fixed point, it is an arbitrary fixed point. Therefore, the equality (2) is valid at any point. Now let's omit the intermediate calculations and write (2) as

$$E_i = \sum_{i=1}^{3} S_i^j e_j. \qquad ...(3)$$

They are strikingly similar, and det S ≠ 0. Formula (3) means that tangent vectors to coordinate lines E_1, E_2, E_3 form a basis (see Fig. 4 .3), matrices are transition matrices to this basis and back to the Cartesian basis.

Figure 4 .3



Despite obvious similarity of the formulas, there is some crucial difference of basis E_1, E_2, E_3 as compared to e_1, e_2, e_3. Vectors E_1, E_2, E_3 are not free. They are bound to that point where derivatives are calculated. And they move when we move this point. For this reason basis E_1, E_2, E_3 is called moving frame of the curvilinear coordinate system. During their motion the vectors of the moving frame E_1, E_2, E_3 are not simply translated from point to point, they can change their lengths and the angles they form with each other. Therefore, in general the moving frame E_1, E_2, E_3 is a skew-angular basis. In some cases vectors E_1, E_2, E_3 can be orthogonal to each other at all points of space. In that case we say that we have an orthogonal curvilinear coordinate system. Most of the well known curvilinear coordinate systems are orthogonal, e.g. spherical, cylindrical, elliptic, parabolic, toroidal, and others. However, there is no curvilinear coordinate system with the moving frame being ONB! We shall not prove this fact since it leads deep into differential geometry.

4 .5 Dynamics of Moving Frame

Thus, we know that the moving frame moves. Let's describe this motion quantitatively. Accordingly the components of matrix S in (3) are functions of the curvilinear coordinates y^1, y^2, y^3. Therefore, differentiating E_i with respect to y^j we should expect to get some nonzero vector $\partial E_i / \partial y^j$. This vector can be expanded back in moving frame E_1, E_2, E_3. This expansion is written as

$$\frac{\partial E_i}{\partial y^j} = \sum_{k=1}^{3} \Gamma_{ij}^{k} E_k. \qquad ...(1)$$

Formula (1) is known as the derivational formula. Coefficients Γ_{ij}^{k} in (1) are called Christoffel symbols or connection components.

Exercise 1.1: Relying upon formula (1) draw the vectors of the moving frame for cylindrical coordinates.

Notes

Task Do the same for spherical coordinates.

Exercise 3.1: Relying upon formula (1) and results of exercise 1.1. Calculate the Christoffel symbols for cylindrical coordinates.

Exercise 4.1: Do the same for spherical coordinates.

Exercise 5.1: Remember formula from which you derive

$$\text{Ei} = \frac{\partial R}{\partial y^i}. \qquad ...(2)$$

Substitute (2) into left hand side of the derivational formula (1) and relying on the properties of mixed derivatives prove that the Christoffel symbols are symmetric with respect to their lower indices: $\Gamma^k_{ij} = \Gamma^k_{ji}$.

Notes Christoffel symbols Γ^k_{ij} form a three-dimensional array with one upper index and two lower indices. However, they do not represent a tensor. We shall not prove this fact since it again leads deep into differential geometry.

4 .6 Formula for Christoffel Symbols

Let's take formula (3) and substitute it into both sides of (1). As a result we get the following equality for Christoffel symbols Γ^k_{ji}:

$$\sum_{q=1}^{3} \frac{\partial S^q_i}{\partial y^j} e_q = \sum_{k=1}^{3}\sum_{q=1}^{3} \Gamma^k_{ij} S^q_k e_q. \qquad ...(1)$$

Cartesian basis vectors e_q do not depend on y^j ; therefore, they are not differentiated when we substitute (3) into (1). Both sides of (1) are expansions in the base e_1, e_2, e_3 of the auxiliary Cartesian coordinate system. Due to the uniqueness of such expansions we have the following equality derived from (1):

$$\frac{\partial S^q_i}{\partial y^j} = \sum_{k=1}^{3} \Gamma^k_{ij} S^q_k. \qquad ...(2)$$

Exercise 1.1: Using concept of the inverse matrix ($T = S^{-1}$) derive the following formula for the Christoffel symbols Γ^k_{ij} from (2):

$$\Gamma^k_{ij} = \sum_{q=1}^{3} T^k_q \frac{\partial S^q_i}{\partial y^j}. \qquad ...(3)$$

Due to this formula (3) can be transformed in the following way:

$$\Gamma^k_{ij} = \sum_{q=1}^{3} T^k_q \frac{\partial S^q_i}{\partial y^j} = \sum_{q=1}^{3} T^k_q \frac{\partial^2 x^q}{\partial y^i \partial y^j} = \sum_{q=1}^{3} T^k_q \frac{\partial S^q_j}{\partial y^i}. \qquad ...(4)$$

Formulas (4) are of no practical use because they express Γ^k_{ij} through an external thing like transition matrices to and from the auxiliary Cartesian coordinate system. However, they will help us below in understanding the differentiation of tensors.

4 .7 Tensor Fields in Curvilinear Coordinates

As we remember, tensors are geometric objects related to bases and represented by arrays if some basis is specified. Each curvilinear coordinate system provides us a numeric representation for points, and in addition to this it provides the basis.

This is the moving frame. Therefore, we can refer tensorial objects to curvilinear coordinate systems, where they are represented as arrays of functions:

$$X^{i_1 \dots i_r}_{j_1 \dots j_s} = X^{i_1 \dots i_r}_{j_1 \dots j_s}(y^1, y^2, y^3). \qquad \text{...(1)}$$

We also can have two curvilinear coordinate systems and can pass from one to another by means of transition functions:

$$\begin{cases} \tilde{y}^1 = \tilde{y}^1(y^1, y^2, y^3), \\ \tilde{y}^2 = \tilde{y}^2(y^1, y^2, y^3), \\ \tilde{y}^3 = \tilde{y}^3(y^1, y^2, y^3), \end{cases} \qquad \begin{cases} y^1 = y^1(\tilde{y}^1, \tilde{y}^2, \tilde{y}^3), \\ y^2 = y^2(\tilde{y}^1, \tilde{y}^2, \tilde{y}^3), \\ y^3 = y^3(\tilde{y}^1, \tilde{y}^2, \tilde{y}^3). \end{cases} \qquad \text{...(2)}$$

If we call $\tilde{y}^1, \tilde{y}^2, \tilde{y}^3$ the new coordinates, and y^1, y^2, y^3 the old coordinates, then transition matrices S and T are given by the following formulas:

$$S^i_j = \frac{\partial y^i}{\partial \tilde{y}^j}, \qquad T^i_j = \frac{\partial \tilde{y}^i}{\partial y^j}. \qquad \text{...(3)}$$

They relate moving frames of two curvilinear coordinate systems:

$$\tilde{E}_i = \sum_{j=1}^{3} S^j_i E_j, \qquad E_j = \sum_{i=1}^{3} T^i_j \tilde{E}_i. \qquad \text{...(4)}$$

Exercise 1.1: Derive (3) from (4) and (2) using some auxiliary Cartesian coordinates with basis e_1, e_2, e_3 as intermediate coordinate system:

$$(E_1, E_2, E_3) \underset{T}{\overset{S}{\rightleftarrows}} (e_1, e_2, e_3) \underset{T}{\overset{\tilde{S}}{\rightleftarrows}} (\tilde{E}_1, \tilde{E}_2, \tilde{E}_3) \qquad \text{...(5)}$$

Transformation formulas for tensor fields for two curvilinear coordinate systems are the same:

$$\tilde{X}^{i_1 \dots i_r}_{j_1 \dots j_s}(\tilde{y}^1, \tilde{y}^2, \tilde{y}^3) = \sum_{\substack{h_1, \dots h_r \\ k_1, \dots ks}}^{3} \dots \sum^{3} T^{i_1}_{h_1} \dots T^{i_r}_{h_r} S^{k_1}_{j_1} \dots S^{k_s}_{j_s} X^{h_1 \dots h_r}_{k_1 \dots k_s}(y^1, y^2, y^3), \qquad \text{...(6)}$$

$$\tilde{X}^{i_1 \dots i_r}_{j_1 \dots j_s}(y_1, y_2, y_3) = \sum_{\substack{h_1, \dots h_r \\ k_1, \dots ks}}^{3} \dots \sum^{3} S^{i_1}_{h_1} \dots S^{i_r}_{h_r} T^{k_1}_{j_1} \dots T^{k_s}_{j_s} \tilde{X}^{h_1 \dots h_r}_{k_1 \dots k_s}(\tilde{y}^1, \tilde{y}^2, \tilde{y}^3). \qquad \text{...(7)}$$

Notes

4 .8 Differentiation of Tensor Fields in Curvilinear Coordinates

We already know how to differentiate tensor fields in Cartesian coordinates (see section 21). We know that operator ∇ produces tensor field of type (r, s + 1) when applied to a tensor field of type (r, s). The only thing we need now is to transform ∇ to a curvilinear coordinate system. In order to calculate tensor ∇X in curvilinear coordinates, let's first transform X into auxiliary Cartesian coordinates, then apply ∇, and then transform ∇X back into curvilinear coordinates:

$$\begin{array}{ccc} X^{h_1\ldots h_r}_{k_1\ldots k_s}(y^1,y^2,y^3) & \xrightarrow{S,T} & X^{h_1\ldots h_r}_{k_1\ldots k_s}(x^1,x^2,x^3) \\ \downarrow^{\nabla p} & & \downarrow^{\Delta q=\partial/\partial x^q} \\ \nabla p X^{i_1\ldots i_r}_{j_1\ldots j_s}(y^1,y^2,y^3) & \xrightarrow{T,S} & \nabla_q X^{h_1\ldots h_r}_{k_1\ldots k_s}(x^1,x^2,x^3) \end{array} \qquad \ldots(1)$$

Matrices are used in (1). We know that the transformation of each index is a separate multiplicative procedure. When applied to the α-th upper index, the whole chain of transformations (1) looks like

$$\nabla_p X^{\ldots i_\alpha \ldots}_{\ldots\ldots\ldots} = \sum_{q=1}^{3} S^q_p \ldots \sum_{h_\alpha=1}^{3} T^{i_\alpha}_{h_\alpha} \ldots \nabla_q \ldots \sum_{m_\alpha=1}^{3} S^{h_\alpha}_{m_\alpha} \ldots X^{\ldots m_\alpha \ldots}_{\ldots\ldots\ldots}. \qquad \ldots(2)$$

Note that $\nabla q = \partial/\partial x^q$ is a differential operator and we have

$$\sum_{q=1}^{3} S^q_p \frac{\partial}{\partial xq} = \frac{\partial}{\partial x^p}. \qquad \ldots(3)$$

Any differential operator when applied to a product produces a sum with as many summands as there were multiplicand in the product. Here is the summand produced by term $S^{h_\alpha}_{m_\alpha}$ in formula (2):

$$\nabla_p X^{\ldots i_\alpha \ldots}_{\ldots\ldots\ldots} = \ldots + \sum_{m_\alpha=1}^{3}\sum_{h_\alpha=1}^{3} T^{i_\alpha}_{h_\alpha} \frac{S^{h_\alpha}_{m_\alpha}}{\partial y^p} X^{\ldots m_\alpha \ldots}_{\ldots\ldots\ldots} + \ldots.. \qquad \ldots(4)$$

We can transform it into the following equality:

$$\nabla_p X^{\ldots i_\alpha \ldots}_{\ldots\ldots\ldots} = \ldots + \sum_{m_\alpha=1}^{3} \Gamma^{i_\alpha}_{pm_\alpha} X^{\ldots m_\alpha \ldots}_{\ldots\ldots\ldots} + \ldots \qquad \ldots(5)$$

Now let's consider the transformation of the ″-th lower index in (1):

$$\nabla_p X^{\ldots\ldots\ldots}_{\ldots j_\alpha \ldots} = \sum_{q=1}^{3} S^q_p \ldots \sum_{k_\alpha=1}^{3} S^{k_\alpha}_{j_\alpha} \ldots \nabla_q \ldots \sum_{n_\alpha=1}^{3} T^{n_\alpha}_{k_\alpha} \ldots X^{\ldots\ldots\ldots}_{\ldots n_\alpha \ldots}. \qquad \ldots(6)$$

Applying (3) to (6) with the same logic as in deriving (4) we get

$$\nabla_p X^{\ldots\ldots\ldots}_{\ldots j_\alpha \ldots} = \ldots + \sum_{n_\alpha=1}^{3}\sum_{k_\alpha=1}^{3} S^{k_\alpha}_{j_\alpha} \frac{T^{n_\alpha}_{k\alpha}}{\partial y^p} X^{\ldots\ldots\ldots}_{\ldots n_\alpha \ldots} + \ldots \qquad \ldots(7)$$

In order to simplify (7) we need the following formula derived :

$$\Gamma^k_{ij} = -\sum_{q=1}^{3} S^q_i \frac{\partial T^k_q}{\partial y^j}. \qquad \ldots(8)$$

Notes

Applying (8) to (7) we obtain

$$\nabla_p X^{\cdots\cdots\cdots}_{\ldots j_\alpha \ldots} = \ldots - \sum_{n_\alpha=1}^{3} G^{n_\alpha}_{pj_\alpha} X^{\cdots\cdots\cdots}_{\ldots n_\alpha \ldots} + \ldots \qquad \ldots(9)$$

Now we should gather (5), (9), and add the term produced when rq in (2) (or equivalently in (4)) acts upon components of tensor X. As a result we get the following general formula for $\nabla_p X^{i_1 \ldots i_r}_{j_1 \ldots j_s}$:

$$\nabla_p X^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \frac{\partial X^{i_1 \ldots i_r}_{j_1 \ldots j_s}}{\partial y^p} + \sum_{\alpha=1}^{r} \sum_{m_\alpha=1}^{3} G^{i_\alpha}_{pm_\alpha} X^{i_1 \ldots m_\alpha \ldots i_r}_{j_1 \ldots\ldots\ldots\ldots j_s} - \sum_{\alpha=1}^{s} \sum_{n_\alpha=1}^{3} \Gamma^{n_\alpha}_{pj_\alpha} X^{i_1 \ldots\ldots\ldots i_r}_{j_1 \ldots\ldots\ldots j_s}. \qquad \ldots(10)$$

The operator ∇p determined by this formula is called the covariant derivative.

Exercise 1.1: Apply the general formula (10) to a vector field and calculate the covariant derivative $\nabla_p X^q$.

Exercise 2.1: Apply the general formula (10) to a covector field and calculate the covariant derivative $\nabla_p X_q$.

Exercise 3.1: Apply the general formula (10) to an operator field and find $\nabla_p F^q_m$. Consider special case when ∇_p is applied to the Kronecker symbol δ^q_m.

Exercise 4.1: Apply the general formula (10) to a bilinear form and find $\nabla_p a_{qm}$.

Exercise 5.1: Apply the general formula (10) to a tensor product a ⊗ x for the case when x is a vector and a is a covector. Verify formula ∇(a ⊗ x) = ∇a ∇ x + a ⊗ ∇x.

Exercise 6.1: Apply the general formula (10) to the contraction C(F) for the case when F is an operator field. Verify the formula ∇C(F) = C(∇F).

4 .9 Concordance of Metric and Connection

Let's remember that we consider curvilinear coordinates in Euclidean space E. In this space, we have the scalar product and the metric tensor.

Exercise 1.1: Transform the metric tensor to curvilinear coordinates using transition matrices and show that here it is given by formula

$$g_{ij} = (E_i, E_j). \qquad \ldots(1)$$

In Cartesian coordinates all components of the metric tensor are constant since the basis vectors e_1, e_2, e_3 are constant. The covariant derivative (10) in Cartesian coordinates reduces to differentiation $\nabla_p = \partial/\partial x^p$. Therefore,

$$\nabla p g_{ij} = 0. \qquad \ldots(2)$$

But ∇g is a tensor. If all of its components in some coordinate system are zero, then they are identically zero in any other coordinate system (explain why). Therefore the identity (2) is valid in curvilinear coordinates as well.

The identity is known as the concordance condition for the metric g_{ij} and connection Γ^k_{ij}. It is very important for general relativity.

Remember that the metric tensor enters into many useful formulas for the gradient, divergency, rotor, and Laplace operator. What is important is that all of these formulas remain valid in curvilinear coordinates, with the only difference being that you should understand that ∇_p is not the partial derivative $\partial/\partial x^p$, but the covariant derivative in the sense of formula (10).

Exercise 2.1: Calculate rot A, div H, grad φ (vectorial gradient) in cylindrical and spherical coordinates.

Exercise 3.1: Calculate the Laplace operator Δφ applied to the scalar field φ in cylindrical and in spherical coordinates.

4 .10 Summary

- What are coordinates, if we forget for a moment about radius-vectors, bases and axes ? What is the pure idea of coordinates? The pure idea is in representing points of space by triples of numbers. This means that we should have one to one map $P \rightleftarrows (y^1, y^2, y^3)$ in the whole space or at least in some domain, where we are going to use our coordinates y^1, y^2, y^3. In Cartesian coordinates this map $P \rightleftarrows (y^1, y^2, y^3)$ is constructed by means of vectors and bases. Arranging other coordinate systems one can use other methods. For example, in spherical coordinates $y^1 = r$ is a distance from the point P to the center of sphere, $y^2 = q$ and $y^3 = φ$ are two angles. By the way, spherical coordinates are one of the simplest examples of curvilinear coordinates. Furthermore, let's keep in mind spherical coordinates when thinking about more general and hence more abstract curvilinear coordinate systems.
- Now we know almost everything about Cartesian coordinates and almost nothing about the abstract curvilinear coordinate system y^1, y^2, y^3 that we are going to study. Therefore, the best idea is to represent each point P by its radius vector r_P in some auxiliary Cartesian coordinate system and then consider a map $r_P \rightleftarrows (y_1, y_2, y_3)$. The radius-vector itself is represented by three coordinates in the basis e_1, e_2, e_3 of the auxiliary Cartesian coordinate system:

  $$r_P = \sum_{i=1}^{3} x^i e_i.$$

  Therefore, we have a one-to-one map $(x^1, x^2, x^3) \rightleftarrows (y^1, y^2, y^3)$. Hurrah! This is a numeric map. We can treat it numerically. In the left direction it is represented by three functions of three variables:

  $$\begin{cases} x^1 = x^1(y^1, y^2, y^3), \\ x^2 = x^2(y^1, y^2, y^3), \\ x^3 = x^3(y^1, y^2, y^3). \end{cases}$$

- Cartesian basis vectors e_q do not depend on y^j ; therefore, they are not differentiated. Both sides are expansions in the base e_1, e_2, e_3 of the auxiliary Cartesian coordinate system.

4 .11 Keywords

Spherical coordinates are one of the simplest examples of curvilinear coordinates.

Vector-function: Then the vector-function

$$R(y^1, y^2, y^3) = r_P = \sum_{i=1}^{3} x^i(y^1, y^2, y^3) e_i$$

is a differentiable function of three variables y^1, y^2, y^3.

Christoffel symbols Γ^k_{ij} form a three-dimensional array with one upper index and two lower indices.

4 .12 Self Assessment

1. .... are one of the simplest examples of curvilinear coordinates.
2. The .... itself is represented by three coordinates in the basis e_1, e_2, e_3 of the auxiliary Cartesian coordinate system $r_P = \sum_{i=1}^{3} x^i e_i$.
3. Coordinate lines taken in whole form a coordinate grid. This is an infinitely dense grid. But usually, when drawing, it is represented as a grid with ....
4. The parallels do not intersect, but the .... one family of coordinate lines do intersect at the North and at South Poles. This means that North and South Poles are singular points for spherical coordinates.
5. .... Γ^k_{ij} form a three-dimensional array with one upper index and two lower indices.
6. .... e_q do not depend on y^j; therefore, they are not differentiated when we substitute $E_i = \sum_{i=1}^{3} S^j_i e_j$ into $\frac{\partial E_i}{\partial y^j} = \sum_{k=1}^{3} \Gamma^k_{ij} E_k$. Both sides of $\sum_{q=1}^{3} \frac{\partial S^q_i}{\partial y^j} e_q = \sum_{k=1}^{3}\sum_{q=1}^{3} \Gamma^k_{ij} S^q_k e_q$ are expansions in the base e_1, e_2, e_3 of the auxiliary Cartesian coordinate system.

4 .13 Review Questions

1. Remember the exact definition of spherical coordinates and find all singular points for them.
2. Relying upon formula $\frac{\partial E_i}{\partial y^j} = \sum_{k=1}^{3} \Gamma^k_{ij} E_k$, calculate the Christoffel symbols for cylindrical coordinates.
3. Remember formula $E_i = \frac{\partial R}{\partial y^i} = \sum_{j=1}^{3} \frac{\partial x^j}{\partial y^i} e_j = \sum_{j=1}^{3} S^j_i e_j$ from which you derive

$$E_i = \frac{\partial R}{\partial y^i}.$$

Answers: Self Assessment

1.	Spherical coordinates	2.	radius-vector
3.	finite density.	4.	meridians forming
5.	Christoffel symbols	6.	Cartesian basis vectors

4 .14 Further Readings

Notes

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 5 : Theory of Space Curves

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define Arc Length
- Discuss Curvature and Fenchel's Theorem
- Explain The Unit Normal Bundle and Total Twist
- Define Moving Frames
- Describe Curves at a Non-inflexional Point and the Frenet Formulas
- Explain Plane Convex Curves and the Four Vertex Theorem

Introduction

In the last unit, you have studied about Curvilinear coordinates. These coordinates may be derived from a set of Cartesian coordinates by using a transformation that is locally invertible (a one-to-one map) at each point. The term curve has several meanings in non-mathematical language as well. For example, it can be almost synonymous with mathematical function or graph of a function. An arc or segment of a curve is a part of a curve that is bounded by two distinct end points and contains every point on the curve between its end points. Depending on how the arc is defined, either of the two end points may or may not be part of it. When the arc is straight, it is typically called a line segment.

5 .1 Arc Length

Notes

A parametrized curve in Euclidean three-space e^3 is given by a vector function

$$x(t) = (x_1(t), x_2(t), x_3(t))$$

that assigns a vector to every value of a parameter t in a domain interval [a, b]. The coordinate functions of the curve are the functions $x_i(t)$. In order to apply the methods of calculus, we suppose the functions $x_i(t)$ to have as many continuous derivatives as needed in the following treatment.

For a curve x(t), we define the first derivative x′(t) to be the limit of the secant vector from x(t) to x(t+h) divided by h as h approaches 0, assuming that this limit exists. Thus,

$$x'(t) = \lim_{h\to 0}\left(\frac{x(t+h)-x(t)}{h}\right).$$

The first derivative vector x′(t) is tangent to the curve at x(t). If we think of the parameter t as representing time and we think of x(t) as representing the position of a moving particle at time t, then x′(t) represents the velocity of the particle at time t. It is straightforward to show that the coordinates of the first derivative vector are the derivatives of the coordinate functions, i.e.

$$x'(t) = (x_1'(t), x_2'(t), x_3'(t)).$$

For most of the curves we will be concerning ourselves with, we will make the "genericity assumption" that x′(t) is non-zero for all t. Lengths of polygons inscribed in x as the lengths of the sides of these polygons tend to zero. By the fundamental theorem of calculus, this limit can be expressed as the integral of the speed s′(t) = |x′(t)| between the parameters of the end-points of the curve, a and b. That is,

$$s(b) - s(a) = \int_a^b |x'(t)|\,dt = \int_a^b \sqrt{\sum_{i=1}^{3} x_i'(t)^2}\,dt.$$

For an arbitrary value t ∈ (a, b), we may define the distance function

$$s(t) - s(a) = \int_a^t |x'(t)|\,dt,$$

which gives us the distance from a to t along the curve.

Notice that this definition of arc length is independent of the parametrization of the curve. If we define a function v(t) from the interval [a, b] to itself such that v(a) = a, v(b) = b and v′(t) > 0, then we may use the change of variables formula to express the arc length in terms of the new parameter v:

$$\int_a^b |x'(t)|\,dt = \int_v (a) = a^v(b) = b\left|x'(v(t))\right| v'(t)dt = \int_a^b \left|x'(v)\right| dv.$$

We can also write this expression in the form of differentials:

$$ds = |x'(t)|\,dt = |x'(v)|\,dv.$$

This differential formalism becomes very significant, especially when we use it to study surfaces and higher dimensional objects, so we will reinterpret results that use integration or differentiation in differential notation as we go along. For example, the statement $s'(t) = \sqrt{\sum_{i=1}^{3} x_i'(t)^2}$ can be rewritten as

$$\left(\frac{ds}{dt}\right)^2 = \sum_{i=1}^{3}\left(\frac{dx_i}{dt}\right)^2,$$

and this may be expressed in the form

$$ds^2 = \sum_{i=1}^{3} dx_i^2,$$

which has the advantage that it is independent of the parameter used to describe the curve. ds is called the element of arc. It can be visualized as the distance between two neighboring points.

One of the most useful ways to parametrize a curve is by the arc length s itself. If we let $s = s(t)$, then we have

$$s'(t) = |x'(t)| = |x'(s)|\,s'(t),$$

from which it follows that $|x'(s)| = 1$ for all s. So the derivative of x with respect to arc length is always a unit vector.

This parameter s is defined up to the transformation $s \to \pm s + c$, where c is a constant. Geometrically, this means the freedom in the choice of initial point and direction in which to traverse the curve in measuring the arc length.

Exercise 1: One of the most important space curves is the circular helix

$$x(t) = (a \cos t, a \sin t, bt),$$

where $a \neq 0$ and b are constants. Find the length of this curve over the interval $[0, 2\pi]$.

Exercise 2: Find a constant c such that the helix

$$x(t) = (a \cos(ct), a \sin(ct), bt)$$

is parametrized by arclength, so that $|x'(t)| = 1$ for all t.

Exercise 3: The astroid is the curve defined by

$$x(t) = (a \cos^3 t, a \sin^3 t, 0),$$

on the domain $[0, 2\pi]$. Find the points at which x(t) does not define an immersion, i.e., the points for which $x'(t) = 0$.

Exercise 4: The trefoil curve is defined by

$$x(t) = ((a + b \cos(3t)) \cos(2t), (a + b \cos(3t)) \sin(2t), b \sin(3t)),$$

where a and b are constants with $a > b > 0$ and $0 \le t \le 2\pi$. Sketch this curve, and give an argument to show why it is knotted, i.e. why it cannot be deformed into a circle without intersecting itself in the process.

Exercise 5: (For the serious mathematician) Two parametrized curves x(t) and y(u) are said to be equivalent if there is a function u(t) such that $u'(t) > 0$ for all $a < t < b$ and such that $y(u(t)) = x(t)$. Show that relation satisfies the following three properties:

1. Every curve x is equivalent to itself
2. If x is equivalent to y, then y is equivalent to x

3. If x is equivalent to y and if y is equivalent to z, then x is equivalent to z

A relation that satisfies these properties is called an equivalence relation. Precisely speaking, a curve is considered be an equivalence class of parametrized curves.

5 .2 Curvature and Fenchel's Theorem

If x is an immersed curve, with $x'(t) \neq 0$ for all t in the domain, then we may define the unit tangent vector T(t) to be $\frac{x'(t)}{|x'(t)|}$. If the parameter is arclength, then the unit tangent vector T(s) is given simply by x′(s). The line through $x(t_0)$ in the direction of $T(t_0)$ is called the tangent line at $x(t_0)$. We can write this line as $y(u) = x(t_0) + uT(t_0)$, where u is a parameter that can take on all real values.

Since $T(t) \cdot T(t) = 1$ for all t, we can differentiate both sides of this expression, and we obtain $2T'(t) \cdot T(t) = 0$. Therefore T′(t) is orthogonal to T(t). The curvature of the space curve x(t) is defined by the condition $\kappa(t) = \frac{T'(t)}{|x'(t)|}$, so $= \kappa(t)s'(t) = |T'(t)|$. If the parameter is arclength, then $x'(s) = T(s)$ and $\kappa(s) = |T'(s)| = |x''(s)|$.

Proposition 1. If κ(t) = 0 for all t, then the curve lies along a straight line.

Proof. Since κ(t) = 0, we have T′(t) = 0 and T(t) = a, a constant unit vector. Then $x'(t) = s'(t)T(t) = s'(t)a$, so by integrating both sides of the equation, we obtain $x(t) = s(t)a + b$ for some constant b. Thus, x(t) lies on the line through b in the direction of a.

Curvature is one of the simplest and at the same time one of the most important properties of a curve. We may obtain insight into curvature by considering the second derivative vector x″(t), often called the acceleration vector when we think of x(t) as representing the path of a particle at time t. If the curve is parametrized by arclength, then $x'(s) \cdot x'(s) = 1$ so $x''(s) \cdot x'(t) = 0$ and $\kappa(s) = |x''(s)|$. For a general parameter t, we have $x'(t) = s'(t)T(t)$ so $x''(t) = s''(t)T(t) + s'(t)T'(t)$. If we take the cross product of both sides with x′(t) then the first term on the right is zero since x′(t) is parallel to T(t). Moreover x′(t) is perpendicular to T′(t) so

$$|T'(t) \times x'(t)| = |T'(t)|\,|x'(t)| = s'(t)^2\kappa(t)\,.$$

Thus,

$$x''(t) \times x'(t) = s'(t)T'(t) \times x'(t)$$

and

$$|x''(t) \times x'(t)| = s'(t)^3\kappa(t)\,.$$

This gives a convenient way of finding the curvature when the curve is defined with respect to an arbitrary parameter. We can write this simply as

$$\kappa(t) = \frac{|x''(t) \times x'(t)|}{|x'(t)x'(t)|^{3/2}}.$$

Notes The curvature κ(t) of a space curve is non-negative for all t. The curvature can be zero, for example at every point of a curve lying along a straight line, or at an isolated point like t = 0 for the curve $x(t) = (t, t^3, 0)$. A curve for which κ(t) > 0 for all t is called non-inflectional.

The unit tangent vectors emanating from the origin form a curve T(t) on the unit sphere called the tangential indicatrix of the curve x. To calculate the length of the tangent indicatrix, we form the integral of $|T'(t)| = \kappa(t)s'(t)$ with respect to t, so the length is $\kappa(t)s'(t)dt = \kappa(s)ds$. This significant integral is called the total curvature of the curve x.

Up to this time, we have concentrated primarily on local properties of curves, determined at each point by the nature of the curve in an arbitrarily small neighborhood of the point. We are now in a position to prove our first result in global differential geometry or differential geometry in the large.

By a closed curve x(t), $a \le t \le b$, we mean a curve such that $x(b) = x(a)$. We will assume moreover that the derivative vectors match at the endpoints of the interval, so $x'(b) = x'(a)$.

Theorem 1 (Fenchel's Theorem): The total curvature of a closed space curve x is greater than or equal to 2π.

$$\kappa(s)ds \ge 2\pi$$

The first proof of this result was found independently by **B. Segre in 1934** and later independently by **H. Rutishauser and H. Samelson in 1948.** The following proof depends on a lemma by **R. Horn in 1971:**

Lemma 1. Let g be a closed curve on the unit sphere with length L < 2. Then there is a point m on the sphere that is the north pole of a hemisphere containing g.

To see this, consider two points p and q on the curve that break g up into two pieces g_1 and g_2 of equal length, therefore both less than π. Then the distance from p to q along the sphere is less than π so there is a unique minor arc from p to q. Let m be the midpoint of this arc. We wish to show that no point of g hits the equatorial great circle with m as north pole. If a point on one of the curves, say g_1, hits the equator at a point r, then we may construct another curve $g_t^{'}$ by rotating g_1 one-half turn about the axis through m, so that p goes to q and q to p while r goes to the antipodal point r′. The curve formed by g_1 and $g_t^{'}$ has the same length as the original curve g, but it contains a pair of antipodal points so it must have length at least 2π, contradicting the hypothesis that the length of g was less than 2π.

From this lemma, it follows that any curve on the sphere with length less than 2π is contained in a hemisphere centered at a point m. However if x(t) is a closed curve, we may consider the differentiable function $f(t) = x(t) . m$. At the maximum and minimum values of f on the closed curve x, we have

$$0 = f'(t) = x'(t) \cdot m = s'(t)T(t) \cdot m$$

so there are at least two points on the curve such that the tangential image is perpendicular to m. Therefore the tangential indicatrix of the closed curve x is not contained in a hemisphere, so by the lemma, the length of any such indicatrix is greater than 2π. Therefore, the total curvature of the closed curve x is also greater than 2π.

Corollary 1. If, for a closed curve x, we have $\kappa(t) \le \frac{1}{R}$ for all t, then the curve has length $L \ge 2\pi R$.

Proof.

$$L = \int ds \ge \int R\kappa(s)ds = R\int \kappa(s)ds \ge 2\pi R$$

Fenchel also proved the stronger result that the total curvature of a closed curve equals 2π if and only if the curve is a convex plane curve.

I. F´ary and J. Milnor proved independently that the total curvature must be greater than 4π for any non-self-intersecting space curve that is knotted (not deformable to a circle without self-intersecting during the process.)

Exercise 6: Let x be a curve with $x'(t_0) \neq 0$. Show that the tangent line at $x(t_0)$ can be written as $y(u) = x(t_0) + ux'(t_0)$ where u is a parameter that can take on all real values.

Exercise 7: The plane through a point $x(t_0)$ perpendicular to the tangent line is called the normal plane at the point. Show that a point y is on the normal plane at $x(t_0)$ if and only if

$$x'(t_0) \cdot y = x'(t_0) \cdot x(t_0)$$

Exercise 8: Show that the curvature κ of a circular helix

$$x(t) = (r \cos(t), r \sin(t), pt)$$

is equal to the constant value $\kappa = \dfrac{|r|}{r^2 + p^2}$. Are there any other curves with constant curvature?

Give a plausible argument for your answer.

Exercise 9: Assuming that the level surfaces of two functions $f(x_1, x_2, x_3) = 0$ and $g(x_1, x_2, x_3) = 0$ meet in a curve, find an expression for the tangent vector to the curve at a point in terms of the gradient vectors of f and g (where we assume that these two gradient vectors are linearly independent at any intersection point.) Show that the two level surfaces $x_2 - x_1^2 = 0$ and $x_3x_1 - x_2^2 = 0$ consists of a line and a "twisted cubic" $x_1(t) = t$, $x_2(t) = t^2$, $x_3(t) = t^3$. What is the line?

> *Tasks* What is the geometric meaning of the function $f(t) = x(t) \cdot m$ used in the proof of Fenchel's theorem?
>
> Let m be a unit vector and let x be a space curve. Show that the projection of this curve into the plane perpendicular to m is given by
>
> $$y(t) = x(t) - (x(t) \cdot m)m.$$
>
> Under what conditions will there be a t_0 with $y'(t_0) = 0$?

5 .3 The Unit Normal Bundle and Total Twist

Consider a curve $x(t)$ with $x'(t) \neq 0$ for all t. A vector z perpendicular to the tangent vector $x'(t_0)$ at $x(t_0)$ is called a normal vector at $x(t_0)$. Such a vector is characterized by the condition $z \cdot x(t_0) = 0$, and if $|z| = 1$, then z is said to be a unit normal vector at $x(t_0)$. The set of unit normal vectors at a point $x(t_0)$ forms a great circle on the unit sphere. The unit normal bundle is the collection of all unit normal vectors at $x(t)$ for all the points on a curve x.

At every point of a parametrized curve $x(t)$ at which $x'(t) \neq 0$, we may consider a frame $E_2(t), E_3(t)$, where $E_2(t)$ and $E_3(t)$ are mutually orthogonal unit normal vectors at $x(t)$. If $\underline{E}_2(t), \underline{E}_3(t)$ is another such frame, then there is an angular function $\phi(t)$ such that

$$E_2(t) = \cos(\phi(t))\underline{E}_2(t) - \sin(\phi(t))\underline{E}_3(t)$$

$$E_3(t) = \sin(\phi(t))\underline{E}_2(t) + \cos(\phi(t))\underline{E}_3(t)$$

Notes

or, equivalently,

$$\underline{E}_2(t) = \cos(\phi(t))E_2(t) + \sin(\phi(t))E_3(t)$$
$$\underline{E}_3(t) = \sin(\phi(t))E_2(t) + \cos(\phi(t))E_3(t)\ .$$

From these two representations, we may derive an important formula:

$$E_2'(t)\ .\ E_3(t) = \underline{E}_2'(t)\ .\ \underline{E}_3(t) - \phi'(t)$$

Expressed in the form of differentials, without specifying parameters, this formula becomes:

$$dE_2E_3 = d\underline{E}_2\underline{E}_3 - d\phi.$$

Since $E_3(t) = T(t) \times E_2(t)$, we have:

$$E_2'(t)\ .\ E_3(t) = [\underline{E}_2'(t), E_2(t), T(t)]$$

or, in differentials:

$$dE_2E_3 = -[dE_2, E_2, T]\ .$$

More generally, if z(t) is a unit vector in the normal space at x(t), then we may define a function $w(t) = -[z'(t), z(t), T(t)]$. This is called the connection function of the unit normal bundle. The corresponding differential form $w = -[dz, z, T]$ is called the connection form of the unit normal bundle.

A vector function z(t) such that $|z(t)| = 1$ for all t and $z(t) \cdot x'(t) = 0$ for all t is called a unit normal vector field along the curve x. Such a vector field is said to be parallel along x if the connection function $w(t) = -[z'(t), z(t), T(t)] = 0$ for all t. In the next section, we will encounter several unit normal vector fields naturally associated with a given space curve. For now, we prove some general theorems about such objects.

Proposition 2. If $E_2(t)$ and $\underline{E}_2(t)$ are two unit normal vector fields that are both parallel along the curve x, then the angle between $E_2(t)$ and $\underline{E}_2(t)$ is constant.

Proof. From the computation above, then:

$$E_2'(t) \cdot (-E_2(t) \times T(t)) = \underline{E}_2'(t) \cdot (-\underline{E}_2(t) \times T(t)) - \phi'(t).$$

But, by hypothesis,

$$E_2'(t) \cdot (-E_2(t) x T(t)) = 0 = \underline{E}_2'(t)(-\underline{E}_2(t) \times T(t))$$

so it follows that $\phi'(t) = 0$ for all t, i.e., the angle $\phi(t)$ between $E_2(t)$ and $\underline{E}_2(t)$ is constant.

Given a closed curve x and a unit normal vector field z with z(b) = z(a),

we define

$$\mu(x, z) = -\frac{1}{2\pi}\int [z'(t), z(t), T(t)]dt = -\frac{1}{2\pi}[dz, z, T].$$

If z is another such field, then

$$\mu(x, z) - \mu(x, z) = -\frac{1}{2\pi}\int [z'(t), z(t), T(t)] - [\underline{z}'(t), \underline{z}(t), T(t)]dt$$

$$= -\frac{1}{2\pi}\int \phi'(t)dt = -\frac{1}{2\pi}[\phi(b) - \phi(a)].$$

Since the angle $\phi(b)$ at the end of the closed curve must coincide with the angle $\phi(a)$ at the beginning, up to an integer multiple of 2π, it follows that the real numbers $\mu(x, z)$ and $\mu(x, \underline{z})$ differ by an integer. Therefore, the fractional part of $\mu(x, z)$ depends only on the curve x and not on the unit normal vector field used to define it. This common value $\mu(x)$ is called the total twist of the curve x. It is a global invariant of the curve.

Proposition 3. If a closed curve lies on a sphere, then its total twist is zero.

Proof. If x lies on the surface of a sphere of radius r centered at the origin, then $|x(t)|^2 = x(t) . x(t) = r^2$ for all t. Thus, $x'(t) \cdot x(t) = 0$ for all t, so x(t) is a normal vector at x(t). Therefore, $z(t) = \frac{x(t)}{r}$ is a unit normal vector field defined along x, and we may compute the total twist by evaluating $\mu(x, z) = -\frac{1}{2\pi}\int [z'(t), z(t), T(t)]dt$.

But

$$[z'(t), z(t), T(t)] = [\frac{x'(t)}{r}, \frac{x(t)}{r}, T(t)] = 0$$

for all t since x′(t) is a multiple of T(t). In differential form notation, we get the same result: $[dz, z, T] = \frac{1}{r^2}[x'(t), x(t), T(t)]dt = 0.$ Therefore, $\mu(x, z) = 0$, so the total twist of the curve x is zero.

Remark 1. W. Scherrer proved that this property characterized a sphere, i.e. if the total twist of every curve on a closed surface is zero, then the surface is a sphere.

Remark 2. T. Banchoff and J. White proved that the total twist of a closed curve is invariant under inversion with respect to a sphere with center not lying on the curve.

Remark 3. The total twist plays an important role in modern molecular biology, especially with respect to the structure of DNA.

Exercise 10: Let x be the circle $x(t) = (r \cos(t), r \sin(t), 0)$, where r is a constant > 1. Describe the collection of points $x(t) + z(t)$ where z(t) is a unit normal vector at x(t).

Exercise 11: Let Σ be the sphere of radius $r > 0$ about the origin. The inversion through the sphere S maps a point x to the point $x = r^2 \frac{x}{|x|^2}$. Note that this mapping is not defined if x = 0, the center of the sphere. Prove that the coordinates of the inversion of $x = (x_1, x_2, x_3)$ through S are given by $\underline{x}_i = \frac{r^2 x_i}{x_1^2 + x_2^2 + x_3^2}$. Prove also that inversion preserves point that lie on the sphere S itself, and that the image of a plane is a sphere through the origin, except for the origin itself.

Exercise 12: Prove that the total twist of a closed curve not passing through the origin is the same as the total twist of its image by inversion through the sphere S of radius r centered at the origin.

5 .4 Moving Frames

In the previous section, we introduced the notion of a frame in the unit normal bundle of a space curve. We now consider a slightly more general notion. By a frame, or more precisely a right-handed rectangular frame with origin, we mean a point x and a triple of mutually orthogonal unit vectors E_1, E_2, E_3 forming a right-handed system. The point x is called the origin of the frame.

Note that $E_i \cdot E_j = 1$ if i = j and 0 if $i \neq j$.

Notes

Moreover,

$$E_1 \times E_2 = E_3,\ E_2 = E_3 \times E_1,\ \text{and}\ E_3 = E_1 \times E_2.$$

In the remainder of this section, we will always assume that small Latin letters run from 1 to 3.

Note that given two different frames, x, E_1, E_2, E_3 and $\underline{x}, E_1, E_2, E_3$, there is exactly one affine motion of Euclidean space taking x to $\underline{x}$ and taking E_i to $\underline{E}_i$. When $x(t), E_1(t), E_2(t), E_3(t)$ is a family of frames depending on a parameter t, we say we have a moving frame along the curve.

Proposition 4. A family of frames x(t), E1(t), $E_2(t)$, $E_3(t)$ satisfies a system of differential equations:

$$x'(t) = \Sigma p_i(t)E_i(t)$$

$$E_t^{'}(t) = \Sigma q_{ij}(t)E_j(t)$$

where $p_i(t) = x'(t) \cdot E_i(t)$ and $q_ij(t) = E_t^{'}(t) \cdot E_j(t)$.

Since $E_i(t) \cdot E_j(t) = 0$ for $i \neq j$, it follows that

$$q_{ij}(t) + q_{ji}(t) = E_t^{'}(t) \cdot E_j(t) + E_i(t) \cdot E_j^{'}(t) = 0$$

i.e. the coefficients $q_{ij}(t)$ are anti-symmetric in i and j. This can be expressed by saying that the matrix $((q_ij(t)))$ is an anti-symmetric matrix, with 0 on the diagonal.

In a very real sense, the function $p_i(t)$ and $q_{ij}(t)$ completely determine the family of moving frames.

Specifically we have:

Proposition 5. If x(t), $E_1(t)$, $E_2(t)$, $E_3(t)$ and $\underline{x}(t)$, $\underline{E}_1(t)$, $\underline{E}_2(t)$, $\underline{E}_3(t)$ are two families of moving frames such that $\underline{p}_i(t) = \underline{p}_i(t)$ and $q_{ij}(t) = \underline{q}_{ij}(t)$ for all t, then there is a single affine motion that takes x(t), E1(t), $E_2(t)$, $E_3(t)$ to $\underline{x}(t), \underline{E}_1(t), \underline{E}_2(t), \underline{E}_3(t))$ for all t.

Proof. Recall that for a specific value t_0, there is an affine motion taking $x(t_0)$, $E_1(t_0)$, $E_2(t_0)$, $E_3(t_0)$ to $\underline{x}(t_0)$, $E_1(t_0)$, $\underline{E}_2(t_0)$, $\underline{E}_3(t_0)$. We will show that this same motion takes x(t), $E_1(t)$, $E_2(t)$, $E_3(t)$ to $\underline{x}(t)$, $\underline{E}_1(t)$, $\underline{E}_2(t)$, $E_3(t)$ for all t. Assume that the motion has been carried out so that the frames x(t0), $E_1(t_0)$, $E_2(t_0)$, $E_3(t_0)$ and $\underline{x}(t_0)$, $\underline{E}_1(t_0)$, $\underline{E}_2(t_0)$, $\underline{E}_3(t_0)$ coincide.

Now consider

$$\begin{aligned}
(\Sigma E_i(t) \cdot E_i(t))' &= \Sigma E_j^{'}(t) \cdot \underline{E}_i(t) + \Sigma Ei(t) \cdot \underline{E}_i^{'}(t) \\
&= \Sigma\Sigma q_{ij}(t)E_j(t) \cdot \underline{E}_i(t) + \Sigma E_i(t) \cdot \Sigma q_{ij}(t)\underline{E}_j(t) \\
&= \Sigma\Sigma q_{ij}(t)E_j(t) \cdot \underline{E}_i(t) + \Sigma\Sigma q_{ij}(t)E_i(t) \cdot \underline{E}_j(t) \\
&= \Sigma\Sigma q_{ij}(t)E_j(t) \cdot \underline{E}_i(t) + \Sigma\Sigma q_{ji}(t)E_j(t) \cdot \underline{E}i(t) \\
&= 0\,.
\end{aligned}$$

It follows that

$$\Sigma E_i(t) \cdot \underline{E}_i(t) = \Sigma E_i(t_0) \cdot \underline{E}_i(t_0) = \Sigma E_i(t_0) \cdot E_i(t_0) = 3$$

for all t. But since $|E_i(t) \cdot \underline{E}_i(t)| \leq 1$ for any pair of unit vectors, we must have $E_i(t) \cdot \underline{E}_i(t) = 1$ for all t. Therefore, $E_i(t) = \underline{E}_i(t)$ for all t.

Next, consider

$$(x(t) - \underline{x}(t))' = \Sigma p_i(t)E_i(t) - \Sigma p_i(t)\underline{E}_i(t) = \Sigma p_i(t)E_i(t) - \Sigma p_i(t)E_i(t) = 0\,.$$

Since the origins of the two frames coincide at the value t0, we have

$$x(t) - \underline{x}(t) = x(t_0) - \underline{x}(t_0) = 0$$

for all t.

This completes the proof that two families of frames satisfying the same set of differential equations differ at most by a single affine motion.

Exercise 13: Prove that the equations $E_i'(t) = \Sigma q_{ij}(t)E_j(t)$ can be written $E_i'(t) = d(t) \times E_i(t)$, where $d(t) = q_{23}(t)E_1(t) + q_{31}(t)E_2(t) + q_{12}(t)E_3(t)$. This vector is called the instantaneous axis of rotation.

Exercise 14: Under a rotation about the x_3-axis, a point describes a circle x(t) = (a cos(t), a sin(t), b). Show that its velocity vector satisfies x′(t) = d × x(t) where d = (0, 0, 1). (Compare with the previous exercise.).

Exercise 15: Prove that (v · v)(w ·w)″(v ·w)2 = 0 if and only if the vectors v and w are linearly dependent.

5 .5 Curves at a Non-inflexional Point and the Frenet Formulas

A curve x is called non-inflectional if the curvature κ(t) is never zero. By our earlier calculations, this condition is equivalent to the requirement that x′(t) and x″(t) are linearly independent at every point x(t), i.e. $x'(t) \times x''(t) \neq 0$ for all t. For such a non-inflectional curve x, we may define a pair of natural unit normal vector fields along x.

Let $b(t) = \dfrac{x'(t) \times x''(t)}{|x'(t) \times x''(t)|}$, called the binormal vector to the curve x(t). Since b(t) is always perpendicular to T(t), this gives a unit normal vector field along x.

We may then take the cross product of the vector fields b(t) and T(t) to obtain another unit normal vector field N(t) = b(t) × T(t), called the principal normal vector. The vector N(t) is a unit vector perpendicular to T(t) and lying in the plane determined by x′(t) and x″(t). Moreover, $x''(t) \cdot N(t) = k(t)s'(t)^2$, a positive quantity.

Note that if the parameter is arclength, then x′(s) = T(s) and x″(s) is already perpendicular to T(s). It follows that x″(s) = k(s)N(s) so we may define $N(s) = \dfrac{x''(s)}{k(s)}$ and then define b(s) = T(s) × N(s).

This is the standard procedure when it happens that the parametrization is by arclength. The method above works for an arbitrary parametrization.

We then have defined an orthonormal frame x(t)T(t)N(t)b(t) called the Frenet frame of the non-inflectional curve x.

By the previous section, the derivatives of the vectors in the frame can be expressed in terms of the frame itself, with coefficients that form an antisymmetric matrix. We already have x′(t) = s′(t)T(t), so

$$p_1(t) = s'(t),\ p_2(t) = 0 = p_3(t)\ .$$

Also T′(t) = k(t)s′(t)N(t), so

$$q_{12}(t) = k(t)s'(t) \text{ and } q_{13}(t) = 0\ .$$

We know that

$$b'(t) = q_{31}(t)T(t) + q_{32}(t)N(t), \text{ and } q_{31}(t) = -q_{13}(t) = 0\ .$$

Thus b′(t) is a multiple of N(t), and we define the torsion w(t) of the curve by the condition

$$b'(t) = -w(t)s'(t)N(t),$$

so $q_{32}(t) = -w(t)s'(t)$ for the Frenet frame. From the general computations about moving frames, it then follows that

$$N'(t) = q_{21}(t)T(t) + q_{23}(t)b(t) = -k(t)s'(t)T(t) + w(t)s'(t)b(t)\ .$$

The formulas for T′(t), N′(t), and b′(t) are called the Frenet formulas for the curve x.

If the curve x is parametrized with respect to arclength, then the Frenet formulas take on a particularly simple form:

$$x'(s) = T(s)$$

$$T'(s) = k(s)N(s)$$

$$N'(s) = -k(s)T(s) + w(s)b(s)$$

$$b'(s) = -w(s)b(s)\ .$$

The torsion function w(t) that appears in the derivative of the binormal vector determines important properties of the curve. Just as the curvature measures deviation of the curve from lying along a straight line, the torsion measures deviation of the curve from lying in a plane. Analogous to the result for curvature, we have:

Proposition 6. If w(t) = 0 for all points of a non-inflectional curve x, then the curve is contained in a plane.

Proof. We have $b'(t) = -w(t)s'(t)N(t) = 0$ for all t so b(t) = a, a constant unit vector. Then, T(t)a = 0 for all t so $(x(t)\cdot a)' = x'(t)\cdot a = 0$ and $x(t)\cdot a = x(a)\cdot a$, a constant. Therefore, $(x(t) - x(a))\cdot a = 0$ and x lies in the plane through x(a) perpendicular to a.

If x is a non-inflectional curve parametrized by arclength, then

$$w(s) = b(s)\cdot N'(s) = [T(s),N(s),N'(s)]\ .$$

Since $N(s) = \dfrac{x''(s)}{k(s)}$, we have,

$$N'(s) = \frac{x'''(s)}{k(s)} + x''(s)\frac{-k'(s)}{k(s)^2},$$

so

$$w(s) = \left[x'(s), \frac{x''(s)}{k(s)}, \frac{x'''(s)}{k(s)} + x''(s)\frac{-k'(s)}{k(s)^2}\right] = \frac{[x'(s),x''(s),x'''(s)]}{2}.$$

We can obtain a very similar formula for the torsion in terms of an arbitrary parametrization of the curve x. Recall that

$$x''(t) = s''(t)T(t) + k(t)s'(t)T'(t) = s''(t)T(t) + k(t)s'(t)^2N(t),$$

so

$$x'''(t) = s'''(t)T(t)+s''(t)s'(t)k(t)N(t) + [k(t)s'(t)^2]'N(t) + k(t)s'(t)^2N'(t)\ .$$

Therefore,

$$x'''(t)b(t) = k(t)s'(t)^2N'(t)b(t) = k(t)s'(t)2w(t)s'(t),$$

and

$$x'''(t) \cdot x'(t) \times x''(t) = k^2(t)s'(t)^6 w(t) .$$

Thus, we obtain the formula

$$w(t) = \frac{x'''(t).x'(t)xx''(t)}{|x'(t) \times x''(t)|^2},$$

valid for any parametrization of x.

Notice that although the curvature k(t) is never negative, the torsion w(t) can have either algebraic sign. For the circular helix x(t) = (r cos(t), r sin(t), pt) for example, we find $w(t) = \frac{p}{r^2 + p^2}$, so the torsion has the same algebraic sign as p. In this way, the torsion can distinguish between a right-handed and a left-handed screw.

Changing the orientation of the curve from s to –s changes T to –T, and choosing the opposite sign for k(s) changes N to –N. With different choices, then, we can obtain four different right-handed orthonormal frames, xTNb, x(–T)N(–b), xT(–N)(–b), and x(–T)(–N)b. Under all these changes of the Frenet frame, the value of the torsion w(t) remains unchanged.

A circular helix has the property that its curvature and its torsion are both constant. Furthermore, the unit tangent vector T(t) makes a constant angle with the vertical axis. Although the circular helices are the only curves with constant curvature and torsion, there are other curves that have the second property. We characterize such curves, as an application of the Frenet frame.

Proposition 7. The unit tangent vector T(t) of a non-inflectional space curve x makes a constant angle with a fixed unit vector a if and only if the ratio $\frac{w(t)}{k(t)}$ is constant.

Proof. If T(t) · a = constant for all t, then differentiating both sides, we

obtain

$$T'(t) \cdot a = 0 = k(t)s'(t)N(t) \cdot a,$$

so a lies in the plane of T(t) and b(t). Thus, we may write a = cos(ϕ)T(t) + sin(ϕ)b(t) for some angle ϕ. Differentiating this equation, we obtain 0 = cos(ϕ)T′(t) + sin(ϕ)b′(t) = cos(ϕ)k(t)s′(t)N(t) – sin(ϕ)w(t)s′(t)N(t), so $\frac{w(t)}{k(t)} = \frac{\sin(\phi)}{\cos(\phi)} = \tan(\phi)$. This proves the first part of the proposition and identifies the constant ratio of the torsion and the curvature.

Conversely, if $\frac{w(t)}{k(t)}$ = constant = tan(ϕ) for some ϕ, then, by the same calculations, the expression cos(ϕ)T(t) + sin(ϕ)b(t) has derivative 0 so it equals a constant unit vector. The angle between T(t) and this unit vector is the constant angle ϕ.

Curves with the property that the unit tangent vector makes a fixed angle with a particular unit vector are called generalized helices. Just as a circular helix lies on a circular cylinder, a generalized helix will lie on a general cylinder, consisting of a collection of lines through the curve parallel to a fixed unit vector. On this generalized cylinder, the unit tangent vectors make a fixed angle with these lines, and if we roll the cylinder out onto a plane, then the generalized helix is rolled out into a straight line on the plane.

We have shown in the previous section that a moving frame is completely determined up to an affine motion by the functions $p_i(t)$ and $q_{ij}(t)$. In the case of the Frenet frame, this means that if two curves x and $\underline{x}$ have the same arclength s(t), the same curvature k(t), and the same torsion w(t), then the curves are congruent, i.e. there is an affine motion of Euclidean three-space taking x(t) to $\underline{x}(t)$ for all t. Another way of stating this result is:

Theorem 2. The Fundamental Theorem of Space Curves. Two curves parametrized by arclength having the same curvature and torsion at corresponding points are congruent by an affine motion.

Exercise 16: Compute the torsion of the circular helix. Show directly that the principal normals of the helix are perpendicular to the vertical axis, and show that the binormal vectors make a constant angle with this axis.

Exercise 17: Prove that if the curvature and torsion of a curve are both constant functions, then the curve is a circular helix (i.e. a helix on a circular cylinder).

Exercise 18: Prove that a necessary and sufficient condition for a curve x to be a generalized helix is that

$$x''(t) \times x'''(t) \cdot xiv(t) = 0 .$$

Exercise 19: Let y(t) be a curve on the unit sphere, so that $|y(t)| = 1$ and $y(t) \cdot y'(t) \times y''(t) \neq 0$ for all t. Show that the curve $x(t) = c\int y(u) \times y''(u)du$ with $c \neq 0$ has constant torsion $\frac{1}{c}$.

Exercise 20: (For students familiar with complex variables) If the coordinate functions of the vectors in the Frenet frame are given by

$$T = (e_{11}, e_{12}, e_{13}),$$
$$N = (e_{21}, e_{22}, e_{23}),$$
$$b = (e_{31}, e_{32}, e_{33}),$$

then we may form the three complex numbers

$$z_j = \frac{e_{1j} + ie_{2j}}{1 - e_{3j}} = \frac{1 + e_{3j}}{e_{1j} - ie_{2j}}.$$

Then the functions z_j satisfy the Riccati equation

$$z_j' = -ik(s)zj + \frac{i}{2}w(s)(-1 + z_j^2).$$

This result is due to S. Lie and G. Darboux.

5 .6 Local Equations of a Curve

We can "see" the shape of a curve more clearly in the neighborhood of a point $x(t_0)$ when we consider its parametric equations with respect to the Frenet frame at the point. For simplicity, we will assume that $t_0 = 0$, and we may then write the curve as

$$x(t) = x(0) + x_1(t)T(0) + x_2(t)N(0) + x_3(t)b(0) .$$

On the other hand, using the Taylor series expansion of x(t) about the point t = 0, we obtain

$$x(t) = x(0) + tx'(0) + \frac{t^2}{2}x''(0) + \frac{t_3}{6}x'''(0) + \text{higher order terms} .$$

From our earlier formulas, we have

$x'(0) = s'(0)T(0),$

$x''(0) = s''(0)T(0) + k(0)s'(0)^2N(0),$

$x'''(0) = s'''(0)T(0) + s''(0)s'(0)k(0)N(0) + (k(0)s'(0)^2)'N(0) + k(0)s'(0)^2(-k(0)s'(0)T(0) + w(0)s'(0)b(0)).$

Substituting these equations in the Taylor series expression, we find:

$$x(t) = x(0) + \left(ts'(0) + \frac{t^2}{2}s''(0) + \frac{t^3}{6}[''' (0) - k(0)^2 s'(0)^3] + ...\right)T(0)$$

$$+ \left(\frac{t^2}{2}k(0)s'(0)^2 + \frac{t^3}{6}[s''(0)s'(0)k(0) + (k(0)s'(0)^2] + ...\right)N(0)$$

$$+ \left(\frac{t^3}{6}k(0)w(0)s'(0)^3 + ...\right)b(0).$$

If the curve is parametrized by arclength, this representation is much simpler:

$$x(s) = x(0) + \left(s - \frac{k(0)^2}{6}s^3 + ...\right)T(0) + \left(\frac{k(0)}{2}s^2 + \frac{k'(0)}{6}s^3 + ...\right)N(0) + \left(\frac{k(0)w(0)}{6}s^3 + ...\right)b(0).$$

Relative to the Frenet frame, the plane with equation $x_1 = 0$ is the normal plane; the plane with $x_2 = 0$ is the rectifying plane, and the plane with $x_3 = 0$ is the osculating plane. These planes are orthogonal respectively to the unit tangent vector, the principal normal vector, and the binormal vector of the curve.

5 .7 Plane Curves and a Theorem on Turning Tangents

The general theory of curves developed above applies to plane curves. In the latter case there are, however, special features which will be important to bring out. We suppose our plane to be oriented. In the plane, a vector has two components and a frame consists of an origin and an ordered set of two mutually perpendicular unit vectors forming a right-handed system. To an oriented curve C defined by x(s) the Frenet frame at s consists of the origin x(s), the unit tangent vector T(s) and the unit normal vector N(s). Unlike the case of space curves, this Frenet frame is uniquely determined, under the assumption that both the plane and the curve are oriented.

The Frenet formulas are

$$x' = T,$$

$$T' = kN, \tag{1}$$

$$N' = -kT .$$

The curvature k(s) is defined with sign. It changes its sign when the orientation of the plane or the curve is reversed.

The Frenet formulas in (1) can be written more explicitly. Let

$$x(s) = (x_1(s), x_2(s)) \qquad ...(2)$$

Then,

$$T(s) = (x_1'(s), x_2'(s)),$$

Notes

$$N(s) = (-x_2'(s), x_1'(s)). \qquad ...(3)$$

Expressing the last two equations of (1) in components, we have

$$x_1'' = -kx_2' \qquad ...(4)$$

$$x_2'' = kx_1'. \qquad ...(5)$$

These equations are equivalent to (1).

Since T is a unit vector, we can put

$$T(s) = (\cos \tau(s), \sin \tau(s)), \qquad ...(6)$$

so that t(s) is the angle of inclination of T with the x_1-axis. Then

$$N(s) = (-\sin \tau(s), \cos \tau(s)), \qquad ...(7)$$

and (1) gives

$$\frac{d\tau}{ds} = k(s) \qquad ...(8)$$

This gives a geometrical interpretation of k(s).

A curve C is called simple if it does not intersect itself. One of the most important theorems in global differential geometry is the theorem on turning tangents:

Theorem 3. For a simple closed plane curve, we have

$$\frac{1}{2\pi}\oint k\, ds = \pm 1.$$

To prove this theorem we give a geometrical interpretation of the integral at the left-hand side of (3). By (8)

$$\frac{1}{2\pi}\oint k\, ds = \frac{1}{2\pi}\oint d\tau.$$

But τ, as the angle of inclination of τ(s), is only defined up to an integral multiple of 2π, and this integral has to be studied with care.

Let O be a fixed point in the plane. Denote by Γ the unit circle about O; it is oriented by the orientation of the plane. The tangential mapping or Gauss mapping

$$g : C \mapsto \Gamma \qquad ...(9)$$

is defined by sending the point x(s) of C to the point T(s) of Γ. In other words, g(P), P ∈ C, is the end-point of the unit vector through O parallel to the unit tangent vector to C at P. Clearly, g is a continuous mapping. If C is closed, it is intuitively clear that when a point goes along C once its image point under g goes along Û a number of times. This integer is called the rotation index of C. It is to be defined rigorously as follows:

We consider O to be the origin of our coordinate system. As above we denote by τ(s) the angle of inclination of T(s) with the x_1-axis. In order to make the angle uniquely determined, we suppose O ≤ τ(s) < 2π. But τ(s) is not necessarily continuous. For in every neighborhood of s_0 at which $\tau(s_c) = 0$, there may be values of τ(s) differing from 2π by arbitrarily small quantities.

We have, however, the following lemma:

Lemma 2. There exists a continuous function $\tilde{\tau}(s)$ such that $\tilde{\tau}(s) \equiv \tau(s) \bmod 2\pi$.

Proof. We suppose C to be a closed curve of total length L. The continuous mapping g is uniformly continuous. There exists therefore a number $\delta > 0$ such that for $|s_1 - s_2| < \delta$, $T(s_1)$ and $T(s_2)$ lie in the same open half-plane. Let $s_0(= O) < s_1 < \cdots < s_i(= L)$ satisfy $|s_i - s_{i-1}| < \delta$ for i = 1, . . ., m. We put $\tilde{\tau}(s_0) = \tau(s_0)$. For $s_0 \le s \le s_1$, we define $\tilde{\tau}(s)$ to be $\tilde{\tau}(s_0)$ plus the angle of rotation from $g(s_0)$ to g(s) remaining in the same half-plane. Carrying out this process in successive intervals, we define a continuous function $\tilde{\tau}(s)$ satisfying the condition in the lemma. The difference $\tilde{\tau}(L) - \tilde{\tau}(O)$ is an integral multiple of 2π. Thus, $\tilde{\tau}(L) - \tilde{\tau}(O) = \gamma 2\pi$. We assert that the integer γ is independent of the choice of the function $\tilde{\tau}$. In fact let $\tilde{\tau}'(s)$ be a function satisfying the same conditions. Then we have $\tilde{\tau}'(s) - \tilde{\tau}(s) = n(s)2p$ where n(s) is an integer. Since n(s) is continuous in s, it must be constant. It follows that $\tilde{\tau}'(L) - \tilde{\tau}(O) = \tilde{t}(L) - \tilde{\tau}(O)$, which proves the independence of γ from the choice of $\tilde{\tau}$. We call the rotation index of C. In performing integration over C we should replace t(s) by $\tilde{\tau}$ in (8). Then we have

$$\frac{1}{2\pi}\oint k\, ds = \frac{1}{2\pi}\oint d\tilde{\tau} = \gamma. \qquad ...(10)$$

We consider the mapping h which sends an ordered pair of points $x(s_1)$, $x(s_2)$, $O \le s_1 \le s_2 \le L$, of C into the end-point of the unit vector through O parallel to the secant joining $x(s_1)$ to $x(s_2)$. These ordered pairs of points can be represented as a triangle Δ in the (s_1, s_2)-plane defined by $O \le s_1 \le s_2 \le L$. The mapping h of Δ into Γ is continuous. Moreover, its restriction to the side $s_1 = s_2$ is the tangential mapping g in (9).

To a point $p \in \Delta$ let $\tau(p)$ be the angle of inclination of $\overline{Oh(p)}$ to the x_1-axis, satisfying $O \le \tau(p) < 2\pi$. Again this function need not be continuous. We shall, however, prove that there exists a continuous function $\tau(\tilde{p}), p \in \Delta$, such that $\tilde{\tau}(p) \equiv \tau(p) \bmod 2\pi$. In fact, let m be an interior point of Δ. We cover Δ by the radii through m. By the argument used in the proof of the above lemma we can define a function $\tilde{\tau}(p), p \in \Delta$, such that $\tilde{\tau}(p) \equiv \tau(p) \bmod 2\pi$, and such that it is continuous on every radius through m. It remains to prove that it is continuous in Δ.

For this purpose let $p_0 \in 4$. Since h is continuous, it follows from the compactness of the segment $\overline{mp_0}$ that there exists a number $\eta = \eta(p_0) > 0$, such that for $q_0 \in \overline{mp_0}$ and for any point $q \in \Delta$ for which the distance $d(q, q_0) < \eta$ the points h(q) and h(q0) are never antipodal. The latter condition can be analytically expressed by

$$\tilde{\tau}(q) \not\equiv \tilde{\tau}(q_0) \quad \bmod \pi. \qquad ...(11)$$

Now let R / > 0, R / < $\frac{\pi}{2}$ be given. We choose a neighborhood U of p_0 such that U is contained in the η-neighborhood of p_0 and such that, for $p \in U$, the angle between $\overline{Oh(p_0)}$ and $\overline{Oh(p)}$ is < R. This is possible, because the mapping h is continuous. The last condition can be expressed in the form

$$\tilde{\tau}(p) - \tilde{t}(p_0) = R' + 2k(p)\pi, \qquad ...(12)$$

Notes

where k(p) is an integer. Let q0 be any point on the segment $\overline{mp_0}$. Draw the segment $\overline{qq_0}$ parallel to $\overline{pp_0}$, with q on $\overline{mp}$. The function $\tilde{\tau}(q) - \tilde{t}(q_0)$ is continuous in q along $\overline{mp}$ and equals O when q coincides with m. Since d(q, q_0) < η, it follows from (11) that $\left|\tilde{\tau}(q) - \tilde{t}(q_0)\right| < \pi$. In particular, for $q_0 = p_0$ this gives $\left|\tilde{\tau}(p) - \tilde{\tau}(p_0)\right| < \pi$. Combining this with (12), we get k(p) = 0. Thus we have proved that $\tilde{\tau}(p)$ is continuous in 4, as asserted above. Since $\tilde{\tau}(p) \equiv \tau(p)$ mod 2π, it is clear that $\tilde{\tau}(p)$ is differentiable.

Now let A(O,O), B(O,L), D(L,L) be the vertices of Δ. The rotation index γ of C is, by (10), defined by the line integral

$$2\pi\gamma = \oint \frac{d\tilde{\tau}}{AD}.$$

Since ˜t(p) is defined in 4, we have

$$\oint_{\overline{AD}} d\tilde{\tau} = \oint_{\overline{AB}} d\tilde{\tau} + \oint_{\overline{BD}} d\tilde{\tau}.$$

To evaluate the line integrals at the right-hand side, we suppose the origin O to be the point x(O) and C to lie in the upper half-plane and to be tangent to the x_1-axis at O. This is always possible for we only have to take x(O) to be the point on C at which the x_2-coordinate is a minimum. Then the x_1-axis is either in the direction of the tangent vector to C at O or opposite to it. We can assume the former case, by reversing the orientation of C if necessary. The line integral along $\overline{AB}$ is then equal to the angle rotated by $\overline{OP}$ as P goes once along C. Since C lies in the upper half-plane, the vector $\overline{OP}$ never points downward. It follows that the integral along $\overline{AB}$ is equal to π. On the other hand, the line integral along $\overline{BD}$ is the angle rotated by $\overline{PO}$ as P goes once along C. Since the vector $\overline{PO}$ never points upward, this integral is also equal to π. Hence, their sum is 2π and the rotation index γ is +1. Since we may have reversed the orientation of C, the rotation index is ±1 in general.

Exercise 21: Consider the plane curve x(t) = (t, f(t)). Use the Frenet formulas in (1) to prove that its curvature is given by

$$k(t) = \frac{\ddot{f}}{(1+\dot{f}^2)^{3/2}}. \qquad \text{...(13)}$$

Exercise 22: Draw closed plane curves with rotation indices 0, –2, +3 respectively.

Exercise 23: The theorem on turning tangents is also valid when the simple closed curve C has “corners.” Give the theorem when C is a triangle consisting of three arcs. Observe that the theorem contains as a special case the theorem on the sum of angles of a rectilinear triangle.

Exercise 24: Give in detail the proof of the existence of η = η(p_0) used in the proof of the theorem on turning tangents. η = η (p_0) .

Notes

5 .8 Plane Convex Curves and the Four Vertex Theorem

A closed curve in the plane is called convex, if it lies at one side of every tangent line.

Proposition 8: A simple closed curve is convex, if and only if it can be so oriented that its curvature k is ≥ 0.

The definition of a convex curve makes use of the whole curve, while the curvature is a local property. The proposition, therefore, gives a relationship between a local property and a global property. The theorem is not true if the closed curve is not simple. Counter examples can be easily constructed.

Let $\tilde{\tau}(s)$ be the function constructed above, so that we have k = $\frac{d\tau}{ds}$. The condition k ≥ O is, therefore, equivalent to the assertion that $\tilde{\tau}(s))$ is a monotone non-decreasing function. We can assume that $\tilde{\tau}(O) = O.$ By the theorem on turning tangents, we can suppose C so oriented that $\tilde{\tau}(L) = 2\pi.$

Suppose $\tilde{\tau}(s),$ O ≤ s ≤ L, be monotone non-decreasing and that C is not convex. There is a point A = $x(s_0)$ on C such that there are points of C at both sides of the tangent to C at A. Choose a positive side of k and consider the oriented perpendicular distance from a point x(s) of C to λ. This is a continuous function in s and attains a maximum and a minimum at the points M and N respectively. Clearly M and N are not on and the tangents to C at M and N are parallel to x. Among these two tangents and k itself there are two tangents parallel in the same sense. Call $s_1 < s_2$ the values of the parameters at the corresponding points of contact. Since $\tilde{\tau}(s)$ is monotone non-decreasing and O ≤ $\tilde{\tau}(s)$ ≤ 2π, this happens only when $\tilde{\tau}(s) = \tilde{\tau}(s_1)$ for all s satisfying $s_1 \leq s \leq s_2$. It follows that the arc $s_1 \leq s \leq s_2$ is a line segment parallel to λ. But this is obviously impossible.

Next let C be convex. To prove that $\tilde{\tau}(s)$ is monotone non-decreasing, suppose $\tau(\tilde{s}_1) = \tau(\tilde{s}_2)$, $s_1 < s_2$. Then the tangents at $x(s_1)$ and $x(s_2)$ are parallel in the same sense. But there exists a tangent parallel to them in the opposite sense. From the convexity of C it follows that two of them coincide.

We are, thus, in the situation of a line λ tangent to C at two distinct points A and B. We claim that the segment $\overline{AB}$ must be a part of C. In fact, suppose this is not the case and let D be a point on $\overline{AB}$ not on C. Draw through D a perpendicular λ to in the half-plane which contains C. Then μ intersects C in at least two points. Among these points of intersection let F be the farthest from λ and G the nearest one, so that F ≠ G. Then G is an interior point of the triangle ABF. The tangent to C at G must have points of C in both sides which contradicts the convexity of C.

It follows that under our assumption, the segment $\overline{AB}$ is a part of C, so that the tangents at A and B are parallel in the same sense. This proves that the segment joining $x(s_1)$ to $x(s_2)$ belongs to C. Hence, $\tilde{\tau}(s)$ remains constant in the interval $s_1 \leq s \leq s_2$. We have, therefore, proved that $\tilde{\tau}(s)$ is monotone and K ≥ O.

A point on C at which k′ = 0 is called a vertex. A closed curve has at least two vertices, e.g., the maximum and the minimum of k. Clearly a circle consists entirely of vertices. An ellipse with unequal axes has four vertices, which are its intersection with the axes.

Notes

Theorem 4 (Four-vertex Theorem.). A simple closed convex curve has at least four vertices.

Remark 4. This theorem was first given by Mukhopadhyaya (1909). The following proof was due to G. Herglotz. It is also true for non-convex curves, but the proof will be more difficult.

5 .9 Isoperimetric Inequality in the Plane

Among all simple closed curves having a given length the circle bounds the largest area, and is the only curve with this property. We shall state the theorem as follows:

Theorem 5. Let L be the length of a simple closed curve C and A be the area it bounds. Then

$$L^2 - 4\pi A \geq 0 . \qquad ...(14)$$

Moreover, the equality sign holds only when C is a circle.

The proof given below is due to E. Schmidt (1939).

We enclose C between two parallel lines g, g′, such that C lies between g, g′ and is tangent to them at the points P, Q respectively. Let s = 0, s_0 be the parameters of P, Q. Construct a circle C tangent to g, g′ at P, Q respectively. Denote its radius by r and take its center to be the origin of a coordinate system. Let x(s) = (x_1(s), x_2(s)) be the position vector of C, so that

$$(x_1(0), x_2(0)) = (x_1(L), x_2(L)) .$$

As the position vector of $\bar{C}$ we take $(\bar{x}_1(s), \bar{x}_2)$, such that

$$\bar{x}_1(s) = x_1(s),$$

$$\bar{x}_2(s) = -\sqrt{r^2 - x_1^2(s)}, 0 \leq s \leq s_0$$

$$= +\sqrt{r^2 - x_1^2(s)}, s_0 \leq s \leq L.$$

Denote by $\bar{A}$ the area bounded by $\bar{C}$. Now the area bounded by a closed curve can be expressed by the line integral

$$A = \int_0^L x_1 x_2' ds = -\int_0^L x_2 x_1' ds = \frac{1}{2}\int_0^L (x_1 x_2' - x_2 x_1') ds.$$

Applying this to our two curves C and $\bar{C}$, we get

$$A = \int_0^L x1 x_2' ds,$$

$$\bar{A} = \pi r^2 = -\int_0^L x_2 \bar{x}_1' ds = -\int_0^L \bar{x}_2 x_1' ds.$$

Adding these two equations, we have

$$A + \pi r^2 = \int_0^L (x_1 x_2' - \bar{x}_2 x_1') ds \leq \int_0^L \sqrt{(x_1 x_2' - \bar{x}_2 x_1')^2} ds$$

$$\leq \int_0^L \sqrt{(x_1^2 - \bar{x}_2^2)(x_1'^2 + x_2'^2)} ds \qquad ...(15)$$

Notes

$$= \int_0^L \sqrt{x_1^2 + \bar{x}_2^2}\,ds = Lr.$$

Since the geometric mean of two numbers is ≤ their arithmetic mean, it follows that

$$\sqrt{A}\sqrt{\pi r^2} \le \frac{1}{2}(A + \pi r^2) \le \frac{1}{2}Lr.$$

This gives, after squaring and cancellation of r, the inequality (14).

Suppose now that the equality sign in (14) holds. A and πr^2 have then the same geometric and arithmetic mean, so that $A = \pi r^2$ and $L = 2\pi r$. The direction of the lines g, g′ being arbitrary, this means that C has the same "width" in all directions. Moreover, we must have the equality sign everywhere in (15). It follows in particular that

$$(x_1, x_2' - \bar{x}_2 x_1')^2 = (x_1^2 + \bar{x}_2^2)(x_1'^2 + x_2'^2),$$

which gives

$$\frac{x_1}{x_2'} = \frac{-\bar{x}_2}{x_1'} = \frac{\sqrt{x_1^2 + \bar{x}_2^2}}{\sqrt{x_1'^2 + x_2'^2}} = \pm r.$$

From the first equality in (15), the factor of proportionality is seen to be r, i.e.,

$$x_1 = r x_2',\ \bar{x}_2 = -r x_1'.$$

This remains true when we interchange x_1 and x_2, so that

$$x_2 = r x_1'.$$

Therefore, we have

$$x_1^2 + x_2^2 = r^2,$$

which means that C is a circle.

5 .10 Summary

- A parametrized curve in Euclidean three-space e^3 is given by a vector function

  $$x(t) = (x_1(t), x_2(t), x_3(t))$$

  that assigns a vector to every value of a parameter t in a domain interval [a, b]. The coordinate functions of the curve are the functions $x_i(t)$. In order to apply the methods of calculus, we suppose the functions $x_i(t)$ to have as many continuous derivatives as needed in the following treatment.
- One of the most useful ways to parametrize a curve is by the arc length s itself. If we let s = s(t), then we have

  $$s'(t) = |x'(t)| = |x'(s)|\,s'(t),$$

  from which it follows that $|x'(s)| = 1$ for all s. So the derivative of x with respect to arc length is always a unit vector.

Notes

- If x is an immersed curve, with $x'(t) \neq 0$ for all t in the domain, then we may define the unit tangent vector T(t) to be $\frac{x'(t)}{|x'(t)|}$. If the parameter is arclength, then the unit tangent vector T(s) is given simply by x′(s). The line through $x(t_0)$ in the direction of $T(t_0)$ is called the tangent line at $x(t_0)$. We can write this line as $y(u) = x(t_0) + uT(t_0)$, where u is a parameter that can take on all real values.

  Since $T(t) \cdot T(t) = 1$ for all t, we can differentiate both sides of this expression, and we obtain $2T'(t) \cdot T(t) = 0$. Therefore T′(t) is orthogonal to T(t). The curvature of the space curve x(t) is defined by the condition $\kappa(t) = \frac{T'(t)}{|x'(t)|}$, so $= \kappa(t)s'(t) = |T'(t)|$. If the parameter is arclength, then $x'(s) = T(s)$ and $\kappa(s) = |T'(s)| = |x''(s)|$.

- The unit tangent vectors emanating from the origin form a curve T(t) on the unit sphere called the tangential indicatrix of the curve x.
- W. Scherrer proved that this property characterized a sphere, i.e. if the total twist of every curve on a closed surface is zero, then the surface is a sphere.
- T. Banchoff and J. White proved that the total twist of a closed curve is invariant under inversion with respect to a sphere with center not lying on the curve.
- The total twist plays an important role in modern molecular biology, especially with respect to the structure of DNA.
- Let x be the circle x(t) = (r cos(t), r sin(t), 0), where r is a constant > 1. Describe the collection of points x(t) + z(t) where z(t) is a unit normal vector at x(t).

5 .11 Keywords

Curvature is one of the simplest and at the same time one of the most important properties of a curve.

Fenchel's Theorem: The total curvature of a closed space curve x is greater than or equal to 2π.

$$\kappa(s)ds \geq 2\pi$$

Non-inflectional: A curve x is called non-inflectional if the curvature κ(t) is never zero. By our earlier calculations, this condition is equivalent to the requirement that x′(t) and x″(t) are linearly independent at every point x(t), i.e. $x'(t) \times x''(t) \neq 0$ for all t.

5 .12 Self Assessment

1. If k(t) = 0 for all t, then the curve lies along a ....
2. The unit tangent vectors emanating from the origin form a curve T(t) on the unit sphere called the .... of the curve x.
3. .... The total curvature of a closed space curve x is greater than or equal to 2π.

$$\kappa(s)ds \geq 2\pi$$

4. A curve x is called .... if the curvature k(t) is never zero. By our earlier calculations, this condition is equivalent to the requirement that x′(t) and x″(t) are linearly independent at every point x(t), i.e. $x'(t) \times x''(t) \neq 0$ for all t.

5. If .... for all points of a non-inflectional curve x, then the curve is contained in a plane.

5 .13 Review Questions

1. One of the most important space curves is the circular helix x(t) = (a cos t, a sin t, bt), where $a \neq 0$ and b are constants. Find the length of this curve over the interval $[0, 2\pi]$.

2. Find a constant c such that the helix x(t) = (a cos(ct), a sin(ct), bt) is parametrized by arclength, so that $|x'(t)| = 1$ for all t.

3. The astroid is the curve defined by $x(t) = (a\cos^3 t, a\sin^3 t, 0)$, on the domain $[0, 2\pi]$. Find the points at which x(t) does not define an immersion, i.e., the points for which $x'(t) = 0$.

4. The trefoil curve is defined by x(t) = ((a + b cos(3t)) cos(2t), (a + b cos(3t)) sin(2t), b sin(3t)), where a and b are constants with $a > b > 0$ and $0 \leq t \leq 2\pi$. Sketch this curve, and give an argument to show why it is knotted, i.e. why it cannot be deformed into a circle without intersecting itself in the process.

5. Let x be a curve with $x'(t_0) \neq 0$. Show that the tangent line at $x(t_0)$ can be written as $y(u) = x(t_0) + ux'(t_0)$ where u is a parameter that can take on all real values.

6. The plane through a point $x(t_0)$ perpendicular to the tangent line is called the normal plane at the point. Show that a point y is on the normal plane at $x(t_0)$ if and only if

$$x'(t_0) \cdot y = x'(t_0) \cdot x(t_0)$$

7. Show that the curvature k of a circular helix

$$x(t) = (r\cos(t), r\sin(t), pt)$$

is equal to the constant value $\kappa = \dfrac{|r|}{r^2 + p^2}$. Are there any other curves with constant curvature? Give a plausible argument for your answer.

8. Assuming that the level surfaces of two functions $f(x_1, x_2, x_3) = 0$ and $g(x_1, x_2, x_3) = 0$ meet in a curve, find an expression for the tangent vector to the curve at a point in terms of the gradient vectors of f and g (where we assume that these two gradient vectors are linearly independent at any intersection point.) Show that the two level surfaces $x_2 - x_1^2 = 0$ and $x_3x_1 - x_2^2 = 0$ consists of a line and a "twisted cubic" $x_1(t) = t$, $x_2(t) = t^2$, $x_3(t) = t^3$. What is the line?

9. What is the geometric meaning of the function $f(t) = x(t) \cdot m$ used in the proof of Fenchel's theorem?

10. Let m be a unit vector and let x be a space curve. Show that the projection of this curve into the plane perpendicular to m is given by

$$y(t) = x(t) - (x(t) \cdot m)m.$$

Under what conditions will there be a t_0 with $y'(t_0) = 0$?

11. W. Scherrer proved that this property characterized a sphere, i.e. if the total twist of every curve on a closed surface is zero, then the surface is a sphere.

12. T. Banchoff and J. White proved that the total twist of a closed curve is invariant under inversion with respect to a sphere with center not lying on the curve.

13. The total twist plays an important role in modern molecular biology, especially with respect to the structure of DNA.

14. Let x be the circle $x(t) = (r \cos(t), r \sin(t), 0)$, where r is a constant > 1. Describe the collection of points $x(t) + z(t)$ where $z(t)$ is a unit normal vector at $x(t)$.

15. Prove that the total twist of a closed curve not passing through the origin is the same as the total twist of its image by inversion through the sphere S of radius r centered at the origin.

16. Prove that the equations $E_i'(t) = \Sigma q_{ij}(t)E_j(t)$ can be written $E_i'(t) = d(t) \times E_i(t)$, where $d(t) = q_{23}(t)E_1(t) + q_{31}(t)E_2(t) + q_{12}(t)E_3(t)$. This vector is called the instantaneous axis of rotation.

17. Under a rotation about the x_3-axis, a point describes a circle $x(t) = (a \cos(t), a \sin(t), b)$. Show that its velocity vector satisfies $x'(t) = d \times x(t)$ where $d = (0, 0, 1)$. (Compare with the previous exercise.).

18. Prove that (v · v)(w ·w)”(v ·w)2 = 0 if and only if the vectors v and w are linearly dependent.

19. Draw closed plane curves with rotation indices 0, –2, +3 respectively.

20. The theorem on turning tangents is also valid when the simple closed curve C has “corners.” Give the theorem when C is a triangle consisting of three arcs. Observe that the theorem contains as a special case the theorem on the sum of angles of a rectilinear triangle.

21. Give in detail the proof of the existence of $\eta = \eta(p_0)$ used in the proof of the theorem on turning tangents. $\eta = \eta\ (p_0)$.

Answers: Self Assessment

1.	straight line	2.	tangential indicatrix
3.	Fenchel’s Theorem	4.	non-inflectional
5.	$w(t) = 0$		

5 .14 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 6 : Serret-Frenet Formulae

Notes

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define serret-frenet formula
- Explain serret frenet formula

Introduction

In the last unit, you have studied about space theory of curve. Depending on how the arc is defined, either of the two end points may or may not be part of it. When the arc is straight, it is typically called a line segment. The derivatives of the vectors t, p, and b can be expressed as a linear combination of these vectors. The formulae for these expressions are called the Frenet-Serret Formulae. This is natural because t, p, and b form an orthogonal basis for a three-dimensional vector space.

6 .1 Serret-Frenet Formulae

Given a curve f:]a, b[$\to E^n$ (or f: [a, b] $\to E^n$) of class C^p, with $p \geq n$, it is interesting to consider families $(e_1(t), \ldots, e_n(t))$ of orthonormal frames. Moreover, if for every k, with $1 \leq k \leq n$, the kth derivative $f^{(k)}(t)$ of the curve f(t) is a linear combination of $(e_1(t), \ldots, e_k(t))$ for every t ∈]a, b[, then such a frame plays the role of a generalized Frenet frame. This leads to the following definition:

Lemma 1. Let f:]a, b[$\to E^n$ (or f : [a, b] $\to E^n$) be a curve of class C^p, with $p \geq n$. A family $(e_1(t), \ldots, e_n(t))$ of orthonormal frames, where each e_i :]a, b[$\to E^n$ is C^{n-i} continuous for i = 1, . . ., n – 1 and en is C^1-continuous, is called a moving frame along f. Furthermore, a moving frame $(e_1(t), \ldots, e_n(t))$ along f so that for every k, with $1 \leq k \leq n$, the kth derivative $f^{(k)}(t)$ of f(t) is a linear combination of $(e_1(t), \ldots, e_k(t))$ for every t ∈]a, b[, is called a Frenet n-frame or Frenet frame.

If $(e_1(t), \ldots, e_n(t))$ is a moving frame, then

$$e_i(t) \cdot e_j(t) = \delta_{ij} \text{ for all } i, j,\ 1 \leq i, j \leq n.$$

Lemma 2. Let f:]a, b[$\to E^n$ (or f: [a, b] $\to E^n$) be a curve of class C^p, with $p \geq n$, so that the derivatives $f^{(1)}(t), \ldots, f^{(n-1)}(t)$ of f(t) are linearly independent for all $t \in$]a, b[. Then, there is a unique Frenet n-frame $(e_1(t), \ldots, e_n(t))$ satisfying the following conditions:

(1) The k-frames $(f^{(1)}(t), \ldots, f^{(k)}(t))$ and $(e_1(t), \ldots, e_k(t))$ have the same orientation for all k, with $1 \leq k \leq n - 1$.

(2) The frame $(e_1(t), \ldots, e_n(t))$ has positive orientation.

Proof. Since $(f^{(1)}(t), \ldots, f^{(n-1)}(t))$ is linearly independent, we can use the Gram-Schmidt orthonormalization procedure to construct $(e_1(t), \ldots, e_{n-1}(t))$ from $(f^{(1)}(t), \ldots, f^{(n-1)}(t))$. We use the generalized cross-product to define e_n, where

$$e_n = e_1 \times \cdots \times e_{n-1}.$$

From the Gram-Schmidt procedure, it is easy to check that ek(t) is C^{n-k} for $1 \leq k \leq n - 1$, and since the components of e_n are certain determinants involving the components of $(e_1, \ldots, e_{n-1})$, it is also clear that en is C^1.

The Frenet n-frame given by Lemma 2 is called the distinguished Frenet n-frame. We can now prove a generalization of the Frenet-Serret formula that gives an expression of the derivatives of a moving frame in terms of the moving frame itself.

Lemma 3. Let f:]a, b[$\to E^n$ (or f: [a, b] $\to E^n$) be a curve of class C^p, with $p \geq n$, so that the derivatives $f^{(1)}(t), \ldots, f^{(n-1)}(t)$ of f(t) are linearly independent for all $t \in$]a, b[. Then, for any moving frame $(e_1(t), \ldots, e_n(t))$, if we write $\omega_{ij}(t) = e_i^{'}(t) \cdot e_j(t)$, we have

$$e_i^{'}(t) = \sum_{j=1}^{n} \omega_{ij}(t) e_j(t),$$

with

$$\omega_{ij}(t) = -\omega_{ij}(t),$$

and there are some functions $\alpha_i(t)$ so that

$$f'(t) = \sum_{i=1}^{n} \alpha_i(t) e_i(t).$$

Furthermore, if $(e_1(t), \ldots, e_n(t))$ is the distinguished Frenet n-frame associated with f, then we also have

$$\alpha_1(t) = \|f'(t)\|,\ \alpha_i(t) = 0 \text{ for } i \geq 2,$$

and

$$\omega_{ij}(t) = 0 \text{ for } j > i + 1.$$

Proof. Since $(e_1(t), \ldots, e_n(t))$ is a moving frame, it is an orthonormal basis, and thus, f′(t) and $e_t^{'}(t)$ are linear combinations of $(e_1(t), \ldots, {}_e n(t))$. Also, we know that

$$e_i^{'}(t) = \sum_{j=1}^{n} (e_i^{'}(t) \cdot e_j(t)) e_j(t),$$

and since $e_i(t) \cdot e_j(t) = \delta_{ij}$, by differentiating, if we write $\omega_{ij}(t) = e_i^{'}(t) \cdot e_j(t)$, we get

$$\omega_{ji}(t) = -\omega_{ij}(t).$$

Notes

Now, if $(e_1(t), \ldots, e_n(t))$ is the distinguished Frenet frame, by construction, $e_i(t)$ is a linear combination of $f^{(1)}(t), \ldots, f^{(i)}(t)$, and thus $e_i'(t)$ is a linear combination of $f^{(2)}(t), \ldots, f^{(i+1)}(t)$, hence, of $(e_1(t), \ldots, e_{i+1}(t))$.

In matrix form, when $(e_1(t), \ldots, e_n(t))$ is the distinguished Frenet frame, the row vector $(e_i'(t),\ldots,e_n'(t))$ can be expressed in terms of the row vector $(e_1(t), \ldots, e_n(t))$ via a skew symmetric matrix ω, as shown below:

$$(e_i'(t),\ldots,e_n'(t)) = -(e_1(t),\ldots,e_n(t))\omega(t)\ ,$$

where

$$\omega = \begin{pmatrix} 0 & \omega_{12} & & & \\ -\omega_{12} & 0 & \omega_{23} & & \\ & -\omega_{23} & 0 & \ddots & \\ & & \ddots & \ddots & \omega_{n-1n} \\ & & & -\omega_{n-1n} & 0 \end{pmatrix}$$

The next lemma shows the effect of a reparametrization and of a rigid motion.

Lemma 4. Let $f:]a, b[\to E^n$ (or $f: [a, b] \to E^n$) be curve of class C^p, with $p \geq n$, so that the derivatives $f^{(1)}(t), \ldots, f^{(n-1)}(t)$ of $f(t)$ are linearly independent for all $t \in]a, b[$. Let $h : E^n \to E^n$ be a rigid motion, and assume that the corresponding linear isometry is R. Let $\tilde{f} = h \circ f$. The following properties hold:

(1) For any moving frame $(e_1(t), \ldots, e_n(t))$, the n-tuple $(\tilde{e}_1(t), \ldots, \tilde{e}_n(t))$, where $\tilde{e}_i(t) = R(e_i(t))$, is a moving frame along ef, and we have

$$\tilde{\omega}_{ij}(t) = \omega_{ij}(t)) \quad \text{and} \quad \left\|\tilde{f}'(t)\right\| = \|f'(t)\|.$$

(2) For any orientation-preserving diffeomorphism $\rho :]c, d[\to]a, b[$ (i.e., $\rho'(t) > 0$ for all $t \in]c, d[$), if we write $\tilde{f} = f \circ r$, then for any moving frame $(e_1(t), \ldots, e_n(t))$ on f, the n-tuple $(\tilde{e}_1(t), \ldots, \tilde{e}_n(t))$, where $\tilde{e}_i(t) = e_i(\rho(t))$, is a moving frame on $\tilde{f}$.

Furthermore, if $\left\|\tilde{f}'(t)\right\| \neq 0$, then

$$\frac{\tilde{\omega}_{ij}(t)}{\left\|\tilde{f}'(t)\right\|} = \frac{\omega_{ij}(\rho(t))}{\|f'(\rho(t))\|}.$$

The proof is straightforward and is omitted.

The above lemma suggests the definition of the curvatures $\kappa_1, \ldots, \kappa_{n-1}$.

Lemma 5. Let $f :]a, b[\to E^n$ (or $f: [a, b] \to E^n$) be a curve of class C^p, with $p \geq n$, so that the derivatives $f^{(1)}(t), \ldots, f^{(n-1)}(t)$ of $f(t)$ are linearly independent for all $t \in]a, b[$. If $(e_1(t), \ldots, e_n(t))$ is the distinguished Frenet frame associated with f, we define the ith curvature, κ_i, of f, by

$$\kappa_i(t) = \frac{\omega_{ii+1}(t)}{\|f'(t)\|},$$

with $1 \leq i \leq n - 1$.

Notes

Observe that the matrix ω(t) can be written as

$$w(t) = ||f'(t)||\kappa(t),$$

where

$$\kappa = \begin{pmatrix} 0 & \kappa_{12} & & & & \\ -\kappa_{12} & 0 & \kappa_{23} & & & \\ & -\kappa_{23} & 0 & \ddots & & \\ & & \ddots & \ddots & \kappa_{n-1n} \\ & & & -k_{n-1n} & 0 \end{pmatrix}.$$

The matrix κ is sometimes called the Cartan matrix.

Lemma 6. Let $f :]a, b[\to E^n$ (or $f : [a, b] \to E^n$) be a curve of class C^p, with $p \geq n$, so that the derivatives $f^{(1)}(t), \ldots, f(n-1)^{(t)}$ of $f(t)$ are linearly independent for all $t \in]a, b[$. Then for every i, with $1 \leq i \leq n - 2$, we have $\kappa_i(t) > 0$.

Proof. Lemma 2 shows that $e_1, \ldots, e_{n-1}$ are expressed in terms of $f^{(1)}, \ldots, f^{(n-1)}$ by a triangular matrix (aij), whose diagonal entries a_{ii} are strictly positive, i.e., we have

$$e_i = \sum_{j=1}^{i} a_{ij} f^{(j)},$$

for $i = 1, \ldots, n - 1$, and thus,

$$f^{(i)} = \sum_{j=1}^{i} b_{ij} e_j,$$

for $i = 1, \ldots, n - 1$, with $b_{i\,i} = a_{ii}^{-1} > 0$. Then, since $e_{i+1} \cdot f^{(j)} = 0$ for $j \leq i$, we get

$$\|f'\|\kappa_i = \omega_{i\,i+1} = e_i' \cdot ei + 1 = a_{i\,i} f^{(i+1)} \cdot e_{i+1} = a_{i\,i} b_{i+1i+1},$$

and since $a_{i\,i} b_{i+1\,i+1} > 0$, we get $\kappa_i > 0$ $(i = 1, \ldots, n - 2)$.

We conclude by exploring to what extent the curvatures $\kappa_1, \ldots, \kappa_{n-1}$ determine a curve satisfying the non-degeneracy conditions of Lemma 2. Basically, such curves are defined up to a rigid motion.

Lemma 7. Let $f :]a, b[\to E^n$ and $\tilde{f} :]a, b[\to E^n$ (or $f : [a, b] \to E^n$ and $\tilde{f} : [a, b] \to E^n$) be two curves of class C^p, with $p \geq n$, and satisfying the non-degeneracy conditions of Lemma 2. Denote the distinguished Frenet frames associated with f and $\tilde{f}$ by $(e_1(t), \ldots, e_n(t))$ and $(\tilde{e}_1(t),\ldots,\tilde{e}_n(t))$.

If $\kappa_i(t) = \tilde{\kappa}_i(t)$ for every i, with $1 \leq i \leq n - 1$, and $\|f'(t)\| = \|\tilde{f}'(t)\|$ for all $t \in]a, b[$, then there is a unique rigid motion h so that

$$\tilde{f} = h \circ f.$$

Proof. Fix $t_0 \in]a, b[$. First of all, there is a unique rigid motion h so that

$$h(f(t_0)) = \tilde{f}(t_0) \text{ and } R(e_i(t_0)) = \tilde{e}_i(t_0),$$

for all i, with $1 \le i \le n$, where R is the linear isometry associated with h (in fact, a rotation). Consider the curve $\bar{f} = h \circ f$. The hypotheses of the lemma and Lemma 4, imply that

Notes

$$\bar{\omega}_{ij}(t) = \tilde{\omega}_{ij}(t) = \omega_{ij}(t),\ \|\bar{f}'(t)\| = \|\tilde{f}'(t)\| = \|f'(t)\|,$$

and, by construction,

$$(\bar{e}_1(t_0), \ldots, \bar{e}_n(t_0)) = (\tilde{e}_1(t_0), \ldots, \tilde{e}_n(t_0)) \text{ and } \bar{f}(t_0) = \tilde{f}(t_0).$$

Let

$$\delta(t) = \sum_{i=1}^{n} (\bar{e}_i(t) - \tilde{e}_i(t) \cdot (\bar{e}_i(t) - \tilde{e}_i(t)).$$

Then, we have

$$\delta'(t) = 2\sum_{i=1}^{n} (\bar{e}_i(t) - \tilde{e}_i(t) \cdot (\bar{e}'_i(t) - \tilde{e}'_i(t))$$

$$= -2\sum_{i=1}^{n} (\bar{e}_i(t) \cdot \tilde{e}'_i(t) + \bar{e}_i(t) \cdot \tilde{e}'_i(t)).$$

Using the Frenet equations, we get

$$\delta'(t) = -2\sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{ij} \bar{e}_i \cdot \tilde{e}_j - 2\sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{ij} \bar{e}_j \cdot \tilde{e}_i$$

$$= -2\sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{ij} \bar{e}_i \cdot \tilde{e}_j - 2\sum_{j=1}^{n}\sum_{i=1}^{n} \omega_{ij} \bar{e}_i \cdot \tilde{e}_j$$

$$= -2\sum_{i=1}^{n}\sum_{j=1}^{n} \omega_{ij} \bar{e}_i \cdot \tilde{e}_j + 2\sum_{j=1}^{n}\sum_{i=1}^{n} \omega_{ij} \bar{e}_i \cdot \tilde{e}_j$$

$$= 0,$$

since ω is skew symmetric. Thus, (t) is constant, and since the Frenet frames at t_0 agree, we get $\delta(t) = 0$.

Then, $\bar{e}_i(t) = \tilde{e}_i(t)$ for all i, and since $\|\bar{f}'(t)\| = \|\tilde{f}'(t)\|$, we

have

$$\bar{f}'(t) = \|\bar{f}'(t)\| \bar{e}_1(t) = \|\tilde{f}'(t)\| \tilde{e}_1(t) = \tilde{f}'(t),$$

so that $\bar{f}(t) - \tilde{f}(t)$ is constant. However, $\bar{f}(t_0) = \tilde{f}(t_0)$, and so, $\bar{f}(t) = \tilde{f}(t)$, and $\tilde{f} = \bar{f} = h \circ f$.

Lemma 8. Let $\kappa_1, \ldots, \kappa_{n-1}$ be functions defined on some open]a, b[containing 0 with κ_i C^{n-i-1} continuous for i = 1, . . ., n – 1, and with $\kappa_i(t) > 0$ for i = 1, . . ., n – 2 and all t ∈]a, b[. Then, there is curve f :]a, b[→ E^n of class C^p, with $p \ge n$, satisfying the non-degeneracy conditions of Lemma 2, so that $\|f'(t)\| = 1$ and f has the n – 1 curvatures $\kappa_1(t), \ldots, \kappa_{n-1}(t)$.

Proof. Let X(t) be the matrix whose columns are the vectors $e_1(t), \ldots, e_n(t)$ of the Frenet frame along f. Consider the system of ODE's,

$$X'(t) = -X(t)\kappa(t),$$

with initial conditions X(0) = I, where κ(t) is the skew symmetric matrix of curvatures. By a standard result in ODE's, there is a unique solution X(t).

We claim that X(t) is an orthogonal matrix. For this, note

that

$$(XX^T)' = X'X^T + X(X^T)' = -X\kappa X^T - X\kappa^T X^T$$
$$= -X\kappa X^T + X\kappa X^T = 0.$$

Since X(0) = I, we get $XX^T = I$. If F(t) is the first column of X(t), we define the curve f by

$$f(s) = \int_0^s F(t)dt,$$

with $s \in]a, b[$. It is easily checked that f is a curve parametrized by arc length, with Frenet frame X(s), and with curvatures κ_i's.

6 .2 Summary

- Lemma 1. Let f: $]a, b[\to E^n$ (or f : $[a, b] \to E^n$) be a curve of class C^p, with $p \geq n$. A family $(e_1(t), \ldots, e_n(t))$ of orthonormal frames, where each $e_i :]a, b[\to E^n$ is C^{n-i} continuous for i = 1, . . ., n – 1 and en is C^1-continuous, is called a moving frame along f. Furthermore, a moving frame $(e_1(t), \ldots, e_n(t))$ along f so that for every k, with $1 \leq k \leq n$, the kth derivative $f^{(k)}(t)$ of f(t) is a linear combination of $(e_1(t), \ldots, e_k(t))$ for every $t \in]a, b[$, is called a Frenet n-frame or Frenet frame.
- Lemma 2. Let f: $]a, b[\to E^n$ (or f: $[a, b] \to E^n$) be a curve of class C^p, with $p \geq n$, so that the derivatives $f^{(1)}(t), \ldots, f^{(n-1)}(t)$ of f(t) are linearly independent for all $t \in]a, b[$. Then, there is a unique Frenet n-frame $(e_1(t), \ldots, e_n(t))$ satisfying the following conditions:
  - The k-frames $(f^{(1)}(t), \ldots, f^{(k)}(t))$ and $(e_1(t), \ldots, e_k(t))$ have the same orientation for all k, with $1 \leq k \leq n - 1$.
  - The frame $(e_1(t), \ldots, e_n(t))$ has positive orientation.
- Let f: $]a, b[\to E^n$ (or f: $[a, b] \to E^n$) be curve of class C^p, with $p \geq n$, so that the derivatives $f^{(1)}(t), \ldots, f^{(n-1)}(t)$ of f(t) are linearly independent for all $t \in]a, b[$. Let $h : E^n \to E^n$ be a rigid motion, and assume that the corresponding linear isometry is R. Let $\tilde{f}$ = h o f. The following properties hold:
  - For any moving frame $(e_1(t), \ldots, e_n(t))$, the n-tuple $(\tilde{e}_1(t), \ldots, \tilde{e}_n(t))$, where $\tilde{e}_i(t) = R(e_i(t))$, is a moving frame along ef, and we have

    $$\tilde{\omega}_{ij}(t) = \omega_{ij}(t)) \quad \text{and} \quad \|\tilde{f}'(t)\| = \|f'(t)\|.$$

  - For any orientation-preserving diffeomorphism $\rho :]c, d[\to]a, b[$ (i.e., $\rho'(t) > 0$ for all $t \in]c, d[$), if we write $\tilde{f} = f \circ r$, then for any moving frame $(e_1(t), \ldots, e_n(t))$ on f, the n-tuple $(\tilde{e}_1(t), \ldots, \tilde{e}_n(t))$, where $\tilde{e}_i(t) = e_i(\rho(t))$, is a moving frame on $\tilde{f}$.

Notes

6 .3 Keywords

Moving frame along f: Let f:]a, b[$\to E^n$ (or f : [a, b] $\to E^n$) be a curve of class C^p, with $p \geq n$. A family $(e_1(t), \ldots, e_n(t))$ of orthonormal frames, where each e_i :]a, b[$\to E^n$ is C^{n-i} continuous for i = 1, . . ., n – 1 and en is C^1-continuous, is called a moving frame along f.

Frenet n-frame or Frenet frame: A moving frame $(e_1(t), \ldots, e_n(t))$ along f so that for every k, with $1 \leq k \leq n$, the kth derivative $f^{(k)}(t)$ of f(t) is a linear combination of $(e_1(t), \ldots, e_k(t))$ for every t $\in$]a, b[, is called a Frenet n-frame or Frenet frame.

Linear isometry: Let f:]a, b[$\to E^n$ (or f: [a, b] $\to E^n$) be curve of class C^p, with $p \geq n$, so that the derivatives $f^{(1)}(t), \ldots, f^{(n-1)}(t)$ of f(t) are linearly independent for all t $\in$]a, b[. Let h : $E^n \to E^n$ be a rigid motion, and assume that the corresponding linear isometry is R. Let $\tilde{f}$ = h o f.

6 .4 Self Assessment

1. Let f:]a, b[$\to E^n$ (or f : [a, b] $\to E^n$) be a curve of class C^p, with $p \geq n$. A family $(e_1(t), \ldots, e_n(t))$ of orthonormal frames, where each e_i :]a, b[$\to E^n$ is C^{n-i} continuous for i = 1, . . ., n – 1 and en is C^1-continuous, is called a ....
2. A moving frame $(e_1(t), \ldots, e_n(t))$ along f so that for every k, with $1 \leq k \leq n$, the kth derivative $f^{(k)}(t)$ of f(t) is a linear combination of $(e_1(t), \ldots, e_k(t))$ for every t $\in$]a, b[, is called a ....
3. Let f:]a, b[$\to E^n$ (or f: [a, b] $\to E^n$) be curve of class C^p, with $p \geq n$, so that the derivatives $f^{(1)}(t), \ldots, f^{(n-1)}(t)$ of f(t) are linearly independent for all t $\in$]a, b[. Let h : $E^n \to E^n$ be a rigid motion, and assume that the corresponding .... is R. Let $\tilde{f}$ = h o f.
4. Let $\kappa_1, \ldots, \kappa_{n-1}$ be functions defined on some open]a, b[containing 0 with κ_i C^{n-i-1} continuous for i = 1, . . ., n – 1, and with $\kappa_i(t) > 0$ for i = 1, . . ., n – 2 and all t $\in$]a, b[. Then, there is curve f :]a, b[$\to E^n$ of class C^p, with $p \geq n$, satisfying the ....
5. Let X(t) be the matrix whose columns are the vectors $e_1(t), \ldots, e_n(t)$ of the Frenet frame along f. Consider the system of ODE's, ...

6 .5 Review Questions

1. Show that the helix

$$\gamma : [0,10] \to E^8 : s \mapsto (2\cos(\frac{s}{\sqrt{5}}), 2\sin(\frac{s}{\sqrt{5}}), \frac{s}{\sqrt{5}})$$

   is a unit speed curve and has constant curvature and torsion.

2. Why do we always have $k[\beta] \geq 0$?
3. For all s, $\beta''(s) \cdot \beta(s) = 0$; so the acceleration is always perpendicular to the acceleration along unit-speed curves. What about $a'(t) \cdot a''(t)$ on arbitrary speed curves?
4. Derive the Frenet-Serret equations for an arbitrary-speed regular curve and show that the following hold for a curve with speed $\sqrt{\alpha' . \alpha'} = \|\alpha'\| = v > 0$: :

$$T = \alpha'/v, \quad N = B \times T, \quad B = \frac{\alpha' \times \alpha''}{\|\alpha' \times \alpha''\|}$$

$$k = \frac{\|\alpha' \times \alpha''\|}{v^3}$$

$$\tau = \frac{\alpha' \times \alpha'' . \alpha'''}{\|\alpha' \times \alpha''\|^2}.$$

5. Viviani's curve[2] is the intersection of the cylinder $(x - a)^2 + y^2 = a^2$ and the sphere $x^2 + y^2 + z^2 = 4a^2$ and has parametric equation:

$$a : [0, 4\pi] \to E^8 : t \mapsto a(1 + \cos t, \sin t, 2\sin\frac{t}{2}).$$

Show that it has curvature and torsion given by $k(t) = \dfrac{\sqrt{13 + 3\cos t}}{a(3 + \cos t)^{\frac{s}{2}}}$ and $\tau(t) = \dfrac{6\cos\frac{t}{2}}{a(13 + 3\cos t)}$.

6. Investigate the following curves for n = 0, 1, 2, 3

$$\gamma : [0, 2\pi\sqrt{6}] \to E^8 : s \mapsto (\sqrt{6}\cos(\frac{s}{\sqrt{6}}), \sqrt{\frac{3}{2}}\sin(\frac{s}{\sqrt{6}}), \frac{\sqrt{3}}{2}\sin(\frac{ns}{\sqrt{6}}))$$

7. Show that for all q ∈ [0, 2π] the matrix $R_s(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$ when applied to the coordinates of a curve in E^8 rotates the curve through angle θ in the (x,y)-plane, that is, round the z-axis. Find a matrix $R_y(\theta)$ representing rotation round the y-axis and hence obtain explicitly the result of rotating the curves in the previous question by 60° round the y-axis.

8. On plane curves, τ = 0 everywhere and we sometimes use the signed curvature $k^2{}_,$ defined by

$$k^2[a](t) = \frac{\alpha''(t).J\alpha'(t)}{\|a'(t)\|^3}, \text{ where J is the linear operator } J : \mathbb{R}^2 \to \mathbb{R}^2 : (p, g) \mapsto (-g, p).$$

We call $1/k^2[\alpha]$ the radius of curvature. Find the radius of curvature of some plane curves.

9. (i) Find two matrices, R_y and R_z from SO(3) which represent, respectively, rotation by π/3 about the y-axis and rotation by π/4 about the z-axis; each rotation must be in a right-hand-screw sense in the positive direction of its axis. Find the product matrix R_yR_z and show that its transpose is its inverse.

   (ii) By considering $(R_yR_z)^{-1}$, or otherwise, show that the curve

$$\gamma : [0, \infty) \to E^8 : t \mapsto (\frac{1}{\sqrt{2}}\cos h\, t/2 + \frac{t}{2\sqrt{2}}, -\sqrt{2}\cosh t/2 + \frac{t}{\sqrt{2}}, -\frac{\sqrt{3}}{\sqrt{2}}\cosh t/2 - \frac{\sqrt{3}t}{2\sqrt{2}})$$

   lies in a plane and find its curvature function and arc length function.

Notes

10. Vertical projection from E^8 onto its xy-plane is given by the map

$$\pi : E^8 \to E^8 : (x, y, z) \mapsto (x, y, 0).$$

A unit speed curve b : [0, L] $\to E^8$ lies above the xy-plane and has vertical projection

$$\pi \circ \beta : [0, L] \to E^8 : s \mapsto (\frac{s}{2}\cos(\log s/2), \frac{s}{2}\sin(\log s/2), 0).$$

Find explicitly a suitable β and for it compute the Frenet-Serret frame, curvature and torsion.

Answers: Self Assessment

1. moving frame along f
2. Frenet n-frame or Frenet frame
3. linear isometry
4. non-degeneracy conditions
5. $X'(t) = -X(t)k(t)$

6 .6 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Richa Nandra, Lovely Professional University

Unit 7 : Curves

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define Preliminaries
- Explain Plane Curves
- Identify Fenchel's Theorem

Introduction

In last unit, you have studied about serret-frenet formula. In this unit, you will read about curves.

7 .1 Preliminaries

Definition 1. A *parametrized curve* is a smooth (C^∞) function $\gamma : I \to \mathbb{R}^n$. A curve is regular if $\gamma' \neq 0$.

When the interval I is closed, we say that γ is C^∞ on I if there is an interval J and a C^∞ function β on J which agrees with γ on I.

Definition 2. Let $\gamma : I \to \mathbb{R}^n$ be a parametrized curve, and let $\beta : J \to \mathbb{R}^n$ be another parametrized curve. We say that β is a reparametrization (orientation preserving reparametrization) of γ if there is a smooth map $\tau : J \to I$ with $\tau' > 0$ such that $\beta = \gamma \circ \tau$.

Notes

> *Notes* The relation β is a reparametrization of γ is an equivalence relation. A *curve* is an equivalence class of parametrized curves. Furthermore, if γ is regular then every reparametrization of γ is also regular, so we may speak of *regular curves.*

Definition 3. Let $\gamma : I \blacklozenge \mathbb{R}^n$ be a regular curve. For any compact interval $[a, b] \subset I$, the arclength of γ over [a, b] is given by:

$$L_\gamma([a,b]) = \int_a^b |\gamma'|\,dt.$$

Note that if β is a reparametrization of γ then γ and β have the same length. More specifically, if $\beta = \gamma \circ \tau$, then

$$L_\gamma([\tau(c), \tau(d)]) = L_\beta([c,d]).$$

Definition 4. Let γ be a regular curve. We say that γ is parametrized by arc length if $|\gamma'| = 1$

Note that this is equivalent to the condition that for all $t \in I = [a, b]$ we have:

$$L_\gamma([a,t]) = t - a.$$

Furthermore, any regular curve can be parametrized by arclength. Indeed, if γ is a regular curve, then the function

$$s(t) = \int_a^t |\gamma'|,$$

is strictly monotone increasing. Thus, s(t) has an inverse function τ(s) function, satisfying:

$$\frac{d\tau}{ds} = \frac{1}{|\gamma'|}.$$

It is now straightforward to check that $\beta = \gamma \circ \tau$ is parametrized by arclength.

7 .2 Local Theory for Curves in $\mathbb{R}^3$

We will assume throughout this section that $\gamma : I \to \mathbb{R}^3$ is a regular curve in $\mathbb{R}^3$ parametrized by arclength and that $\gamma'' \neq 0.$ Note that $\gamma' \cdot \gamma'' = 0.$

Definition 5. Let $\gamma : I \to \mathbb{R}^3$ be a curve in $\mathbb{R}^3$. The unit vector $T = \gamma'$ is called the unit tangent of γ. The curvature κ is the scalar $\kappa = |\gamma''|.$ The unit vector $N = k^{-1} T'$ is called the principal normal. The binormal is the unit vector $B = T \times N$. The positively oriented orthonormal frame (T, N, B) is called the Frenet frame of γ.

It is not difficult to see that $N' + \kappa T$ is perpendicular to both T and N, hence, we can define the torsion τ of γ by: $N' + \kappa T = \tau B$. Note that the torsion, unlike the curvature, is signed. Finally, it is easy to check that $B' = -\tau N$. Let X denote the 3 × 3 matrix whose columns are (T, N, B). We will call X also the Frenet frame of γ. Define the rotation matrix of γ:

$$\omega := \begin{pmatrix} 0 & \kappa & 0 \\ -\kappa & 0 & \tau \\ 0 & -\tau & 0 \end{pmatrix} \qquad ...(1)$$

Notes

Proposition 1 (Frenet frame equations). The Frenet frame X = (T, N, B) of a curve in $\mathbb{R}^3$ satisfies:

$$X' = X\omega. \qquad ...(2)$$

The Frenet frame equations, Equation (2), form a system of nine linear ordinary differential equations.

Definition 6. A rigid motion of $\mathbb{R}^3$ is a function of the form $R(x) = x_0 + Qx$ where Q is orthonormal with det Q = 1.

Note that if X is the Frenet frame of γ and $R(x) = x_0 + Qx$ is a rigid motion of $\mathbb{R}^3$, then QX is the Frenet frame of $R \circ \gamma$. This follows easily from the fact that Q is preserves the inner product and orientation of $\mathbb{R}^3$.

Theorem 1 (Fundamental Theorem). Let $\kappa > 0$ and τ be smooth scalar functions on the interval [0, L]. Then there is a regular curve γ parametrized by arclength, unique up to a rigid motion of $\mathbb{R}^3$, whose curvature is κ and torsion is τ.

Proof. Let ω be given by (1). The initial value problem

$$X' = X\omega,$$
$$X(0) = I$$

can be solved uniquely on [0,L]. The solution X is an orthogonal matrix with det X = 1 on [0, L]. Indeed, since ω is anti-symmetric, the matrix $A = XX^t$ is constant. Indeed,

$$A' = X\omega X^t + X\omega^t X^t = X(\omega + \omega^t)X^t = 0,$$

and since A(0) = I, we conclude that $A \equiv I$, and X is orthogonal. Furthermore, detX is continuous, and detX(0) = 1, so detX = 1 on [0,L]. Let (T,N,B) be the columns of X, and let $\gamma = \int T$, then (T, N, B) is orthonormal and positively oriented on [0, L]. Thus, γ is parametrized by arclength, $\gamma' = T$, and $N = k^{-1} T'$ is the principal normal of γ. Similarly, B is the binormal, and consequently, κ is the curvature of γ and τ its torsion.

Now suppose that $\tilde{\gamma}$ is another curve with curvature κ and torsion τ, and let $\tilde{X}$ be its Frenet frame. Then there is a rigid motion $R(x) = Qx + x_0$ of $\mathbb{R}^3$ such that $R\gamma(0) = \tilde{\gamma}(0)$, and $QX(0) = \tilde{X}(0)$. By the remark preceding the theorem, QX is the Frenet frame of the curve $R \circ \gamma$, and thus, both QX and $\tilde{X}$ satisfy the initial value problem:

$$Y' = Y\omega,$$
$$Y(0) = QX(0).$$

By the uniqueness of solutions of the initial value problem, it follows that $QX = \tilde{X}$. In particular, $(R \circ \gamma)' = \tilde{\gamma}'$, and since $R \circ \gamma(0) = \tilde{\gamma}(0)$ we conclude $R \circ \gamma = \tilde{\gamma}$.

Assuming $\gamma(0) = 0$, the Taylor expansion of γ of order 3 at s = 0 is:

$$\gamma(s) = \gamma'(0)s + \frac{1}{2}\gamma''(0)s^2 + \frac{1}{6}\gamma'''(0)s^3 + O(s^4).$$

Denote $T_0 = T(0)$, $N_0 = N(0)$, $B_0 = B(0)$, $\kappa_0 = \kappa(0)$, and $\tau_0 = \tau(0)$. We have $\gamma'(0) = T_0, \gamma''(0) = \kappa_0 N_0$, and $\gamma'''(0) = \kappa'(0)N_0 + \kappa_0(-k_0T_0 + \tau_0 B_0)$. Substituting these into the equation above, decomposing into T, N, and B components, and retaining only the leading order terms, we get:

$$\gamma(s) = \left(s + O\left(s^3\right)\right)T + \left(\frac{\kappa}{2}s^2 + O\left(s^3\right)\right)N + \left(\frac{\tau}{6}s^3 + O\left(s^4\right)\right)B$$

Notes

The planes spanned by pairs of vectors in the Frenet frame are given special names:

(1) T and N — the osculating plane;

(2) N and B — the normal plane;

(3) T and B — the rectifying plane.

We see that to second order the curve stays within its osculating plane, where it traces a parabola $y = (k/2)\, s^2$. The projection onto the normal plane is a cusp to third order: $x = \left((3\tau/2)y\right)^{2/3}$. The projection onto the rectifying plane is to second order a line, whence its name.

Here are a few simple applications of the Frenet frame.

Theorem 2. Let γ be a regular curve with $k \equiv 0$. Then γ is a straight line.

Proof. Since $|T'| = k = 0$, it follows that T is constant and γ is linear.

Theorem 3. Let γ be a regular curve with $k > 0$, and $\tau = 0$. Then γ is planar.

Proof. Since $B' = 0$, B is constant. Thus the function $\xi = (\gamma - \gamma(0))$. B vanishes identically:

$$\xi(0) = 0, \quad \xi' = T \cdot B = 0.$$

It follows that γ remains in the plane through $\gamma(0)$ perpendicular to B.

Theorem 4. Let γ be a regular curve with k constant and $\tau = 0$. Then γ is a circle.

Proof. Let $b = \gamma + k^{-1} N$. Then

$$\beta' = T + \frac{1}{k}(-kT + \tau B) = 0.$$

Thus, β is constant, and $|\gamma - \beta| = k^{-1}$. It follows that γ lies in the intersection between a plane and a sphere, thus γ is a circle.

7 .3 Plane Curves

7 .3.1 Local Theory

Let $\gamma : [a,b] \to \mathbb{R}^2$ be a regular plane curve parametrized by arclength, and let k be its curvature. Note that k is signed, and in fact changes sign (but not magnitude) when the orientation of γ is reversed. The Frenet frame equations are:

$$e_1' = ke_2, \quad e_2' = -ke_1$$

Proposition 2. Let $\gamma : [a,b] \to \mathbb{R}^2$ be a regular curve with $|\gamma'| = 1$. Then there exists a differentiable function $\theta : [a,b] \to \mathbb{R}$ such that

$$e_1 = (\cos\theta, \sin\theta). \qquad ...(3)$$

Moreover, θ is unique up to a constant integer multiple of 2π, and in particular $\theta(b) - \theta(a)$ is independent of the choice of θ. The derivative of θ is the curvature: $\theta' = k$.

Notes

Proof. Let $a = t_0 < t_1 < \cdots < t_n = b$ be a partition of [a, b] so that the diameter of $e_1([t_{i-1}, t_i])$ is less than 2, i.e., e_1 restricted to each subinterval maps into a semi-circle. Such a partition exists since e_1 is uniformly continuous on [a, b]. Choose $\theta(a)$ so that (3) holds at a, and proceed by induction on i: if θ is defined at t_i then there is a unique continuous extension so that (3) holds. If ψ is any other continuous function satisfying (3), then $k = (1/2\pi)(\theta - \psi)$ is a continuous integer-valued function, hence is constant. Finally, $e_2 = (-\sin\theta, \cos\theta)$ hence

$$e_1' = ke_2 = \theta'(-\sin\theta, \cos\theta),$$

and we obtain $\theta' = k$.

7 .3.2 Global Theory

Definition 7. A curve $\gamma : [a,b] \to \mathbb{R}^n$ is closed if $\gamma^{(k)}(a) = \gamma^{(k)}(b).$ A closed curve $\gamma : [a,b] \to \mathbb{R}^n$ is simple if $\gamma\big|_{(a,b)}$ is one-to-one. The rotation number of a smooth closed curve is:

$$n_\gamma = \frac{1}{2\pi}\big(\theta(a) - \theta(b)\big), \quad \text{...(4)}$$

where θ is the function defined in Proposition 2.

We note that the rotation number is always an integer. For reference, we also note that the rotation number of a curve is the winding number of the map e_1. Finally, in view of the last statement in Proposition 2, we have:

$$n_\gamma = \frac{1}{2\pi}\int_{[0,L]} k\, ds.$$

Theorem 5 (Rotation Theorem). Let $\gamma : [0,L] \to \mathbb{R}^2$ be a smooth, regular, simple, closed curve. Then $n_\gamma = \pm 1.$ In particular,

$$\frac{1}{2\pi}\int_{[0,L]} k\, ds = \pm 1.$$

For the proof, we will need the following technical lemma. We say that a set $\Delta \subset \mathbb{R}^n$ is star-shaped with respect to $x_0 \in \Delta$ if for every $y \in \Delta$ the line segment $\overline{x_0 y}$ lies in Δ.

Lemma 1. Let $\Delta \subset \mathbb{R}^n$ Rn be star-shaped with respect to $x_0 \in \Delta,$ and let $e : \Delta \to \mathbb{S}^1$ be a continuous function. Then there exists a continuous function $\theta : \Delta \to \mathbb{R}$ such that:

$$e = (\cos\theta, \sin\theta). \quad \text{...(5)}$$

Moreover, if ψ is another continuous function satisfying (5), then $\theta - \psi = 2\pi k$ where k is a constant integer.

In fact, it is sufficient to assume that Δ is simply connected, but we will not prove this more general result here.

Proof. Define $\theta(x_0)$ so that (5) holds at x_0. For each $x \in \Delta$, define θ continuously along the line segment $\overline{x_0 x}$ as in the proof of Proposition 2. Since Δ is star-shaped with respect to x_0, this

Notes

defines θ everywhere in Δ. It remains to show that θ is continuous. Let $y_0 \in \Delta$. Since $\overline{x_0 y_0}$ is compact, it is possible to choose δ small enough that the following holds: $y' \in \overline{x_0 y_0}$ and $|y - y'| < \delta$ implies $|e(y) - e(y')| < 2$ or equivalently e(y) and e(y′) are not antipodal. Let 0 < ∈ < π. Then there exists a neighborhood $U \subset B_\delta(y_0)$ of y_0 such that y ∈ U implies $\theta(y) - \theta(y_0) = 2\pi k(y) + \in'(y)$ where $|\in'(y)| < \in$ and $k(y)$ is integer-valued. It remain to prove that k ≡ 0. Let y ∈ U and consider the continuous function:

$$\phi(s) = \theta\left(x_0 + s(y - x_0)\right) - \theta\left(x0 + s(y_0 - x_0)\right), \quad 0 \le s \le 1.$$

Since

$$\left|\left(x_0 + s(y - x_0)\right) - (x_0 + s(y_0 - x_0))\right| = |s(y - y_0)| < \delta,$$

it follows from our choice of δ that $e\left(x_0 + s(y - x_0)\right)$ and $e\left(x_0 + s(y_0 - x_0)\right)$ are not antipodal. Thus, $\phi(s) \ne \pi$ for all $0 \le s \le 1$, and since ϕ(0) = 0 we conclude that $|\phi| < \pi$. In particular,

$$|2\pi k(y) + \in'(y)| = |\theta(y) - \theta(y_0)| = |\phi(1)| < \pi,$$

and it follows that

$$|2\pi k(y)| \le |2\pi k(y) + \in'(y)| + |\in'(y)| < 2\pi.$$

Since k(y) is integer-valued this implies k(y) = 0.

Proof of the Rotation Theorem. Pick a line which intersects the curve γ and pick a last point *p* on this line, i.e., a point with the property that one ray of the line from *p* has no other intersection points with γ. Let *h* be the unit vector pointing in the direction of that ray. We assume without loss of generality that γ is parametrized by arclength, $\gamma(0) = \gamma(L) = 0$. Now, let $\Delta = \left\{(t_1, t_2) \in \mathbb{R}^2 : 0 \le t_1 \le t_2 \le L\right\}$, and note that Δ is star-shaped. Define the $\mathbb{S}^1$-valued function:

$$e(t_1, t_2) = \begin{cases} \gamma'(t_1) & \text{if } t_1 = t_2; \\ -\gamma'(0) & \text{if } (t_1, t_2) = (0, L); \\ \dfrac{\gamma(t_2) - \gamma(t_1)}{|\gamma(t_2) - \gamma(t_1)|} & \text{otherwise.} \end{cases}$$

It is straightforward to check that e is continuous on Δ. By the Lemma, there is a continuous function $\theta : \Delta \to \mathbb{R}$ such that e = (cos θ, sin θ). We claim that θ(L,L) – θ(0, 0) = ±2π which proves the theorem, since θ(t, t) is a continuous function satisfying (3) in Proposition 6, and thus can be used on the right-hand side of (4) to compute the rotation number.

To prove this claim, note that, for any 0 < t < L, the unit vector

$$e(0, t) = \frac{\gamma(t) - \gamma(0)}{|\gamma(t) - \gamma(0)|}$$

is never equal to h. Hence, there is some value α such that θ(0, t) – θ(0, 0) ≠ α + 2πk for any integer k. Thus, $|\theta(0, t) - \theta(0, 0)| < 2\pi$, and since e(0,L) = –e(0, 0) it follows that θ(0,L) – θ(0, 0) = ±π.

Since the curves e(0, t) and e(t, L) are related via a rigid motion, i.e., e(t, L) = Re(0, t) where R is rotation by π, it follows that $\psi(t) = (\theta(t,L) - \theta(0,L)) - (\theta(0,t) - \theta(0,0))$ is a constant. Since clearly $\psi(0) = 0$, we get θ(0, L) – θ(0, 0) = θ(L,L) – θ(0,L), and we conclude:

$$\theta(L,L) - \theta(0,0) = (\theta(t,L) - \theta(0,L)) + (\theta(0,t) - \theta(0,0)) = \pm 2\pi.$$

Definition 8. A piecewise smooth curve is a continuous function $\gamma : [a,b] \to \mathbb{R}^n$ such that there is a partition of [a, b]:

$$a = a_0 < a_1 < \cdots < b_n = b$$

such that for each $1 \le j \le n$ the curve segment $\gamma_j = \gamma\big|[a_{j-1}, a_j]$ is smooth. The points $\gamma(a_j)$ are called the corners of γ. The directed angle $-\pi < \psi_j \le \pi$ from $\gamma'(a_j-)$ to $\gamma'(a_j+)$ is called the exterior angle at the j-th corner. Define $\theta_j : [a_{j-1}, a_j] \to \mathbb{R}$ as in Proposition 2, i.e., so that $\gamma'_j = (\cos\theta_j, \sin\theta_j)$. The rotation number of γ is given by:

$$n_\gamma = \frac{1}{2\pi}\sum_{j=1}^{n}(\theta_j(a_j) - \theta_j(a_{j-1})) + \frac{1}{2\pi}\sum_{j=1}^{n}\psi_j.$$

Again, n_γ is an integer, and we have:

$$n_\gamma = \frac{1}{2\pi}\int_{[a,b]} k\, ds + \frac{1}{2\pi}\sum_{j=1}^{n}\psi_j.$$

The Rotation Theorem can be generalized to piecewise smooth curves provided corners are taken into account.

Theorem 6. Let $\gamma : [0,L] \to \mathbb{R}^2$ be a piecewise smooth, regular, simple, closed curve, and assume that none of the exterior angles are equal to π. Then $n_\gamma = \pm 1$.

7 .3.3 Convexity

Definition 9. Let $\gamma : [0,L] \to \mathbb{R}^2$ be a regular closed plane curve. We say that γ is convex if for each $t_0 \in [0, L]$ the curve lies on one side only of its tangent at t_0, i.e., if one of the following inequality holds:

$$(\gamma - \gamma(t0)) \cdot e_2 \le 0,$$

$$(\gamma - \gamma(t0)) \cdot e_2 \ge 0,$$

Theorem 7. Let $\gamma : [0,L] \to \mathbb{R}^2$ be a regular simple closed plane curve, and let k be its curvature. Then γ is convex if and only if either $k \ge 0$ or $k \le 0$.

We note that an orientation reversing reparametrization of γ changes $k \ge 0$ into $k \le 0$ and vice versa. Thus, ignoring orientation, those two conditions are equivalent. We also note that the theorem fails if γ is not assumed simple.

Proof. We may assume without loss of generality that $|\gamma'| = 1$. Let $\theta : [0,L] \to \mathbb{R}$ be the continuous function given in Proposition 2 satisfying:

$$e_1 = (\cos\theta, \sin\theta),$$

and $\theta' = k$.

Suppose that γ is convex. We will show that θ is weakly monotone, i.e., if $t_1 < t_2$ and $\theta(t_1) = (t_2)$ then θ is constant on $[t_1, t_2]$. First, we note that since γ is simple, we have $n_\gamma = \pm 1$ by the Rotation Theorem, and it follows that e_1 is onto $\mathbb{S}^1$, see Exercise 5. Thus, there is $t_3 \in [0,L]$ such that

$$e_1(t_3) = -e_1(t_1) = -e_1(t_2).$$

By convexity, the three parallel tangents at t_1, t_2, and t_3 cannot be distinct, hence at least two must coincide. Let $p_1 = \gamma(s_1)$ and $p_2 = \gamma(s_2), s_1 < s_2$ denote these two points, then the line $\overline{p_1p_2}$ is contained in γ. Otherwise, if q is a point on $\overline{p_1p_2}$ not on γ, then the line through q perpendicular to $\overline{p_1p_2}$ intersects γ in at least two points r and s, which by convexity must lie on one side of $\overline{p_1p_2}$. Without loss of generality, assume that r is the closer of the two to $\overline{p_1p_2}$. Then r lies in the interior of the triangle p_1p_2s. Regardless of the inclination of the tangent at r, the three points p_1, p_2 and s, all belonging to γ, cannot all lie on one side of the tangent, in contradiction to convexity. If $\overline{p_1p_2} \neq \{\gamma(s) : s_1 \le s \le s_2\}$, then $\overline{p_1p_2} = \{\gamma(s) : s_2 \le s \le L\} \cap \{\gamma(s) : 0 \le s \le s_1\}$. However, in that case, we would have $\theta(s_2) - \theta(s_1) = \theta(L) - \theta(0) = 2\pi$, a contradiction. Thus, we have

$$\overline{p_1p_2} = \{\gamma(s) : s_1 \le s \le s_2\} = \{\gamma(t) : t_1 \le t \le t_2\}.$$

In particular $\theta(t) = \theta(t_1) = \theta(t_2)$.

Conversely, suppose that γ is not convex. Then, there is $t_0 \in [0, L]$ such that the function $\phi = (\gamma - \gamma(t_0)) \cdot e_2$ changes sign. We will show that θ' also changes sign. Let t_+, $t_- \in [0,L]$ be such that

$$\min_{[0,L]} \phi = \phi(t_-) < 0 = \phi(t_0) = \phi(t+) = \max_{[0,L]} \phi.$$

Note that the three tangents at t_-, t_+ and t_0 are parallel but distinct. Since $\phi'(t_-) = \phi'(t_+) = 0$, we have that $e_1(t_-)$ and $e_1(t_+)$ are both equal to $\pm e_1(t_0)$.

Thus, at least two of these vectors are equal. We may assume, after reparametrization, that there exists $0 < s < L$ such that $e_1(0) = e_1(s)$. This implies that

$$\theta(s) - \theta(0) = 2\pi k, \quad \theta(L) - \theta(s) = 2\pi k'$$

with $k, k' \in \mathbb{Z}$. By the Rotation Theorem, $n_\gamma = k + k' = \pm 1$. Since $\gamma(0)$ and $\gamma(s)$ do not lie on a line parallel to $e_1(t_0)$, it follows that is not constant on either [0, s] or [0, L]. If $k = 0$ then θ' changes sign on [0, s], and similarly if $k' = 0$ then θ' changes sign on [s, L]. If $kk' \neq 0$, then since $k + k' = \pm 1$, it follows that $kk' < 0$ and θ' changes sign on [0, L].

Definition 8. Let $\gamma : [0,L] \to \mathbb{R}^2$ be a regular plane curve. A vertex of γ is a critical point of the curvature k.

Theorem 11 (The Four Vertex Theorem). A regular simple convex closed curve has at least four vertices.

Proof. Clearly, k has a maximum and minimum on [0,L], hence γ has at least two vertices. We will assume, without loss of generality, that γ is parametrized by arclength, has its minimum at

$t = 0$, its maximum at $t = t_0$ where $0 < t_0 < L$, that $\gamma(0)$ and $\gamma(t_0)$ lie on the x-axis, and that γ enters the upper-half plane in the interval $[0, t_0]$. All these properties can be achieved by reparametrizing and rotating γ.

We now claim that $p = \gamma(0)$ and $q = \gamma(t_0)$ are the only points of γ on the x-axis. Indeed, suppose that there is another point $r = \gamma(t_1)$ on the x-axis, then one of these points lies between the other two, and the tangent at that point must, by convexity, contain the other two. Thus, by the argument used in the proof of Theorem 10 the segment between the outer two is contained in γ, and in particular $\overline{pq}$ is contained in γ. If follows that $k \equiv 0$ at p and q where k has its minimum and maximum, hence $k \equiv 0$, a contradiction since then γ is a line and cannot be closed. We conclude that γ remains in the upper half-plane in the interval $[0, t_0]$ and remains in the lower half-plane in the interval $[t_0, L]$.

Suppose now by contradiction that $\gamma(0)$ and $\gamma(t_0)$ are the only vertices of γ. Then it follows that:

$$k' \geq 0 \text{ on } [0, t_0], \quad k' \leq 0 \text{ on } [t_0, L].$$

Thus, if we write $\gamma = (x, y)$, then we have $k'y \geq 0$ on $[0, L]$, and $x'' = -ky'$, hence:

$$0 = \int_0^L x''\,ds = -\int_0^L -ky'\,ds = \int_0^L k'y\,ds.$$

Since the integrand in the last integral is non-negative, we conclude that $k'y \equiv 0$, hence $y \equiv 0$, again a contradiction.

It follows that k has another point where k' changes sign, i.e., an extremum.

Since extrema come in pairs, k has at least four extrema.

7 .4 Fenchel's Theorem

We will use without proof the fact that the shortest path between two points on a sphere is always an arc of a great circle. We also use the notation $\gamma_1 + \gamma_2$ to denote the curve γ_1 followed by the curve γ_2.

Definition 11. Let $\gamma : [0,L] \to \mathbb{R}^n$ be a regular curve parametrized by arclength. The spherical image of γ is the curve $\gamma' : [0,L] \to \mathbb{S}^{n-1}$. The total curvature of $\gamma : [0,L] \to \mathbb{R}^n$ is:

$$K_\gamma = \int_I |\gamma''|\,ds.$$

We note that the total curvature is simply the length of the spherical image.

Theorem 9. Let γ be a regular simple closed curve in $\mathbb{R}^n$ parametrized by arclength. Then the total curvature of γ is at least 2π:

$$K_\gamma \geq 2\pi,$$

with equality if and only if γ is planar and convex.

The proof will follow from two lemmata which are interesting in their own right.

Lemma 2. Let $\gamma : [0,L] \to \mathbb{R}^n$ be a regular closed curve parametrized by arclength. Then the spherical image of γ cannot map into an open hemisphere. If γ' maps into a closed hemisphere, then γ maps into an equator.

Proof. Suppose, by contradiction, that there is $\upsilon \in \mathbb{S}^{n-1}$ such that $\gamma' \cdot \upsilon > 0$.

Then

$$0 = \gamma \cdot \upsilon|L - \gamma \cdot \upsilon|_0 = \int_0^L \gamma' \cdot \upsilon \, ds > 0.$$

If $\gamma' \cdot \upsilon \geq 0$, then the same inequality shows that $\gamma' \cdot \upsilon \equiv 0$, hence, γ lies in the plane perpendicular to υ through $\gamma(0)$.

Lemma 3. Let $n \geq 3$, and let $\gamma : [0,L] \to \mathbb{S}^{n-1}$ be a regular closed curve on the unit sphere parametrized by arclength.

1. If the arclength of γ is less than 2π then γ is contained in an open hemisphere.
2. If the arclength of γ is equal to 2π then γ is contained in a closed hemisphere.

Proof

1. First observe that no piecewise smooth curve of arclength less than 2π contains two antipodal points. Otherwise the two segments of the curve between p and q would each have length at least π, and hence, the length of the curve would have to be at least 2π. Now pick a point p on γ and let q on γ be chosen so that the two segments γ_1 and γ_2 from p to q along γ have equal length. Note that p and q cannot be antipodal. Let υ be the midpoint along the shorter of the two segments of the great circle between p and q. Suppose that γ_1 intersects the equator, the great circle $\upsilon \cdot x = 0$. Let $\tilde{\gamma}_1$ be the reflection of γ with respect to υ, then the length of $\gamma_1 + \tilde{\gamma}_1$ is the same as the length of γ hence is less than 2π. But $\gamma_1 + \tilde{\gamma}_1$ contains two antipodal points, a contradiction. Thus, γ_1 cannot intersect the equator. Similarly, γ_2 cannot intersect the equator, and we conclude γ stays in the open hemisphere $\upsilon \cdot x > 0$.
2. If the arclength of γ is 2π, we refine the above argument. If p and q are antipodal, then both γ_1 and γ_2 are great semi-circle, thus, γ stays in a closed hemisphere.[1] So we can assume that p and q are not antipodal and proceed as before, defining υ to be the midpoint on the shorter arc of the great circle between p and q. Now, if γ_1 crosses the equator, then $\gamma_1 + \tilde{\gamma}_1$ contains two antipodal points on the equator, and the two segments joining these points enter both hemispheres. Thus, these segments are not semi-circle, and consequently both have arclength strictly greater than π. Thus the arclength of $\gamma_1 + \tilde{\gamma}_1$ is strictly larger than 2π a contradiction. Similarly, γ_2 does not cross the equator, and we conclude that γ stays in the closed hemisphere $\upsilon \cdot x \geq 0$.

Proof of Fenchel's Theorem. Note that the total curvature is simply the arclength of the spherical image of γ. By Lemma 2 γ' is not contained in an open hemisphere, so by Lemma 3

$$K_\gamma = \int_I |\gamma''| \, ds \geq 2\pi.$$

If the arclength of γ' is 2π, then by Lemma 3, γ' is contained in a closed hemisphere, and by Lemma 2, γ maps into an equator. If $n > 3$, we may proceed by induction until we obtain that is planar. Once we have that γ is planar, the Rotation Theorem gives $n_\gamma = \pm 1$. Without loss of generality,[2] we may assume that $n_\gamma = 1$. Hence,

---

1 In fact, since γ is smooth, γ_1 and γ_1 are contained in the same great circle, and hence γ is itself a great circle.

2 Reversing the orientation of γ if necessary

Notes

$$0 \le \int_I \left(|k| - k\right) ds = K_\gamma - 2\pi = 0,$$

and it follows that $k = |k| \ge 0,$ which by Theorem 7 implies that γ is convex.

7 .5 Summary

- ***Definition 1.*** A *parametrized curve* is a smooth (C^∞) function $\gamma : I \blacklozenge \mathbb{R}^n$. A curve is regular if $\gamma' \quad 0.$

  When the interval I is closed, we say that γ is C^∞ on I if there is an interval J and a C^∞ function β on J which agrees with γ on I.

- ***Definition 2.*** Let $\gamma : I \blacklozenge \mathbb{R}^n$ be a parametrized curve, and let $\beta : J \blacklozenge \mathbb{R}^n$ be another parametrized curve. We say that β is a reparametrization (orientation preserving reparametrization) of γ if there is a smooth map $\tau : J \blacklozenge I$ with $\tau' > 0$ such that $\beta = \gamma \circ \tau.$

- ***Definition 4.*** Let γ be a regular curve. We say that γ is parametrized by arclength if $|\gamma'| = 1$

  Note that this is equivalent to the condition that for all t ∈ I = [a, b] we have:

  $$L_\gamma\left([a, t]\right) = t - a.$$

  Furthermore, any regular curve can be parametrized by arclength. Indeed, if γ is a regular curve, then the function

  $$s(t) = \int_a^t |\gamma'|,$$

- ***Definition 5.*** Let $\gamma : I \to \mathbb{R}^3$ be a curve in $\mathbb{R}^3$. The unit vector $T = \gamma'$ is called the unit tangent of γ. The curvature κ is the scalar $\kappa = |\gamma''|.$ The unit vector $N = k^{-1} T'$ is called the principal normal. The binormal is the unit vector B = T × N. The positively oriented orthonormal frame (T, N, B) is called the Frenet frame of γ.

- ***Theorem 3.*** Let γ be a regular curve with k ≡ 0. Then γ is a straight line.

  Proof. Since $|T'| = k = 0,$ it follows that T is constant and γ is linear.

- ***Theorem 4.*** Let γ be a regular curve with k > 0, and τ = 0. Then γ is planar.

  Proof. Since B′ = 0, B is constant. Thus the function $\xi = \left(\gamma - \gamma(0)\right).$ B vanishes identically:

  $$\xi(0) = 0, \qquad \xi' = T \cdot B = 0.$$

  It follows that γ remains in the plane through γ(0) perpendicular to B.

- ***Theorem 5.*** Let γ be a regular curve with k constant and τ = 0. Then γ is a circle.

  Proof. Let $b = \gamma + k^{-1} N$. Then

  $$\beta' = T + \frac{1}{k}(-kT + \tau B) = 0.$$

  Thus, β is constant, and $|\gamma - \beta| = k^{-1}.$ It follows that γ lies in the intersection between a plane and a sphere, thus γ is a circle.

7 .6 Keywords

Parametrized curve: A parametrized curve is a smooth (C^∞) function $\gamma : I$ ♦ $\mathbb{R}^n$. A curve is regular if γ' 0.

Frenet frame equations. The Frenet frame X = (T, N, B) of a curve in $\mathbb{R}^3$ satisfies: (1.2) X′ = Xω. The Frenet frame equations, form a system of nine linear ordinary differential equations.

Fundamental Theorem: Let κ > 0 and τ be smooth scalar functions on the interval [0, L]. Then there is a regular curve γ parametrized by arclength, unique up to a rigid motion of $\mathbb{R}^3$, whose curvature is κ and torsion is τ.

7 .7 Self Assessment

1. A .... is a smooth (C^∞) function $\gamma : I$ ♦ $\mathbb{R}^n$. A curve is regular if γ' 0.
2. Let $\gamma : I \to \mathbb{R}^3$ be a curve in $\mathbb{R}^3$. The unit vector $T = \gamma'$ is called the .... of γ. The curvature κ is the scalar $\kappa = |\gamma''|$.
3. The Frenet frame X = (T, N, B) of a curve in $\mathbb{R}^3$ satisfies: (1.2) X′ = Xω. The ...., Equation (1.2), form a system of nine linear ordinary differential equations.
4. A rigid .... is a function of the form $R(x) = x_0 + Qx$ where Q is orthonormal with det Q = 1.

7 .8 Review Questions

1. A regular space curve $\gamma : [a,b] \to \mathbb{R}^3$ is a helix if there is a fixed unit vector $u \in \mathbb{R}^3$ such that $e_1 \cdot u$ is constant. Let k and τ be the curvature and torsion of a regular space curve γ, and suppose that k ≠ 0. Prove that γ is a helix if and only if τ = ck for some constant c.
2. Let $\gamma : I \to \mathbb{R}^4$ be a smooth curve parameterized by arclength such that $\gamma', \gamma'', \gamma'''$ are linearly independent. Prove the existence of a Frenet frame, i.e., a positively oriented orthonormal frame $X = (e_1, e_2, e_3, e_4)$ satisfying $e_1 = \gamma'$, and X′ = Xw, where w is anti-symmetric, tri-diagonal, and $w_{i,i+1} > 0$ for i ≤ n – 2. The curvatures of γ are the three functions $k_i = w_{i,i+1}$, i = 1, 2, 3. Note that $k_1, k_2 > 0$, but k_3 has a sign.
3. Prove the Fundamental Theorem for curves in $\mathbb{R}^4$: Given functions k_1, k_2, k_3 on I with $k_1, k_2 > 0$, there is a smooth curve γ parameterized by arclength on I such that k_1, k_2, k_3 are the curvatures of γ. Furthermore, γ is unique up to a rigid motion of $\mathbb{R}^4$.
4. Let $\gamma : [a,b] \to \mathbb{R}^2$ be a regular plane curve with non-zero curvature k ≠ 0, and let $\beta = \gamma + k^{-1} N$ be the locus of the centers of curvature of γ.
   (a) Prove that β is regular provided that k′ = 0.
   (b) Prove that each tangent ℓ of β intersects γ at a right angle.
   (c) Prove that each regular plane curve $\gamma : [a,b] \to \mathbb{R}^2$ has at most one evolute.

5. A convex plane curve $\gamma:[a,b]\to\mathbb{R}^2$ is strictly convex if $k \neq 0$. Prove that if $\gamma:[a,b]\to\mathbb{R}^2$ is a strictly convex simple closed curve, then for every $\upsilon \in \mathbb{S}^1$, there is a unique $t \in [a,b]$ such that $e_1(t) = \upsilon$.

6. Let $\gamma:[0,L]\to\mathbb{R}^2$ be a strictly convex simple closed curve. The width w(t) of γ at $t \in [0,L]$ is the distance between the tangent line at $\gamma(t)$ and the tangent line at the unique point $\gamma(t')$ satisfying $e_1(t') = -e_1(t)$. A curve has constant width if w is independent of t. Prove that if γ has constant width then:

   (a) The line between $\gamma(t)$ and $\gamma(t')$ is perpendicular to the tangent lines at those points.

   (b) The curve γ has length $L = \pi w$.

7. Let $\gamma:[0,L]\to\mathbb{R}^2$ be a simple closed curve. By the Jordan Curve Theorem, the complement of γ has two connected components, one of which is bounded. The area enclosed by γ is the area of this component, and according to Green's Theorem, it is given by:

$$A = \int_\gamma x\,dy = \int_\gamma xy'\,dt,$$

   where the orientation is chosen so that the normal e_2 points into the bounded component. Let L be the length of γ, and let β be a circle of width 2r equal to some width of γ. Prove:

   (a) $A = \frac{1}{2}\int_\gamma \left(xy' - yx'\right)dt.$

   (b) $A + \pi r^2 \leq Lr$.

   (c) The isoperimetric inequality: $4\pi A \leq L^2$.

   (d) If equality holds in (3) then γ is a circle.

8. Prove that if a convex simple closed curve has four vertices, then it cannot meet any circle in more than four points.

Answers: Self Assessment

1. parametrized curve
2. unit tangent
3. Frenet frame equations
4. motion of $\mathbb{R}^3$

7 .9 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Notes

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 8 : New Spherical Indicatrices and their Characterizations

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define Preliminaries
- Explain Tangent Bishop Spherical Images
- Identify M_1 and M_2 Bishop

Introduction

In the existing literature, it can be seen that, most of classical differential geometry topics have been extended to Lorentzian manifolds. In this process, generally, researchers used standard moving Frenet-Serret frame. Using transformation matrix among derivative vectors and frame vectors, some of kinematical models were adapted to this special moving frame. Researchers aimed to have an alternative frame for curves and other applications. Bishop frame, which is also called alternative or parallel frame of the curves, was introduced by L.R. Bishop in 1975 by means of parallel vector fields.

Spherical images of a regular curve in the Euclidean space are obtained by means of Frenet-Serret frame vector fields, so this classical topic is a well-known concept in differential geometry of the curves. In the light of the existing literature, this unit aims to determine new spherical images of regular curves using Bishop frame vector fields. We shall call such curves, respectively, Tangent, M_1 and M_2 Bishop spherical images of regular curves. Considering classical methods, we investigated relations among Frenet-Serret invariants of spherical images in terms of Bishop invariants. Additionally, two examples of Bishop spherical indicatrices are presented.

8 .1 Preliminaries

Notes

To meet the requirements in the next sections, here, the basic elements of the theory of curves in the space E^3 are briefly presented; a more complete elementary treatment can be found further.

The Euclidean 3-space E^3 provided with the standard flat metric given by

$$\langle\ ,\ \rangle = dx_1^2 + dx_2^2 + dx_3^2,$$

where (x_1, x_2, x_3) is a rectangular coordinate system of E^3. Recall that, the norm of an arbitrary vector $a \in E^3$ is given by $\|a\| = \sqrt{\langle a, a\rangle}$. φ is called a unit speed curve if velocity vector v of φ satisfies $\|v\| = 1$. For vectors v, $w \in E^3$ it is said to be orthogonal if and only if $\langle v, w\rangle = 0$. Let $v = v(s)$ be a regular curve in E^3. If the tangent vector of this curve forms a constant angle with a fixed constant vector U, then this curve is called a general helix or an inclined curve. The sphere of radius $r > 0$ and with center in the origin in the space E^3 is defined by

$$S^2 = \{p = (p_1; p_2; p_3) \in E^3 : \langle p; p\rangle = r^2\}.$$

Denote by {T, N, B} the moving Frenet-Serret frame along the curve φ in the space E^3. For an arbitrary curve φ with first and second curvature, K and T in the space E^3, the following Frenet-Serret formulae are written under matrix form

$$\begin{bmatrix} T' \\ N' \\ B' \end{bmatrix} = \begin{bmatrix} 0 & K & 0 \\ -K & 0 & T \\ 0 & -T & 0 \end{bmatrix} \begin{bmatrix} T \\ N \\ B \end{bmatrix},$$

where

$$\langle T, T\rangle = \langle N, N\rangle = \langle B, B\rangle = 1,$$

$$\langle T, N\rangle = \langle T, B\rangle = \langle T, N\rangle = \langle N, B\rangle = 0.$$

Here, curvature functions are defined by $K = K(s) = \|T'(s)\|$ and $T(s) = \langle N, B\rangle$.

Let $u = (u_1, u_2, u_3)$, $v = (v_1, v_2, v_3)$ and $w = (w_1, w_2, w_3)$ be vectors in E^3 and e_1, e_2, e_3 be positive oriented natural basis of E^3. Cross product of u and v is defined by

$$u \times v = \begin{vmatrix} e_1 & e_2 & e_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}.$$

Mixed product of u; v and w is defined by the determinant

$$[u, v, w] = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}.$$

Torsion of the curve φ is given by the aid of the mixed product

$$T = \frac{[\varphi', \varphi'', \varphi''']}{K^2}.$$

The Bishop frame or parallel transport frame is an alternative approach to defining a moving frame that is well defined even when the curve has vanishing second derivative. One can express parallel transport of an orthonormal frame along a curve simply by parallel transporting each component. The tangent vector and any convenient arbitrary basis for the remainder of the frame are used.

$$\begin{bmatrix} T' \\ M_1' \\ M_2' \end{bmatrix} = \begin{bmatrix} 0 & k_1 & k_2 \\ -k_1 & 0 & 0 \\ -k_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} T \\ M_1 \\ M_2 \end{bmatrix}. \tag{1}$$

Here, we shall call the set {T, M_1, M_2} as Bishop trihedra and k_1 and k_2 as Bishop curvatures. The relation matrix may be expressed as

$$\begin{bmatrix} T \\ N \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta(s) & \sin\theta(s) \\ 0 & -\sin\theta(s) & \cos\theta(s) \end{bmatrix} \begin{bmatrix} T \\ M_1 \\ M_2 \end{bmatrix},$$

where $\theta(s) = \arctan \dfrac{k_2}{k_1}$, $T(s) = \theta'(s)$ and $k(s) = \sqrt{k_1^2 + k_2^2}$. Here, Bishop curvatures are defined by

$$\begin{cases} k_1 = K \cos \mu(s) \\ k_2 = K \sin \mu(s) \end{cases}.$$

Izumiya and Takeuchi have introduced the concept of slant helix in the Euclidean 3-space E^3 saying that the normal lines makes a constant angle with a fixed direction. They characterized a slant helix by the condition that the function

$$\frac{k^2}{(k^2 + t^2)^{3/2}} \left(\frac{T}{K}\right)'$$

is constant. In further researches, spherical images, the tangent and the binormal indicatrix and some characterizations of such curves are presented. In the same space, the authors defined and gave some characterizations of slant helices according to Bishop frame with the following definition and theorem:

Definition 1. A regular curve $\gamma : I \to E^3$ is called a slant helix according to Bishop frame provided the unit vector $M_1(s)$ of γ has constant angle θ with some fixed unit vector u; that is,

$$\langle M_1, u \rangle = \cos\theta$$

for all $s \in I$.

Theorem 1. Let $\gamma : I \to E^3$ be a unit speed curve with nonzero natural curvatures. Then γ is a slant helix if and only if

$$\frac{k_1}{k_2} = \text{constant}.$$

To separate a slant helix according to Bishop frame from that of Frenet-Serret frame, in the rest of the paper, we shall use notation for the curve defined above as "B-slant helix".

Notes

It is well-known that for a unit speed curve with non-vanishing curvatures the following propositions hold:

Proposition 1. Let $\varphi = \varphi(s)$ be a regular curve with curvatures K and T. The curve φ lies on the surface of a sphere if and only if

$$\frac{T}{K} + \left[\frac{1}{T}\left(\frac{1}{K}\right)'\right]' = 0.$$

Proposition 2. Let $\varphi = \varphi(s)$ be a regular curve with curvatures K and T. φ is a general helix if and only if

$$\frac{K}{T} = \text{constant}.$$

Proposition 3. Let $\varphi = \varphi(s)$ be a regular curve with curvatures K and T. φ is a slant helix if and only if

$$\sigma(s) = \left[\frac{k^2}{(K^2 + T^2)^{\frac{3}{2}}}\left(\frac{T}{K}\right)'\right] = \text{constant}.$$

8 .2 Tangent Bishop Spherical Images of a Regular Curve

Definition 2. Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the first (tangent) vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image $\xi = \xi(s_\xi)$. This curve is called tangent Bishop spherical image or indicatrix of the curve $\gamma = \gamma(s)$.

Let $\xi = \xi(s_\xi)$ be tangent Bishop spherical image of a regular curve $\gamma = \gamma(s)$. One can differentiate of ξ with respect to s :

$$\xi' = \frac{d\xi}{ds_\xi}\frac{ds_\xi}{ds} = k_1 M_1 + K_2 M_2.$$

Here, we shall denote differentiation according to s by a dash, and differentiation according to s_ξ by a dot. In terms of Bishop frame vector fields, we have the tangent vector of the spherical image as follows:

$$T_\xi = \frac{k_1 M_1 + k_2 M_2}{\sqrt{k_1^2 + k_2^2}},$$

where

$$\frac{ds_\xi}{ds} = \sqrt{k_1^2 + k_2^2} = k(s).$$

In order to determine the first curvature of ξ, we write

$$\dot{T}_\xi = -T + \frac{k_2^3}{(k_1^2 + k_2^2)^2}\left(\frac{k_1}{k_2}\right)' M_1 + \frac{k_1^3}{(k_1^2 + k_2^2)^2}\left(\frac{k_2}{k_1}\right)' M_2.$$

Since, we immediately arrive at

$$K_\xi = \|\dot{T}_\xi\| = \sqrt{1 + \left[\frac{k_2^3}{(k_1^2+k_2^2)^2}\left(\frac{k_1}{k_2}\right)'\right]^2 + \left[\frac{k_1^3}{(k_1^2+k_2^2)^2}\left(\frac{k_2}{k_1}\right)'\right]^2}. \qquad (2)$$

Therefore, we have the principal normal

$$N_\xi = \frac{1}{K_\xi}\left\{-T + \frac{k_2^3}{(k_1^2+k_2^2)^2}\left(\frac{k_1}{k_2}\right)' M_1 + \frac{k_1^3}{(k_1^2+k_2^2)^2}\left(\frac{k_2}{k_1}\right)' M_2\right\}.$$

By the cross product of $T_\xi \times N_\xi$, we obtain the binormal vector field

$$B_\xi = \frac{1}{K_\xi}\left\{\left[\frac{k_1^4}{(k_1^2+k_2^2)^{\frac{5}{2}}}\left(\frac{k_2}{k_1}\right)' - \frac{k_2^4}{(k_1^2+k_2^2)^{\frac{5}{2}}}\left(\frac{k_1}{k_2}\right)'\right]T - \left[\frac{k_2}{\sqrt{k_1^2+k_2^2}}\right]M_1 + \left[\frac{k_1}{\sqrt{k_1^2+k_2^2}}\right]M_2\right\}.$$

By means of obtained equations, we express the torsion of the tangent Bishop spherical image

$$T_\xi = \frac{(-k_1\{3k_2'(k_1k_1'+k_2k_2') - (k_1^2+k_2^2)[k_2'' - k_2(k_1^2+k_2^2)]\} + k_2\{3k_1'(k_1k_1'+k_2k_2') - (k_1^2+k_2^2)[k_1'' - k_1(k_1^2+k_2^2)\}}{\left[k_1^2\left(\frac{k_2}{k_1}\right)'\right]^2 + (k_1^2+k_2^2)^3} \qquad (3)$$

Consequently, we determined Frenet-Serret invariants of the tangent Bishop spherical indicatrix according to Bishop invariants.

Corollary 1. Let $\xi = \xi(s_\xi)$ be the tangent Bishop spherical image of a regular curve $\gamma = \gamma(s)$. If $\gamma = \gamma(s)$ is a B-slant helix, then the tangent spherical indicatrix ξ is a circle in the osculating plane.

Proof. Let $\xi = \xi(s_\xi)$ be the tangent Bishop spherical image of a regular curve $\gamma = \gamma(s)$: If $\gamma = \gamma(s)$ is a B-slant helix, then Theorem 1 holds. So, $\frac{k_1}{k_2} = \text{constant}.$ Substituting this to equations (2) and (3), we have K_ξ = constant and $T_\xi = 0$, respectively. Therefore, ξ is a circle in the osculating plane.

Remark 1. Considering $\theta_\xi = \int_0^{s_\xi} T_\xi ds_\xi$ and using the transformation matrix, one can obtain the Bishop trihedra $\{T_\xi, M_{1\xi}, M_{2\xi}\}$ of the curve $\xi = \xi(s_\xi)$.

Here, one question may come to mind about the obtained tangent spherical image, since, Frenet-Serret and Bishop frames have a common tangent vector field. Images of such tangent images are the same as we shall demonstrate in subsequent section. But, here we are concerned with the tangent Bishop spherical image's Frenet-Serret apparatus according to Bishop invariants.

8 .3 M_1 Bishop Spherical Images of a Regular Curve

Definition 3. Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the second vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image $\delta = \delta(s_\delta)$. This curve is called M_1 Bishop spherical image or indicatrix of the curve $\gamma = \gamma(s)$.

Notes

Let $\delta = \delta(s_\delta)$ be M_1 Bishop spherical image of a regular curve $\gamma = \gamma(s)$. We follow the same procedure to investigate the relations among Bishop and Frenet-Serret invariants. Thus, we differentiate

$$\delta' = \frac{d\delta}{ds_\delta}\frac{ds_\delta}{ds} = -k_1 T.$$

First, we have

$$T_\delta = T \text{ and } \frac{ds_\delta}{ds} = -k_1. \tag{4}$$

So, one can calculate

$$T_\delta' = \dot{T}_\delta \frac{ds_\delta}{ds} = k_1 M_1 + k_2 M_2$$

or

$$T_\delta' = -M_1 - \frac{k_2}{k_1} M_2.$$

Since, we express

$$K_\delta = \left\|\dot{T}_\delta\right\| = \sqrt{1 + \left(\frac{k_2}{k_1}\right)^2} \tag{5}$$

and

$$N_\delta = -\frac{M_1}{K_\delta} - \frac{k_2}{k_1 K_\delta} M_2.$$

Cross product of $T_\delta \times N_\delta$ gives us the binormal vector field of M_1 spherical image of $\gamma = \gamma(s)$

$$B_\delta = \frac{k_2}{k_1 K_\delta} M_1 - \frac{1}{K_\delta} M_2.$$

Using the formula of the torsion, we write

$$T_\delta = -\frac{k_1\left(\frac{k_2}{k_1}\right)'}{k_1^2 + k_2^2}. \tag{6}$$

Considering equations (5) and (6) by the Theorem 1, we get:

Corollary 2. Let $\delta = \delta(s_\delta)$ be the M_1 Bishop spherical image of the curve $\gamma = \gamma(s)$. If $\gamma = \gamma(s)$ is a B-slant helix, then, the M_1 Bishop spherical indicatrix $\delta(s_\delta)$ is a circle in the osculating plane.

Theorem 2. Let $\delta = \delta(s_\delta)$ be the M_1 Bishop spherical image of a regular curve $\gamma = \gamma(s)$. There exists a relation among Frenet-Serret invariants $\delta(s_\delta)$ and Bishop curvatures of $\gamma = \gamma(s)$ as follows:

$$\frac{k_2}{k_1} = \int_0^{s_\delta} K_\delta^2 T_\delta ds_\delta. \tag{7}$$

Notes

Proof. Let $\delta = \delta(s_\delta)$ be M_1 Bishop spherical image of a regular curve $\gamma = \gamma(s)$. Then, the equations (4) and (6) hold. Using (4) in (6), we have

$$T_\delta = -\frac{k_1 \dfrac{d}{ds_\delta}\left(\dfrac{k_2}{k_1}\right)\dfrac{ds_\delta}{ds}}{k_1^2 + k_2^2}. \tag{8}$$

Substituting (5) to (8) and integrating both sides, we have (7) as desired.

In the light of the Propositions 2 and 3, we state the following theorems without proofs:

Theorem 3. Let $\delta = \delta(s_\delta)$ be M_1 Bishop spherical image of a regular curve $\gamma = \gamma(s)$. If δ is a general helix, then, Bishop curvatures of γ satisfy

$$\frac{k_1^2\left(\dfrac{k_2}{k_1}\right)'}{(k_1^2 + k_2^2)^{\frac{3}{2}}} = \text{constant}.$$

Theorem 4. Let $\delta = \delta(s_\delta)$ be the M_1 Bishop spherical image of a regular curve $\gamma = \gamma(s)$. If δ is a slant helix, then, the Bishop curvatures of γ satisfy

$$\left[\frac{k_1^2\left(\dfrac{k_2}{k_1}\right)'}{(k_1^2 + k_2^2)^{\frac{3}{2}}}\right]' \frac{(k_1^2 + k_2^2)^4}{k_1^3\left[\left(\dfrac{k_2}{k_1}\right)'^2 + (k_1^2 + k_2^2)^3\right]^{\frac{3}{2}}} = \text{constant}.$$

We know that δ is a spherical curve, so, by the Proposition 3 one can prove.

Theorem 5. Let δ be the M_1 Bishop spherical image of a regular curve $\gamma = \gamma(s)$. The Bishop curvatures of the regular curve $\gamma = \gamma(s)$ satisfy the following differential equation

$$\frac{k_1^2\left(\dfrac{k_2}{k_1}\right)'}{(k_1^2 + k_2^2)^{\frac{3}{2}}} - \left[\frac{k_1 k_2}{\sqrt{k_1^2 + k_2^2}}\right]' = \text{constant}.$$

Remark 2. Considering $\theta_\delta = \int_0^{s_\delta} T_\delta ds_\delta$ and using the transformation matrix, one can obtain the Bishop trihedra $\{T_\delta, M_{1\delta}, M_{2\delta}\}$ of the curve $\delta = \delta(s_\delta)$.

8 .4 M_2 Bishop Spherical Images of a Regular Curve

Definition 11. Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the third vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image of $\psi = \psi(s_\psi)$. This curve is called the M_2 Bishop spherical image or the indicatrix of the curve $\gamma = \gamma(s)$.

Let $\psi = \psi(s_\psi)$ be M_2 spherical image of the regular curve $\gamma = \gamma(s)$: We can write

$$\psi' = \frac{d\psi}{ds_\psi}\frac{ds_\psi}{ds} = -k_2 T.$$

Notes

Similar to the M_1 Bishop spherical image, one can have

$$T_\psi = T \text{ and } \frac{ds_\psi}{ds} = -k_2. \tag{9}$$

So, by differentiating of the formula (9), we get

$$T_\psi^{'} = \dot{T}\psi \frac{ds_\psi}{ds} = k_1 M_1 + k_2 M_2$$

or, in another words,

$$T_\psi^{'} = -\frac{k_1}{k_2} M_1 - M_2,$$

since, we express

$$k_\psi = \left\| \dot{T}_y \right\| = \sqrt{1 + \left(\frac{k_1}{k_2}\right)^2} \tag{10}$$

and

$$N_\psi = -\frac{k_1}{k_2 K_\psi} M_1 - \frac{M_2}{K_\psi}.$$

The cross product $T_\psi \times N_\psi$ gives us

$$B_\psi = -\frac{1}{K_\psi} M_1 - \frac{k_1}{k_2 K_\psi} M_2.$$

By the formula of the torsion, we have

$$T_\psi = \frac{k_2 \left(\frac{k_1}{k_2}\right)^{'}}{k_1^2 + k_2^2}. \tag{11}$$

In terms of equations (10) and (11) and by the Theorem 2, we may obtain:

Corollary 3. Let $\psi = \psi(s_\psi)$ be the M_2 spherical image of a regular curve $\gamma = \gamma(s)$. If $\gamma = \gamma(s)$ is a B-slant helix, then the M_2 Bishop spherical image $\psi(s_\psi)$ is a circle in the osculating plane.

Theorem 6. Let $\psi = \psi(s_\psi)$ be the M_2 spherical image of a regular curve $\gamma = \gamma(s)$. Then, there exists a relation among Frenet-Serret invariants of $\psi(s_\psi)$ and the Bishop curvatures of $\gamma = \gamma(s)$ as follows:

$$\frac{k_1}{k_2} + \int_0^{s_\psi} k_\psi^2 T_\psi ds_\psi = 0.$$

Proof. Similar to proof of the theorem 6, above equation can be obtained by the equations (9), (10) and (11).

In the light of the propositions 4 and 5, we also give the following theorems for the curve $\psi = \psi(s_\psi)$:

Theorem 7. Let $\psi = \psi(s_\psi)$ be the M_2 Bishop spherical image of a regular curve $\gamma = \gamma(s)$. If ψ is a general helix, then, Bishop curvatures of γ satisfy

$$\frac{k_2^2\left(\frac{k_1}{k_2}\right)'}{(k_1^2+k_2^2)^{\frac{3}{2}}} = \text{constant}.$$

Theorem 8. Let $\psi = \psi(s_\psi)$ be the M_2 Bishop spherical image of a regular curve $\gamma = \gamma(s)$. If ψ is a slant helix, then, the Bishop curvatures of γ satisfy

$$\left[\frac{k_2^2\left(\frac{k_1}{k_2}\right)'}{(k_1^2+k_2^2)^{\frac{3}{2}}}\right]' \frac{(k_1^2+k_2^2)^4}{k_2^3\left[\left(\frac{k_1}{k_2}\right)'^2 + (k_1^2+k_2^2)^3\right]^{\frac{3}{2}}} = \text{constant}.$$

We also know that ψ is a spherical curve. By the Proposition 3, it is safe to report the following theorem:

Theorem 9. Let $\psi = \psi(s_\psi)$ be the M_2 Bishop spherical image of a regular curve $\gamma = \gamma(s)$. The Bishop curvatures of the regular curve $\gamma = \gamma(s)$ satisfy the following differential equation

$$\frac{k_2^2\left(\frac{k_1}{k_2}\right)'}{(k_1^2+k_2^2)^{\frac{3}{2}}} + \left[\frac{k_1 k_2}{\sqrt{k_1^2+k_2^2}}\right]' = \text{constant}.$$

Remark 3. Considering $\theta_\psi = \int_0^{s_\psi} T_\psi ds_\psi$ and using the transformation matrix, one can obtain the Bishop trihedra $\{T_\psi, M_{1\psi}, M_{2\psi}\}$ of the curve $\psi = \psi(s_\psi)$.

Example 1: In this section, we give two examples of Bishop spherical images.

First, let us consider a unit speed circular helix by

$$\beta = \beta(s) = \left(a\cos\frac{s}{c}, a\sin\frac{s}{c}, \frac{bs}{c}\right), \tag{12}$$

where c = $\sqrt{a^2+b^2} \in R$. One can calculate its Frenet-Serret apparatus as the following:

$$\begin{cases} K = \dfrac{a}{c^2} \\ T = \dfrac{b}{c^2} \\ T = \dfrac{1}{c}(-a\sin\dfrac{s}{c}, a\cos\dfrac{s}{c}, b) \\ N = (-\cos\dfrac{s}{c}, -\sin\dfrac{s}{c}, 0) \\ B = \dfrac{1}{c}(b\sin\dfrac{s}{c}, -b\cos\dfrac{s}{c}, a) \end{cases}$$

Notes

In order to determine the Bishop frame of the curve $\beta = \beta(s)$, let us form

$$\theta(s) = \int_0^s \frac{b}{c^2} ds = \frac{bs}{c^2}.$$

Since, we can write the transformation matrix

$$\begin{bmatrix} T \\ N \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\frac{bs}{c^2} & \sin\frac{bs}{c^2} \\ 0 & -\sin\frac{bs}{c^2} & \cos\frac{bs}{c^2} \end{bmatrix} \begin{bmatrix} T \\ M_1 \\ M_2 \end{bmatrix},$$

by the method of Cramer, one can obtain the Bishop trihedra as follows:

The tangent:

$$T = \frac{1}{c}(-a\sin\frac{s}{c}, a\cos\frac{s}{c}, b) \qquad (13)$$

The M_1:

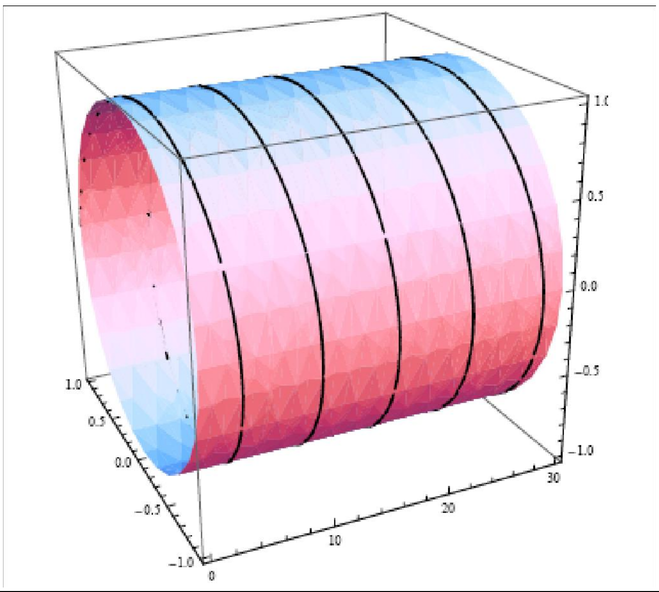
$$M_1 = (-\cos\frac{s}{c}\cos\frac{bs}{c^2} - \frac{b}{c}\sin\frac{s}{c}\sin\frac{bs}{c^2}, \frac{b}{c}\cos\frac{s}{c}\sin\frac{bs}{c^2} - \sin\frac{s}{c}\cos\frac{bs}{c^2}, -\frac{a}{c}\sin\frac{bs}{c^2})$$

The M_2:

$$M_2 = (\frac{b}{c}\sin\frac{s}{c}\cos\frac{bs}{c^2} - \cos\frac{s}{c}\sin\frac{bs}{c^2}, -\frac{b}{c}\cos\frac{s}{c}\cos\frac{bs}{c^2} - \sin\frac{s}{c}\sin\frac{bs}{c^2}, \frac{a}{c}\cos\frac{bs}{c^2})$$

We may choose a = 12, b = 5 and c = 13 in the equations (12–15). Then, one can see the curve at the Figure 21.1. So, we can illustrate spherical images see Figure 21.2.

Figure 8 .1: Circular Helix b = b(s) for a = 12; b = 5 and c = 13



Notes

Example 2: Next, let us consider the following unit speed curve $\gamma(s) = (\gamma_1, \gamma_2, \gamma_3)$:

$$\begin{cases} \gamma_1 = \dfrac{9}{208}\sin 16s - \dfrac{1}{117}\sin 36s \\ \gamma_2 = -\dfrac{9}{208}\cos 16s + \dfrac{1}{117}\cos 36s \\ \gamma_3 = \dfrac{6}{65}\sin 10s \end{cases}$$

It is rendered in Figure 8 .2. And, this curve's curvature functions are expressed as in [12]:

$$\begin{cases} K(s) = -24\sin 10s \\ T(s) = 24\cos 10s \end{cases}$$

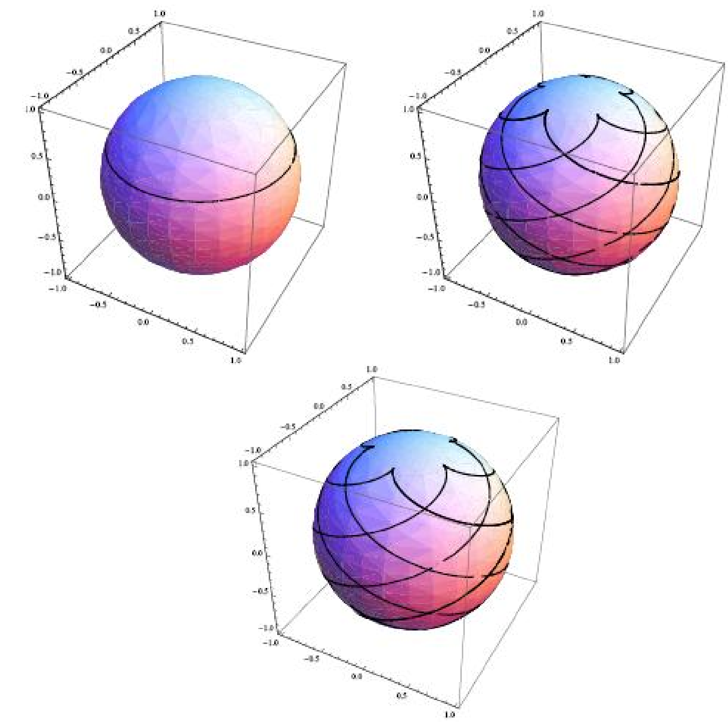
It is an easy problem to calculate Frenet-Serret apparatus of the unit speed curve $\gamma = \gamma(s)$. We also need

$$\theta(s) = \int_0^s 24\cos(10s)ds = \frac{24}{10}\sin(10s).$$

The transformation matrix for the curve $\gamma = \gamma(s)$ has the form

$$\begin{bmatrix} T \\ N \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\frac{24}{10}\sin 10s) & \sin(\frac{24}{10}\sin 10s) \\ 0 & -\sin(\frac{24}{10}\sin 10s) & \cos(\frac{24}{10}\sin 10s) \end{bmatrix} \begin{bmatrix} T \\ M_1 \\ M_2 \end{bmatrix}$$

Figure 8 .2: Tangent, M_1 and M_2 Bishop Spherical Images of [redacted] = [redacted](s) for a = 12; b = 5 and c = 13.



8 .5 Summary

Notes

- If the tangent vector of this curve forms a constant angle with a fixed constant vector U, then this curve is called a general helix or an inclined curve.
- A regular curve $\gamma : I \to E^3$ is called a slant helix according to Bishop frame provided the unit vector $M_1(s)$ of γ has constant angle θ with some fixed unit vector u; that is,

$$\langle M_1, u\rangle = \cos\theta$$

  for all $s \in I$.
- Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the first (tangent) vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image $\xi = \xi(s_\xi)$. This curve is called tangent Bishop spherical image or indicatrix of the curve $\gamma = \gamma(s)$.
- Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the second vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image $\delta = \delta(s_\delta)$. This curve is called M_1 Bishop spherical image or indicatrix of the curve $\gamma = \gamma(s)$.
- Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the third vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image of $\psi = \psi(s_\psi)$. This curve is called the M_2 Bishop spherical image or the indicatrix of the curve $\gamma = \gamma(s)$.

8 .6 Keywords

General helix: If the tangent vector of this curve forms a constant angle with a fixed constant vector U, then this curve is called a general helix or an inclined curve.

Slant helix: A regular curve $\gamma : I \to E^3$ is called a slant helix according to Bishop frame provided the unit vector $M_1(s)$ of γ has constant angle θ with some fixed unit vector u; that is,

$$\langle M_1, u\rangle = \cos\theta$$

for all $s \in I$.

Tangent Bishop spherical image: Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the first (tangent) vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image $\xi = \xi(s_\xi)$. This curve is called tangent Bishop spherical image or indicatrix of the curve $\gamma = \gamma(s)$.

M_1 Bishop spherical image: Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the second vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image $\delta = \delta(s_\delta)$. This curve is called M_1 Bishop spherical image or indicatrix of the curve $\gamma = \gamma(s)$.

M_2 Bishop spherical image: Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the third vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image of $\psi = \psi(s_\psi)$. This curve is called the M_2 Bishop spherical image or the indicatrix of the curve $\gamma = \gamma(s)$.

8 .7 Self Assessment

1. A regular curve $\gamma : I \to E^3$ is called a .... according to Bishop frame provided the unit vector $M_1(s)$ of γ has constant angle θ with some fixed unit vector u; that is,

$$\langle M_1, u\rangle = \cos\theta$$

   for all $s \in I$.

2. If the tangent vector of this curve forms a constant angle with a fixed constant vector U, then this curve is called a .... or an inclined curve

3. Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the first (tangent) vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image $\xi = \xi(s_\xi)$. This curve is called .... or indicatrix of the curve $\gamma = \gamma(s)$.

4. Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the second vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image $\delta = \delta(s_\delta)$. This curve is called .... or indicatrix of the curve $\gamma = \gamma(s)$.

5. Let $\gamma = \gamma(s)$ be a regular curve in E^3. If we translate of the third vector field of Bishop frame to the center O of the unit sphere S^2, we obtain a spherical image of $\psi = \psi(s_\psi)$. This curve is called the .... or the indicatrix of the curve $\gamma = \gamma(s)$.

8 .8 Review Questions

1. Find Relation Matrix $\begin{bmatrix} T \\ N \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta(s) & \sin\theta(s) \\ 0 & -\sin\theta(s) & \cos\theta(s) \end{bmatrix} \begin{bmatrix} T \\ M_1 \\ M_2 \end{bmatrix}$.

2. Let $\xi = \xi(s_\xi)$ be the tangent Bishop spherical image of a regular curve $\gamma = \gamma(s)$. If $\gamma = \gamma(s)$ is a B-slant helix, then the tangent spherical indicatrix ξ is a circle in the osculating plane.

3. Next, let us consider the following unit speed curve $\gamma(s) = (\gamma_1, \gamma_2, \gamma_3)$:

$$\begin{cases} \gamma_1 = \dfrac{9}{208}\sin 16s - \dfrac{1}{116}\sin 36s \\ \gamma_2 = -\dfrac{9}{327}\cos 16s + \dfrac{1}{117}\cos 36s \\ \gamma_3 = \dfrac{6}{65}\sin 10s \end{cases}$$

Answers: Self Assessment

1. slant helix
2. general helix
3. tangent Bishop spherical image
4. M_1 Bishop spherical image
5. M_2 Bishop spherical image

8 .9 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Notes

Unit 9 : Bertrand Curves

CONTENTS

Objectives

After studying this unit, you will be able to:

- Discuss Special Frenet Curves in E^n
- Describe Bertrand Curves in E^n
- Explain (1, 3)-Bertrand Curves in E^4

Introduction

We denote by E^3 a 3-dimensional Euclidean space. Let C be a regular C^∞- curve in E^3, that is, a C^∞-mapping $c : L \to E^3$ $(s \mapsto c(s))$. Here $L \subset R$ is some interval, and s ($\in$ L) is the arc-length parameter of C. Following Wong and Lai [7], we call a curve C a C^∞-special Frenet curve if there exist three C^∞-vector fields, that is, the unit tangent vector field t, the unit principal normal vector field n, the unit binormal vector field b, and two C^∞-scalar functions, that is, the curvature function k(> 0), the torsion function T($\neq$ 0). The three vector fields t, n and b satisfy the Frenet equations. A C^∞-special Frenet curve C is called a Bertrand curve if there exist another C^∞-special Frenet curve $\bar{C}$ and a C^∞-mapping $\varphi : C \to \bar{C}$ such that the principal normal line of C at c(s) is collinear to the principal normal vector n(s). It is a well-known result that a C^∞-special Frenet curve C in E^3 is a Bertrand curve if and only if its curvature function K and torsion function T satisfy the condition aK(s) + bT(s) = 1 for all s $\in$ L, where a and b are constant real numbers.

In an n-dimensional Euclidean space E^n, let C be a regular C^∞-curve, that is a C^∞-mapping $c : L \to E^n$ $(s \mapsto c(s))$, where s is the arc-length parameter of C. Then we can define a C^∞-special Frenet curve C. That is, we define $t(s) = c'(s)$, $n_1(s) = (1/||c''(s)||).\ c''(s)$, and we inductively define $n_k(s)$ (k = 2, 3, ..., n – 1) by the higher order derivatives of c (see next section, in detail). The n vector fields t, n_1, ..., n_{n-1} along C satisfy the Frenet equations with positive curvature functions k_1, ...,

Notes

k_{n-2} of C and positive or negative curvature function k_{n-1} of C. We call n_j the Frenet j-normal vector field along C, and the Frenet j-normal line of C at c(s) is a line generated by $n_j(s)$ through c(s) (j = 1, 2, ..., n – 1). The Frenet (j, k)-normal plane of C at c(s) is a plane spanned by $n_j(s)$ and $n_k(s)$ through c(s) (j, k = 1, 2, ... , n – 1; j ≠ k). A C^∞-special Frenet curve C is called a Bertrand curve if there exist another C^∞-special Frenet $\bar{C}$ and a C^∞-mapping $\varphi : C \to \bar{C}$ such that the Frenet 1-normal lines of C and $\bar{C}$ at corresponding points coincide. Then we obtain

Theorem A. If n ≥ 4, then no C^∞-special Frenet curve in E^n is a Bertrand curve.

This is claimed with different viewpoint, thus we prove the above Theorem.

We will show an idea of generalized Bertrand curve in E^4. A C^∞-special Frenet curve C in E^4 is called a (1, 3)-Bertrand curve if there exist another C^∞-special Frenet curve $\bar{C}$ and a C^∞-mapping $\varphi : C \to \bar{C}$ such that the Frenet (1, 3)-normal planes of C and $\bar{C}$ at corresponding points coincide. Then we obtain

Theorem B. Let C be a C^∞-special Frenet curve in E^4 with curvature functions k_1, k_2, k_3. Then C is a (1, 3)-Bertrand curve if and only if there exist constant real numbers α, β, γ, δ satisfying

$$\alpha k_2(s) - \beta k_3(s) \neq 0 \qquad \text{(a)}$$

$$\alpha k_1(s) + \gamma\{\alpha k_2(s) - \beta k_3(s)\} = 1 \qquad \text{(b)}$$

$$\gamma k_1(s) - k_2(s) = \delta k_3(s) \qquad \text{(c)}$$

$$(\gamma^2 - 1)k_1(s)k_2(s) + \gamma\{(k_1(s))^2 - (k_2(s))^2 - (k_3(s))^2\} \neq 0 \qquad \text{(d)}$$

for all s ∈ L.

This Theorem is proved in subsequent section.

We remark that if the Frenet j-normal vector fields of C and $\bar{C}$ are not vector fields of same meaning then we can not consider coincidence of the Frenet 1-normal lines or the Frenet (1, 3)-normal planes of C and $\bar{C}$. Then we consider only special Frenet curves.

Give an example of (1, 3)-Bertrand curve.

We shall work in C^∞-category.

9 .1 Special Frenet Curves in E^n

Let E^n be an n-dimensional Euclidean space with Cartesian coordinates $(x^1, x^2,..., x^n)$. By a parametrized curve C of class C^∞, we mean a mapping c of a certain interval I into E^n given by

$$c(t) = \begin{bmatrix} x^1(t) \\ x^2(t) \\ \vdots \\ x^n(t) \end{bmatrix} \qquad \forall t \in I.$$

If $\left\|\frac{dc(t)}{dt}\right\| = \left\langle \frac{dc(t)}{dt}, \frac{dc(t)}{dt} \right\rangle^{\frac{1}{2}} \neq 0$ for all t ∈ I, then C is called a regular curve in En. Here $\langle .,. \rangle$ denotes the Euclidean inner product on E^n.

A regular curve C is parametrized by the arc-length parameter s, that is, $c : L \to En$ ($L \ni s \mapsto c(s) \in En$)([1]). Then the tangent vector field $\frac{dc}{ds}$ along C has unit length, that is, $\left\|\frac{dc(s)}{ds}\right\| = 1$ for all $s \in L$.

Hereafter, curves considered are regular C^∞-curves in E^n parametrized by the arc-length parameter.

Let C be a curve in E^n, that is, $c(s) \in E^n$ for all $s \in L$. Let $t(s) = \frac{dc(s)}{ds}$ for all $s \in L$. The vector field t is called a unit tangent vector field along C, and we assume that the curve C satisfies the following conditions $(C_1) \sim (C_{n-1})$:

$$(C_1) : k_1(s) = \left\|\frac{dt(s)}{ds}\right\| = \left\|\frac{d^2c(s)}{ds^2}\right\| > 0 \quad \text{for all } s \in L.$$

Then we obtain a well-defined vector field n_1 along C, that is, for all $s \in L$,

$$n_1(s) = \frac{1}{k1(s)} \cdot \frac{dt(s)}{ds},$$

and we obtain,

$$\langle t(s), n_1(s)\rangle = 0, \qquad \langle n_1(s), n_1(s)\rangle = 1.$$

$$(C_2) : k_2(s) = \left\|\frac{dn_1(s)}{ds} + k_1(s).t(s)\right\| > 0 \qquad \text{for all } s \in L.$$

Then we obtain a well-defined vector field n_2 along C, that is, for all $s \in L$,

$$n_2(s) = \frac{1}{k_2(s)} \cdot \left(\frac{dn_1(s)}{ds} + k_1(s).t(s)\right),$$

and we obtain, for i, j = 1, 2,

$$\langle t(s), n_i(s)\rangle = 0, \qquad \langle n_i(s), n_j(s)\rangle = \delta_{ij},$$

where δ_{ij} denotes the Kronecker's symbol.

By an inductive procedure, for $\ell = 3, 4, ..., n - 2$,

$$(C_\ell) : k_\ell(s) = \left\|\frac{dn_{\ell-1}(s)}{ds} + k_{\ell-1}(s).n_{\ell-2}(s)\right\| > 0 \text{ for all } s \in L.$$

Then we obtain, for $\ell = 3, 4, ..., n - 2$, a well-defined vector field n_ℓ along C, that is, for all $s \in L$

$$n_\ell(s) = \frac{1}{k_\ell(s)} \cdot \left(\frac{dn_{\ell-1}(s)}{ds} + k_{\ell-1}(s).n_{\ell-2}(s)\right),$$

and for i, j = 1, 2, ..., n – 2

$$\langle t(s), n_i(s)\rangle = 0, \qquad \langle n_i(s), n_j(s)\rangle = \delta_{ij}.$$

And

$$(C_{n-1}) : k_{n-1}(s) = \left\langle \frac{dn_{n-2}(s)}{d_s}, n_{n-1}(s) \right\rangle \neq 0 \text{ for all } s \in L,$$

where the unit vector field n_{n-1} along C is determined by the fact that the frame $\{t, n_1, ..., n_{n-1}\}$ is of orthonormal and of positive orientation. We remark that the functions $k_1, ..., k_{n-2}$ are of positive and the function k_{n-1} is of non-zero. Such a curve C is called a special Frenet curve in E^n. The term "special" means that the vector field n_{i+1} is inductively defined by the vector fields n_i and n_{i-1} and the positive functions k_i and k_{i-1}. Each function k_i is called the i-curvature function of C (i = 1, 2, ... , n – 1). The orthonormal frame $\{t, n_1, ..., n_{n-1}\}$ along C is called the special Frenet frame along C.

Thus, we obtain the Frenet equations

$$\frac{dt(s)}{ds} = k_1(s) . n_1(s)$$

$$\frac{dn_1(s)}{ds} = -k_1(s) . t(s) + k_2(s) . n_2(s)$$

$$...$$

$$\frac{dn_\ell(s)}{ds} = -k_\ell(s) . n_{\ell-1} + k_{\ell+1}(s) . n_{\ell+1}(s)$$

$$...$$

$$\frac{dn_{n-2}(s)}{ds} = -k_{n-2}(s) . n_{n-3}(s) + k_{n-1}(s) . n_{n-1}(s)$$

$$\frac{dn_{n-1}(s)}{ds} = -k_{n-1}(s) . n_{n-2}(s)$$

for all $s \in L$. And, for j = 1, 2, ... , n – 1, the unit vector field n_j along C is called the Frenet j-normal vector field along C. A straight line is called the Frenet j-normal line of C at c(s) (j = 1, 2, ... n – 1 and $s \in L$), if it passes through the point c(s) and is collinear to the j-normal vector $n_j(s)$ of C at c(s).

Remark. In the case of Euclidean 3-space, the Frenet 1-normal vector fields n1 is already called the principal normal vector field along C, and the Frenet 1-normal line is already called the principal normal line of C at c(s).

For each point c(s) of C, a plane through the point c(s) is called the Frenet (j, k)-normal plane of C at c(s) if it is spanned by the two vectors $n_j(s)$ and $n_k(s)$ (j, k = 1, 2, ... , n – 1; j < k).

Remark. In the case of Euclidean 3-space, 1-curvature function k_1 is called the curvature of C, 2-curvature function k_2 is called the torsion of C, and (1, 2)-normal plane is already called the normal plane of C at c(s).

9 .2 Bertrand Curves in E^n

A C^∞-special Frenet curve C in E^n ($c : L \to E^n$) is called a Bertrand curve if there exist a C^∞-special Frenet curve $\bar{C}$ ($\bar{c} : \bar{L} \to E^n$), distinct from C, and a regular C^∞-map $\varphi : L \to \bar{L}$ ($\bar{s} = \varphi(s), \frac{d\varphi(s)}{ds} \neq 0$ for all $s \in L$) such that curves C and $\bar{C}$ have the same 1-normal line at each pair of corresponding points c(s) and $\bar{c}(\bar{s}) = \bar{c}(\varphi(s))$ under φ. Here s and $\bar{s}$ arc-length parameters of C and $\bar{C}$ respectively. In this case, $\bar{C}$ is called a Bertrand mate of C. The following results are well-known:

Theorem (the case of n = 2). Every C^∞-plane curve is a Bertrand curve.

Theorem (the case of n = 3). A C^∞-special Frenet curve in E^3 with 1-curvature function k_1 and 2-curvature function k_2 is a Bertrand curve if and only if there exists a linear relation

$$ak_1(s) + bk_2(s) = 1$$

for all s ∈ L, where a and b are nonzero constant real numbers.

The typical example of Bertrand curves in E^3 is a circular helix. A circular helix has infinitely many Bertrand mates.

We consider the case of n ≥ 4. Then we obtain Theorem A.

Proof of Theorem A. Let C be a Bertrand curve in En (n ≥ 4) and $\bar{C}$ a Bertrand mate of C. $\bar{C}$ is distinct from C. Let the pair of c(s) and $\bar{c}(\bar{s}) = \bar{c}(\varphi(s))$ be of corresponding points of C and $\bar{C}$. Then the curve $\bar{C}$ is given by

$$\bar{c}(\bar{s}) = \bar{c}(\varphi(s)) = c(s) + \alpha(s) \,.\, n_1(s) \tag{1}$$

where α is a C^∞-function on L. Differentiating (1) with respect to s, we obtain

$$j'(s) \,.\, \left.\frac{d\bar{c}(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)} = c'(s) + \alpha'(s) \,.\, n_1(s) + \alpha(s) \,.\, n_1'(s).$$

Here and hereafter, the prime denotes the derivative with respect to s. By the Frenet equations, it holds that

$$\varphi'(s) \,.\, \bar{t}(\varphi(s)) = (1 - \alpha(s)k_1(s)) \,.\, t(s) + \alpha'(s) \,.\, n_1(s) + \alpha(s)k_2(s) \,.\, n_2(s).$$

Since $\langle \bar{t}(\varphi(s)), \bar{n}_1(\varphi(s)) \rangle = 0$ and $\bar{n}_1(\varphi(s)) = \pm n_1(s)$, we obtain, for all s ∈ L,

$$\alpha'(s) = 0,$$

that is, α is a constant function on L with value a (we can use the same letter without confusion). Thus, (1) are rewritten as

$$\bar{c}(\bar{s}) = \bar{c}(\varphi(s)) = c(s) + \alpha \,.\, n_1(s), \tag{1$'$}$$

and we obtain

$$\varphi'(s) \,.\, \bar{t}(\varphi(s)) = (1 - \alpha k_1(s)) \,.\, t(s) + \alpha k_2(s) \,.\, n_2(s) \tag{2}$$

for all s ∈ L. By (2), we can set

$$\bar{t}(\varphi(s)) = (\cos\theta(s)) \,.\, t(s) + (\sin\theta(s)) \,.\, n_2(s), \tag{3}$$

where θ is a C^∞-function on L and

$$\cos\theta(s) = (1 - \alpha k_1(s))/\varphi'(s) \tag{4.1}$$

$$\sin\theta(s) = \alpha k_2(s)/\varphi'(s). \tag{4.2}$$

Differentiating (3) and using the Frenet equations, we obtain

$$\bar{k}_1(\varphi(s))\varphi'(s).\bar{n}_1(\varphi(s)) = \frac{d\cos\theta(s)}{ds}.\,t(s) + (k_1(s)\cos\theta(s) - k_2(s)\sin\theta(s).n_1(s) + \frac{d\sin q(s)}{ds}.n_2(s) + k_3(s)\sin\theta(s).n_3(s).$$

Since $\bar{n}_1(\varphi(s)) = \pm n_1(s)$ for all $s \in L$, we obtain

$$k_3(s)\sin q(s) \equiv 0. \tag{5}$$

By $k_3(s) \neq 0$ ($\forall s \in L$) and (5), we obtain that $\sin\theta(s) \equiv 0$. Thus, by k2(s) > 0($\forall s \in L$) and (4.2), we obtain that $\alpha = 0$. Therefore, (1)′ implies that $\bar{C}$ coincides with C. This is a contradiction. This completes the proof of Theorem A.

9 .3 (1, 3)-Bertrand Curves in E^4

By the results in the previous section, the notion of Bertrand curve stands only on E^2 and E^3. Thus, we will try to get the notion of generalization of Bertrand curve in E^n ($n \geq 4$).

Let C and $\bar{C}$ be C^∞-special Frenet curves in E^4 and $\varphi : L \to \bar{L}$ a regular C^∞-map such that each point c(s) of C corresponds to the point $\bar{c}(\bar{s}) = \bar{c}(\varphi(s))$ of $\bar{C}$ for $s \in L$. Here s and $\bar{s}$ arc-length parameters of C and $\bar{C}$ respectively. If the Frenet (1, 3)-normal plane at each point c(s) of C coincides with the Frenet (1, 3)-normal plane at each point c(s) of C coincides with the Frenet (1, 3)-normal plane at corresponding point $\bar{c}(\bar{s}) = \bar{c}(\varphi(s))$ of $\bar{C}$ for all $s \in L$, then C is called a (1, 3)-Bertrand curve in E^4 and $\bar{C}$ is called a (1, 3)-Bertrand mate of C. We obtain a characterization of (1, 3)-Bertrand curve, that is, we obtain Theorem B.

Proof of Theorem B. (i) We assume that C is a (1, 3)-Bertrand curve parametrized by arc-length s. The (1, 3)-Bertrand mate $\bar{C}$ is given by

$$\bar{c}(\bar{s}) = \bar{c}(\varphi(s)) = c(s) + \alpha(s)\,.\,n_1(s) + \beta(s)\,.\,n_3(s) \tag{1}$$

for all $s \in L$. Here α and β are C^∞-functions on L, and $\bar{s}$ is the arc-length parameter of $\bar{C}$. Differentiating (1) with respect to s, and using the Frenet equations, we obtain

$$\varphi'(s)\,.\,\bar{t}(\varphi(s)) = (1 - \alpha(s)k_1(s))\,.\,t(s) + \alpha'(s)\,.\,n_1(s) + (\alpha(s)k_2(s) - \beta(s)k_3(s))\,.\,n_2(s) + \beta'(s)\,.\,n_3(s)$$

for all $s \in L$.

Since the plane spanned by $n_1(s)$ and $n_3(s)$ coincides with the plane spanned by $\bar{n}_1(\varphi(s))$ and $\bar{n}_3(\varphi(s))$, we can put

$$\bar{n}_1(\varphi(s)) = (\cos\theta(s))\,.\,n_1(s) + (\sin\theta(s))\,.\,n_3(s) \tag{2.1}$$

$$\bar{n}_3(\varphi(s)) = (-\sin\theta(s))\,.\,n_1(s) + (\cos\theta(s))\,.\,n_3(s) \tag{2.2}$$

and we notice that $\sin\theta(s) \neq 0$ for all $s \in L$. By the following facts

$$0 = \left\langle \varphi'(s)\,.\,\bar{t}(\varphi(s)), \bar{n}_1(\varphi(s)) \right\rangle = \alpha's\,.\,(\cos\theta(s) + \beta'(s)\,.\,(\sin\theta(s))$$

$$0 = \left\langle \varphi'(s)\,.\,\bar{t}(\varphi(s)), \bar{n}_3(\varphi(s)) \right\rangle = -\alpha's\,.\,(\sin\theta(s) + \beta'(s)\,.\,(\cos\theta(s)),$$

we obtain

$$\alpha'(s) \equiv 0,\ \beta'(s) \equiv 0,$$

that is, α and β are constant functions on L with values a and b, respectively. Therefore, for all $s \in L$, (1) is rewritten as

$$\bar{c}(\bar{s}) = \bar{c}(\varphi(s)) = c(s) + \alpha \,.\, n_1(s) + \beta \,.\, n_3(s), \tag{1$'$}$$

and we obtain

$$\varphi'(s) \,.\, \bar{t}(\varphi(s)) = (1 - \alpha k_1(s)) \,.\, t(s) + (\alpha k_2(s) - \beta k_3(s)) \,.\, n_2(s). \tag{3}$$

Here, we notice that

$$(\varphi'(s))^2 = (1 - \alpha k_1(s))^2 + (\alpha k_2(s) - \beta k_3(s))^2 \neq 0 \tag{4}$$

for all $s \in L$. Thus, we can set

$$\bar{t}(\varphi(s)) = (\cos T(s)) \,.\, t(s) + (\sin T(s) \,.\, n_2(s) \tag{5}$$

$$\cos T(s) = (1 - \alpha k_1(s))/(\varphi'(s))$$

$$\sin T(s) = (\alpha k_2(s) - \beta k_3(s))/(\varphi'(s))$$

where T is a C^∞-function on L. Differentiating (5) with respect to s and using the Frenet equations, we obtain

$$\varphi'(s)\bar{k}_1(\varphi(s)) \,.\, n_1(\varphi(s)) = \frac{d\cos(T(s))}{ds}.t(s) + \{k_1(s)\cos(T(s)) - k_2(s)\sin(T(s))\}.n_1(s)$$

$$+\frac{d\sin(T(s))}{ds}.n_2(s) + k_3(s)\sin(T(s)).n_3(s).$$

Since $\bar{n}_1(\varphi(s))$ is expressed by linear combination of $n_1(s)$ and $n_3(s)$, it holds that

$$\frac{d\cos T(s)}{ds} \equiv 0, \quad \frac{d\sin T(s))}{ds} \equiv 0,$$

that is, T is a constant function on L with value T_0. Thus, we obtain

$$\bar{t}(\varphi(s)) = (\cos T_0) \,.\, t(s) + (\sin T_0) \,.\, n_2(s) \tag{5$'$}$$

$$\varphi'(s) \cos T_0 = 1 - \alpha k_1(s) \tag{6.1}$$

$$\varphi'(s) \sin T_0 = \alpha k_2(s) - \beta k_3(s) \tag{6.2}$$

for all $s \in L$. Therefore, we obtain

$$(1 - \alpha k_1(s)) \sin T_0 = (\alpha k_2(s) - \beta k_3(s)) \cos T_0 \tag{7}$$

for all $s \in L$.

If $\sin T_0 = 0$, then it holds $\cos T_0 = \pm 1$. Thus, (5)′ implies that $\bar{t}(\varphi(s)) = \pm t(s)$. Differentiating this equality, we obtain

$$\varphi'(s)\bar{k}_1(\varphi(s)) \,.\, \bar{n}_1(\varphi(s)) = \pm k_1(s) \,.\, n_1(s),$$

that is,

$$\bar{n}_1(\varphi(s)) = \pm n_1(s),$$

Notes

for all $s \in L$. By theorem A, this fact is a contradiction. This we must consider only the case of sin $T_0 \neq 0$. Then (6.2) implies

$$\alpha k_2(s) - \beta k_3(s) \neq 0 \ (s \in L),$$

that is, we obtain the relation (a).

The fact sin $T_0 \neq 0$ and (7) simply

$$\alpha k_1(s) + \{(\cos T_0)\ (\sin T_0)^{-1}\}\ (\alpha k_2(s) - \beta k_3(s)) = 1.$$

From this, we obtain

$$\alpha k_1(s) + \gamma(\alpha k_2(s) - \beta k_3(s) = 1$$

for all $s \in L$, where $\gamma = (\cos T_0)\ (\sin T_0)^{-1}$ is a constant number. Thus we obtain the relation (b).

Differentiating (5)′ with respect to s and using the Frenet equations, we obtain

$$\varphi'(s)\bar{k}_1(\varphi(s))\ .\ \bar{n}_1(\varphi(s)) = (k_1(s)\cos T_0 - k_2(s)\sin T_0).n_1(s) + k_3(s)\sin T_0.n_3(s)$$

for all $s \in L$. From the above equality, (6.1), (6.2) and (b), we obtain

$$\{\varphi'(s)\bar{k}_1(\varphi(s))\}^2$$

$$= \{k_1(s)\cos T_0 - k_2(s)\sin T_0\}^2 + \{k_3(s)\sin T_0\}^2$$

$$= (\alpha k_2(s) - \beta k_3(s))^2\,[(\gamma k_1(s) - k_2(s))^2 + (k_3(s))^2]\ (\varphi'(s))^{-2}.$$

for all $s \in L$. From (4) and (b), it holds

$$(\varphi'(s))^2 = (\gamma^2 + 1)\ (\alpha k^2(s) - \beta k_3(s))^2.$$

Thus we obtain

$$\{\varphi'(s)\bar{k}_1(\varphi(s))\}^2 = \frac{1}{\gamma^2+1}\{(\gamma k_1(s) - k_2(s) + k_3(s))^2\} \tag{8}$$

By (6.1), (6.2) and (b), we can set

$$\bar{n}_1(\varphi(s)) = (\cos\eta(s)).n_1(s) + (\sin\eta(s).n_3(s), \tag{9}$$

where

$$\cos\eta(s) = \frac{(\alpha k_2(s) - \beta k_3(s))(\gamma k_1(s) - k_2(s))}{\bar{k}_1(\varphi(s))(\varphi'(s))^2} \tag{10.1}$$

$$\sin\eta(s) = \frac{(\alpha k_2(s) - \beta k_3(s))k_3(s)}{\bar{k}_1(\varphi(s))(\varphi'(s))^2} \tag{10.2}$$

for all $s \in L$. Here, h is a C^∞-function on L.

Differentiating (9) with respect to s and using the Frenet equations, we obtain

$$-j'(s)\bar{k}_1(\varphi(s)).\bar{t}(j(s)) + j'(s)\bar{k}_2(j(s)).\bar{n}_2(\varphi(s))$$

$$= \frac{d\cos\eta(s)}{ds}.n_1(s) + \frac{d\sin\eta(s)}{ds}.n_3(s) - k_1(s)(\cos\eta(s))\ .\ t(s) + (k_2(s)(\cos\eta(s)) - k_3(s)(\sin\eta(s)).n_2(s)$$

for all s ∈ L. From the above fact, it holds

$$\frac{d\cos\eta(s)}{ds} \equiv 0, \quad \frac{d\sin\eta(s)}{ds} \equiv 0,$$

that is, η is a constant function on L with value η_0. Let $\delta = (\cos\eta_0)(\sin\eta_0)^{-1}$ be a constant number. Then (10.1) and (10.2) imply

$$\gamma k_1(s) - k_2(s) = \delta k_3(s) \quad (\forall s \in L),$$

that is, we obtain the relation (c).

Moreover, we obtain

$$-\varphi'(s)\bar{k}_1(\varphi(s)) \,.\, \bar{t}(\varphi(s)) + \varphi'(s)\bar{k}_2(\varphi(s)) \,.\, \bar{n}_2(\varphi(s))$$

$$= -k_1(s)(\cos\eta(s) \,.\, t(s) + \{k_2(s)(\cos\eta(s)) - k_3(s)(\sin\eta(s))\} \,.\, n_2(s)$$

By the above equality and (3), we obtain

$$-\varphi'(s)\bar{k}_2(\varphi(s)) \,.\, \bar{n}_2(\varphi(s)) = \varphi'(s)\bar{k}_1(\varphi(s)) \,.\, \bar{t}(\varphi(s))$$

$$-k_1(s)(\cos\eta_0) \,.\, t(s) + \{k_2(s)(\cos\eta_0) - k_3(s)(\sin\eta_0)\} \,.\, n_2(s)$$

$$= (\varphi'(s))^{-2}\{\bar{k}_1(\varphi(s))\}^{-1} \,.\, \{A(s) \,.\, t(s) + B(s) \,.\, n_2(s)\},$$

where

$$A(s) = \{\varphi'(s)\bar{k}_1(\varphi(s))\}^2(1 - \alpha k_1(s)) - k_1(s)(\alpha k_2(s) - \beta k_3(s))(\gamma k_1(s) - k_2(s))$$

$$B(s) = \{\varphi'(s)\bar{k}_1(\varphi(s))\}^2(\alpha k_2(s)) - \beta k_3(s)) + (\alpha k_2(s) - \beta k_3(s))(\gamma k_1(s) - k_2(s))k_2(s)$$

$$-(\alpha k_2(s) - \beta k_3(s))(k_3(s))^2$$

for all s ∈ L. From (b) and (8), A(s) and B(s) are rewritten as:

$$A(s) = -(\gamma^2 + 1)^{-1}(\alpha k_2(s) - \beta k_3(s)) \times [(\gamma^2 - 1)k_1(s)k_2(s) + \gamma\{(k_1(s))^2 - (k_2(s))^2 - (k_3(s))^2\}]$$

$$B(s) = \gamma(\gamma^2 + 1)^{-1}(\alpha k_2(s) - \beta k_3(s)) \times [(\gamma^2 - 1)k_1(s)k_2(s) + \gamma\{(k_1(s))^2 - (k_2(s))^2 - (k_3(s))^2\}].$$

Since $\varphi'(s)\bar{k}_2(\varphi(s)) \,.\, \bar{n}_2(\varphi(s)) \neq 0$ for all s ∈ L, it holds

$$(\gamma^2 - 1)k_1(s)k_2(s) + \gamma\{(k_1(s))^2 - (k_2(s))^2 - (k_3(s))^2\} \neq 0$$

for all s ∈ L. Thus, we obtain the relation (d).

(ii) We assume that C ($c: L \to E^4$) is a C^∞-special Frenet curve in E^4 with curvature functions k_1, k_2 and k_3 satisfying the relation (a), (b), (c) and (d) for constant numbers a, b, g and d. Then we define a C^∞-curve $\bar{C}$ by

$$\bar{c}(s) = c(s) + \alpha \,.\, n_1(s) + \beta \,.\, n_3(s) \tag{11}$$

for all s ∈ L, where s is the arc-length parameter of C. Differentiating (11) with respect to s and using the Frenet equations, we obtain

$$\frac{d\bar{c}(s)}{ds} = (1 - \alpha k_1(s)) \,.\, t(s) + (\alpha k_2(s) - \beta k_3(s)) \,.\, n_2(s)$$

Notes

for all $s \in L$. Thus, by the relation (b), we obtain

$$\frac{d\bar{c}(s)}{ds} = (\alpha k_2(s) - \beta k_3(s)) \,.\, (\gamma \,.\, t(s) + n_2(s)) \tag{12}$$

for all $s \in L$. Since the relation (a) holds, the curve $\bar{C}$ is a regular curve. Then there exists a regular map $\varphi : L \to \bar{L}$ defined by

$$\bar{s} = \varphi(s) = \int_0^s \left\| \frac{d\bar{c}(t)}{dt} \right\| dt \qquad (\forall s \in L),$$

where $\bar{s}$ denotes the arc-length parameter of $\bar{C}$, and we obtain

$$\varphi'(s) = \varepsilon\sqrt{\gamma^2 + 1}\,(\alpha k_2(s) - \beta k_3(s)) > 0, \tag{13}$$

where $\varepsilon = 1$ if $\alpha k_2(s) - \beta k_3(s) > 0$, and $\varepsilon = -1$ if $\alpha k_2(s) - \beta k_3(s) < 0$, for all $s \in L$. Thus the curve $\bar{C}$ is rewritten as

$$\begin{aligned} \bar{c}(\bar{s}) &= \bar{c}(\varphi(s)) \\ &= c(s) + \alpha \,.\, n_1(s) + \beta \,.\, n_3(s) \end{aligned}$$

for all $s \in L$. Differentiating the above equality with respect to s, we obtain

$$\varphi'(s) \,.\, \left.\frac{d\bar{c}(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)} = (\alpha k_2(s) - \beta k_3(s)) \,.\, \{\gamma \,.\, t(s) + n_2(s)\}. \tag{14}$$

We can define a unit vector field $\bar{t}$ along $\bar{C}$ by $\bar{t}(\bar{s}) = d\bar{c}(\bar{s})/d\bar{s}$ for all $\bar{s} \in \bar{L}$. By (13) and (14), we obtain

$$\bar{t}(\varphi(s)) = \varepsilon(\gamma^2 + 1)^{-1/2} \,.\, \{\gamma \,.\, t(s) + n_2(s)\} \tag{15}$$

for all $s \in L$. Differentiating (15) with respect to s and using the Frenet equations, we obtain

$$\varphi'(s) \,.\, \left.\frac{d\bar{t}(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)} = \varepsilon(\gamma^2 + 1)^{-1/2} \,.\, \{(\gamma k_1(s) - k_2(s)) \,.\, n_1(s) + k_3(s) \,.\, n_3(s)\}$$

and

$$\left\| \left.\frac{d\bar{t}(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)} \right\| = \frac{\sqrt{(\gamma k_1(s) - k_2(s))^2 + (k_3(s))^2}}{\varphi'(s)\sqrt{\gamma^2 + 1}}.$$

By the fact that $k_3(s) > 0$ for all $s \in L$, we obtain

$$\bar{k}_1(\varphi(s)) = \left\| \left.\frac{d\bar{t}(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)} \right\| > 0 \tag{16}$$

for all $s \in L$. Then we can define a unit vector field $\bar{n}_1$ along $\bar{C}$ by

$$\bar{n}_1(\bar{s}) = \bar{n}_1(\varphi(s))$$

Notes

$$= \frac{1}{\bar{k}_1(\varphi(s))} \cdot \frac{d\bar{t}(\bar{s})}{d\bar{s}}\bigg|_{\bar{s}=\varphi(s)}$$

$$= \frac{1}{\varepsilon\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2} \cdot \{(\gamma k_1(s)-k_2(s) \cdot n_1(s)+k_3(s) \cdot n_3(s)\}}$$

for all $s \in L$. Thus, we can put

$$\bar{n}_1(\varphi(s)) = (\cos\xi(s)) \cdot n_1(s) + (\sin\xi(s)) \cdot n_3(s), \tag{17}$$

where

$$\cos\xi(s) = \frac{\gamma k_1(s)-k_2(s)}{\varepsilon\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}} \tag{18.1}$$

$$\sin\xi(s) = \frac{k_3(s)}{\varepsilon\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}} > 0 \tag{18.2}$$

for all $s \in L$. Here, ξ is a C^∞-function on L. Differentiating (17) with respect to s and using the Frenet equations, we obtain

$$\varphi'(s) \cdot \frac{d\bar{n}_1(\bar{s})}{d\bar{s}}\bigg|_{\bar{s}=\varphi(s)} = -k_1(s)(\cos\xi(s)) \cdot t(s) + \frac{d\cos\xi(s)}{ds} \cdot n_1(s)$$

$$+ \{k_2(s)(\cos\xi(s)) - k_3(s)(\sin\xi(s))\} \cdot n_2(s) + \frac{d\sin\xi(s)}{ds} \cdot n_3(s)$$

Differentiating (c) with respect to s, we obtain

$$(gk_1'(s)-k_2'(s)k_3(s)-(\gamma k_1(s)-k_2(s))k_3'(s) \equiv 0. \tag{19}$$

Differentiating (18.1) and (18.2) with respect to s and using (4.19), we obtain

$$\frac{d\cos\xi(s)}{ds} \equiv 0, \qquad \frac{d\sin\xi(s)}{ds} \equiv 0,$$

that is, ξ is a constant function on L with value ξ_0. Thus, we obtain

$$\frac{\gamma k_1(s)-k_2(s)}{\varepsilon\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}} = \cos\xi_0, \tag{18.1$'$}$$

$$\frac{k_3(s)}{\varepsilon\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}} = \sin\xi_0 > 0. \tag{18.2$'$}$$

From (17), it holds

$$\bar{n}_1(\varphi(s)) = (\cos\xi_0) \cdot n_1(s) + (\sin\xi_0) \cdot n_3(s). \tag{20}$$

Thus we obtain, by (15) and (16),

$$\bar{k}_1(\varphi(s)) \cdot \bar{t}(\varphi(s)) = \frac{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}{\varepsilon\varphi'(s)(\gamma^2+1)\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}} \cdot (\gamma \cdot t(s)+n_2(s)),$$

Notes

and by (18.1)′, (18.2)′ and (20),

$$\left.\frac{d\bar{n}_1(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)} = \frac{-k_1(s)(\gamma k_1(s)-k_2(s))}{\varepsilon\varphi'(s)\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}}\,.\,t(s)+\frac{k_2(s)(\gamma k_1(s)-k_2(s)-(k_3(s))^2}{\varepsilon\varphi'(s)\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}}\,.\,n_2(s),$$

for all $s \in L$. By the above equalities, we obtain

$$\left.\frac{d\bar{n}_1(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)}+\bar{k}_1(\varphi(s))\,.\,\bar{t}(\varphi(s))=\frac{P(s)}{R(s)}\,.\,t(s)+\frac{Q(s)}{R(s)}\,.\,n_2(s),$$

where

$P(s) = -[\gamma\{(k_1(s))^2 - (k_2(s))^2 - (k_3(s))^2\} + (\gamma^2 - 1)k_1(s)k_2(s)]$

$Q(s) = \gamma[\gamma\{(k_1(s))^2 - (k_2(s))^2 - (k_3(s))^2\} + (\gamma^2 - 1)k_1(s)k_2(s)]$

$R(s) = \varepsilon(\gamma^2 + 1)\,\varphi'(s)\,\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2} \neq 0$

for all $s \in L$. We notice that, by (c), $P(s) \neq 0$ for all $s \in L$. Thus we obtain

$\bar{k}_2(\varphi(s))$

$$= \left\|\left.\frac{d\bar{n}_1(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)}+\bar{k}_1(\varphi(s)\,.\,\bar{t}(\varphi(s))\right\|$$

$$= \frac{\left|\gamma\{(k1(s))^2-(k_2(s))^2-(k_3(s))^2+(g2-1)k_1(s)k_2(s)\right|}{\varphi'(s)\sqrt{\gamma^2+1}\sqrt{\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}} > 0$$

for all $s \in L$. Thus, we can define a unit vector field $\bar{n}_2(\bar{s})$ along $\bar{C}$ by

$$\begin{aligned}\bar{n}_2(\bar{s}) &= \bar{n}_2(\varphi(s))\\ &= \frac{1}{\bar{k}_2(\varphi(s))}\,.\left(\left.\frac{d\bar{n}_1(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)}+\bar{k}_1(\varphi(s))\,.\,\bar{t}(\varphi(s))\right),\end{aligned}$$

that is,

$$\bar{n}_2(\varphi(s)) = \frac{1}{\varepsilon\sqrt{\gamma^2+1}}\,.\,(-t(s)+\gamma\,.\,n_2(s)) \tag{21}$$

for all $s \in L$. Next we can define a unit vector field $\bar{n}_3$ along $\bar{C}$ by

$$\begin{aligned}\bar{n}_3(\bar{s}) &= \bar{n}_3(j(s))\\ &= \frac{1}{\varepsilon\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}\,.\,\{-k_3(s)\,.\,n_1(s)+(\gamma k_1(s)-k_2(s))\,.\,n_3(s)\},}\end{aligned}$$

that is,

$$\bar{n}_3(\varphi(s)) = -(\sin\xi_0)\,.\,n_1(s) + (\cos\xi_0)\,.\,n_3(s) \tag{22}$$

Notes

for all $s \in L$. Now we obtain, by (15), (20), (21) and (22),

$$\det[\bar{t}(\varphi(s)), \bar{n}_1(\varphi(s), \bar{n}_2(\varphi(s)), \bar{n}_3(j(s))] = \det[t(s), n_1(s), n_2(s), n_3(s)] = 1$$

for all $s \in L$. And we obtain

$$\left\langle \bar{t}(\varphi(s), \bar{n}_i(\varphi(s)) \right\rangle = 0, \quad \left\langle \bar{n}_i(\varphi(s)), \bar{n}_j(\varphi(s)) \right\rangle = \delta_{ij}$$

for all $s \in L$ and i, j = 1, 2, 3. Thus the frame $\{\bar{t}, \bar{n}_1, \bar{n}_2, \bar{n}_3\}$ along $\bar{C}$ is of orthonormal and of positive. And we obtain

$$\bar{k}_3(\varphi(s)) = \left\langle \left.\frac{d\bar{n}_2(\bar{s})}{d\bar{s}}\right|_{\bar{s}=\varphi(s)}, \bar{n}_3(\varphi(s)) \right\rangle$$

$$= \frac{\sqrt{\gamma^2+1}\,k_1(s)k_3(s)}{\varphi'(s)\sqrt{(\gamma k_1(s)-k_2(s))^2+(k_3(s))^2}}$$

$$> 0$$

for all $s \in L$. Thus curve $\bar{C}$ is a C^∞-special Frenet curve in E^4. And it is trivial that the Frenet (1, 3)-normal plane at each point c(s) of C coincides with the Frenet (1, 3)-normal plane at corresponding point $\bar{c}(\bar{s}) = \bar{c}(\varphi(s))$ of $\bar{C}$. Therefore C is a (1, 3)-Bertrand curve in E^4.

Thus (i) and (ii) complete the proof of theorem B.

9 .4 An Example of (1, 3)-Bertrand Curve

Let a and b be positive numbers, and let r be an integer greater than 1. We consider a C^∞-curve C in E^4 defined by $c : L \to E^4$;

$$c(s) = \begin{bmatrix} a\cos\left(\frac{r}{\sqrt{r^2a^2+b^2}}s\right) \\ a\sin\left(\frac{r}{\sqrt{r^2a^2+b^2}}s\right) \\ b\cos\left(\frac{r}{\sqrt{r^2a^2+b^2}}s\right) \\ b\sin\left(\frac{r}{\sqrt{r^2a^2+b^2}}s\right) \end{bmatrix}$$

for all $s \in L$. The curve C is a regular curve and s is the arc-length parameter of C. Then C is a special Frenet curve in E^4 and its curvature functions are as follows:

$$k_1(s) = \frac{\sqrt{r^4a^2+b^2}}{r^2a^2+b^2},$$

$$k_2(s) = \frac{r(r^2-1)ab}{(r^2a^2+b^2)\sqrt{r^4a^2+b^2}},$$

$$k_3(s) = \frac{r}{\sqrt{r^4a^2 + b^2}}.$$

We take constant α, β, γ and δ defined by

$$\alpha = \frac{-(r^2aA + bB) + (r^2a^2 + b^2)}{\sqrt{r^4a^2 + b^2}},$$

$$\beta = \frac{-(r^2aB - bA) + (r^2 - 1)ab}{\sqrt{r^4a^2 + b^2}},$$

$$\gamma = \frac{r^2aA + bB}{r(aB - bA)},$$

$$\delta = \frac{r^4aA + bB}{r^2(aB - bA)}.$$

Here, A and B are positive numbers such that aB ≠ bA. Then it is trivial that (a), (b), (c) and (d) hold. Therefore, the curve C is a Bertrand curve in E^4, and its Bertrand mate curve $\bar{C}$ in E^4 $(\bar{c} : \bar{L} \to E^4)$ is given by

$$\bar{c}(\bar{s}) = \begin{bmatrix} A\cos\left(\dfrac{r}{\sqrt{r^2A^2 + B^2}}\bar{s}\right) \\ A\sin\left(\dfrac{r}{\sqrt{r^2A^2 + B^2}}\bar{s}\right) \\ B\cos\left(\dfrac{1}{\sqrt{r^2A^2 + B^2}}\bar{s}\right) \\ B\sin\left(\dfrac{r}{\sqrt{r^2A^2 + B^2}}\bar{s}\right) \end{bmatrix}$$

for all $\bar{s} \in \bar{L}$, where $\bar{s}$ is the arc-length parameter of $\bar{C}$, and a regular C^∞-map $\varphi : L \to \bar{L}$ is given by

$$\bar{s} = \varphi(s) = \frac{\sqrt{r^2A^2 + B^2}}{\sqrt{r^2a^2 + b^2}}s \qquad (\forall s \in L).$$

Remark: If $a^2 + b^2 = 1$, then the curve C in E^4 is a leaf of Hopf r-foliation on S^3 ([6], [8]).

9 .5 Summary

- **Theorem A.** If $n \geq 4$, then no C^∞-special Frenet curve in E^n is a Bertrand curve.
- Let E^n be an n-dimensional Euclidean space with Cartesian coordinates $(x^1, x^2, \ldots, x^n)$. By a parametrized curve C of class C^∞, we mean a mapping c of a certain interval I into E^n given by

$$c(t) = \begin{bmatrix} x^1(t) \\ x^2(t) \\ \vdots \\ x^n(t) \end{bmatrix} \qquad \forall t \in I.$$

If $\left\|\frac{dc(t)}{dt}\right\| = \left\langle \frac{dc(t)}{dt}, \frac{dc(t)}{dt} \right\rangle^{\frac{1}{2}} \neq 0$ for all $t \in I$, then C is called a regular curve in E^n. Here $\langle .,. \rangle$ denotes the Euclidean inner product on E^n. We refer to[2] for the details of curves in E^n.

- In the case of Euclidean 3-space, the Frenet 1-normal vector fields n1 is already called the principal normal vector field along C, and the Frenet 1-normal line is already called the principal normal line of C at c(s).
- A C^∞-special Frenet curve C in E^n $(c : L \to E^n)$ is called a Bertrand curve if there exist a C^∞-special Frenet curve $\bar{C}\ (\bar{c} : \bar{L} \to E^n)$, distinct from C, and a regular C^∞-map $\varphi : L \to \bar{L}$ $(\bar{s} = \varphi(s), \frac{d\varphi(s)}{ds} \neq 0$ for all $s \in L)$ such that curves C and $\bar{C}$ have the same 1-normal line at each pair of corresponding points c(s) and $\bar{c}(\bar{s}) = \bar{c}(\varphi(s))$ under φ. Here s and $\bar{s}$ arc-length parameters of C and $\bar{C}$ respectively. In this case, $\bar{C}$ is called a Bertrand mate of C. The following results are well-known:

  Theorem (the case of $n = 2$). Every C^∞-plane curve is a Bertrand curve.

  Theorem (the case of $n = 3$). A C^∞-special Frenet curve in E^3 with 1-curvature function k_1 and 2-curvature function k_2 is a Bertrand curve if and only if there exists a linear relation

$$ak_1(s) + bk_2(s) = 1$$

  for all $s \in L$, where a and b are nonzero constant real numbers.

- Let C and $\bar{C}$ be C^∞-special Frenet curves in E^4 and $\varphi : L \to \bar{L}$ a regular C^∞-map such that each point c(s) of C corresponds to the point $\bar{c}(\bar{s}) = \bar{c}(\varphi(s))$ of $\bar{C}$ for $s \in L$. Here s and $\bar{s}$ arc-length parameters of C and $\bar{C}$ respectively. If the Frenet (1, 3)-normal plane at each point c(s) of C coincides with the Frenet (1, 3)-normal plane at each point c(s) of C coincides with the Frenet (1, 3)-normal plane at corresponding point $\bar{c}(\bar{s}) = \bar{c}(\varphi(s))$ of $\bar{C}$ for all $s \in L$, then C is called a (1, 3)-Bertrand curve in E4 and $\bar{C}$ is called a (1, 3)-Bertrand mate of C. We obtain a characterization of (1, 3)-Bertrand curve, that is, we obtain Theorem B.

9 .6 Keyword

Bertrand curve: If $n \geq 4$, then no C^∞-special Frenet curve in E^n is a Bertrand curve.

9 .7 Self Assessment

1. If $n \geq 4$, then no C^∞-special Frenet curve in E^n is a ....
1. In the case of .... the Frenet 1-normal vector fields n1 is already called the principal normal vector field along C, and the Frenet 1-normal line is already called the principal normal line of C at c(s).
3. A C^∞-special Frenet curve in E^3 with 1-curvature function k_1 and 2-curvature function k_2 is a Bertrand curve if and only if there exists a linear relation .... for all $s \in L$, where a and b are nonzero constant real numbers.
4. Let C be a Bertrand curve in En ($n \geq 4$) and $\bar{C}$ a .... of C. $\bar{C}$ is distinct from C.

9 .8 Review Questions

1. Discuss Special Frenet Curves in E^n
2. Describe Bertrand Curves in E^n
3. Explain (1, 3)-Bertrand Curves in E^4

Answers: Self Assessment

1. Bertrand curve.
2. Euclidean 3-space,
3. $ak_1(s) + bk_2(s) = 1$
4. Bertrand mate

2 .9 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 10: Developable Surface Fitting to Point Clouds

CONTENTS

Objectives

After studying this unit, you will be able to:

- Discuss The Blaschke Model of Oriented Planes in R^3
- Explain Incidence of Point and Plane
- Define Tangency of Sphere and Plane

- Discuss the Classification of Developable Surfaces according to their Image on B
- Describe Cones and Cylinders of Revolution
- Explain Recognition of Developable Surfaces from Point Clouds
- Describe Reconstruction of Developable Surfaces from Measurements

Introduction

Given a cloud of data points p_i in $\mathbb{R}^3$, we want to decide whether p_i are measurements of a cylinder or cone of revolution, a general cylinder or cone or a general developable surface. In case, where this is true we will approximate the given data points by one of the mentioned shapes. In the following, we denote all these shapes by developable surfaces. To implement this we use a concept of classical geometry to represent a developable surface not as a two-parameter set of points but as a one-parameter set of tangent planes and show how this interpretation applies to the recognition and reconstruction of developable shapes.

Points and vectors in $\mathbb{R}^3$ or $\mathbb{R}^4$ are denoted by boldface letters, p, v. Planes and lines are displayed as italic capital letters, T,L. We use Cartesian coordinates in $\mathbb{R}^3$ with axes x, y and z. In $\mathbb{R}^4$, the axes of the Cartesian coordinate system are denoted by $u_1, . . . , u_4$.

Developable surfaces shall briefly be introduced as special cases of ruled surfaces. A ruled surface R carries a one parameter family of straight lines L. These lines are called generators or generating lines. The general parametrization of a ruled surface R is

$$x(u, v) = c(u) + ve(u), \tag{1}$$

where c(u) is called directrix curve and e(u) is a vector field along c(u). For fixed values u, this parametrization represents the straight lines L(u) on R.

The normal vector n(u, v) of the ruled surface x(u, v) is computed as cross product of the partial derivative vectors x_u and x_v, and we obtain

$$n(u, v) = \dot{c}(u) \times e(u) + v\dot{e}(u) \times e(u). \tag{2}$$

For fixed $u = u_0$, the normal vectors $n(u_0, v)$ along $L(u_0)$ are linear combinations of the vectors $\dot{c}(u_0) \times e(u_0)$ and $\dot{e}(u_0) \times e(u_0)$. The parametrization x(u, v) represents a developable surface D if for each generator L all points $x \in L$ have the same tangent plane (with exception of the singular point on L). This implies that the vectors $\dot{c} \times e$ and $\dot{e} \times e$ are linearly dependent which is expressed equivalently by the following condition

$$\det(\dot{c}, e, \dot{e}) = 0. \tag{3}$$

Any regular generator L(u) of a developable surface D carries a unique singular point s(u) which does not possess a tangent plane in the above defined sense, and $s(u) = x(u, v_s)$ is determined by the parameter value

$$v_s = -\frac{(\dot{c} \times e) \cdot (\dot{e} \times e)}{(e \times e)^2}. \tag{4}$$

If e and $\dot{e}$ are linearly dependent, the singular point s is at infinity, otherwise it is a proper point. In Euclidean space $\mathbb{R}^3$, there exist three different basic classes of developable surfaces:

(1) ***Cylinder:*** the singular curve degenerates to a single point at infinity.

(2) ***Cone:*** the singular curve degenerates to a single proper point, which is called vertex.

(3) Surface consisting of the tangent lines of a regular space curve s(u), which is the singular curve of the surface.

In all three cases, the surface D can be generated as envelope of its one parameter family of tangent planes. This is called the dual representation of D. A cylinder of revolution is obtained by rotating a plane around an axis which is parallel to this plane. A cone of revolution is obtained by rotating a plane around a general axis, but which is not perpendicular to this plane. Further, it is known that smooth developable surfaces can be characterized by vanishing Gaussian curvature. In applications surfaces appear which are composed of these three basic types.

There is quite a lot of literature on modeling with developable surfaces and their references. B-spline representations and the dual representation are well-known. The dual representation has been used for interpolation and approximation of tangent planes and generating lines. Pottmann and Wallner study approximation of tangent planes, generating lines and points. The treatment of the singular points of the surface is included in the approximation with relatively little costs. To implement all these tasks, a local coordinate system is used for the representation of developable surfaces such that their tangent planes T(t) are given by T(t) : $e_4(t) + e_1(t)x + ty - z = 0$. This concept can be used for surface fitting too, but the representation is a bit restrictive.

We note a few problems occurring in surface fitting with developable B-spline surfaces. In general, for fitting a B-spline surface

$$b(u, v) = \sum N_i(u)N_j(v)b_{ij}$$

with control points bij to a set of unorganized data points p_k, one estimates parameter values (u_i, v_j) corresponding to p_k. The resulting approximation leads to a linear problem in the unknown control points b_{ij}. For surface fitting with ruled surfaces we might choose the degrees n and 1 for the B-spline functions $N_i(u)$ and $N_j(v)$ over a suitable knot sequence. There occur two main problems in approximating data points by a developable B-spline surface:

- For fitting ruled surfaces to point clouds, we have to estimate in advance the approximate direction of the generating lines of the surface in order to estimate useful parameter values for the given data. To perform this, it is necessary to estimate the asymptotic lines of the surface in a stable way.
- We have to guarantee that the resulting approximation b(u, v) is developable, which is expressed by equation (3). Plugging the parametrization b(u, v) into this condition leads to a highly non-linear side condition in the control points b_{ij} for the determination of the approximation b(u, v).

10.1 Contribution of the Article

To avoid above mentioned problems, we follow another strategy. The reconstruction of a developable surface from scattered data points is implemented as reconstruction of a one-parameter family of planes which lie close to the estimated tangent planes of the given data points. Carrying out this concept, we can automatically guarantee that the approximation is developable. This concept avoids the estimation of parameter values and the estimation of the asymptotic curves. The reconstruction is performed by solving curve approximation techniques in the space of planes.

The proposed algorithm can also be applied to approximate nearly developable surfaces (or better slightly distorted developable surfaces) by developable surfaces. The test implementation has been performed in Matlab and the data has been generated by scanning models of developable surfaces with an optical laser scanner. Some examples use data generated by simulating a scan of mathematical models.

Notes

The basic properties concerning the Blaschke image (Blaschke model) of the set of planes in R3 which is relevant for the implementation of the intended reconstruction. Section 23.3 tells about a classification, and Section 23.4 discusses the recognition of developable surfaces in point clouds using the Blaschke image of the set of estimated tangent planes of the point set. Section 23.5 describes the concept of reconstruction of these surfaces from measurements. Finally, we present some examples and discuss problems of this approach and possible solutions.

10.2 The Blaschke Model of Oriented Planes in R^3

Describing points x by their Cartesian coordinate vectors x = (x, y, z), an oriented plane E in Euclidean space $\mathbb{R}^3$ can be written in the Hesse normal form,

$$E : n_1x + n_2y + n_3z + d = 0,\ n_1^2 + n_2^2 + n_3^2 = 1. \tag{5}$$

We note that $n_1x + n_2y + n_3z + d = \text{dist}(x, E)$ is the signed distance between the point x and the plane E. In particular, d is the origin's distance to E. The vector $n = (n_1, n_2, n_3)$ is the unit normal vector of E. The vector n and the distance d uniquely define the oriented plane E and we also use the notation $E : n . x + d = 0$.

The interpretation of the vector (n_1, n_2, n_3, d) as point coordinates in $\mathbb{R}^4$, defines the Blaschke mapping

$$b : E \mapsto b(E) = (n_1, n_2, n_3, d) = (n, d). \tag{6}$$

In order to carefully distinguish between the original space $\mathbb{R}^3$ and the image space $\mathbb{R}^4$, we denote Cartesian coordinates in the image space $\mathbb{R}^4$ by (u_1, u_2, u_3, u_4). According to the normalization $n^2 = 1$ and (6), the set of all oriented planes of $\mathbb{R}^3$ is mapped to the entire point set of the so-called Blaschke cylinder,

$$B :\ u_1^2 + u_2^2 + u_3^2 = 1. \tag{7}$$

Thus, the set of planes in $\mathbb{R}^3$ has the structure of a three-dimensional cylinder, whose cross sections with planes u_4 = const. are copies of the unit sphere S^2 (Gaussian sphere). Any point $U \in B$ is image point of an oriented plane in $\mathbb{R}^3$. Obviously, the Blaschke image b(E) = (n, d) is nothing else than the graph of the support function d (distance to the origin) over the Gaussian image point n.

Let us consider a pencil (one-parameter family) of parallel oriented planes $E(t) : n . x + t = 0$. The Blaschke mapping (6) implies that the image points b(E(t)) = (n, t) lie on a generating line of B which is parallel to the u_4-axis.

10.2.1 Incidence of Point and Plane

We consider a fixed point $p = (p_1, p_2, p_3)$ and all planes $E : n . x + d = 0$ passing through this point. The incidence between p and E is expressed by

$$p_1n_1 + p_2n_2 + p_3n_3 + d = p . n + d = 0, \tag{8}$$

and therefore, the image points $b(E) = (n_1, n_2, n_3, d)$ in $\mathbb{R}^4$ of all planes passing through p lie in the three-space

$$H : p_1u_1 + p_2u_2 + p_3u_3 + u_4 = 0, \tag{9}$$

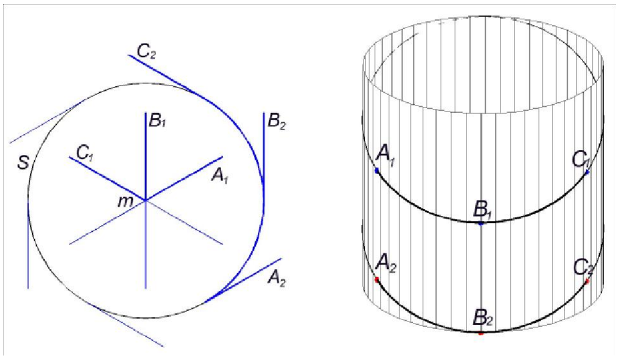
passing through the origin of $\mathbb{R}^4$. The intersection $H \cap B$ with the cylinder B is an ellipsoid and any point of $H \cap B$ is image of a plane passing through p. Fig. 23.1 shows a 2D illustration of this property.

10.2.2 Tangency of Sphere and Plane

Let S be the oriented sphere with center m and signed radius r, S : $(x - m)^2 - r^2 = 0$. The tangent planes T_S of S are exactly those planes, whose signed distance from m equals r. Therefore, they satisfy

$$T_S : n_1m_1 + n_2m_2 + n_3m_3 + d = n \cdot m + d = r. \quad (10)$$

Figure 10.1: Blaschke Images of a Pencil of Lines and of Lines Tangent to an or. Circle



Their Blaschke image points b(TS) thus lie in the three-space

$$H : m_1u_1 + m_2u_2 + m_3u_3 + u_4 - r = 0, \quad (11)$$

and $b(T_S)$ are the points of the intersection $H \cap B$, which is again an ellipsoid.

This also follows from the fact that S is the offset surface of m at signed distance r. The offset operation, which maps a surface $F \subset \mathbb{R}^3$ (as set of tangent planes) to its offset F_r at distance r, appears in the Blaschke image B as translation by the vector (0, 0, 0, r), see Fig. 10.1.

Conversely, if points $q = (q_1, q_2, q_3, q_4) \in B$ satisfy a linear relation

$$H : a_0 + u_1a_1 + u_2a_2 + u_3a_3 + u_4a_4 = 0,$$

q = b(T) are Blaschke images of planes T which are tangent to a sphere in case $a_4 \neq 0$. Center and radius are determined by

$$m = \frac{1}{a_4}(a_1, a_2, a_3), \qquad r = \frac{-a_0}{a_4}.$$

If $a_0 = 0$, the planes b(T) pass through the fixed point m. If $a_4 = 0$, the planes T form a constant angle with the direction vector $a = (a_1, a_2, a_3)$ because of $a \cdot n = -a_0$, with $n = (u_1, u_2, u_3)$.

Here it would lead to far to explain more about Laguerre geometry, the geometry of oriented planes and spheres in $\mathbb{R}^3$.

10.2.3 The Tangent Planes of a Developable Surface

Let T(u) be a one-parameter family of planes

$$T(u) : n_4(u) + n_1(u)x + n_2(u)y + n_3(u)z = 0$$

with arbitrary functions ni, i = 1, . . . , 4. The vector $n(u) = (n_1, n_2, n_3)(u)$ is a normal vector of T(u). Excluding degenerate cases, the envelope of T(u) is a developable surface D, whose generating lines L(u) are

$$L(u) = T(u) \cap \dot{T}(u),$$

Notes

where $\dot{T}(u)$ denotes the derivative with respect to u. The generating lines themselves envelope the singular curve s(u) which is the intersection

$$s(u) = T(u) \cap \dot{T}(u) \cap \ddot{T}(u).$$

Taking the normalization $n_1^2 + n_2^2 + n_3^2 = n(u)^2 = 1$ into account, the Blaschke image b(T(u)) = b(D) of the developable surface D is a curve on the Blaschke cylinder B. This property will be applied later to fitting developable surfaces to point clouds.

10.3 The Classification of Developable Surfaces according to their Image on B

This section will characterize cylinders, cones and other special developable surfaces D by studying their Blaschke images b(D).

Cylinder: D is a general cylinder if all its tangent planes T(u) are parallel to a vector a and thus its normal vectors n(u) satisfy n . a = 0. This implies that the image curve b(T(u)) = b(D) is contained in the three-space

$$H : a_1u_1 + a_2u_2 + a_3u_3 = 0. \tag{1}$$

Cone: D is a general cone if all its tangent planes T(u) pass through a fixed point p = (p_1, p_2, p_3). This incidence is expressed by $p_1n_1 + p_2n_2 + p_3n_3 + n_4 = 0$. Thus, the Blaschke image curve b(T(u)) = b(D) is contained in the three space

$$H : p_1u_1 + p_2u_2 + p_3u_3 + u_4 = 0. \tag{2}$$

There exist other special types of developable surfaces. Two of them will be mentioned here.

The surface D is a developable of constant slope, if its normal vectors n(u) form a constant angle ϕ with a fixed direction vector a. Assuming $\|a\| = 1$, we get $\cos(\phi) = a \, . \, n(u) = \gamma$ = const. This implies that the Blaschke images of the tangent planes of D are contained in the three-space

$$H : -\gamma + a_1u_1 + a_2u_2 + a_3u_3 = 0. \tag{3}$$

The developable surface D is tangent to a sphere with center m and radius r, if the tangent planes T(u) of D satisfy $n_4 + n_1m_1 + n_2m_2 + n_3m_3 - r = 0$, according to (11). Thus, the image curve b(D) is contained in the three-space

$$H : -r + u_1m_1 + u_2m_2 + u_3m_3 + u_4 = 0. \tag{4}$$

10.3.1 Cones and Cylinders of Revolution

For applications, it is of particular interest if a developable surface D is a cone or cylinder of revolution.

Let D be a *cylinder of revolution* with axis A and radius r. The tangent planes T of D are tangent to all spheres of radius r, whose centers vary on A. Let

$$S_1 : (x - p)^2 - r^2 = 0, S_2 : (x - q)^2 - r^2 = 0$$

be two such spheres with centers p, q. According to (11), the images b(T) of the tangent planes T satisfy the relations

$$H_1 : -r + u_1p_1 + u_2p_2 + u_3p_3 + u_4 = 0, \tag{5}$$

$$H_2 : -r + u_1q_1 + u_2q_2 + u_3q_3 + u_4 = 0.$$

Since $p \neq q$, the image curve b(D) lies in the plane $P = H_1 \cap H_2$ and b(D) is a conic.

Cones of revolution D can be obtained as envelopes of the common tangent planes of two oriented spheres S_1, S_2 with different radii $r \neq s$. Thus, b(D) is a conic contained in the plane $P = H_1 \cap H_2$ which is defined by

$$H_1 : -r + u_1p_1 + u_2p_2 + u_3p_3 + u_4 = 0, \tag{6}$$

$$H_2 : -s + u_1q_1 + u_2q_2 + u_3q_3 + u_4 = 0.$$

Conversely, if the Blaschke image b(D) of a developable surface is a planar curve $\subset$ P, how can we decide whether D is a cone or cylinder of revolution?

Let b(D) = b(T(u)) be a planar curve $\subset$ P and let P be given as intersection of two independent three-spaces H_1, H_2, with

$$H_i : h_{i0} + h_{i1}u_1 + h_{i2}u_2 + h_{i3}u_3 + h_{i4}u_4 = 0. \tag{7}$$

Using the results of Section 23.2.2, the incidence relation $b(T(u)) \subset H_1$ implies that T(u) is tangent to a sphere, or is passing through a point ($h_{10} = 0$), or encloses a fixed angle with a fixed direction ($h_{14} = 0$). The same argumentation holds for H_2.

Thus, by excluding the degenerate case $h_{14} = h_{24} = 0$, we can assume that $P = H_1 \cap H_2$ is the intersection by two three-spaces H_1, H_2 of the form (5) or (6).

(1) Let the plane $P = H_1 \cap H_2$ be given by equations (5). Then, the developable surface D is a cylinder of revolution. By subtracting the equations (5) it follows that the normal vector n(u) of T(u) satisfies

$$n \,.\, (p - q) = 0.$$

Thus, the axis A of D is given by $a = p - q$ and D's radius equals r.

(2) Let the plane $P = H_1 \cap H_2$ be given by equations (6). The pencil of three spaces $\lambda H_1 + \mu H_2$ contains a unique three-space H, passing through the origin in $\mathbb{R}^4$, whose equation is

$$H : \sum_{i=1}^{3} u_i(sp_i - rq_i) + u_4(s - r) = 0.$$

Thus, the tangent planes of the developable surface D are passing through a fixed point corresponding to H, and D is a cone of revolution. Its vertex v and the inclination angle ϕ between the axis A : $a = p - q$ and the tangent planes T(u) are

$$V = \frac{1}{s-r}(sp - rq), \text{ and } \sin\phi = \frac{s-r}{\|q-p\|}$$

10.4 Recognition of Developable Surfaces from Point Clouds

Given a cloud of data points p_i, this section discusses the recognition and classification of developable surfaces according to their Blaschke images. The algorithm contains the following steps:

(1) Estimation of tangent planes T_i at data points p_i and computation of the image points $b(T_i)$.

(2) Analysis of the structure of the set of image points $b(T_i)$.

(3) If the set $b(T_i)$ is curve-like, classification of the developable surface which is close to p_i.

10.4.1 Estimation of Tangent Planes

Notes

We are given data points p_i, i = 1, . . . , N, with Cartesian coordinates x_i, y_i, z_i in $\mathbb{R}^3$ and a triangulation of the data with triangles t_j. The triangulation gives topological information about the point cloud, and we are able to define adjacent points q_k for any data point p.

The estimated tangent plane T at p shall be a plane best fitting the data points q_k. T can be computed as minimizer (in the l_1 or l_2-sense) of the vector of distances dist(q_k, T) between the data points q_k and the plane T. This leads to a set of N estimated tangent planes T_i corresponding to the data points p_i. For more information concerning reverse engineering.

Assuming that the original surface with measurement points p_i is a developable surface D, the image points b(T_i) of the estimated tangent planes T_i will form a curve-like region on B. To check the property 'curvelike', neighborhoods with respect to a metric on B will be defined. Later, we will fit a curve c(t) to the curve-like set of image points b(T_i), and this fitting is implemented according to the chosen metric.

10.4.2 A Euclidean Metric in the Set of Planes

Now we show that the simplest choice, namely the canonical Euclidean metric in the surrounding space R^4 of the Blaschke cylinder B, is a quite useful metric for data analysis and fitting. This says that the distance dist(E, F) between two planes E, F

$$E : e_1x_1 + e_2x_2 + e_3x_3 + e_4 = 0,\ F : f_1x_1 + f_2x_2 + f_3x_3 + f_4 = 0,$$

with normalized normal vectors e = (e_1, e_2, e_3) and f = (f_1, f_2, f_3) $(\|e\| = \|f\| = 1)$ is defined to be the Euclidean distance of their image points b(E) and b(F). Thus, the squared distance between E and F is defined by

$$\text{dist}(E, F)^2 = (e_1 - f_1)^2 + (e_2 - f_2)^2 + (e_3 - f_3)^2 + (e_4 - f_4)^2. \quad (1)$$

To illustrate the geometric meaning of dist$(E, F)^2$ between two planes E and F we choose a fixed plane M(= F) in $\mathbb{R}^3$ as x – m = 0. Its Blaschke image is b(M) = (1, 0, 0, –m). All points of the Blaschke cylinder, whose Euclidean distance to b(M) equals r, form the intersection surface S of B with the three-dimensional sphere $(u_1 - 1)^2 + u_2^2 + u_3^2 + (u_4 + m)^2 = r^2$. Thus, S is an algebraic surface of order 4 in general. Its points are Blaschke images b(E) of planes E in $\mathbb{R}^3$ which have constant distance r from M and their coordinates ei satisfy

$$(e_1 - 1)^2 + e_2^2 + e_3^2 + (e_4 + m)^2 - r^2 = 0. \quad (2)$$

The coefficients e_i satisfy the normalization $e_1^2 + e_2^2 + e_3^2 = 1$. If we consider a general homogeneous equation E : $w_1x_1 + w_2x_2 + w_3x_3 + w_4 = 0$ of E, these coefficients wi are related to e_i by

$$e_i = \frac{w_i}{\sqrt{w_1^2 + w_2^2 + w_3^2}},\ i = 1, 2, 3, 4.$$

We plug this into (2) and obtain the following homogeneous relation of degree four in plane coordinates w_i,

$$[(2 - r^2 + m^2)(w_1^2 + w_2^2 + w_3^2) + w_4^2]^2 = 4(w_1 - mw_4)^2(w_1^2 + w_2^2 + w_3^2). \quad (3)$$

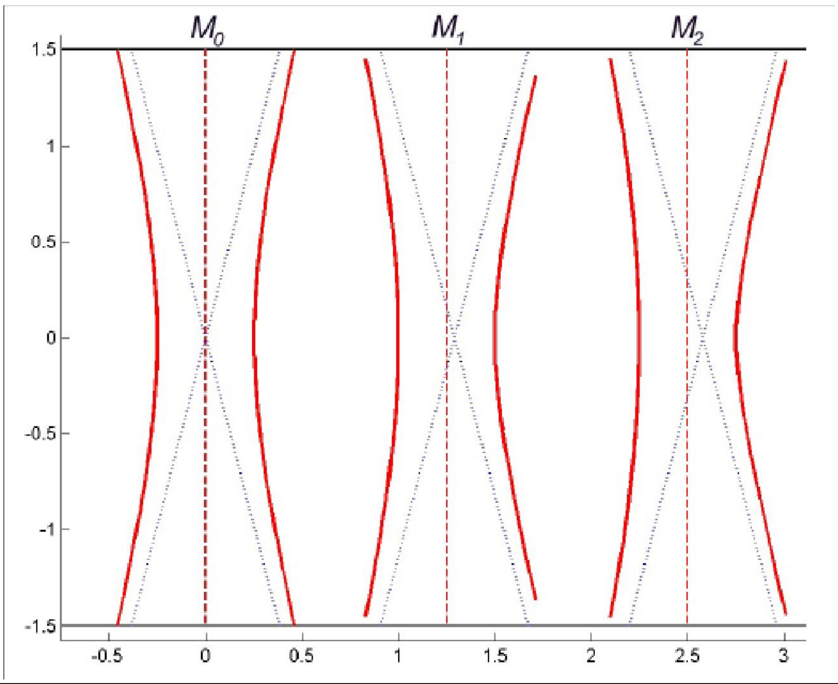
Hence, all planes E, having constant distance dist(E, M) = r from a fixed plane M, form the tangent planes of an algebraic surface $b^{-1}(S)$ = U of class 4, and U bounds the tolerance region of the plane M. If a plane E deviates from a plane M in the sense, that b(E) and b(M) have at most distance r, then the plane E lies in a region of $\mathbb{R}^3$, which is bounded by the surface U (3).

For visualization, we choose the 2D-case. Figure 10.2 shows the boundary curves of tolerance regions of lines M : x = m, for values m = 0, 1.25, 2.5 and radius r = 0.25. The lines M_i are drawn dashed. The largest perpendicular distance of E(| | M) and M within the tolerance regions is r. The largest angle of E and M is indicated by the asymptotic lines (dotted style) of the boundary curves. For m = 0, the intersection point of the asymptotic lines lies on M_0, but for increasing values of |m|, this does not hold in general and the tolerance regions will become asymmetrically. For large values of |m|, this intersection point might even be outside the region, and the canonical Euclidean metric in $\mathbb{R}^4$ is then no longer useful for the definition of distances between planes.

The tolerance zone of an oriented plane M is rotationally symmetric with respect to the normal n of M passing through the origin. In the planes through *n* there appears the 2D-case, so that the 2D-case is sufficient for visualization.

The introduced metric is not invariant under all Euclidean motions of the space $\mathbb{R}^3$. The metric is invariant with respect to rotations about the origin, but this does not hold for translations. If the distance $d = m$ of the plane

Figure 10.2: Boundary Curves of the Tolerance Regions of the Center Lines M_i.



M to the origin changes, then the shape of the tolerance region changes, too. However, within an area of interest around the origin (e.g. |m| < 1), these changes are small and thus the introduced metric is useful.

In practice, we uniformly scale the data in a way that the absolute values of all coordinates x_i, y_i, z_i are smaller than $c = 1/\sqrt{3}$. Then the object is contained in a cube, bounded by the planes x = ±c, y = ±c, z = ±c and the maximum distance of a data point p_i to the origin is 1. Considering planes passing through the data points p_i, the maximum distance dist(O, E) of a plane E to the origin is also 1.

Notes

According to the normalization $e_1^2 + e_2^2 + e_3^2 = 1,$ the distance of the Blaschke image b(E) to the origin in $\mathbb{R}^4$ is bounded by 1. This is also important for a discretization of the Blaschke cylinder which we discuss in the following.

10.4.3 A Cell Decomposition of the Blaschke Cylinder

For practical computations on B, we use a cell decomposition of B to define neighborhoods of image points b(T) of (estimated) tangent planes T. We recall that B's equation is $u_1^2 + u_2^2 + u_3^2 = 1.$ Any cross section with a plane u_4 = const. is a copy of the unit sphere S^2 in $\mathbb{R}^3$. In order to obtain a cell decomposition of B, we start with a triangular decomposition of S^2 and lift it to B.

A tessellation of S^2 can be based on the net of a regular icosahedron. The vertices v_i, i = 1, . . . , 12, with $\|v_i\| = 1$ of a regular icosahedron form twenty triangles t_j and thirty edges. All edges have same arc length. This icosahedral net is subdivided by computing the midpoints of all edges (geodesic circles). Any triangle t_j is subdivided into four new triangles. The inner triangle has equal edge lengths, the outer three have not, but the lengths of the edges to not vary too much. By repeated subdivision, one obtains a finer tessellation of the unit sphere.

The cell decomposition of the Blaschke cylinder consists of triangular prismatic cells which are lifted from the triangular tessellation of S^2 in u_4-direction. Since we measure distances according to (1), the height of a prismatic cell has to be approximately equal to the edge length of a triangle. When each triangle of the tessellation is subdivided into four children, each interval in u_4-direction is split into two subintervals.

According to the scaling of the data points p_i, the coordinates of the image points b(E) on B are bounded by ±1. We start with 20 triangles, 12 vertices and 2 intervals in u_4-direction. The test-implementation uses the resolution after three subdivision steps with 1280 triangles, 642 vertices and 16 intervals in u_4-direction. In addition to the cell structure on B, we store adjacency information of these cells.

Remark concerning the visualization: It is easy to visualize the spherical image (first three coordinates) on S^2, but it is hard to visualize the Blaschke image on B. We confine ourself to plot the spherical image on S^2, and if necessary, we add the fourth coordinate (support function) in a separate figure. This seems to be an appropriate visualization of the geometry on the Blaschke cylinder, see Figures 10.3 to Figure 10.7.

10.4.4 Analysis and Classification of the Blaschke Image

Having computed estimates T_i of the tangent planes of the data points and their images $b(T_i)$, we check whether the Blaschke image of the considered surface is curve-like. According to Section 23.4.3, the interesting part of the Blaschke cylinder B is covered by 1280 × 16 cells C_k. We compute the memberships of image points $b(T_i)$ and cells C_k and obtain a binary image on the cell structure C of B. Let us recall some basic properties of the Blaschke image of a surface.

(1) If the data points p_i are contained in a single plane P, the image points $b(T_i)$ of estimated tangent planes T_i form a point-like cluster around b(P) on B.

(2) If the data points p_i are contained in a developable surface, the image points $b(T_i)$ form a curve-like region in B.

(3) If the data points p_i are contained in a doubly curved surface S, the image points $b(T_i)$ cover a two-dimensional region on B.

(4) If the data points p_i are contained in a spherical surface S, the image points $b(T_i)$ cover a two-dimensional region on B which is contained in a three-space.

In the following, we assume that the data comes from a smooth developable surface. Since the estimation of tangent planes gives bad results on the boundary of the surface patch and near measurement errors, there will be outliers in the Blaschke image. To find those, we search for cells C_k carrying only a few image points. These cells and image points are not considered for the further computations. The result is referred to as cleaned Blaschke image. In addition, a thinning of the Blaschke image can be performed.

After having analyzed and cleaned the Blaschke image from outliers we are able to decide whether the given developable surface D is a general cone or cylinder, a cone or cylinder of revolution, another special developable or a general developable surface.

So, let T_i, i = 1, . . . ,M be the reliable estimated tangent planes of D after the cleaning and let $b(T_i) = b_i$ be their Blaschke images. As we have worked out in Section 10.4.3 we can classify the type of the developable surface D in the following way.

To check if the point cloud bi on B can be fitted well by a hyperplane H,

$$H : h_0 + h_1u_1 + \ldots + h_4u_4 = 0,\ h_1^2 + \ldots + h_4^2 = 1. \tag{4}$$

we perform a principal component analysis on the points b_i. This is equivalent to computing the ellipsoid of inertia of the points bi. It is known that the best fitting hyperplane passes through the barycenter $c = (\sum b_i)/M$ of the M data points b_i. Using c as new origin, the coordinate vectors of the data points are $q_i = b_i - c$ and the unknown three-space H has vanishing coefficient, $h_0 = 0$. The signed Euclidean distance $d(b_i, H)$ of a point qi and the unknown three space H is

$$d(q_i, H) = h_1q_i,1 + \ldots + h_4q_{i,4} = h \,.\, q_i, \tag{5}$$

where $h = (h_1, \ldots, h_4)$ denotes the unit normal vector of H. The minimization of the sum of squared distances,

$$F(h_1, h_2, h_3, h_4) = \frac{1}{M}\sum_{i=1}^{M} d^2(q_i, H) = \frac{1}{M}\sum_{i=1}^{M}(q_i \,.\, h_i)^2. \tag{6}$$

with respect to $h^2 = 1$ is an ordinary eigenvalue problem. Using a matrix notation with vectors as columns, it is written as

$$F(h) = h^T \,.\, C \,.\, h, \text{ with } C := \frac{1}{M}\sum_{i=1}^{M} q_i \,.\, q_i^T. \tag{7}$$

The symmetric matrix C is known as covariance matrix in statistics and as inertia tensor in mechanics. Let i be an eigenvalue of C and let v_i be the corresponding normalized eigenvector $(v_i^2 = 1)$. Then, $\lambda_i = F_2(v_i)$ holds and thus the best fitting three-space V_1 belongs to the smallest eigenvalue λ_1. The statistical standard deviation of the fit with V_1 is

$$\sigma_1 = \sqrt{\lambda_1/(n-4)}. \tag{8}$$

The distribution of the eigenvalues $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_4$ of the covariance matrix C (and the corresponding standard deviations $\sigma_1 \le \cdots \le \sigma_4$) gives important information on the shape of the surface D:

(1) Two small eigenvalues λ_1, λ_2 and different coefficients h_{10}, h_{20}, $(|h_{10} - h_{20}| > \varepsilon)$: The surface D can be well approximated by a cone of revolution, compare 6 in section 10.3.1. The vertex and the inclination angle are computed according to Section 10.3.

(2) Two small eigenvalues λ_1, λ_2 but nearly equal coefficients h_{10}, h_{20}, ($|h_{10} - h_{20}| \le \varepsilon$): The surface D can be well approximated by a cylinder of revolution, compare with (5) in section 10.3.1. The axis and the radius are computed according to Section 10.3.

(3) One small eigenvalue λ_1 and small coefficient h_{10} : The surface D is a general cone and its vertex is

$$v = \frac{1}{h_{14}}(h_{11}, h_{12}, h_{13}).$$

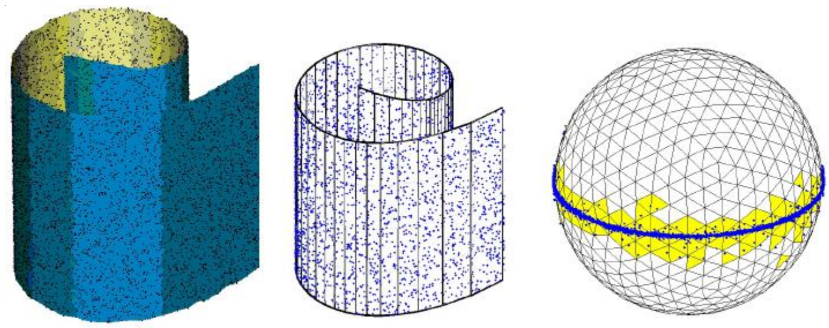
(4) One small eigenvalue λ_1 and small coefficients h_{10} and h_{14} : The surface D is a general cylinder and its axis is parallel to the vector

$$a = (h_{11}, h_{12}, h_{13}).$$

(5) One small eigenvalue λ_1 and small coefficient h_{14} : The surface D is a developable of constant slope. The tangent planes of D form a constant angle with respect to an axis. The angle and the axis are found according to formula (3) in section 10.3. An example is displayed in Figure 10.4.

(6) One small eigenvalue λ_1 characterizes a developable surface D whose tangent planes T_i are tangent to a sphere (compare with (4)) in section 10.3

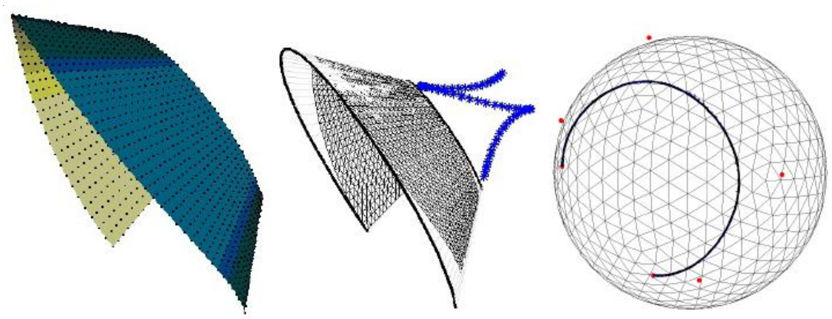
Figure 10.3: ***Left:*** **General cylinder.** ***Middle:*** **Triangulated data points and approximation.** ***Right:*** **Original Blaschke image (projected onto S^2).**



Its centre and radius are:

$$m = \frac{1}{h_{14}}(h_{11}, h_{12}, h_{13}), \qquad r = \frac{-h_{10}}{h_{14}}.$$

Figure 10.4: ***Left:*** **Developable of constant slope (math. model).** ***Middle:*** **Triangulated data points and approximation. Stars represent the singular curve.** ***Right:*** **Spherical image of the approximation with control points.**



For this classification, we need to fix a threshold ξ, to decide what small means. This value depends on the accuracy of the measurement device, the number of data points per area unit and the accuracy of the object. Some experience is necessary to choose this value for particular applications.

10.5 Reconstruction of Developable Surfaces from Measurements

In this section we describe the construction of a best-fitting developable surface to data points p_i or to estimated tangent planes T_i. In addition we address some problems, in particular the control of the singular curve of the approximation. First we note some general demands on the surface D to be approximated.

(1) D is a smooth surface not carrying singular points. D is not necessarily exactly developable, but one can run the algorithm also for nearly developable surfaces (one small principal curvature).

(2) The density of data points pi has to be approximately the same everywhere.

(3) The image $b(T_i)$ of the set of (estimated) tangent planes T_i has to be a simple, curve-like region on the Blaschke cylinder which can be injectively parameterized over an interval.

According to the made assumptions, the reconstruction of a set of measurement point p_i of a developable surface D can be divided into the following tasks:

(1) Fitting a curve $c(t) \subset B$ to the curve-like region formed by the data points $b(T_i)$.

(2) Computation of the one-parameter family of planes E(t) in $\mathbb{R}^3$ and of the generating lines L(t) of the developable D* which approximates measurements p_i.

(3) Computation of the boundary curves of D* with respect to the domain of interest in $\mathbb{R}^3$.

10.5.1 Curve fitting on the Blaschke cylinder B

We are given a set of unorganized data points $b(T_i) \in B$ and according to the made assumptions these points form a curve-like region on the Blaschke cylinder B. The aim is to fit a parametrized curve $c(t) \subset B$ to these points. In order to satisfy the constraint $c(t) \subset B$ we have to guarantee that

$$c_1(t)^2 + c_2(t)^2 + c_3(t)^2 = 1, \tag{9}$$

which says that the projection $c'(t) = (c_1, c_2, c_3)(t)$ of $c(t) = (c_1, c_2, c_3, c_4)(t)$ to $\mathbb{R}^3$ is a spherical curve (in S^2). The computation of a best fitting curve to unorganized points is not trivial, but there are several methods around. Estimation of parameter values or sorting the points are useful ingredients to simplify the fitting. We do not go into detail here but refer to the moving least squares method to estimate parameter values and to the approach by Lee [11] who uses a minimum spanning tree to define an ordering of the points. These methods apply also to thinning of the curve-like point cloud.

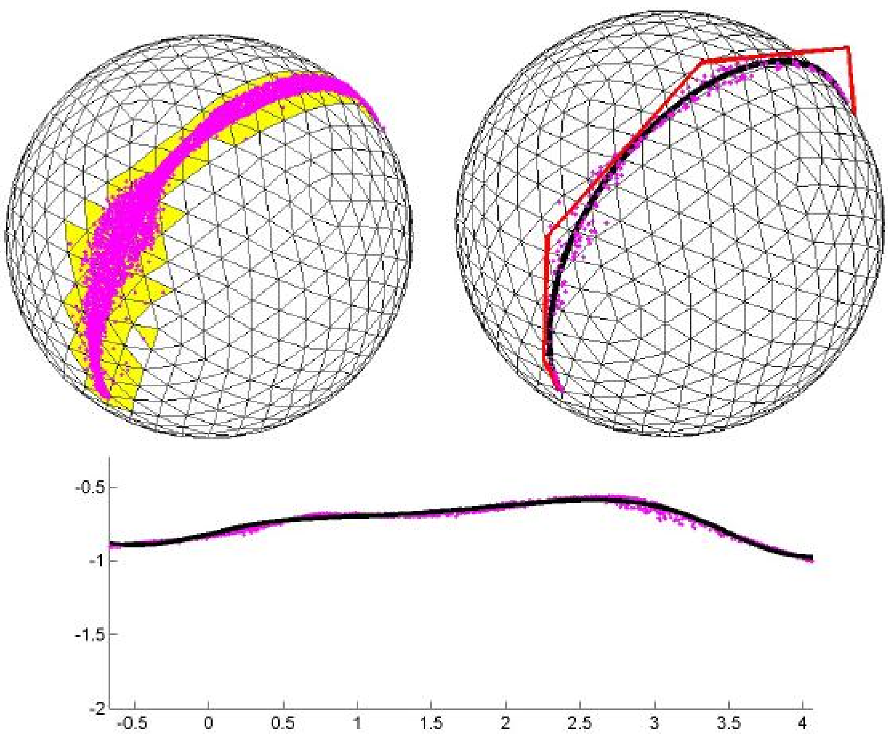
After this preparation we perform standard curve approximation with Baselines and project the solution curve to the Blaschke cylinder B in order to satisfy the constraint (9). If the projection $c' \subset S^2$ of c(t) is contained in a hemisphere of S^2 and if additionally the fourth coordinate $c_4(t)$ does not vary to much, it is appropriate to perform a stereographic projection so that we finally end up with a rational curve c(t) on B. For practical purposes it will often be sufficient to apply a projection to B with rays orthogonally to u_4, the axis of B.

Figure 23.5 shows a curve-like region in S^2 with varying width, an approximating curve $c'(t)$ to this region and the approximation $c_4(t)$ of the support function to a set of image points $b(T_i) \subset B$.

Notes

We mention here that the presented curve fitting will fail in the case when inflection generators occur in the original developable shape, because inflection generators correspond to singularities of the Blaschke image. Theoretically, we have to split the data set at an inflection generator and run the algorithm for the parts separately and join the partial solutions. In practice, however, it is not so easy to detect this particular situation and it is not yet implemented.

Figure 10.5: Blaschke image (left) (projected onto S^2), approximating curve to thinned point cloud (right) and support function (fourth coordinate)



10.5.2 Biarcs in the Space of Planes

We like to mention an interesting relation to biarcs. Biarcs are curves composed of circular arcs with tangent continuity and have been studied at first in the plane. It is known that the G^1-Hermite interpolation problem of Hermite elements (points plus tangent lines) P_1, V_1 and P_2, V_2 possesses a one-parameter solution with biarcs which can be parameterized over the projective line. Usually one can expect that suitable solutions exist, but for some configurations there are no solutions with respect to a given orientation of the tangent lines V_j .

The construction of biarcs can be carried out on quadrics too, in particular on the sphere S^2 or on the Blaschke cylinder B. If we consider a biarc (elliptic) $c = b(D) \subset B$ then the corresponding developable surface D in $\mathbb{R}^3$ is composed of cones or cylinders of revolution with tangent plane continuity along a common generator. To apply this in our context, we sample Hermite elements P_j , V_j , $j = 1, \ldots, n$ from an approximation $c(t) \subset B$ of the set $b(T_i)$. Any pair of Hermite elements P_j , V_j and P_{j+1}, V_{j+1} is interpolated by a pair of elliptic arcs on B with tangent continuity. Applying this concept, the final developable surface is composed of smoothly joined cones of revolution. This has the advantage that the development (unfolding) of the surface is elementary.

Notes

10.5.3 A Parametrization of the Developable Surface

Once we have computed a curve $c(t) \subset B$ that approximates the image points $b(T_i)$ well, the one-parameter family $E(t)$ determining the approximating developable surface D^* is already given by

$$E(t) : c_4(t) + c_1(t)x + c_2(t)y + c_3(t)z = 0.$$

The generating lines $L(t)$ of D^* are the intersection lines $E(t) \cap \dot{E}(t)$. We assume that there exist two bounding planes H_1 and H_2 of the domain of interest in a way that all generating lines $L(t)$ intersect H_1 and H_2 in proper points. The intersection curves $f_i(t)$ of $L(t)$ and H_i, i = 1, 2 are computed by

$$f_i(t) = E(t) \cap \dot{E}(t) \cap H_i, \qquad (10)$$

and the final point representation of D^* is

$$x(t, u) = (1 - u)f_1(t) + uf_2(t). \qquad (11)$$

Figures 10.3, 10.4, 10.6 and 10.7 show developable surfaces which approximate data points (displayed as dots).

The deviation or distance between the given surface D and the approximation D^* can be defined according to distances between estimated planes T_i, i = 1, . . . , N (with corresponding parameter values t_i) and the approximation $E(t)$ by

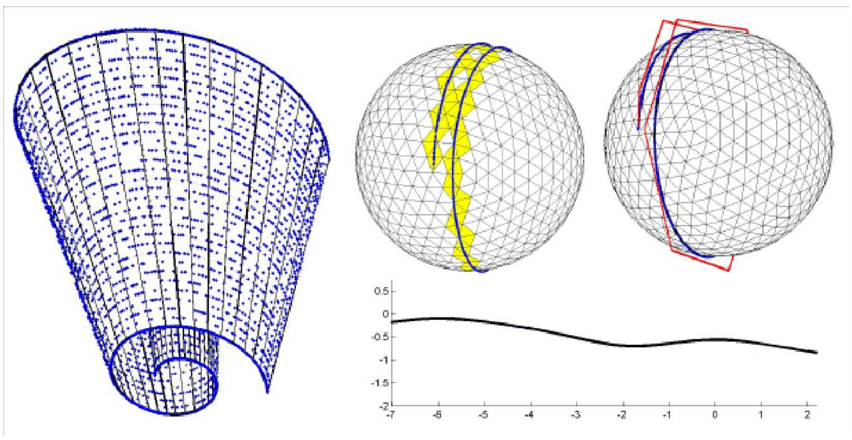
$$d^2(D, D^*) = \frac{1}{N}\sum_i \mathrm{dist}^2\,(T_i, E(t_i)). \qquad (12)$$

If more emphasis is on the deviation of the measurements p_i from the developable D^*, one can use

$$d^2(D, D^*) = \frac{1}{N}\sum_i \mathrm{dist}^2\,(p_i, E(t_i)). \qquad (13)$$

with respect to orthogonal distances between points p_i and planes $E(t_i)$.

Figure 10.6: ***Left:*** **Developable surface approximating the data points.** ***Right:*** **Projection of the Blaschke image onto S^2, approximating curve with control polygon and support function.**



10.5.4 Fitting Developable Surfaces to nearly Developable Shapes

Notes

The proposed method can be applied also to fit a developable surface to data which comes from a nearly developable shape. Of course, we have to specify what nearly developable means in this context. Since the fitting is performed by fitting a one-parameter family of tangent planes, we will formulate the requirements on the data pi in terms of the Blaschke image of the estimated tangent planes T_i.

If the data points p_i are measurements of a developable surface D and if the width in direction of the generators does not vary too much, the Blaschke image b(D) = R will be a tubular-like (curve-like) region on B with nearly constant thickness. Its boundary looks like a pipe surface.

Putting small distortions to D, the normals of D will have a larger variation near these distortions. The Blaschke image b(D) possesses a larger width locally and will look like a canal surface. As long as it is still possible to compute a fitting curve to b(D), we can run the algorithm and obtain a developable surface approximating D. Figures 10.5 and 10.7 illustrate the projection of b(D) onto the unit sphere S^2.

The analysis of the Blaschke image b(D) gives a possibility to check whether D can be approximated by a developable surface or not. By using the cell structure of the cleaned Blaschke image b(D), we pick a cell C and an appropriately chosen neighborhood U of C. Forming the intersection $R = U \cap b(D)$, we compute the ellipsoid of inertia (or a principal component analysis) of R. The existence of one significantly larger eigenvalue indicates that R can be approximated by a curve in a stable way. Thus, the point set corresponding to R can be fitted by a developable surface.

Since approximations of nearly developable shapes by developable surfaces are quite useful for practical purposes, this topic will be investigated in more detail in the future.

10.5.5 Singular Points of a Developable Surface

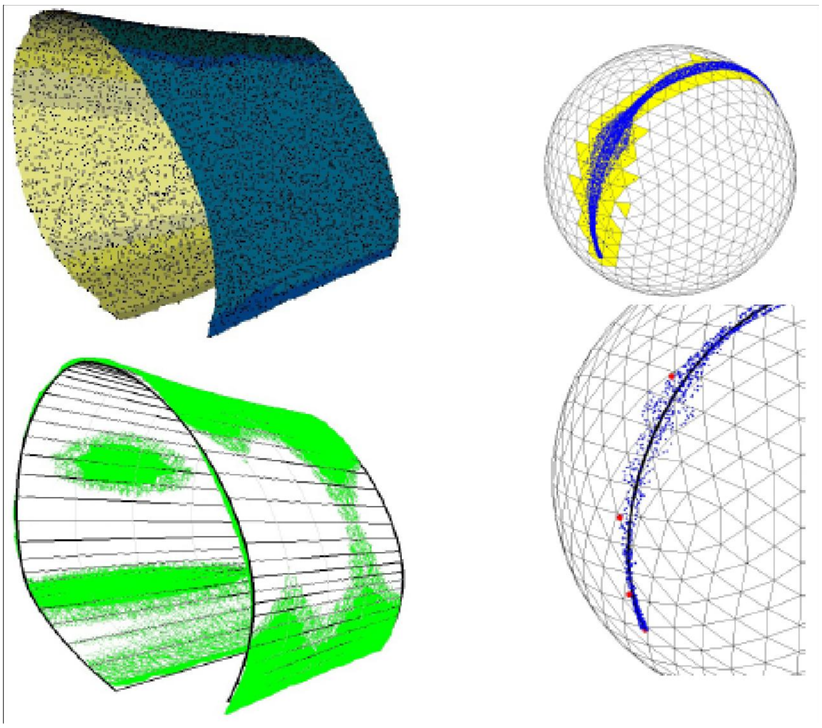
So far, we did not pay any attention to singular points of D*. The control and avoidance of the singular points within the domain of interest is a complicated topic because the integration of this into the curve fitting is quite difficult.

If the developable surface D* is given by a point representation, formula (4) represents the singular curve s(t). If D* is given by its tangent planes E(t), the singular curve s(t) is the envelope of the generators L(t) and so it is computed by

$$s(t) = E(t) \cap \dot{E}(t) \cap \ddot{E}(t). \qquad (14)$$

Thus, the singular curve s(t) depends in a highly nonlinear way on the coordinate functions of E(t).

Figure 10.7: *Left:* Nearly developable surface and developable approximation. *Right:* Projection of the original Blaschke image onto S^2 and thinned Blaschke image with approximating curve.



In order to compute the singular curve s(t), let $n = c \wedge \dot{c} \wedge \ddot{c}$, where ^ denotes the vector product in $\mathbb{R}^4$. The Cartesian coordinates of the singular curve are then found by

$$s(t) = \frac{1}{n_4(t)}(n_1(t), n_2(t), n_3(t)). \tag{15}$$

Zeros of the function n_4 correspond to points at infinity of s(t). In Section 10.4.2, we have assumed that all coordinates of data points are bounded by ±c such that we have $\|p_i\| \leq 1$. In order to approximate the data with singularity-free developable surfaces, we have to guarantee

$$\|s(t)\| > 1, \tag{16}$$

when we fit curves c(t) to image points $b(T_i) \in B$ of estimated tangent planes T_i. Since the data comes from a developable surface without singularities, we

Assuming that the curve $c(t) \in B$ fitted to the data $b(T_i)$ is composed of biarcs, we obtain the following: For two consecutive Hermite elements P_j, V_j and P_{j+1}, V_{j+1} there exists a one-parameter family of interpolating pairs of arcs and condition (16) leads to a quadratic inequality. Thus, solutions can be computed explicitly. However, as we have mentioned in Section 10.5.2, there is no guarantee that feasible solutions exist and the construction clearly depends on the choice of the Hermite elements which have been sampled from an initial solution of the curve fitting.

Notes

10.5.6 Conclusion

We have proposed a method for fitting a developable surface to data points coming from a developable or a nearly developable shape. The approach applies curve approximation in the space of planes to the set of estimated tangent planes of the shape. This approach has advantages compared to usual surface fitting techniques, like

- avoiding the estimation of parameter values and direction of generators,
- guaranteeing that the approximation is developable.

The detection of regions containing inflection generators, and the avoidance of singular points on the fitted developable surface have still to be improved. The approximation of nearly developable shapes by developable surface is an interesting topic for future research. In particular we will study the segmentation of a non-developable shape into parts which can be well approximated by developable surfaces. This problem is relevant in certain applications (architecture, ship hull manufacturing), although one cannot expect that the developable parts will fit together with tangent plane continuity.

Acknowledgements This research has been supported partially by the innovative project '3D Technology' of Vienna University of Technology.

10.6 Summary

- For fitting ruled surfaces to point clouds, we have to estimate in advance the approximate direction of the generating lines of the surface in order to estimate useful parameter values for the given data. To perform this, it is necessary to estimate the asymptotic lines of the surface in a stable way.
- We have to guarantee that the resulting approximation b(u, v) is developable, which is expressed by equation. Plugging the parametrization b(u, v) into this condition leads to a highly non-linear side condition in the control points b_{ij} for the determination of the approximation b(u, v).
- We consider a fixed point $p = (p_1, p_2, p_3)$ and all planes $E : n \,.\, x + d = 0$ passing through this point. The incidence between p and E is expressed by

  $$p_1n_1 + p_2n_2 + p_3n_3 + d = p \,.\, n + d = 0,$$

  and therefore the image points $b(E) = (n_1, n_2, n_3, d)$ in $\mathbb{R}^4$ of all planes passing through p lie in the three-space

  $$H : p_1u_1 + p_2u_2 + p_3u_3 + u_4 = 0,$$

  passing through the origin of $\mathbb{R}^4$. The intersection $H \cap B$ with the cylinder B is an ellipsoid and any point of $H \cap B$ is image of a plane passing through p. shows a 2D illustration of this property.
- Let D be a cylinder of revolution with axis A and radius r. The tangent planes T of D are tangent to all spheres of radius r, whose centers vary on A. Let

  $$S_1 : (x - p)^2 - r^2 = 0, S_2 : (x - q)^2 - r^2 = 0$$

be two such spheres with centers p, q. According to (11), the images b(T) of the tangent planes T satisfy the relations

$$H_1 : -r + u_1p_1 + u_2p_2 + u_3p_3 + u_4 = 0,$$

$$H_2 : -r + u_1q_1 + u_2q_2 + u_3q_3 + u_4 = 0.$$

Since $p \neq q$, the image curve b(D) lies in the plane $P = H_1 \cap H_2$ and b(D) is a conic.

- Given a cloud of data points p_i, this section discusses the recognition and classification of developable surfaces according to their Blaschke images. The algorithm contains the following steps:
  - Estimation of tangent planes T_i at data points p_i and computation of the image points $b(T_i)$.
  - Analysis of the structure of the set of image points $b(T_i)$.
  - If the set $b(T_i)$ is curve-like, classification of the developable surface which is close to p_i.
- D is a smooth surface not carrying singular points. D is not necessarily exactly developable, but one can run the algorithm also for nearly developable surfaces (one small principal curvature).
- The density of data points pi has to be approximately the same everywhere.
- The image $b(T_i)$ of the set of (estimated) tangent planes T_i has to be a simple, curve-like region on the Blaschke cylinder which can be injectively parameterized over an interval.
- According to the made assumptions, the reconstruction of a set of measurement point p_i of a developable surface D can be divided into the following tasks:
  - Fitting a curve $c(t) \subset B$ to the curve-like region formed by the data points $b(T_i)$.
  - Computation of the one-parameter family of planes E(t) in $\mathbb{R}^3$ and of the generating lines L(t) of the developable D* which approximates measurements p_i.
  - Computation of the boundary curves of D* with respect to the domain of interest in $\mathbb{R}^3$.

10.7 Keywords

Cylinder: D is a general cylinder if all its tangent planes T(u) are parallel to a vector a and thus its normal vectors n(u) satisfy n . a = 0. This implies that the image curve b(T(u)) = b(D) is contained in the three-space $H : a_1u_1 + a_2u_2 + a_3u_3 = 0$.

Cone: D is a general cone if all its tangent planes T(u) pass through a fixed point $p = (p_1, p_2, p_3)$. This incidence is expressed by $p_1n_1 + p_2n_2 + p_3n_3 + n_4 = 0$. Thus, the Blaschke image curve b(T(u)) = b(D) is contained in the three space $H : p_1u_1 + p_2u_2 + p_3u_3 + u_4 = 0$.

10.8 Self Assessment

1. Let S be the oriented sphere with center m and signed radius r, ... = 0. The tangent planes T_S of S are exactly those planes, whose signed distance from m equals r.
2. The surface D is a .... of constant slope, if its normal vectors n(u) form a constant angle ϕ with a fixed direction vector a.
3. Let D be a cylinder of revolution with axis A and radius r. The .... T of D are tangent to all spheres of radius r, whose centers vary on A.

4. Let the plane $P = H_1 \cap H_2$ be given by equations. The pencil of three spaces $\lambda H_1 + \mu H_2$ contains a unique three-space H, passing through the origin in $\mathbb{R}^4$, whose equation is ....

5. D is a smooth surface not carrying singular points. D is not necessarily exactly developable, but one can run the .... also for nearly developable surfaces (one small principal curvature).

6. Computation of the boundary curves of D* with respect to the domain of interest in ....

10.9 Review Questions

1. Discuss The Blaschke Model of Oriented Planes in R^3.
2. Explain Incidence of Point and Plane.
3. Define Tangency of Sphere and Plane.
4. Discuss the Classification of Developable Surfaces according to their Image on B.
5. Describe Cones and Cylinders of Revolution.
6. Explain Recognition of Developable Surfaces from Point Clouds.
7. Describe Reconstruction of Developable Surfaces from Measurements.

Answers: Self Assessment

1.	$S : (x - m)^2 - r^2$	2.	developable
3.	tangent planes	4.	$H : \sum_{i=1}^{3} u_i(sp_i - rq_i) + u_4(s - r) = 0.$
5.	algorithm	6.	$\mathbb{R}^3$.

10.10 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 11: Two Fundamental Form

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define surfaces
- Explain the first fundamental form
- Describe the second fundamental form
- Discuss some example related to fundamental forms

Introduction

In last unit, you have studied about development surfaces. In mathematics, specifically in topology, a surface is a two-dimensional topological manifold. The most familiar examples are those that arise as the boundaries of solid objects in ordinary three-dimensional Euclidean space R^3 - for example, the surface of a ball. There are surfaces, such as the Klein bottle, that cannot be embedded in three-dimensional Euclidean space without introducing singularities or self-intersections.

11.1 Surfaces

Definition 1. A parametric surface patch is a smooth mapping:

$$X : U \to \mathbb{R}^3,$$

where $U \subset \mathbb{R}^2$ is open, and the Jacobian dX is non-singular.

Notes

Write $X = (x^1, x^2, x^3)$, and each $x^i = x^i(u^1, u^2)$, then the Jacobian has the matrix representation:

$$dX = \begin{pmatrix} x_1^1 & x_2^1 \\ x_1^2 & x_2^2 \\ x_1^3 & x_2^3 \end{pmatrix}$$

where we have used the notation $f_i = f_{u^i} = \partial f / \partial u^i$. According to the definition, we are requiring that this matrix has rank 2, or equivalently that the vectors $X_1 = \left(x_2^1, x_1^2, x_1^3\right)$ and $X_2 = \left(x_2^1, x_2^2, x_2^3\right)$ are linearly independent. Another equivalent requirement is that $dX : \mathbb{R}^2 \to \mathbb{R}^3$ is injective.

Example: Let $U \subset \mathbb{R}^2$ be open, and suppose that $f : U \to \mathbb{R}$ is smooth. Define the graph of f as the parametric surface $X(u^1, u^2) = (u^1, u^2, f(u^1, u^2))$. To verify that X is indeed a parametric surface, note that:

$$dX = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ f_1 & f_2 \end{pmatrix}$$

so that clearly X is non-singular.

A diffeomorphism between open sets U, $V \subset \mathbb{R}^2$ is a map $\phi : U \to V$ which is smooth, one-to-one, and whose inverse is also smooth. If $\det(d\phi) > 0$, then we say that ϕ is an orientation-preserving diffeomorphism.

Definition 2. Let $X : U \to \mathbb{R}^3$, and $\tilde{X} : \tilde{U} \to \mathbb{R}^3$ be parametric surfaces. We say that $\tilde{X}$ is reparametrization of X if $\tilde{X} = X \circ \phi$, where $\phi : \tilde{U} \to U$ is a diffeomorphism. If ϕ is an orientation-preserving diffeomorphism, then $\tilde{X}$ is an orientation-preserving reparametrization.

Clearly, the inverse of a diffeomorphism is a diffeomorphism. Thus, if $\tilde{X}$ is a reparametrization of X, then X is a reparametrization of $\tilde{X}$.

Definition 3. The tangent space $T_u \mid X$ of the parametric surface $X : U \to \mathbb{R}^3$ at $u \in U$ is the 2-dimensional linear subspace of $\mathbb{R}^3$ spanned by the two vectors X_1 and X_2.[1]

If $Y \in T_u X$, then it can be expressed as a linear combination in X_1 and X_2:

$$Y = y^1 X_1 + y^2 X_2 = \sum_{i=1}^{2} y^i X_i,$$

where $y^i \in \mathbb{R}$ are the components of the vector Y in the basis X_1, X_2 of $T_u X$. We will use the *Einstein Summation Convention:* every index which appears twice in any product, once as a subscript (covariant) and once as a superscript (contravariant), is summed over its range. For example, the above equation will be written $Y = y^i X_i$. The next proposition show that the tangent space is invariant under reparametrization, and gives the law of transformation for the components of

[1] Note that the tangent plane to the surface X(U) at u is actually the affine subspace $X(u) + T_u X$. However, it will be very convenient to have the tangent space as a linear subspace of $\mathbb{R}^3$.

Notes

a tangent vector. Note that covariant and contravariant indices have different transformation laws, cf. (1) and (2).

Proposition 1. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let $\tilde{X} = X \circ \phi$ be a reparametrization of X. Then $T_{\phi(\tilde{u})}X = T_{\tilde{u}}\tilde{X}$. Furthermore, if $Z \in T_{\phi(\tilde{u})}\tilde{X}$, and $Z = z^i X_i = \tilde{z}^j \tilde{X}_j$, then:

$$z^i = \tilde{z}^j \frac{\partial u^i}{\partial \tilde{u}^j}, \qquad ...(1)$$

where $d\phi = \left(\partial u^i / \partial \tilde{u}^j\right)$.

Proof. By the chain rule, we have:

$$\tilde{X}_j = \frac{\partial u^i}{\partial \tilde{u}^j} X_i. \qquad ...(2)$$

Thus $T_{\tilde{u}}\tilde{X} \subset T_{\phi(\tilde{u})}X$, and since we can interchange the roles of X and $\tilde{X}$, we conclude that $T_{\tilde{u}}\tilde{X} = T_{\phi(\tilde{u})}X$. Substituting (2) in $\tilde{z}^j \tilde{X}_j$, we find:

$$\tilde{z}^i X_i = \tilde{z}^j \frac{\partial u^i}{d\tilde{u}^j} X_i,$$

and (1) follows.

Definition 4. A vector field along a parametric surface $X : U \to \mathbb{R}^3$, is a smooth mapping $Y : U \to \mathbb{R}^3$.[2] A vector field Y is tangent to X if $Y(u) \in T_u X$ for all $u \in U$. A vector field Y is normal to X if $Y(u) \perp T_u X$ for all $u \in U$.

Example: The vector fields X_1 and X_2 are tangent to the surface. The vector field $X_1 \times X_2$ is normal to the surface.

We call the unit vector field

$$N = \frac{X_1 \times X_2}{|X_1 \times X_2|}$$

the unit normal. Note that the triple (X_1, X_2, N), although not necessarily orthonormal, is positively oriented. In particular, we can see that the choice of an orientation on X, e.g., $X_1 \to X_2$, fixes a unit normal, and vice-versa, the choice of a unit normal fixes the orientation. Here we chose to use the orientation inherited from the orientation $u^1 \to u^2$ on U.

Definition 5. We call the map $N : U \to \mathbb{S}^2$ the Gauss map.

The Gauss map is invariant under orientation-preserving reparametrization.

Proposition 2. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let $N : U \to \mathbb{S}^2$ be its Gauss map. Let $\tilde{X} = X \circ \phi$ be an orientation-preserving reparametrization of X. Then the Gauss map of $\tilde{X}$ is $N \circ \phi$.

[2] We often visualize Y(u) as being attached at X(u), i.e. belonging to the tangent space of $\mathbb{R}^3$ at X(u); cf. see footnote 1.

Proof. Let $\upsilon \in V$. The unit normal $\tilde{N}(\upsilon)$ of $\tilde{X}$ at υ is perpendicular to $T_\upsilon\tilde{X}$. By Proposition 1, we have $T_{\phi(\upsilon)}X = T_\upsilon\tilde{X}$. Thus, $\tilde{N}(\upsilon)$ is perpendicular to $T_{\phi(\upsilon)}X$, as is $N(\phi(\upsilon))$. It follows that the two vectors are co-linear, and hence $\tilde{N}(\upsilon) = \pm N(\phi(\upsilon))$. But since ϕ is orientation preserving, the two pairs (X_1, X_2) and $(\tilde{X}_1, \tilde{X}_2)$ have the same orientation in the plane $T_\upsilon\tilde{X}$. Since also, the two triples $(X1(\phi(\upsilon)), X2(\phi(\upsilon)), N(\phi(\upsilon)))$ and $(\tilde{X}_1(\upsilon), \tilde{X}_2(\upsilon), N(\upsilon))$ have the same orientation in $\mathbb{R}^3$, it follows that $N(\phi(\upsilon)) = \tilde{N}(\upsilon)$.

11.2 The First Fundamental Form

Definition 6. A symmetric bilinear form on a vector space V is function $B : V \times V \to \mathbb{R}$ satisfying:

1. $B(aX + bY,Z) = aB(X,Z) + bB(Y,Z)$, for all $X, Y \in V$ and $a, b \in R$.
2. $B(X, Y) = B(Y, X)$, for all $X, Y \in V$.

The symmetric bilinear form B is positive definite if $B(X, X) \geq 0$, with equality if and only if $X = 0$.

With any symmetric bilinear form B on a vector space, there is associated a quadratic form $Q(X) = B(X, X)$. Let V and W be vector spaces and let $T : V \to W$ be a linear map. If B is a symmetric bilinear form on W, we can define a symmetric bilinear form $T^* Q$ on V by $T^* Q(X, Y) = Q(TX, TY)$. We call $T^* Q$ the pull-back of Q by T. The map T is then an isometry between the inner-product spaces $(V, T^* Q)$ and (W,Q).

Example: Let $V = \mathbb{R}^3$ and define $B(X, Y) = X \cdot Y$, then B is a positive definite symmetric bilinear form. The associated quadratic form is $Q(X) = |X|^2$.

Example: Let A be a symmetric 2×2 matrix, and let $B(X, Y) = AX \cdot Y$, then B is a symmetric bilinear form which is positive definite if and only if the eigenvalues of A are both positive.

Definition 7. Let $X : U \to \mathbb{R}^3$ be a parametric surface. The first fundamental form is the symmetric bilinear form g defined on each tangent space $T_u X$ by:

$$g(Y, Z) = Y \cdot Z, \quad \forall Y, Z \in T_u X.$$

Thus, g is simply the restriction of the Euclidean inner product in above Example to each tangent space of X. We say that g is induced by the Euclidean inner product.

Let $g_{ij} = g(X_i, X_j)$, and let $Y = y^i X_i$ and $Z = z^i X_i$ be two vectors in $T_u X$, then

$$g(Y,Z) = g_{ij}\, y^i z^j. \qquad \text{...(3)}$$

Thus, the so-called coordinate representation of g is at each point $u_0 \in U$ an instance of above example. In fact, if $A = (g_{ij})$, and $B(\xi, \eta) = \xi \cdot A\eta$ for $\xi, \eta \in \mathbb{R}^2$ as in the above example, then B is the pull-back by $dX_u : \mathbb{R}^2 \to T_u X$ of the restriction of the Euclidean inner product on $T_u X$.

The classical (Gauss) notation for the first fundamental form is $g_{11} = E$, $g_{12} = g_{21} = F$, and $G = g_{22}$, i.e.,

$$\left(g_{ij}\right) = \begin{pmatrix} E & F \\ F & G \end{pmatrix}$$

Clearly, $F^2 < EG$, and another condition equivalent to the condition that X_1 and X_2 are linearly independent is that $\det(g_{ij}) = EG - F^2 > 0$. The first fundamental form is also sometimes written:

$$ds^2 = g_{ij}\, du^i\, du^j = E\,(du^1)^2 + 2F\, du^1\, du^2 + G(du^2)^2.$$

Note that the g_{ij}'s are functions of u. The reason for the notation ds^2 is that the square root of the first fundamental form can be used to compute length of curves on X. Indeed, if $\gamma : [a,b] \to \mathbb{R}^3$ is a curve on X, then $\gamma = X \circ \beta$, where β is a curve in U. Let $\beta(t) = \left(\beta^1(t), \beta^2(t)\right)$, and denote time derivatives by a dot, then

$$L\gamma\left([a,b]\right) = \int_a^b |\dot{\gamma}|\, dt \int_a^b \sqrt{g_{ij}\,\dot{\beta}^i\,\dot{\beta}^j}\, dt.$$

Accordingly, ds is also called the line element of the surface X.

Note that g contains all the intrinsic geometric information about the surface X. The distance between any two points on the surface is given by:

$$d(p,q) = \inf\left\{L_\gamma : \gamma \text{ is a curve on X between p and q}\right\}.$$

Also the angle between two vectors $Y, Z \in T_x X$ is given by:

$$\cos\theta = \frac{g(Y,Z)}{\sqrt{g(Y,Y)\, g(Z,Z)}},$$

and the angle between two curves β and γ on X is the angle between their tangents $\dot{\beta}$ and $\dot{\gamma}$. Intrinsic geometry is all the information which can be obtained from the three functions g_{ij} and their derivatives.

Clearly, the first fundamental form is invariant under reparametrization. The next proposition shows how the g_{ij}'s change under reparametrization.

Proposition 3. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let $\tilde{X} = X \circ \phi$ be a reparametrization of X. Let g_{ij} be the coordinate representation of the first fundamental form of X, and let $\tilde{g}_{ij}$ be the coordinate representation of the first fundamental form of $\tilde{X}$. Then, we have:

$$\tilde{g}_{ij} = g_{kl}\,\frac{\partial u^k}{\partial \tilde{u}^i}\frac{\partial u^l}{\partial \tilde{u}^j}, \qquad ...(4)$$

where $d\phi = \left(\partial u^i / \partial \tilde{u}^j\right)$.

Proof. In view of (2), we have:

$$\overline{g}_{ij} = g\left(\tilde{X}_i, \tilde{X}_j\right) = g\left(\frac{\partial u^k}{\partial \tilde{u}^i}X_k, \frac{\partial u^l}{\partial \tilde{u}^j}X_l\right) = \frac{\partial u^k}{d\tilde{u}^i}\frac{\partial u^l}{d\tilde{u}^j}\, g\left(X_k, X_l\right) = g_{kl}\,\frac{\partial u^k}{\partial \tilde{u}^i}\frac{\partial u^l}{\partial \tilde{u}^j}.$$

11.3 The Second Fundamental Form

Notes

We now turn to the second fundamental form. First, we need to prove a technical proposition. Let Y and Z be vector fields along X, and suppose that $Y = y^i X_i$ is tangential. We define the directional derivative of Z along Y by:

$$\partial_Y Z = y^i Z_i = y^i \frac{\partial Z}{\partial u^i}.$$

Note that the value of $\partial_Y Z$ at u depends only on the value of Y at u, but depends on the values of Z in a neighborhood of u. In addition, $\partial_Y Z$ is reparametrization invariant, but even if Z is tangent, it is not necessarily tangent. Indeed, if we write $Y = \tilde{y}^i \tilde{X}_i$, then we see that:

$$\tilde{y}^i \tilde{\partial}_i Z = \tilde{y}^i \frac{\partial Z}{\partial \tilde{u}^i} = y_j \frac{\partial \tilde{u}^i}{\partial u^j} \frac{\partial Z}{\partial u^k} \frac{\partial u^k}{\partial \tilde{u}^i} = y^i \partial_j Z.$$

The commutator of Y and Z can now be defined as the vector field:

$$[Y,Z] = \partial_Y Z - \partial_Z Y.$$

Proposition 4. Let $X : U \to \mathbb{R}^3$ be a surface, and let N be its unit normal.

(1) If Y and Z are tangential vector fields then $[Y,Z] \in T_u X$.

(2) If $Y, Z \in T_u X$ then $\partial_Y N \cdot Z = \partial_Z N \cdot Y$.

Proof. Note first that since X is smooth, we have $X_{ij} = X_{ji}$, where we have used the notation $X_{ij} = \partial^2 X/\partial u^i u^j$. Now, write $Y = y^i X_i$ and $Z = z^j X_j$, and compute:

$$\partial_Y Z - \partial_Z Y = y^i z^j X_{ji} + y^i \partial_i z^j X_j - y^i z^j X_{ij} - z^j \partial_j y^i X_i$$

$$= \left(y^i \partial_i z^j - z^i \partial_i y^j\right) X_j.$$

To prove (2), extend Y and Z to be vector fields in a neighborhood of u, and use (1):

$$\partial_Y N \cdot Z - \partial_Z N \cdot Y = -N \cdot (\partial_Y Z - \partial_Z Y) = 0.$$

Note that while proving the proposition, we have established the following formula for the commutator:

$$[Y,Z] = \left(y^i \partial_i z^j - z^i \partial_i y^j\right) X_j \qquad \text{...(5)}$$

Definition 8. Let $X : U \to \mathbb{R}^3$ be a surface, and let $N : U \to \mathbb{S}^2$ be its unit normal. The second fundamental form of X is the symmetric bilinear form k defined on each tangent space $T_u X$ by:

$$k(Y, Z) = -\partial_Y N \cdot Z. \qquad \text{...(6)}$$

We remark that since $N \cdot N = 1$, we have $\partial_Y N \cdot N = 0$, hence $\partial_Y N$ is tangential. Thus, according to (6), the second fundamental form is minus the tangential directional derivative of the unit normal, and hence measures the turning of the tangent plane as one moves about on the surface. Note that part (2) of the proposition guarantees that k is indeed a symmetric bilinear form. Note that it is not necessarily positive definite. Furthermore, if we set $k_{ij} = k(X_i, X_j)$ to be the coordinate representation of the second fundamental form, then we have:

$$k_{ij} = X_{ij} \cdot N. \qquad \text{...(7)}$$

This equation leads to another representation. Consider the Taylor expansion of X at a point, say $0 \in U$:

$$X(u) = X(0) + X_i(0)u^i + \frac{1}{2}\partial_{ij}\, X(0)u^i u^j + O\left(|u|^3\right)$$

Thus, the elevation of X above its tangent plane at u is given up to second-order terms by:

$$\left(X(u) - X(0) - X_i(0)u^i\right) \cdot N = \frac{1}{2}k_{ij}(0)u^i u^j + O\left(|u|^3\right).$$

The paraboloid on the right-hand side of the equation above is called the osculating paraboloid. A point u of the surface is called elliptic, hyperbolic, parabolic, or planar, depending on whether this paraboloid is elliptic, hyperbolic, cylindrical, or a plane.

In classical notation, the second fundamental form is:

$$\left(k_{ij}\right) = \begin{pmatrix} L & M \\ M & N \end{pmatrix}.$$

Clearly, the second fundamental form is invariant under orientation-preserving reparametrizations. Furthermore, the k_{ij}'s, the coordinate representation of k, changes like the first fundamental form under orientation-preserving reparametrization:

$$\tilde{k}_{ij} = k\left(\tilde{X}_u, \tilde{X}_j\right) = k_{ml}\frac{\partial u^m}{\partial \tilde{u}^i}\frac{\partial u^l}{\partial \tilde{u}^j},$$

Yet another interpretation of the second fundamental form is obtained by considering curves on the surface. The following theorem is essentially due to Euler.

Theorem 1. Let $\gamma = X \circ \beta : [a,b] \to \mathbb{R}^3$ be a curve on a parametric surface $X : U \to \mathbb{R}^3$, where $\beta : [a,b] \to U$. Let k be the curvature of γ, and let θ be the angle between the unit normal N of X, and the principal normal e_2 of γ. Then:

$$k\cos\theta = k\left(\dot{\gamma}, \dot{\gamma}\right). \qquad \text{...(8)}$$

Proof. We may assume that γ is parametrized by arclength. We have:

$$\dot{\gamma} = \dot{\beta}^i X_i,$$

and

$$ke_2 = \ddot{\gamma} = \ddot{\beta}^i X_i + \dot{\beta}^i\dot{\beta}^j X_{ij}.$$

The theorem now follows by taking inner product with N, and taking (7) into account.

The quantity k cos θ is called the normal curvature of γ. It is particularly interesting to consider normal sections, i.e., curves γ on X which lie on the intersection of the surface with a normal plane. We may always orient such a plane so that the normal e_2 to γ in the plane coincide with the unit normal N of the surface. In that case, we obtain the simpler result:

$$k = k\left(\dot{\gamma}, \dot{\gamma}\right).$$

Thus, the second fundamental form measures the signed curvature of normal sections in the normal plane equipped with the appropriate orientation.

Definition 9. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let k be its second fundamental form. Denote the unit circle in the tangent space at u by $S_uX = \{Y \in T_uX : |Y| = 1\}$. For u $\in$ U, define the principal curvatures of X at u by:

$$k_1 = \min_{Y \in S_uX} k(Y, Y), \quad k_2 = \max_{Y \in SsX} k(Y, Y).$$

The unit vectors $Y \in S_uX$ along which the principal curvatures are achieved are called the principal directions. The mean curvature H and the Gauss curvature K of X at u are given by:

$$H = \frac{1}{2}(k_1 + k_2), \quad K = k_1k_2.$$

If we consider the tangent space T_uX with the inner product g and the unique linear transformation $\ell : T_uX \to T_uX$ satisfying:

$$g(\ell(Y), Z) = k(Y, Z), \quad \forall Z \in T_uX, \qquad ...(9)$$

then $k_1 \le k_2$ are the eigenvalues of ℓ and the principal directions are the eigenvectors of ℓ. If $k_1 = k_2$ then $k = \lambda g$ and every direction is a principal direction. A point where this holds is called an umbilical point. Otherwise, the principal directions are perpendicular. We have that H is the trace and K the determinant of ℓ. Let(g^{ij}) be the inverse of the 2 × 2 matrix (g^{ij}):

$$g^{im}g_{mj} = \delta^i_j.$$

Set $\ell(X_i) = \ell^j_i$ Xj, then since $k_{ij} = g(\ell(X_i), X_j) = \ell^m_i g_{mj}$, we find:

$$\ell^j_i = k_{im}g^{mj}.$$

It is customary to say that g raises the index of k and to write the new object $k^j_i = k_{im}g^{mj}$. Here since k_{ij} is symmetric, it is not necessary to keep track of the position of the indices, and hence we write: $\ell^j_i = k^j_i$. In particular, we have:

$$H = \frac{1}{2}k^i_i, \quad K = \frac{\det(k_{ij})}{\det(g_{ij})}. \qquad ...(10)$$

Now, $k^{ij} = g^{im}g^{il}k_{lm}$, and we have

$$|k|^2 = k_{ij}k^{ij} = \mathrm{tr}\,\ell^2 = k_1^2 + k_2^2 = 4H^2 - 2K.$$

Hence, we conclude

$$K = 2H^2 - \frac{1}{2}|k|^2 \qquad ...(11)$$

11.4 Examples

In this section, we use $u^1 = u$, and $u^2 = \upsilon$ in order to simplify the notation.

24.4.1. Planes. Let $U \subset \mathbb{R}^2$ be open, and let $X : U \to \mathbb{R}^3$ be a linear function:

$$X(u, \upsilon) = Au + B\upsilon,$$

with $A, B \in \mathbb{R}^3$ linearly independent. Then X is a plane. After reparametrization, we may assume that A and B are orthonormal. In that case, the first fundamental form is:

$$ds^2 = du^2 + d\upsilon^2.$$

Furthermore, $|A \times B| = 1$, and $N = A \times B$ is constant, hence k = 0. In particular, all the points of X are planar, and we have for the mean and Gauss curvatures: H = K = 0.

It is of interest to note that if all the points of a parametric surface are planar, then X(U) is contained in a plane. We will later prove a stronger result: X has a reparametrization which is linear.

Proposition 5. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and suppose that its second fundamental form k = 0. Then, there is a fixed vector A and a constant b such that $X \cdot A = b$, i.e., X is contained in a plane.

Proof. Let A be the unit normal N of X. Let $1 \le i \le 2$, and note that N_i is tangential. Indeed, $N \cdot N = 1$, and differentiating along u^i, we get $N \cdot N_i = 0$. However, since k = 0 it follows from (2.6) that $N_i \cdot X_j = -k_{ij} = 0$. Thus, $N_i = 0$ for i = 1, 2, and we conclude that N is constant. Consequently, $(X \cdot N)_i = X_i \cdot N = 0$, and $X \cdot N$ is also constant, which proves the proposition.

11.4.2. Spheres. Let $U = (0, \pi) \times (0, 2\pi) \subset \mathbb{R}^2$, and let $X : U \to \mathbb{R}^2$ be given by:

$$X(u, v) = (\sin u \cos v, \sin u \sin v, \cos u).$$

The surface X is a parametric representation of the unit sphere. A straightforward calculation shows that the first fundamental form is:

$$ds^2 = du^2 + \sin^2 u\, dv^2,$$

and the unit normal is N = X. Thus, $N_i = X_i$, and consequently $k_{ij} = -N_i \cdot X_j = -X_i \cdot X_j = -g_{ij}$, i.e., k = –g. In particular, the principal curvatures are both equal to –1 and all the points are umbilical. We have for the mean and Gauss curvatures:

$$H = -1, \quad K = 1$$

Proposition 6. Let $X : U \to \mathbb{R}^3$ be a parametric surface and suppose that all the points of X are umbilical. Then, X(U) is either contained in a plane or a sphere.

Proof. By hypothesis, we have

$$N_i = \lambda X_i. \qquad \text{...(12)}$$

We first show that is a constant. Differentiating, we get $N_{ij} = \lambda_j X_i + \lambda X_{ij}$. Interchanging i and j, subtracting these two equations, and taking into account $N_{ij} - N_{ji} = X_{ij} - X_{ji} = 0$, we obtain $\lambda_i X_j - \lambda_j X_i = 0$, e.g.,

$$\lambda_1 X_2 - \lambda_2 X_1 = 0.$$

Since X_1 and X_2 are linearly independent, we conclude that $\lambda_1 = \lambda_2 = 0$ and it follows that λ is constant. Now, if $\lambda = 0$ then all points are planar, and by Proposition 6, X is contained in a plane. Otherwise, let $A = X - \lambda^{-1} N$, then A is constant:

$$A_i = X_i - \lambda^{-1} N_i = 0,$$

and $|X - A| = |\lambda|^{-1}$ is also constant, hence X is contained in a sphere.

11.4.3. Ruled Surfaces. A ruled surface is a parametric surface of the form:

$$X(u, v) = \gamma(u) + vY(u)$$

Notes

for a curve $\gamma:[a,b]\to\mathbb{R}^3$, and a vector field $Y:[a,b]\to\mathbb{R}^3$ along γ. The curve γ is the directrix, and the lines $\gamma(u)+tY(u)$ for u fixed are the generators of X. We may assume that Y is a unit vector field. Provided $\dot{Y}\neq 0.$ We will also assume that $\dot{Y}\neq 0.$ In this case, it is possible to arrange by reparametrization that $\dot{\gamma}\cdot\dot{Y}=0,$ in which case γ is said to be a line of striction. Indeed, if this is not the case, then we can set $\phi=\left(\dot{\gamma}\cdot\dot{Y}\right)/\left|\dot{Y}\right|^2$, and note that the curve

$$\alpha = \gamma + \phi Y$$

lies on the surface X, and satisfies $\dot{\alpha}\cdot\dot{Y}=0.$ Consequently, the surface:

$$\tilde{X}(s,t)=\alpha(s)+tY(s)$$

is a reparametrization of X. Furthermore, there is only one line of striction on X. Indeed, if β and γ are two lines of striction, then since both β is a curve on X we may write β = γ + ϕY for some function ϕ and consequently:

$$\dot{\beta}=\dot{\gamma}+\dot{\phi}Y+\phi\dot{Y}.$$

Taking inner product with $\dot{Y}$ and using the fact that Y is a unit vector, we obtain $\phi\left|\dot{Y}\right|^2=0$ which implies that ϕ = 0 and thus, β = γ.

We have $X_u=\dot{\gamma}+_v\dot{Y}, X_v=Y,$ and X_{vv} = 0. Thus, the first fundamental is:

$$\left(g_{ij}\right)=\begin{pmatrix}1+v^2\left|\dot{Y}\right|^2 & \dot{\gamma}\cdot Y\\ \dot{\gamma}\cdot Y & 1\end{pmatrix}$$

and

$$\det\left(g_{ij}\right)=1+v^2\left|\dot{Y}\right|^2+\left(\dot{\gamma}\cdot Y\right)^2\geq v^2\left|\dot{Y}\right|^2.$$

Hence, dX is non-singular except possibly on the line of striction. Furthermore, $k_{vv} = N \cdot X_{vv} = 0$, hence $\det\left(k_{ij}\right)=-k_{uv}^2$ and if $\det\left(k_{ij}\right)=0$ then $N_v\cdot X_u=N_v\cdot X_v=0,$ is constant along generators. We have proved the following proposition.

Proposition 7. Let X be a ruled surface. Then X has non-positive Gauss curvature K ≤ 0, and K(u) = 0 if and only if N is constant along the generator through u.

11.4.3.1. Cylinders. Let $\gamma:[a,b]\to\mathbb{R}^3$ be a planar curve, and A be a unit normal to the plane which contains γ. Define $X:[a,b]\times\mathbb{R}\to\mathbb{R}^3$ by:

$$X(u,v)=\gamma(u)+vA.$$

The surface X is a cylinder. The first fundamental form is:

$$ds^2 = du^2 + dv^2,$$

and we see that for a cylinder dX is always non-singular. After possibly reversing the orientation of A, the unit normal is N = e_2. Clearly, N_v = 0, and $N_u = -ke_1$.

Thus, the second fundamental form is:

$$kdu^2$$

The principal curvatures are 0 and k. We have for the mean and Gauss curvatures:

$$H = -\frac{1}{2}k, \quad K = 0.$$

A surface on which K = 0 is called developable.

11.4.3.2. Tangent Surfaces. Let $\gamma : [a,b] \to \mathbb{R}^3$ be a curve with nonzero curvature $k \neq 0$. Its tangent surface is the ruled surface:

$$X(u,v) = \gamma(u) + v\dot{\gamma}(u).$$

Since $\dot{\gamma} \cdot \ddot{\gamma} = 0$, the curve γ is the line of striction of its tangent surface. We have $X_u = e_1 + vk\, e_2$ and $X_v = e_1$, hence the first fundamental form is:

$$\left(g_{ij}\right) = \begin{pmatrix} 1 + v^2k^2 & 1 \\ 1 & 1 \end{pmatrix}.$$

The unit normal is $N = -e_3$, and clearly $N_v = 0$. Thus,

11.4.3.3. Hyperboloid. Let $\gamma : (0, 2\pi) \to \mathbb{R}^3$ be the unit circle in the $x^1 x^2$-plane: $\gamma(t) = \left(\cos(t), \sin(t), 0\right)$.

Define a ruled surface $X : (0, 2\pi) \times \mathbb{R} \to \mathbb{R}^3$ by:

$$X(u,v) = \gamma(u) + v\left(\dot{\gamma}(u) + e_3\right) = \left(\cos(u) - v\sin(u), \sin(u) + v\cos(u), v\right).$$

Note that $(x^1)^2 + (x^2)^2 - (x^3)^3 = 1$ so that X(U) is a hyperboloid of one sheet. A straightforward calculation gives:

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

and

$$\left|N_v\right|^2 = \frac{2}{1 + 4v^2 + 4v^4}.$$

It follows from Proposition 7 that X has Gauss curvature K < 0.

11.5 Summary

- A parametric surface patch is a smooth mapping:

  $$X : U \to \mathbb{R}^3,$$

  where $U \subset \mathbb{R}^2$ is open, and the Jacobian dX is non-singular.

  Write $X = (x^1, x^2, x^3)$, and each $x^i = x^i(u^1, u^2)$, then the Jacobian has the matrix representation:

  $$dX = \begin{pmatrix} x_1^1 & x_2^1 \\ x_1^2 & x_2^2 \\ x_1^3 & x_2^3 \end{pmatrix}$$

where we have used the notation $f_i = f_{u^i} = \partial f / \partial u^i$. According to the definition, we are requiring that this matrix has rank 2, or equivalently that the vectors $\left(x_2^1, x_1^2, x_1^3\right)$ and $X_2 = \left(x_2^1, x_2^2, x_2^3\right)$ are linearly independent. Another equivalent requirement is that $dX : \mathbb{R}^2 \to \mathbb{R}^3$ is injective.

- Let $X : U \to \mathbb{R}^3$, and $\tilde{X} : \tilde{U} \to \mathbb{R}^3$ be parametric surfaces. We say that $\tilde{X}$ is reparametrization of X if $\tilde{X} = X \circ \phi$, where $\phi : \tilde{U} \to U$ is a diffeomorphism. If ϕ is an orientation-preserving diffeomorphism, then $\tilde{X}$ is an orientation-preserving reparametrization.
- If $Y \in T_u X$, then it can be expressed as a linear combination in X_1 and X_2:

$$Y = y^1 X_1 + y^2 X_2 = \sum_{i=1}^{2} y^i X_i,$$

where $y^i \in \mathbb{R}$ are the components of the vector Y in the basis X_1, X_2 of $T_u X$. We will use the Einstein Summation Convention: every index which appears twice in any product, once as a subscript (covariant) and once as a superscript (contravariant), is summed over its range.

- A vector field along a parametric surface $X : U \to \mathbb{R}^3$, is a smooth mapping $Y : U \to \mathbb{R}^3$.[2] A vector field Y is tangent to X if Y (u) $\in T_u X$ for all u $\in$ U. A vector field Y is normal to X if $Y(u) \perp T_u X$ for all u $\in$ U.
- A symmetric bilinear form on a vector space V is function $B : V \times V \to \mathbb{R}$ satisfying:
  - B(aX + bY,Z) = aB(X,Z) + bB(Y,Z), for all X, Y $\in$ V and a, b $\in$ R.
  - B(X, Y) = B(Y, X), for all X, Y $\in$ V.

  The symmetric bilinear form B is positive definite if B(X, X) $\geq$ 0, with equality if and only if X = 0.

  With any symmetric bilinear form B on a vector space, there is associated a quadratic form Q(X) = B(X, X). Let V and W be vector spaces and let $T : V \to W$ be a linear map. If B is a symmetric bilinear form on W, we can define a symmetric bilinear form T* Q on V by T* Q(X, Y) = Q(TX, TY). We call T* Q the pull-back of Q by T. The map T is then an isometry between the inner-product spaces (V, T* Q) and (W,Q).

11.6 Keywords

Diffeomorphism: A diffeomorphism between open sets U, $V \subset \mathbb{R}^2$ is a map $\phi : U \to V$ which is smooth, one-to-one, and whose inverse is also smooth. If det($d\phi$) > 0, then we sa that ϕ is an orientation-preserving diffeomorphism.

Einstein Summation Convention: every index which appears twice in any product, once as a subscript (covariant) and once as a superscript (contravariant), is summed over its range.

Gauss map: The Gauss map is invariant under orientation-preserving reparametrization.

11.7 Self Assessment

1. Let $X : U \to \mathbb{R}^3$, and $\tilde{X} : \tilde{U} \to \mathbb{R}^3$ be parametric surfaces. We say that $\tilde{X}$ is reparametrization of X if $\tilde{X} = X \circ \phi$, where $\phi : \tilde{U} \to U$ is a ....
2. The tangent space $T_u X$ of the parametric surface .... at $u \in U$ is the 2-dimensional linear subspace of $\mathbb{R}^3$ spanned by the two vectors X_1 and X_2.[1]
3. A .... Y is tangent to X if $Y(u) \in T_u X$ for all $u \in U$. A vector field Y is normal to X if $Y(u) \perp T_u X$ for all $u \in U$.
4. The .... is invariant under orientation-preserving reparametrization.
5. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let $N : U \to \mathbb{S}^2$ be its Gauss map. Let $\tilde{X} = X \circ \phi$ be an orientation-preserving .... of X. Then the Gauss map of $\tilde{X}$ is $N \circ \phi$.

11.8 Review Questions

1. Let $X : U \to \mathbb{R}^3$ and $\tilde{X} : \tilde{U} \to \mathbb{R}^3$ be two parametric surfaces. The angle θ between them is the angle between their unit normals: $\cos\theta = N \cdot \tilde{N}$. Let γ be a regular curve which lies on both X and $\tilde{X}$, and suppose that the angle between X and $\tilde{X}$ is constant along γ. Show that γ is a line of curvature of X if and only if it is a line of curvature of $\tilde{X}$.
2. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let γ be an asymptotic line with curvature $k \neq 0$, and torsion τ. Show that $|\tau| = \sqrt{-K}$
3. Denote by SO(n) the set of orthogonal n × n matrices, and by D(n) the set of n × n diagonal matrices. Let $A : (a,b) \to S^{n\times n}$ be a C^k function, and suppose that A maps into the set of matrices with distinct eigenvalues. Show that there exist C^k functions Q: $(a, b) \to SO(n)$ and $\Lambda : (a, b) \to D(n)$ such that $Q^{-1} AQ = \Lambda$. Conclude the matrix function A has C^k eigenvector fields $e_1, \ldots, e_n : (a,b) \to \mathbb{R}^n, Ae_j = \lambda_j e_j$. Give a counter-example to show that this last conclusion can fail the eigenvalues of A are allowed to coincide.
4. Let $M^{n\times n}$ be the space of all n × n matrices, and let B: $(a, b) \to M^{n\times n}$ be continuously differentiable. Prove that:

$$(\det B)' = \operatorname{tr}(B^* B'),$$

where B^* is the matrix of co-factors of B.

5. Two harmonic surfaces $X, Y : U \to \mathbb{R}^3$ are called conjugate, if they satisfy the Cauchy-Riemann Equations:

$$X_u = Y_v, \quad X_v = -Y_u,$$

where (u, v) denote the coordinates in U. Prove that if X is conformal then Y is also conformal. Let X and Y be conformal conjugate minimal surfaces. Prove that for any t:

$$Z = X \cos t + Y \sin t$$

is also a minimal surface. Show that all the surfaces Z above have the same first fundamental form.

Answers: Self Assessment

1. diffeomorphism
2. $X : U \to \mathbb{R}^3$
3. vector field
4. Gauss map
5. reparametrization

11.9 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 12: Curvature

CONTENTS

Objectives

After studying this unit, you will be able to:

- Discuss the Curvature of plane curves
- Explain the Curvature of a graph
- Define Signed curvature
- Describe Curvature of space curves
- Explain Curves on surfaces

Introduction

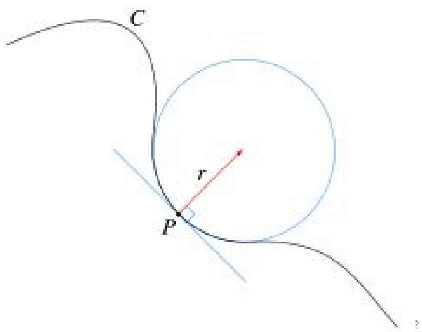
In mathematics, curvature refers to any of a number of loosely related concepts in different areas of geometry. Intuitively, curvature is the amount by which a geometric object deviates from being flat, or straight in the case of a line, but this is defined in different ways depending on the context. There is a key distinction between extrinsic curvature, which is defined for objects embedded in another space (usually a Euclidean space) in a way that relates to the radius of curvature of circles that touch the object, and intrinsic curvature, which is defined at each point in a Riemannian manifold. This unit deals primarily with the first concept. The canonical example of extrinsic curvature is that of a circle, which everywhere has curvature equal to the reciprocal of its radius. Smaller circles bend more sharply, and hence have higher curvature. The curvature of a smooth curve is defined as the curvature of its osculating circle at each point.

In a plane, this is a scalar quantity, but in three or more dimensions it is described by a curvature vector that takes into account the direction of the bend as well as its sharpness. The curvature of more complex objects (such as surfaces or even curved n-dimensional spaces) is described by more complex objects from linear algebra, such as the general Riemann curvature tensor. The remainder of this article discusses, from a mathematical perspective, some geometric examples of curvature: the curvature of a curve embedded in a plane and the curvature of a surface in Euclidean space. See the links below for further reading.

12.1 Curvature of Plane Curves

Cauchy defined the center of curvature *C* as the intersection point of two infinitely close normals to the curve, the radius of curvature as the distance from the point to *C*, and the curvature itself as the inverse of the radius of curvature.

Let *C* be a plane curve (the precise technical assumptions are given below). The curvature of *C* at a point is a measure of how sensitive its tangent line is to moving the point to other nearby points. There are a number of equivalent ways that this idea can be made precise.



One way is geometrical. It is natural to define the curvature of a straight line to be identically zero. The curvature of a circle of radius *R* should be large if *R* is small and small if *R* is large. Thus, the curvature of a circle is defined to be the reciprocal of the radius:

$$k=\frac{1}{R}.$$

Given any curve C and a point P on it, there is a unique circle or line which most closely approximates the curve near P, the osculating circle at P. The curvature of C at P is then defined to be the curvature of that circle or line. The radius of curvature is defined as the reciprocal of the curvature.

Another way to understand the curvature is physical. Suppose that a particle moves along the curve with unit speed. Taking the time s as the parameter for C, this provides a natural parametrization for the curve. The unit tangent vector T (which is also the velocity vector, since the particle is moving with unit speed) also depends on time. The curvature is then the magnitude of the rate of change of T. Symbolically,

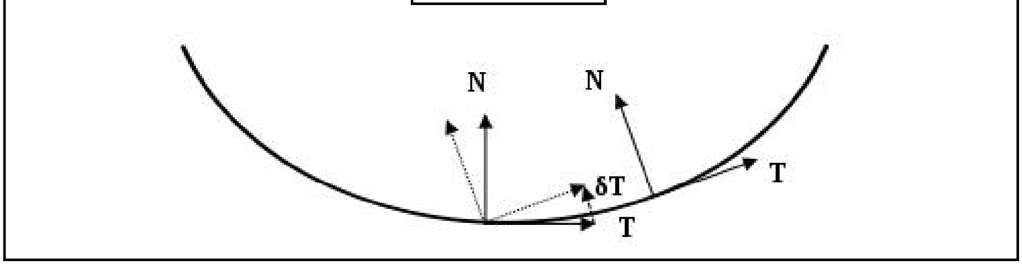
$$k=\left\|\frac{dT}{ds}\right\|.$$

This is the magnitude of the acceleration of the particle. Geometrically, this measures how fast the unit tangent vector to the curve rotates. If a curve keeps close to the same direction, the unit tangent vector changes very little and the curvature is small; where the curve undergoes a tight turn, the curvature is large.

Notes

These two approaches to the curvature are related geometrically by the following observation. In the first definition, the curvature of a circle is equal to the ratio of the angle of an arc to its length. Likewise, the curvature of a plane curve at any point is the limiting ratio of dθ, an infinitesimal angle (in radians) between tangents to that curve at the ends of an infinitesimal segment of the curve, to the length of that segment ds, i.e., dθ/ds. If the tangents at the ends of the segment are represented by unit vectors, it is easy to show that in this limit, the magnitude of the difference vector is equal to dθ, which leads to the given expression in the second definition of curvature.

Figure 12.1



In figure, *T and N vectors at two points on a plane curve, a translated version of the second frame (dotted), and the change in T: äT′. äs is the distance between the points. In the limit* $\frac{dT}{ds}$ *will be in the direction N and the curvature describes the speed of rotation of the frame.*

Suppose that C is a twice continuously differentiable immersed plane curve, which here means that there exists parametric representation of C by a pair of functions ã(t) = (x(t), y(t)) such that the first and second derivatives of x and y both exist and are continuous, and

$$\|\gamma'\|^2 = x'(t)^2 + y'(t)^2 \neq 0$$

throughout the domain. For such a plane curve, there exists a reparametrization with respect to arc length s. This is a parametrization of C such that

$$\|\gamma'\|^2 = x'(s)^2 + y'(s)^2 = 1.$$

The velocity vector T(s) is the unit tangent vector. The unit normal vector N(s), the curvature ê(s), the oriented or signed curvature k(s), and the radius of curvature R(s) are given by

$$T(s) = \gamma'(s),\ \ T'(s) = k(s)N(s),\ \ k(s) = \|T'(s)\| = \|\gamma''(s)\| = |k(s)|,\ \ R(s) = \frac{1}{k(s)}.$$

Expressions for calculating the curvature in arbitrary coordinate systems are given below.

12.1.1 Signed Curvature

The sign of the signed curvature k indicates the direction in which the unit tangent vector rotates as a function of the parameter along the curve. If the unit tangent rotates counterclockwise, then $k > 0$. If it rotates clockwise, then $k < 0$.

The signed curvature depends on the particular parametrization chosen for a curve. For example the unit circle can be parametrised by (cos(θ), sin(θ)) (counterclockwise, with $k > 0$), or by (cos(–θ), sin(–θ)) (clockwise, with $k < 0$). More precisely, the signed curvature depends only on the choice of orientation of an immersed curve. Every immersed curve in the plane admits two possible orientations.

12.1.2 Local Expressions

Notes

For a plane curve given parametrically in Cartesian coordinates as ã(t) = (x(t),y(t)), the curvature is

$$k = \frac{|x'y'' - y'x''|}{(x'^2 + y'^2)^{3/2}},$$

where primes refer to derivatives with respect to parameter t. The signed curvature k is

$$k = \frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}}.$$

These can be expressed in a coordinate-independent manner via

$$k = \frac{\det(\gamma', \gamma'')}{\|\gamma'\|^3}, \quad k = \frac{|\det(\gamma', \gamma'')|}{\|\gamma'\|^3}$$

12.1.3 Curvature of a Graph

For the less general case of a plane curve given explicitly as y = f(x), and now using primes for derivatives with respect to coordinate x , the curvature is

$$k = \frac{|y''|}{(1 + y'^2)^{3/2}},$$

and the signed curvature is

$$k = \frac{y''}{(1 + y'^2)^{3/2}}.$$

This quantity is common in physics and engineering; for example, in the equations of bending in beams, the 1D vibration of a tense string, approximations to the fluid flow around surfaces (in aeronautics), and the free surface boundary conditions in ocean waves. In such applications, the assumption is almost always made that the slope is small compared with unity, so that the approximation:

$$k \approx \left| \frac{d^2 y}{dx^2} \right|$$

may be used. This approximation yields a straightforward linear equation describing the phenomenon, which would otherwise remain intractable.

If a curve is defined in polar coordinates as r(θ), then its curvature is

$$k(\theta) = \frac{|r^2 + 2r'^2 - rr''|}{(r^2 + r'^2)^{3/2}}$$

where here the prime now refers to differentiation with respect to θ.

Example: Consider the parabola $y = x^2$. We can parametrize the curve simply as $\gamma(t) = (t, t^2) = (x, y)$. If we use primes for derivatives with respect to parameter t , then

$$x' = 1, \quad x'' = 0, \quad y' = 2t, \quad y'' = 2.$$

Substituting and dropping unnecessary absolute values, get

$$k(t) = \left|\frac{x'y'' - y'x''}{(x'^2 + y'^2)^{3/2}}\right| = \frac{1.2 - (2t)(0)}{(1+(2t)2)^{3/2}} = \frac{2}{(1+4t^2)^{3/2}}.$$

12.2 Curvature of Space Curves

As in the case of curves in two dimensions, the curvature of a regular space curve C in three dimensions (and higher) is the magnitude of the acceleration of a particle moving with unit speed along a curve. Thus if $\gamma(s)$ is the arc length parametrization of C then the unit tangent vector T(s) is given by

$$T(s) = \gamma'(s)$$

and the curvature is the magnitude of the acceleration:

$$k(s) = \|T'(s)\| = \|\gamma''(s)\|.$$

The direction of the acceleration is the unit normal vector N(s), which is defined by

$$N(s) = \frac{T'(s)}{\|T'(s)\|}.$$

The plane containing the two vectors T(s) and N(s) is called the osculating plane to the curve at $\gamma(s)$. The curvature has the following geometrical interpretation. There exists a circle in the osculating plane tangent to $\gamma(s)$ whose Taylor series to second order at the point of contact agrees with that of $\gamma(s)$. This is the osculating circle to the curve. The radius of the circle R(s) is called the radius of curvature, and the curvature is the reciprocal of the radius of curvature:

$$k(s) = \frac{1}{R(s)}.$$

The tangent, curvature, and normal vector together describe the second-order behavior of a curve near a point. In three-dimensions, the third order behavior of a curve is described by a related notion of torsion, which measures the extent to which a curve tends to perform a corkscrew in space. The torsion and curvature are related by the Frenet–Serret formulas (in three dimensions) and their generalization (in higher dimensions).

12.2.1 Local Expressions

For a parametrically defined space curve in three-dimensions given in Cartesian coordinates by ã(t) = (x(t),y(t),z(t)), the curvature is

$$k = \frac{\sqrt{(z''y' - y''z')^2 + (x''z' - z''x')^2 + (y''x' - x''y')^2}}{(x'^2 + y'^2 + z'^2)^{3/2}}.$$

where the prime denotes differentiation with respect to time t. This can be expressed independently of the coordinate system by means of the formula

$$k = \frac{|\gamma' \times \gamma''|}{|\gamma'|^3}$$

where × is the vector cross product. Equivalently,

$$k = \frac{\sqrt{\det((\gamma',\gamma'')^t(\gamma',\gamma''))}}{\|\gamma'\|^3}.$$

Notes

Here the t denotes the matrix transpose. This last formula is also valid for the curvature of curves in a Euclidean space of any dimension.

12.2.2 Curvature from Arc and Chord Length

Given two points P and Q on C, let s(P,Q) be the arc length of the portion of the curve between P and Q and let d(P,Q) denote the length of the line segment from P to Q. The curvature of C at P is given by the limit

$$k(P) = \lim_{Q \to P} \sqrt{\frac{24(s(P,Q) - d(P,Q))}{s(P,Q)^3}}$$

where the limit is taken as the point Q approaches P on C. The denominator can equally well be taken to be $d(P,Q)^3$. The formula is valid in any dimension. Furthermore, by considering the limit independently on either side of P, this definition of the curvature can sometimes accommodate a singularity at P. The formula follows by verifying it for the osculating circle.

12.3 Curves on Surfaces

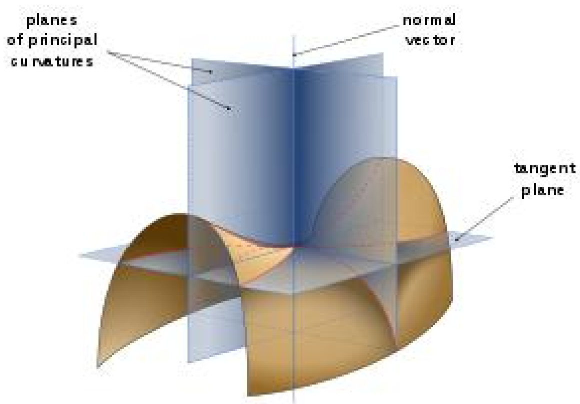
When a one dimensional curve lies on a two dimensional surface embedded in three dimensions R^3, further measures of curvature are available, which take the surface's unit-normal vector, u into account. These are the normal curvature, geodesic curvature and geodesic torsion. Any non-singular curve on a smooth surface will have its tangent vector T lying in the tangent plane of the surface orthogonal to the normal vector. The normal curvature, k_n, is the curvature of the curve projected onto the plane containing the curve's tangent T and the surface normal u; the geodesic curvature, k_g, is the curvature of the curve projected onto the surface's tangent plane; and the geodesic torsion (or relative torsion), $\hat{o}_r$, measures the rate of change of the surface normal around the curve's tangent.

Let the curve be a unit speed curve and let t = u × T so that T, u, t form an orthonormal basis: the Darboux frame. The above quantities are related by:

$$\begin{pmatrix} T' \\ t' \\ u' \end{pmatrix} = \begin{pmatrix} o & K_g & K_n \\ -K_g & o & T_n \\ -K_n & -T_\gamma & o \end{pmatrix} \begin{pmatrix} T \\ t \\ u \end{pmatrix}$$

Figure 12.2: Saddle Surface with Normal Planes in Directions of Principal Curvatures

Principal curvature



Principal Curvature

All curves with the same tangent vector will have the same normal curvature, which is the same as the curvature of the curve obtained by intersecting the surface with the plane containing **T** and **u**. Taking all possible tangent vectors then the maximum and minimum values of the normal curvature at a point are called the **principal curvatures**, k_1 and k_2, and the directions of the corresponding tangent vectors are called **principal directions**.

This is explained in detail in Unit 27 of this book.

12.4 Summary

- Curvature refers to any of a number of loosely related concepts in different areas of geometry. Intuitively, curvature is the amount by which a geometric object deviates from being flat, or straight in the case of a line, but this is defined in different ways depending on the context. There is a key distinction between extrinsic curvature, which is defined for objects embedded in another space (usually a Euclidean space) in a way that relates to the radius of curvature of circles that touch the object, and intrinsic curvature, which is defined at each point in a Riemannian manifold. This article deals primarily with the first concept. Cauchy defined the center of curvature C as the intersection point of two infinitely close normals to the curve, the radius of curvature as the distance from the point to C, and the curvature itself as the inverse of the radius of curvature.
- Let C be a plane curve (the precise technical assumptions are given below). The curvature of C at a point is a measure of how sensitive its tangent line is to moving the point to other nearby points. There are a number of equivalent ways that this idea can be made precise. The sign of the signed curvature k indicates the direction in which the unit tangent vector rotates as a function of the parameter along the curve. If the unit tangent rotates counterclockwise, then $k > 0$. If it rotates clockwise, then $k < 0$. The signed curvature depends on the particular parametrization chosen for a curve. For example the unit circle can be parametrised by $(\cos(\theta), \sin(\theta))$ (counterclockwise, with $k > 0$), or by $(\cos(-\theta), \sin(-\theta))$ (clockwise, with $k < 0$). As in the case of curves in two dimensions, the curvature of a regular space curve C in three dimensions (and higher) is the magnitude of the acceleration of a particle moving with unit speed along a curve. The tangent, curvature, and normal vector together describe the second-order behavior of a curve near a point. In three-dimensions, the third order behavior of a curve is described by a related notion of torsion, which measures the extent to which a curve tends to perform a corkscrew in space.

12.5 Keywords

Curvature: Curvature refers to any of a number of loosely related concepts in different areas of geometry.

Extrinsic curvature: Extrinsic curvature, which is defined for objects embedded in another space (usually a Euclidean space) in a way that relates to the radius of curvature of circles that touch the object

Intrinsic curvature: Intrinsic curvature, which is defined at each point in a Riemannian manifold. This article deals primarily with the first concept.

Notes

12.6 Self Assessment

1. .... refers to any of a number of loosely related concepts in different areas of geometry.
2. ...., which is defined at each point in a Riemannian manifold.
3. .... defined the center of curvature *C* as the intersection point of two infinitely close normals to the curve, the radius of curvature as the distance from the point to *C*, and the curvature itself as the inverse of the radius of curvature.
4. The .... and normal vector together describe the second-order behavior of a curve near a point.

12.7 Review Question

1. Discuss the concept of Curvature of plane curves.
2. Explain the Curvature of a graph.
3. Define Signed curvature and discuss it in detail.
4. Describe Curvature of space curves.
5. Explain Curves on surfaces.

Answers: Self Assessment

1. Curvature
2. Intrinsic curvature
3. Cauchy
4. tangent, curvature

12.8 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Richa Nandra, Lovely Professional University

Unit 13: Lines of Curvature

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define lines of curvature
- Explain the examples of lines of curvature
- Describe the surface area and Bernstein's theorem

Introduction

In last unit, you have studied about curvature. In general, there are two important types of curvature: extrinsic curvature and intrinsic curvature. The extrinsic curvature of curves in two- and three-space was the first type of curvature to be studied historically, culminating in the Frenet formulas, which describe a space curve entirely in terms of its "curvature," torsion, and the initial starting point and direction. This unit will explains the concept of lines of curvature.

13.1 Lines of Curvature

Definition 1. A curve γ on a parametric surface X is called a line of curvature if $\dot{\gamma}$ is a principal direction.

The following proposition, due to Rodriguez, characterizes lines of curvature as those curves whose tangents are parallel to the tangent of their spherical image under the Gauss map.

Notes

Proposition 1. Let γ be a curve on a parametric surface X with unit normal N, and let $\beta = N \circ \gamma$ be its spherical image under the Gauss map. Then γ is a line of curvature if and only if

$$\dot{\beta} + \lambda\dot{\gamma} = 0. \qquad ...(1)$$

Proof. Suppose that (1) holds, then we have:

$$\partial_{\dot{\gamma}} N + \lambda\dot{\gamma} = 0.$$

Let ℓ be the linear transformation on $T_u X$ associated with k. Then, we have for every $Y \in T_u X$:

$$g\left(\ell(\dot{\gamma}), Y\right) = k(\dot{\gamma}, Y) = -\partial_{\dot{\gamma}} N \cdot Y = \lambda g(\lambda\dot{\gamma}, Y).$$

Thus, $\ell\left(\dot{\gamma}\right) = \lambda\dot{\gamma}$, and $\dot{\gamma}$ is a principal direction. The proof of the converse is similar.

It is clear from the proof that λ in (1) is the associated principal curvature. The coordinate curves of a parametric surface X are the two family of curves $\gamma_c(t) = X(t, c)$ and $\beta_c(t) = X(c, t)$. A surface is parametrized by lines of curvature if the coordinate curves of X are lines of curvature. We will now show that any non-umbilical point has a neighborhood in which the surface can be reparametrized by lines of curvature. We first prove the following lemma which is also of independent interest.

Lemma 1. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let Y_1 and Y_2 be linearly independent vector fields. The following statements are equivalent:

1. Any point $u_0 \in U$ has a neighborhood U_0 and a reparametrization $\phi : V_0 \to U_0$ such that if $\tilde{X} = X \circ \phi$ then $\tilde{X}_i = Y_i \circ \phi$.
2. $\left[Y_1, Y_2\right] = 0.$

Proof. Suppose that (1) holds. Then Equation shows that $\left[\tilde{X}_1, \tilde{X}_2\right] = 0.$ However, since the commutator is invariant under reparametrization, it follows that $\left[Y_1, Y_2\right] = 0.$

Conversely, suppose that $\left[Y_1, Y_2\right] = 0.$ Express $X_i = a_i^j Y_j$ and $Y_i = b_i^j X_j$, note that $\left(b_i^j\right)$ is the inverse of $\left(a_i^j\right)$. We now calculate:

$$\begin{aligned} 0 &= \left[X_i, X_j\right] \\ &= \left[a_i^k Y_k, a_j^l Y_l\right] \\ &= \left(a_i^l \partial Y_l a_j^k - a_j^l \partial Y_l a_i^k\right) Y_k + a_i^k a_j^l \left[Y_k, Y_l\right] \\ &= \left(a_i^l b_l^m \partial_m a_j^k - a_j^l b_l^m \partial_m a_i^k\right) Y_k \\ &= \left(\partial_i a_j^k - \partial_j a_i^k\right) Y_k. \end{aligned}$$

Since Y_1 and Y_2 are linearly independent, we conclude that:

$$\partial_i a_j^k - \partial_j a_i^k = 0. \qquad ...(2)$$

Now, fix $1 \le k \le 2$, and consider the over-determined system:

$$\frac{\partial \tilde{u}^k}{\partial u^i} = a_i^k, \quad i = 1,2.$$

The integrability condition for this system is exactly (2), hence there is a solution in a neighborhood of u_0. Furthermore, since the Jacobian of the map $\psi(u^1,u^2) = (\tilde{u}^1,\tilde{u}^2)$ is $d\psi = (a_i^k)$, and $\det(a_i^k) \neq 0$, it follows from the inverse function theorem, that perhaps on yet a smaller neighborhood, ψ is a diffeomorphism. Let $\phi = \psi^{-1}$, then ϕ is a diffeomorphism on a neighborhood V_0 of $\psi(u_0)$, and if we set $\tilde{X} = X \circ \phi$, then:

$$\tilde{X}_i = X_j \frac{\partial u^j}{\partial u^i} = X_j b_i^j = Y_i.$$

Proposition 2. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let Y_1 and Y_2 be linearly independent vector fields. Then for any point $u_0 \in U$ there is a neighborhood of u_0 and a reparametrization $\tilde{X} = X \circ \phi$ such that $\tilde{X}_i = f_i Y_i \circ \phi$ for some functions f_i.

Proof. By Lemma 1 is suffices to show that there are function f_i such that f_1 Y_1 and f_2 Y_2 commute. Write $[Y_1, Y_2] = a_1 Y_1 - a_2 Y_2$, and compute:

$$[f_1Y_1, f_2Y_2] = f_1f_2(a_1Y_1 - a_2Y_2) + f_1(\partial Y_1 f_2)Y_2 - f_2(\partial Y_2 f_1)Y_1.$$

Thus, the commutator $[f_1Y_1, f_2Y_2]$ vanishes if and only if the following two equations are satisfied:

$$\partial Y_2 f_1 - a_1 f_1 = 0$$

$$\partial Y_1 f_2 - a_2 f_2 = 0.$$

We can rewrite those as:

$$\partial Y_2 \log f_1 = a_1$$

$$\partial Y_1 \log f_2 = a_2.$$

Each of those equation is a linear first-order partial differential equation, and can be solved for a positive solution in a neighborhood of u_0.

In a neighborhood of a non-umbilical point, the principal directions define two orthogonal unit vector fields. Thus, we obtain the following Theorem as a corollary to the above proposition.

Theorem 1. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let u_0 be a non-umbilical point. Then there is neighborhood U_0 of u_0 and a diffeomorphism $\phi : \tilde{U}_0 \to U_0$ such that $\tilde{X} = X \circ \phi$ is parametrized by lines of curvature.

If X is parametrized by lines of curvature, then the second fundamental form has the coordinate representation:

$$(k_{ij}) = \begin{pmatrix} k_1 g_{11} & 0 \\ 0 & k_2 g_{22} \end{pmatrix}$$

Definition 2. A curve γ on a parametric surface X is called an asymptotic line if it has zero normal curvature, i.e., $k(\dot{\gamma}, \dot{\gamma}) = 0$.

The term asymptotic stems from the fact that those curve have their tangent $\dot{\gamma}$ along the asymptotes of the Dupin indicatrix, the conic section $k_{ij}\xi^i\xi^j = 1$ in the tangent space. Since the Dupin indicatrix has no asymptotes when K > 0, we see that the Gauss curvature must be non-positive along any asymptotic line.

The following Theorem can be proved by the same method as used above to obtain Theorem 1.

Theorem 2. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let u_0 be a hyperbolic point. Then there is neighborhood U_0 of u_0 and a diffeomorphism $\phi : \tilde{U}_0 \to U_0$ such that $\tilde{X} = X \circ \phi$ is parametrized by asymptotic lines.

13.2 Examples

A surface of revolution is a parametric surface of the form:

$$X(u,v) = (f(u)\cos(v), f(u)\sin(v), g(u)),$$

where $(f(t), g(t))$ is a regular curve, called the generator, which satisfies $f(t) \neq 0$. Without loss of generality, we may assume that $f(t) > 0$. The curves

$$\gamma_v(t) = (f(t)\cos(v), f(t)\sin(v), g(t)), \quad \text{v fixed.}$$

are called meridians and the curves

$$\beta_u(t) = (f(u)\cos(t), f(u)\sin(t), g(u)), \quad \text{u fixed.}$$

are called parallels. Note that every meridian is a planar curve congruent to the generator and is furthermore also a normal section, and every parallel is a circle of radius f(u). It is not difficult to see that parallels and meridians are lines of curvature. Indeed, let γ_v be a meridian, then choosing as in the paragraph following the correct orientation in the plane of γ_v, its spherical image under the Gauss map is $\sigma_v = N \circ \gamma_v = e_2$, and by the Frenet equations, $\dot{\sigma}_v = -ke_1 = -k\dot{\gamma}_v$. Thus, using Proposition 1 and the comment immediately following it, we see that γ_v is a line of curvature with associated principal curvature k. Since the parallels β_u are perpendicular to the meridians γ_v, it follows immediately that they are also lines of curvature. We derive this also follows from Proposition 1 and furthermore obtain the associated principal curvature. A straightforward computation gives that the spherical image of β_u under the Gauss map is:

$$\tau_u = N \circ \beta_u = c\beta_u + B$$

where $B \in \mathbb{R}^3$ and $c \in \mathbb{R}$ are constants. Thus, $\dot{\tau}_u = c\dot{\beta}_u$ and β_u is a line of curvature with associated principal curvature c.

The plane, the sphere, the cylinder, and the hyperboloid are all surfaces of revolution. We discuss one more example.

The catenoid is the parametric surface of revolution obtained from the generating curve $(\cosh(t), t)$:

$$X(u,v) = (\cosh(u), \cos(v), \cosh(u), \sin(v), u).$$

The normal N is easily calculated:

$$N(u,v)=\left(\frac{-\cos(v)}{\cosh(u)},\frac{-\sin(v)}{\cosh(u)},\frac{\sinh(u)}{\cosh(u)}\right)$$

If $\gamma_v(t)$ is a meridian, then $\sigma_v(t) = N(t, v)$ is its spherical image under the Gauss map, and differentiating with respect to t, we get the principal curvature associated with meridians: $k(u, v) = -1/\cosh(u)$. Similarly, the principal curvature associated with parallels is: $1/\cosh(u)$. Thus, we conclude that

$$H=0,\quad K=-\frac{1}{\cosh(u)^2}.$$

Definition 3. A parametric surface X is minimal if it has vanishing mean curvature $H = 0$.

For example, the catenoid is a minimal surface. The justification for the terminology will be given in the next section.

Proposition 3. Let X be a minimal surface. Then X has non-positive Gauss curvature $K \le 0$, and $K(u) = 0$ if and only if u is a planar point.

We will set out to construct a large class of minimal surfaces. We will use the Weierstrass Representation.

Definition 4. A parametric surface X is conformal if the first fundamental form satisfies $g_{11} = g_{22}$ and $g_{12} = 0$. A parametric surface X is harmonic if $\Delta X = X_{11} + X_{22} = 0$.

Proposition 4. Let $X : U \to \mathbb{R}^3$ be a parametric surface which is both conformal and harmonic. Then X is a minimal surface.

Proof. We can write the first fundamental form $\left(g_{ij}\right)$, its inverse $\left(g^{ij}\right)$, and the second fundamental form $\left(k^{ij}\right)$ as:

$$\left(g_{ij}\right)=\begin{pmatrix}\lambda & 0\\ 0 & \lambda\end{pmatrix},\left(g^{ij}\right)=\begin{pmatrix}\lambda^{-1} & 0\\ 0 & \lambda^{-1}\end{pmatrix},\left(k_{ij}\right)=\begin{pmatrix}X_{11}\cdot N & X_{12}\cdot N\\ X_{12}\cdot N & X_{22}\cdot N\end{pmatrix}.$$

Thus, the mean curvature vanishes:

$$H=g^{ij}k_{ij}=\lambda^{-1}\left(X_{11}+X_{22}\right)\cdot N=0.$$

In order to construct parametric surfaces which are both conformal and harmonic, we will use complex analysis in the domain U. Let $\zeta = u+iv$ where i denotes $\sqrt{-1}$, and let $f(\zeta)$ and $h(\zeta)$ be two complex analytic functions on U. Define

$$F_1 = f^2 - h^2,\quad F_2=i\left(f^2+h^2\right),\quad F_3 = 2fh.$$

We have:

$$\left(F_1\right)^2+\left(F_2\right)^2+\left(F_3\right)^2=0.$$

If we write $F_j = \xi_j + in_j$, then this can be written as:

$$\sum_{j=1}^{3}\left[\left(\xi_j\right)^2-\left(n_j\right)^2\right]^2+2i\sum_{j=1}^{3}\xi_j n_j=0.$$

Notes

Now, in any simply connected subset of U, we can always find analytic functions $G_j = x_j + iy_j$ satisfying $(G_j)_\zeta = F_j$. We let $X = (x_1, x_2, x_3)$. Then X is conformal and harmonic. Indeed, x_j being the real parts of complex analytic functions, are harmonic, and hence X is harmonic. Furthermore, we have $(x_j)_u = \xi_j$, and by the Cauchy-Riemann equations $(x_j)_v = -(y_j)_u = -n_j$. Thus, we see that

$$X_u \cdot X_u - X_v \cdot X_v = \sum_{j=1}^{3}\left[(\xi_j)^2 - (\eta_j)^2\right]^2 = 0,$$

and

$$X_u \cdot X_v = -\sum_{j=1}^{3} \xi_j \eta_j = 0,$$

and hence, X is conformal.[3] Since X is real analytic, the zeroes of $\det(X_i \cdot X_j)$ are isolated. Removing the set Z of those zeroes from U, we get that $X : U \setminus Z \to \mathbb{R}^3$ is a harmonic and conformal parametric surface, hence, X is a minimal surface[4].

If we carry out this procedure starting with the complex analytic functions $f(\zeta) = 1$ and $h(\zeta) = 1/\zeta$, then X is another parametrization of the catenoid.

13.3 Surface Area

In this section, we will give interpretations of the Gauss curvature and the mean curvature. Both of these involve the concept of surface area. Before introducing the definition, we first prove a proposition which will show that the definition is reparametrization invariant.

Proposition 5. Let $X : U \to \mathbb{R}^3$ be a parametric surface with first fundamental form (g_{ij}), and $V \subset U$. Let $\tilde{X} : \tilde{U} \to \mathbb{R}^3$ be a reparametrization of X, let $\tilde{V} = \phi^{-1}(V)$, and let $(\tilde{g}_{ij})$ be the coordinate representation of the first fundamental form of $\tilde{X}$. Then, we have:

$$\int_{\tilde{V}} \sqrt{\det(\tilde{g}_{ij})}\, d\tilde{u}^1\, d\tilde{u}^2 = \int_V \sqrt{\det(g_{ij})}\, du^1\, du^2. \qquad ...(3)$$

Proof. Now, we have

$$\sqrt{\det(\tilde{g}_{ij})} = \sqrt{\det(g_{ij})}\left|\det(\phi^i_j)\right|$$

where $\phi^i_j = \partial u^i / \partial \tilde{u}^j$. Thus, for any open subset $V \subset U$, and $\tilde{V} = \phi^{-1}(V)$, we have:

$$\int_{\tilde{V}} \sqrt{\det(\tilde{g}_{ij})}\, d\tilde{u}^1\, d\tilde{u}^2 = \int_{\tilde{V}} \sqrt{\det(g_{ij})}\left|\det(\phi^i_j)\right| d\tilde{u}^1\, d\tilde{u}^2 = \int_V \sqrt{\det(g_{ij})}\, du^1\, du^2$$

Thus, the integral on the right-hand side of (3) is reparametrization invariant. This justifies the following definition.

[3] Of course, $Y = (y_1, y_2, y_3)$ is also conformal.

[4] X is also said to be a branched minimal surface on U. The zeroes of $\det(g_{ij})$ are called branched points.

Definition 5. Let $X : U \to \mathbb{R}^3$ be a parametric surface and let $\left(g_{ij}\right)$ be its first fundamental form. The surface area element of X is:

$$dA = \sqrt{\det\left(g_{ij}\right)}\, du^1 du^2.$$

If $V \subset U$ is open then the surface area of X over V is:

$$AX(V) = \int_V dA = \int_V \sqrt{\det\left(g_{ij}\right)}\, du^1\, du^2 \qquad ...(4)$$

By Proposition 5, the surface area of X over V is reparametrization invariant, and we can thus speak of the surface area of X(V).

Definition 6. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let $V \subset U$ be open. The total curvature of X over V is:

$$K_X(V) = \int_V K\, dA.$$

It is easy to show, as in the proof of Proposition 5 that the total curvature of X over V is invariant under reparametrization. We now introduce the signed surface area, a variant of Definition 5 which allows for smooth maps Y into a surface X, with Jacobian dY not necessarily everywhere non-singular, and which also accounts for multiplicity.

Definition 7. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let $X : U \to X(U)$ be a smooth map. Define σ(u) to be 1, –1, or 0, according to whether the pair $Y_1(u)$, $Y_2(u)$ has the same orientation as the pair $X_1(u)$, $X_2(u)$, the opposite orientation, or is linearly dependent, and let $h_{ij} = Y_i \cdot Y_j$. If $V \subset U$ is open then the signed surface area of Y over V is:

$$\hat{A}_Y(V) = \int_V \sigma \sqrt{\det\left(h_{ij}\right)}\, du^1\, du^2$$

For a regular parametric surface, this definition reduces to Definition 5. Next, we prove that the total curvature of a surface X over an open set U is the area of the image of U under the Gauss map counted with multiplicity.

Theorem 3. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let $V \subset U$ be open. Let $N : U \to \mathbb{S}^2$ be the Gauss map of X, then:

$$K_X(V) = \hat{A}_N(V).$$

Proof. We first derive a formula which is of independent interest:

$$N_i = -k_i^j X_j \qquad ...(5)$$

To verify this formula, it suffices to check that the inner product of both sides with the three linearly independent vectors X_1,X_2,N are equal. Since $N \cdot N = 1$, we have $N \cdot N_i = 0 = -k_i^j X_j \cdot N = 0$, and $-k_i^j X_j \cdot X_l = -k_i^j g_{jl} = -k_{ij} = -N_i \cdot X_k$. In particular, if $h_{ij} = N_i \cdot N_j$, then we find:

$$h_{ij} = \left(k_l^m X_m\right) \cdot \left(k_j^n X_n\right) = k_i^m k_j^n g_{mn} = k_{im} k_{jn} g^{mn}.$$

In particular,

$$\det(h_{ij}) = \frac{(\det(k_{ij}))^2}{\det(g_{ij})}$$

Note also that Equation (5) implies that the pair N_1, N_2 has the same orientation as X_1, X_2 if and only if $\det(k_{ij}) > 0$. Furthermore, since N(u) is also the outward normal to the unit sphere at N(u), and since X_1, X_2, N is positively oriented in $\mathbb{R}^3$, it follows that $X_1(u)$, $X_2(u)$ also gives the positive orientation on the tangent space to the $\mathbb{S}^2$ at N(u). Thus, we deduce that sign $\det(k_{ij}) = \sigma$. Consequently, we obtain:

$$\sigma\sqrt{\det(h_{ij})} = \frac{\text{sign}\,\det(k_{ij})\left|\det(h_{ij})\right|}{\sqrt{\det(g_{ij})}} = K\sqrt{\det(g_{ij})}$$

The proposition follows by integrating over V.

We now turn to an interpretation of the mean curvature. Let X: U → be a parametric surface. A variation of X is a smooth family $F(u;t) : U \times (-\varepsilon, \varepsilon) \to \mathbb{R}^3$ such that F(u; 0) = X. Note that since dF(u; 0) is non-singular, the same is true of dF(u; t_0) for any fixed u_0, perhaps after shrinking the interval (–ε, ε). Thus, all the maps F(u; t_0) for t_0 close enough to 0 are parametric surfaces. The generator of the variation is the vector field dF/dt(u; 0). The variation is compactly supported if F(u; t) = X(u) outside a compact subset of U. The smallest such compact set is called the support of the variation F. Clearly, if a variation is compactly supported, then the support of its generator is compact in U. We say that a variation is tangential if the generator is tangential; we say it is normal if the generator is normal. Suppose now that the closure $\bar{V}$ is compact in U. We consider the area AF (V) of F(u; t) as a function of t. The next proposition shows that the derivative of this function depends only on the generator, and in fact is a linear functional in the generator.

Proposition 6. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let F(u; t) be a variation with generator Y. Then:

$$\left.\frac{dA_F(V)}{dt}\right|_{t=0} = \int_V g^{ij}\, X_i \cdot Y_j\, dA \qquad \text{...(6)}$$

We first need the following lemma from linear algebra. We denote by $S^{n\times n}$ the space of n × n symmetric matrices, and by $S_+^{n\times n}$ the subset of those which are positive definite.

Lemma 2. Let $B : (a,b) \to S_+^{n\times n}$ be continuously differentiable. Then we have:

$$(\log\det B)' = \text{tr}(B - 1\,B'). \qquad \text{...(7)}$$

Proof. First note that (7) follows directly if we assume that B is diagonal. Next, suppose that B is symmetric with distinct eigenvalues. Then there is a continuously differentiable orthogonal matrix Q such that $B = Q^{-1} DQ$, where D is diagonal. Note that $dQ^{-1}/dt = -Q^{-1}(dQ/dt)Q$, hence:

$$B^{-1} B' = -Q^{-1} D^{-1} Q' Q^{-1} DQ + Q^{-1} D^{-1} D' Q + Q^{-1} Q',$$

and in view of tr(AB) = tr(BA), we obtain:

$$\text{tr}(B^{-1} B') = \text{tr}(D^{-1} D').$$

Notes

We also have that detB = detD. Thus taking into the account that (7) holds for D:

$$(\log\det B)' = (\log\det D)' = \mathrm{tr}(D^{-1}D)' = \mathrm{tr}(B^{-1}B').$$

In order to prove the general case, it is more convenient to look at the equivalent identity:

$$(\det B)' = \mathrm{tr}((\det B)B^{-1}B'). \qquad ...(8)$$

Note that by Kramer's rule, the matrix $(\det B)B^{-1}$ is the matrix of co-factors of B, hence its components being determinants of minors of B, are multivariate polynomials in the components of B. Thus, both sides of the identity (8) are linear polynomials

$$p(B';B) = \sum_{i,j=1}^{n} pij(B)b'_{ij}, \quad q(B';B) = \sum_{i,j=1}^{n} qij(B)b'_{ij},$$

in the components b'_{ij} of B′, whose coefficients $p_{ij}(B)$ and $q_{ij}(B)$ are themselves multivariate polynomials in the components b_{ij} of B. Since the set of matrices with distinct eigenvalues is an open set $U \subset S_{+}^{n\times n}$, we have already proved that $p(B';B) = q(B';B)$ holds for all values of B′, and all $B \in U$. For each such $B \in U$ the equality $p(B';B) = q(B';B)$ for all B′ implies that $p_{ij}(B) = q_{ij}(B)$ for i, j = 1, …, n. Since this holds for all B in an open set, we conclude that $p_{ij} = q_{ij}$, and hence p = q.

We remark that the more general identity (8) in fact holds, as easily shown, for all square matrices B. An immediate consequence of the proposition is that:

$$\left(\sqrt{\det B}\right)' = \frac{1}{2}\mathrm{tr}(B^{-1}B')\sqrt{\det B}, \qquad ...(9)$$

for any continuously differentiable family of symmetric positive definite matrices B. We are now ready to prove the proposition.

Proof of Proposition 6. Differentiating the area (4) under the integral sign, and using (9), we get:

$$\frac{dA_F(V)}{dt} = \frac{1}{2}\int_V g^{ij}\frac{dg_{ij}}{dt}\sqrt{\det(g_{ij})}\,du^1\,du^2 = \frac{1}{2}\int_V g^{ij}\frac{dg_{ij}}{dt}dA.$$

Since Y is smooth, we have at t = 0 that $dF_i/dt = (dF/dt)_i = Y_i$, and thus

$$g^{ij}\frac{dg_{ij}}{dt} = g^{ij}(Y_i \cdot X_j + X_i \cdot X_j) = 2g^{ij}\,X_i \cdot Y_j.$$

This completes the proof of the proposition.

Since the variation of the area $dA_F(V)/dt$ is a linear functional in the generator dF/dt of the variation, it is possible to decompose any variation into tangential and normal components. We begin by showing that the area doesn't change under a tangential variation. This is simply the infinitesimal version of Proposition (5).

Proposition 7. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let F(u; t) be a compactly supported tangential variation. If $V \subset U$ is open with $\bar{V}$ compact in U, and the support of F contained in V, then $dA_F(V)/dt = 0$.

Notes

Proof. Let Y be the generator of F(u; t). We will show that there is a smooth family of diffeomorphisms $\phi : U \times (-\delta, \delta) \to U$ such that Y is also the generator of the variation G = X ∘ f. This proves the proposition since Proposition 5 gives that $A_G(U)$ is constant. Since Y is tangential, we can write $Y = y^i X_i$. Consider the initial value problem:

$$\frac{dv^i}{dt} = y^i(v), \quad v^i(0) = u^i.$$

Since the y^i's are compactly supported, a solution v = v(u; t) exists for all t. Defining φ(u; t) = v(u; t), then an application of the inverse function theorem shows that φ(u; t) is a diffeomorphism for t in some small interval (–δ, δ). Finally, we see that:

$$\frac{dX \circ \phi}{dt} = X^i \frac{dv^i}{dt} = X_i y^i = Y.$$

Our next theorem gives an interpretation of the mean curvature as a measure of surface area variation under normal perturbations.

Theorem 4. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let F(u; t) be a compactly supported variation with generator Y . If $V \subset U$ is open with $\bar{V}$ compact in U, and the support of F contained in V , then

$$\frac{dAF(V)}{dt} = -2\int_V (Y \cdot N) H \, dA. \qquad ...(10)$$

Proof. By Propositions 6 and 7, it suffices to consider normal variations with generator Y = fN. In that case, we find that $Y_j = f_j N + f N_j$, so that $g^{ij} X_i \cdot Y_j = f g^{ij} X_i \cdot N_j = -fk_i^i = -2fH$. The theorem follows by substituting into (6).

Definition 8. A parametric surface X is area minimizing if $A_X(U) \le A_{\tilde{X}(U)}$ for any parametric surface $\tilde{X}$ such that $\tilde{X} = X$ on the boundary of U. A parametric surface $X : U \to \mathbb{R}^3$ is locally area minimizing if for any compactly supported variation F(u; t), the area $A_F(U)$ has a local minimum at t = 0.

Clearly, an area-minimizing surface is locally area-minimizing. The following theorem is an immediate corollary of Theorem 4.

Theorem 5. A locally area minimizing surface is a minimal surface.

Note that in general a minimal surface is only a stationary point of the area functional.

13.4 Bernstein's Theorem

In this section, we prove Bernstein's Theorem: A minimal surface which is a graph over an entire plane must itself be a plane. We say that a surface X is a graph over a plane $Y : \mathbb{R}^2 \to \mathbb{R}^3$, where Y is linear, if there is a function $f : \mathbb{R}^2 \to \mathbb{R}$ such that X = Y + fN where N is the unit normal of Y.

Theorem 6 (Bernstein's Theorem). Let X be a minimal surface which is a graph over an entire plane. Then X is a plane.

Notes

We may without loss of generality assume that X is a graph over the plane Y (u, v) = (u, v, 0), i.e. $X(u,v) = (u, v, f(u,v))$. It is then straightforward to check that X is a minimal surface if and only if f satisfies the non-parametric minimal surface equation:

$$(1+q^2)p_u - 2pqp_v + (1+p^2)q_v = 0, \qquad ...(11)$$

where we have used the classical notation: $p = f_u$, $q = f_v$. We say that a solution of a partial differential equation defined on the whole (u, v)-plane is entire. Thus, to prove Bernstein's Theorem, it suffices to prove that any entire solution of (11) is linear.

Proposition 8. Let f be an entire solution of (11). Then f is a linear function.

If f satisfies (11), then p and q satisfy the following equations:

$$\frac{\partial}{\partial u}\left(\frac{1+q^2}{\sqrt{1+p^2+q^2}}\right) = \frac{\partial}{\partial v}\left(\frac{pq}{\sqrt{1+p^2+q^2}}\right), \qquad ...(12)$$

$$\frac{\partial}{\partial u}\left(\frac{pq}{\sqrt{1+p^2+q^2}}\right) = \frac{\partial}{\partial v}\left(\frac{1+p^2}{\sqrt{1+p^2+q^2}}\right). \qquad ...(13)$$

Since the entire plane is simply connected, Equation (13) implies that there exists a function ξ satisfying:

$$\xi_u = \frac{1+p^2}{\sqrt{1+p^2+q^2}}, \quad \xi_v = \frac{pq}{\sqrt{1+p^2+q^2}},$$

and Equation (12) implies that there exists a function satisfying:

$$\eta_u = \frac{pq}{\sqrt{1+p^2+q^2}}, \quad \eta_v = \frac{1+q^2}{\sqrt{1+p^2+q^2}}.$$

Furthermore, $\xi_v = \eta_u$, hence there is a function h so that $h_u = \xi$, $h_v = \eta$. The Hessian of the function h is:

$$(h_{ij}) = \begin{pmatrix} h_{uu} & h_{uv} \\ h_{vu} & h_{vv} \end{pmatrix} = \begin{pmatrix} \xi_u & \xi_v \\ \eta_u & \eta_v \end{pmatrix},$$

hence, h satisfies the Monge-Ampère equation:

$$\det(h_{ij}) = 1. \qquad ...(14)$$

In addition, $h_{11} > 0$, thus (h_{ij}) is positive definite, and we say that h is convex. Proposition 15 now follows from the following result due to Nitsche.

Proposition 9. Let $h \in C^2(\mathbb{R}^2)$ be an entire convex solution of the Monge-Ampère Equation (14). Then h is a quadratic function.

Proof. The proof uses the following transformation introduced by H. Lewy:

$$\varphi = (u, v) \mapsto (\xi, \eta) = (u + p, v + q)$$

where $p = h_u$, and $q = h_v$. Clearly, φ is continuously differentiable, and its Jacobian is:

$$d\varphi = \begin{pmatrix} 1+r & s \\ s & 1+t \end{pmatrix},$$

where $r = h_{uu}$, $s = h_{uv}$, and $t = h_{vv}$. Since $\det(d\varphi) = 2 + r + t > 0$, it follows from the inverse function theorem that φ is a local diffeomorphism, i.e., each point has a neighborhood on which φ is a diffeomorphism. In particular, φ is open.

In view of the convexity of the function h, we have :

$$(u_2 - u_1)(\xi_2 - \xi_1) + (v_2 - v_1)(\eta_2 - \eta_1)$$

$$= (u_2 - u_1)^2 + (v_2 - v_1)^2 + (u_2 - u_1)(p_2 - p_1) + (v_2 - v_1)(q_2 - q_1)$$

$$\geq (u_2 - u_1)^2 + (v_2 - v_1)^2,$$

and therefore:

$$(u_2 - u_1)^2 + (v_2 - v_1)^2 \leq (\xi_2 - \xi_1)^2 + (\eta_2 - \eta_1)^2,$$

i.e., φ is an expanding map. This implies immediately that φ is one-to-one. Thus, φ has an inverse $(u, v) = \varphi^{-1}(\xi, \eta)$ which is also a diffeomorphism. Consider now the function

$$f(\xi + i\eta) = u - p - i(v - q) = 2u - \xi + i(-2v + \eta),$$

where $i = \sqrt{-1}$. In view of

$$d\varphi^{-1} = \begin{pmatrix} u_\xi & u_\eta \\ v_\xi & v_\eta \end{pmatrix} = \frac{1}{2+r+t}\begin{pmatrix} 1+t & -s \\ -s & 1+r \end{pmatrix},$$

it is straightforward to check that f satisfies the Cauchy-Riemann equations, and consequently f is analytic. In fact, f is an entire functions and so is f′. Furthermore,

$$f'(\sigma) = \frac{(t-r) + 2is}{2+r+t}, \quad |f'(\sigma)|^2 = 1 - \frac{4}{2+r+t} < 1,$$

and Liouville's Theorem gives that f′ is constant. Finally, the relations:

$$f = \frac{|1-f'|^2}{1-|f'|^2}, \quad s = \frac{-i(f' - \overline{f}')}{1-|f'|^2}, \quad t = \frac{|1+f'|^2}{1-|f'|^2},$$

show that r, s, t are constants.

13.5 Theorema Egregium

In this section, we prove that the Gauss curvature can be computed in terms of the first fundamental form and its derivatives. We then prove the Fundamental Theorem for surfaces in $\mathbb{R}^3$, analogous

Notes

to curves, which states that a parametric surface is uniquely determined by its first and second fundamental form. Partial derivatives with respect to u^i will be denoted by a subscript i following a comma, unless there is no ambiguity in which case the comma may be omitted.

Proposition 10. Let $X : U \to \mathbb{R}^3$ be a parametric surface. Then the following equations hold:

$$X_{ij} = \Gamma^m_{ij} X_m + k_{ij} N, \qquad ...(15)$$

where,

$$\Gamma^m_{ij} = \frac{1}{2} g^{mn} \left(g_{ni,j} + g_{nj,i} - g_{ij,n} \right), \qquad ...(16)$$

and $\left(g_{ij}\right)$ and $\left(k_{ij}\right)$ are the coordinate representations of its first and second fundamental form.

Proof. Clearly, X_{ij} can be expanded in the basis X_1, X_2, N of $\mathbb{R}^3$. We already saw, that the component of X_{ij} along N is k_{ij}, hence Equation (15) holds with the coefficients Γ^m_{ij} given by

$$X_{ij} \cdot X_m = \Gamma^n_{ij} g_{mn}.$$

In order to derive (16), we differentiate $g_{ij} = X_i \cdot X_j$, and substitute the above equation to obtain:

$$g_{ij,m} = \Gamma^n_{im} g_{nj} + \Gamma^n_{jm} g_{ni}. \qquad ...(17)$$

Now, permute cyclically the indices i, j,m, add the first two equations and subtract the last one:

$$g_{ij,m} + g_{mi,j} - g_{im,i} = 2\Gamma^n_{jm} g_{ni}.$$

Multiplying by g^{il} and dividing by 2 yields (16).

The coefficients Γ^m_{ij} are called the Christoffel symbols of the second kind.[5] It is important to note that the Christofell symbols can be computed from the first fundamental form and its first derivatives. Furthermore, they are not invariant under reparametrization.

Theorem 6. Let $X : U \to \mathbb{R}^3$ be a parametric surface. Then the following equations hold:

$$\Gamma^m_{ij,l} - \Gamma^m_{il,j} + \Gamma^n_{ij}\Gamma^m_{nl} - \Gamma^n_{il}\Gamma^m_{nj} = g^{mn}\left(k_{ij}k_{ln} - k_{il}k_{jn}\right), \qquad ...(18)$$

$$k_{ij,l} - k_{il,j} + \Gamma^m_{ij} k_{lm} - \Gamma^m_{il} k_{jm} = 0. \qquad ...(19)$$

Proof. If we differentiate (15), we get:

$$X_{ijl} = \left(\Gamma^m_{ij} X_m\right)_l + \left(k_{ij} N\right)_l = \Gamma^m_{ij,l} X_m + \Gamma^m_{ij} X_{ml} + k_{ij,l} N + k_{ij} N_l.$$

Substituting X_{ml} from (15) and N_l from (5), and decomposing into tangential and normal components, we obtain:

$$X_{ijl} = A^m_{ijl} X_m + B_{ijl} N,$$

where:

$$A^m_{ijl} = \Gamma^m_{ij,l} + \Gamma^n_{ij}\Gamma^m_{nl} - g^{mn} k_{ij} k_{ln},$$

$$B_{ijl} = k_{ij,l} + \Gamma^m_{ij} k_{lm}.$$

Notes

Taking note of the fact that $X_{ijl} = X_{ilj}$, we now interchange j and l and subtract to obtain (18) and (19).

Equation (18) is called the Gauss Equation, and Equation (19) is called the Codazzi Equation. The Gauss Equation has the following corollary which has been coined Theorema Egregium. It's discovery marked the beginning of intrinsic geometry, the geometry of the first fundamental form.

Corollary 1. Let $X : U \to \mathbb{R}^3$ be a parametric surface. Then the Gauss curvature K of X can be computed in terms of only its first fundamental form $\left(g_{ij}\right)$ and its derivatives up to second order:

$$K = \frac{1}{2} g^{ij}\left(\Gamma^m_{ij,m} - \Gamma^m_{im,j} + \Gamma^n_{ij}\Gamma^m_{nm} - \Gamma^n_{im}\Gamma^m_{nj}\right),$$

where Γ^m_{ij} are the Christoffel symbols of the first kind.

Proof.

We now show, in a manner quite analogous to Theorem 6, that provided they satisfy the Gauss-Codazzi Equations, the first and second fundamental form uniquely determine the parametric surface up to rigid motion.

Theorem 7 (Fundamental Theorem). Let $U \subset \mathbb{R}^2$ be open and simply connected, let $\left(g_{ij}\right) : U \to S^{2\times 2}_+$ and $\left(k_{ij}\right) : U \to S^{2\times 2}$ be smooth, and suppose that they satisfy the Gauss-Codazzi Equations (18)–(19). Then there is a parametric surface $X : U \to \mathbb{R}^3$ such that $\left(g_{ij}\right)$ and $\left(k_{ij}\right)$ are its first and second fundamental forms. Furthermore, X is unique up to rigid motion: if $\tilde{X}$ is another parametric surface with the same first and second fundamental forms, then there is a rigid motion R of $\mathbb{R}^3$ such that $\tilde{X} = R \circ X$.

Proof. We consider the following over-determined system of partial differential equations for X_1, X_2, N:[6]

$$X_{i,j} = \Gamma^m_{ij} X_m + k_{ij} N, \qquad ...(20)$$

$$N_i = -k_{ij} g^{im} X_m, \qquad ...(21)$$

where Γ^m_{ij} is defined in terms of $\left(g_{ij}\right)$ by (16). The integrability conditions for this system are:

$$\left(\Gamma^m_{ij} X_m + k_{ij} N\right)_l = \left(\Gamma^m_{il} X_m + k_{il} N\right)_j \qquad ...(22)$$

$$\left(k_{ij} g^{jm} X_m\right)_l = \left(k_{lj} g^{jm} X_m\right)_i. \qquad ...(23)$$

The proof of Theorem 18 also shows that the Gauss-Codazzi Equations (18)–(19) imply (22) if Xi and N satisfy (20) and (21). We now check that (19) also implies (23). First note that since Γ^m_{ij} is defined by (16), we have

$$\Gamma^m_{ij} g_{mn} = \frac{1}{2}\left(g_{ni,j} + g_{nj,i} - g_{ij,n}\right).$$

Interchanging n and i and adding, we get (17). Now, differentiate (21), and taking into account that $g^{ij}_{,l} = -g^{ia}g_{ab,l}g^{bj}$, substitute (17) to get:

$$N_{i,l} = -k_{ij,l}g^{jm}X_m + k_{ij}g^{ja}\left(\Gamma^n_{al}g_{nb} + \Gamma^n_{bl}g_{na}\right)g^{bm}X_m$$

$$-k_{ij}g^{jm}\left(\Gamma^a_{ml}X_a + k_{ml}g^{jm}N\right) = \left(-k_{ij,l} + k_{in}\Gamma^n_{jl}\right)g^{jm}X_m + k_{ij}k_{ml}g^{jm}N.$$

Note that the last term is symmetric in i and l so that interchanging i and l, and subtracting, we get:

$$N_{i,l} - N_{l,i} = \left(-k_{ij,l} + k_{il,j} - \Gamma^n_{ij}k_{ln} + \Gamma^n_{il}k_{jn}\right)g^{jm}X_m$$

which vanishes by (19). Thus, it follows that (23) is satisfied. We conclude that given values for X_1, X_2, N at a point $u_0 \in U$ there is a unique solution of (20)–(21) in U. We can choose the initial values to that $X_i \cdot X_j = g_{ij}$, $N \cdot X_i = 0$, and $N \cdot N = 1$ at u_0. Using (20) and (21), it is straightforward to check that the functions $h_{ij} = X_i \cdot X_j$, $p_i = N \cdot X_i$ and $q = N \cdot N$, satisfy the differential equations:

$$h_{ij,l} = \Gamma^n_{il}h_{nj} + \Gamma^n_{jl}h_{ni} + k_{il}p_j + k_{jl}p_i,$$

$$p_{i,j} = -k_{jl}g^{lm}h_{mi} + \Gamma^m_{il}p_m + k_{ij}q,$$

$$q_i = -2k_{ij}g^{jm}p_m.$$

However, the functions $h_{ij} = g_{ij}, p_i = 0$ and $q = 1$ also satisfy these equations, as well as the same initial conditions as $h_{ij} = X_i \cdot X_j, p_i = N \cdot X_i$ and $q = N \cdot N$ at u_0. Thus, by the uniqueness statement mentioned above, it follows that $X_i \cdot N_j = g_{ij}, N \cdot N_i = 0$, and $N \cdot N = 1$. Clearly, in view of (20) we have $X_{i,j} = X_{j,i}$, hence there is a function $X : U \to \mathbb{R}^3$ whose partial derivatives are X_i. Since $\left(g_{ij}\right)$ is positive definite we have that X_1, X_2 are linearly independent, hence X is a parametric surface with first fundamental form $\left(g_{ij}\right)$. Furthermore, it is easy to see that the unit normal of X is N, and $X_i \cdot X_j = -N \cdot X_{ij} = -k_{ij}$, hence, the second fundamental form of X is k_{ij}. This completes the proof of the existence statement.

Assume now that $\tilde{X}$ is another surface with the same first and second fundamental forms. Since X and $\tilde{X}$ have the same first fundamental form, it follows that there is a rigid motion R(x) = Qx + y with $Q \in SO(n;\mathbb{R})$ such that $R\left(X\left(u_0\right)\right) = \tilde{X}\left(u_0\right), QX_i\left(u_0\right) = \tilde{X}_i\left(u_0\right), QN\left(u_0\right) = \tilde{N}\left(u_0\right)$. Let $\hat{X} = R \circ X$. Since the two triples $\left(\tilde{X}_1, \tilde{X}_2, \tilde{N}\right)$ and $\left(\hat{X}_1, \hat{X}_2, \hat{N}\right)$ both satisfy the same partial differential equations (20) and (21), it follows that they are equal everywhere, and consequently $\tilde{X} = \hat{X} = R \circ X$.

13.6 Summary

- A curve γ on a parametric surface X is called a line of curvature if $\dot{\gamma}$ is a principal direction.

  The following proposition, due to Rodriguez, characterizes lines of curvature as those curves whose tangents are parallel to the tangent of their spherical image under the Gauss map.

- Let γ be a curve on a parametric surface X with unit normal N, and let $\beta = N \circ \gamma$ be its spherical image under the Gauss map. Then γ is a line of curvature if and only if

$$\dot{\beta} + \lambda\dot{\gamma} = 0.$$

- Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let Y_1 and Y_2 be linearly independent vector fields. The following statements are equivalent:

  - Any point $u_0 \in U$ has a neighborhood U_0 and a reparametrization $\phi : V_0 \to U_0$ such that if $\tilde{X} = X \circ \phi$ then $\tilde{X}_i = Y_i \circ \phi$.

  - $[Y_1, Y_2] = 0$.

- Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let Y_1 and Y_2 be linearly independent vector fields. Then for any point $u_0 \in U$ there is a neighborhood of u_0 and a reparametrization $\tilde{X} = X \circ \phi$ such that $\tilde{X}_i = f_i Y_i \circ \phi$ for some functions f_i.

- Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let u_0 be a hyperbolic point. Then there is neighborhood U_0 of u_0 and a diffeomorphism $\phi : \tilde{U}_0 \to U_0$ such that $\tilde{X} = X \circ \phi$ is parametrized by asymptotic lines.

13.7 Keywords

Line of curvature: A curve γ on a parametric surface X is called a line of curvature if $\dot{\gamma}$ is a principal direction.

Bernstein's Theorem: Let X be a minimal surface which is a graph over an entire plane. Then X is a plane.

13.8 Self Assessment

1. A curve γ on a parametric surface X is called a .... if $\dot{\gamma}$ is a principal direction.
2. Let γ be a curve on a parametric surface X with unit normal N, and let $\beta = N \circ \gamma$ be its spherical image under the Gauss map. Then γ is a line of curvature if and only if ....
3. Let $X : U \to \mathbb{R}^3$ be a ...., and let Y_1 and Y_2 be linearly independent vector fields.
4. A curve γ on a parametric surface X is called an .... if it has zero normal curvature, i.e., $k(\dot{\gamma}, \dot{\gamma}) = 0$.
5. A parametric surface X is minimal if it has vanishing mean ....

Notes

6. Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let $V \subset U$ be open. Let $N : U \to \mathbb{S}^2$ be the Gauss map of X, then ....

7. .... Let X be a minimal surface which is a graph over an entire plane. Then X is a plane.

13.9 Review Questions

1. Prove that setting $f(\zeta) = 1, g(\zeta) = 1/\zeta$ in the Weierstrass representation, we get the catenoid. Find the conjugate harmonic surface of the catenoid.

2. Let $U \subset \mathbb{R}^2$, let $f : U \to \mathbb{R}$ be a smooth function, and let $X : U \to \mathbb{R}^3$ be given by (u, v, f(u, v)), where (u, v) denote the variables in U. Show that X is a minimal surface if and only if it satisfies the non-parametric minimal surface equation:

$$\left(1+q^2\right)p_u - 2pqp_u + \left(1+p^2\right)q_v = 0,$$

where we have used the classical notation: $p = f_u$, $q = f_v$. Show that if f satisfies the equation above then the following equations are also satisfied:

$$\frac{\partial}{\partial u}\left(\frac{1+q^2}{\sqrt{1+p^2+q^2}}\right) = \frac{\partial}{\partial v}\left(\frac{pq}{\sqrt{1+p^2+q^2}}\right),$$

$$\frac{\partial}{\partial u}\left(\frac{pq}{\sqrt{1+p^2+q^2}}\right) = \frac{\partial}{\partial v}\left(\frac{1+p^2}{\sqrt{1+p^2+q^2}}\right).$$

3. Let $f \in C^2(U)$ be a convex function defined on a convex open set U, and let $\Delta f = (p, q) : U \to \mathbb{R}^2$ denote the gradient of f. Prove that for any u_1, $u_2 \in U$ the following inequality holds:

$$\left(u_2 - u_1\right) \cdot \left(\Delta f\left(u_2\right) - \Delta f\left(u_1\right)\right) \geq 0.$$

4. Let $U \subset \mathbb{R}^n$ be open. A map ': U ! Rn is expanding if $|x - y| \leq |\varphi(x) - \varphi(y)|$ for all x, y ∈ U. Let $\varphi : U \to \mathbb{R}^n$ be an open expanding map. Show that the image of the ball $B_R(x_0)$ of radius R centered at $x_0 \in U$ contains the disk $B_R\left(\varphi\left(x_0\right)\right)$ of radius R centered at $\varphi\left(x_0\right)$. Conclude that if $U = \mathbb{R}^n$, then φ is onto $\mathbb{R}^n$.

Answers: Self Assessment

1.	line of curvature	2.	$\dot{\beta} + \lambda\dot{\gamma} = 0.$
3.	parametric surface	4.	asymptotic line
5.	curvature H = 0.	6.	$K_X(V) = \hat{A}_N(V).$
7.	Bernstein's Theorem		

13.10 Further Readings

Notes

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati,T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Notes

Unit 14: Principal Curvatures

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define Principal curvatures
- Discuss the concept of two dimensions: Curvature of surfaces
- Explain the Mean curvature
- Explain the Higher dimensions: Curvature of space

Introduction

In last unit, you have studied about the meaning and concept of curvature. Gaussian curvature, sometimes also called total curvature, is an intrinsic property of a space independent of the coordinate system used to describe it. The Gaussian curvature of a regular surface in $\mathbb{R}^3$ at a point p is formally defined as (S(p)) where S is the shape operator and det denotes the determinant K(p) = det. This unit will provides you information related to principal curvatures.

14.1 Principal Curvatures

All curves with the same tangent vector will have the same normal curvature, which is the same as the curvature of the curve obtained by intersecting the surface with the plane containing T and u.

Taking all possible tangent vectors then the maximum and minimum values of the normal curvature at a point are called the principal curvatures, k_1 and k_2, and the directions of the corresponding tangent vectors are called principal directions.

14.2 Two Dimensions: Curvature of Surfaces

14.2.1 Gaussian Curvature

In contrast to curves, which do not have intrinsic curvature, but do have extrinsic curvature (they only have a curvature given an embedding), surfaces can have intrinsic curvature, independent of an embedding. The Gaussian curvature, named after Carl Friedrich Gauss, is equal to the product of the principal curvatures, k_1k_2. It has the dimension of $1/\text{length}^2$ and is positive for spheres, negative for one-sheet hyperboloids and zero for planes. It determines whether a surface is locally convex (when it is positive) or locally saddle (when it is negative).

This definition of Gaussian curvature is extrinsic in that it uses the surface's embedding in R^3, normal vectors, external planes etc. Gaussian curvature is, however, in fact an intrinsic property of the surface, meaning it does not depend on the particular embedding of the surface; intuitively, this means that ants living on the surface could determine the Gaussian curvature. For example, an ant living on a sphere could measure the sum of the interior angles of a triangle and determine that it was greater than 180 degrees, implying that the space it inhabited had positive curvature. On the other hand, an ant living on a cylinder would not detect any such departure from Euclidean geometry, in particular the ant could not detect that the two surfaces have different mean curvatures which is a purely extrinsic type of curvature.

Formally, Gaussian curvature only depends on the Riemannian metric of the surface. This is Gauss's celebrated Theorema Egregium, which he found while concerned with geographic surveys and map-making.

An intrinsic definition of the Gaussian curvature at a point P is the following: imagine an ant which is tied to P with a short thread of length r. She runs around P while the thread is completely stretched and measures the length C(r) of one complete trip around P. If the surface were flat, she would find C(r) = 2π. On curved surfaces, the formula for C(r) will be different, and the Gaussian curvature K at the point P can be computed by the Bertrand–Diquet–Puiseux theorem as

$$K = \lim_{r \to 0}(2\pi r - C(r)) \cdot \frac{3}{\pi r^3}.$$

The integral of the Gaussian curvature over the whole surface is closely related to the surface's Euler characteristic.

The discrete analog of curvature, corresponding to curvature being concentrated at a point and particularly useful for polyhedra, is the (angular) defect; the analog for the Gauss-Bonnet theorem is Descartes' theorem on total angular defect.

Because curvature can be defined without reference to an embedding space, it is not necessary that a surface be embedded in a higher dimensional space in order to be curved. Such an intrinsically curved two-dimensional surface is a simple example of a Riemannian manifold.

14.2.2 Mean Curvature

The mean curvature is equal to half the sum of the principal curvatures, $(k_1+k_2)/2$. It has the dimension of 1/length. Mean curvature is closely related to the first variation of surface area, in particular a minimal surface such as a soap film, has mean curvature zero and a soap bubble has constant mean curvature. Unlike Gauss curvature, the mean curvature is extrinsic and depends on the embedding, for instance, a cylinder and a plane are locally isometric but the mean curvature of a plane is zero while that of a cylinder is non-zero.

14.2.3 Second Fundamental Form

The intrinsic and extrinsic curvature of a surface can be combined in the second fundamental form. This is a quadratic form in the tangent plane to the surface at a point whose value at a particular tangent vector X to the surface is the normal component of the acceleration of a curve along the surface tangent to X; that is, it is the normal curvature to a curve tangent to X. Symbolically,

$$II(X,X) = N \cdot (\nabla_X X)$$

where N is the unit normal to the surface. For unit tangent vectors X, the second fundamental form assumes the maximum value k_1 and minimum value k_2, which occur in the principal directions u_1 and u_2, respectively. Thus, by the principal axis theorem, the second fundamental form is

$$II(X,X) = k_1(X \cdot u_1)^2 + k_2(X \cdot u_2)^2.$$

Thus, the second fundamental form encodes both the intrinsic and extrinsic curvatures.

A related notion of curvature is the shape operator, which is a linear operator from the tangent plane to itself. When applied to a tangent vector X to the surface, the shape operator is the tangential component of the rate of change of the normal vector when moved along a curve on the surface tangent to X. The principal curvatures are the eigenvalues of the shape operator, and in fact the shape operator and second fundamental form have the same matrix representation with respect to a pair of orthonormal vectors of the tangent plane. The Gauss curvature is, thus, the determinant of the shape tensor and the mean curvature is half its trace.

14.3 Higher Dimensions: Curvature of Space

By extension of the former argument, a space of three or more dimensions can be intrinsically curved; the full mathematical description is described at curvature of Riemannian manifolds. Again, the curved space may or may not be conceived as being embedded in a higher-dimensional space.

After the discovery of the intrinsic definition of curvature, which is closely connected with non-Euclidean geometry, many mathematicians and scientists questioned whether ordinary physical space might be curved, although the success of Euclidean geometry up to that time meant that the radius of curvature must be astronomically large. In the theory of general relativity, which describes gravity and cosmology, the idea is slightly generalised to the "curvature of space-time"; in relativity theory space-time is a pseudo-Riemannian manifold. Once a time coordinate is defined, the three-dimensional space corresponding to a particular time is generally a curved Riemannian manifold; but since the time coordinate choice is largely arbitrary, it is the underlying space-time curvature that is physically significant.

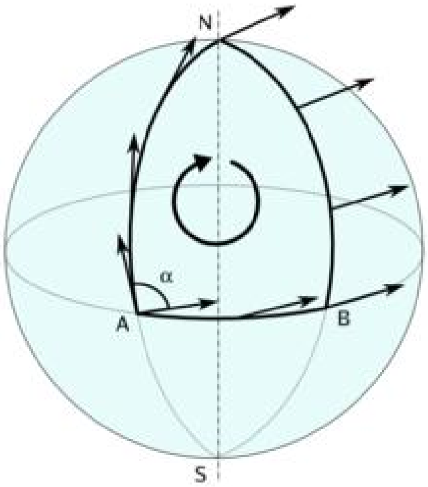
Although an arbitrarily-curved space is very complex to describe, the curvature of a space which is locally isotropic and homogeneous is described by a single Gaussian curvature, as for a surface; mathematically these are strong conditions, but they correspond to reasonable physical assumptions (all points and all directions are indistinguishable). A positive curvature corresponds to the inverse square radius of curvature; an example is a sphere or hypersphere. An example of negatively curved space is hyperbolic geometry. A space or space-time with zero curvature is called flat. For example, Euclidean space is an example of a flat space, and Minkowski space is an

Notes

example of a **flat** space-time. There are other examples of flat geometries in both settings, though. A torus or a cylinder can both be given flat metrics, but differ in their topology. Other topologies are also possible for curved space.

14.4 Generalizations

Figure 14.1



In above figure, *Parallel transporting a vector from A → N → B yields a different vector. This failure to return to the initial vector is measured by the holonomy of the surface.*

The mathematical notion of curvature is also defined in much more general contexts. Many of these generalizations emphasize different aspects of the curvature as it is understood in lower dimensions.

One such generalization is kinematic. The curvature of a curve can naturally be considered as a kinematic quantity, representing the force felt by a certain observer moving along the curve; analogously, curvature in higher dimensions can be regarded as a kind of tidal force (this is one way of thinking of the sectional curvature). This generalization of curvature depends on how nearby test particles diverge or converge when they are allowed to move freely in the space.

Another broad generalization of curvature comes from the study of parallel transport on a surface. For instance, if a vector is moved around a loop on the surface of a sphere keeping parallel throughout the motion, then the final position of the vector may not be the same as the initial position of the vector. This phenomenon is known as holonomy. Various generalizations capture in an abstract form this idea of curvature as a measure of holonomy. A closely related notion of curvature comes from gauge theory in physics, where the curvature represents a field and a vector potential for the field is a quantity that is in general path-dependent: it may change if an observer moves around a loop.

Two more generalizations of curvature are the scalar curvature and Ricci curvature. In a curved surface such as the sphere, the area of a disc on the surface differs from the area of a disc of the same radius in flat space. This difference (in a suitable limit) is measured by the scalar curvature. The difference in area of a sector of the disc is measured by the Ricci curvature. Each of the scalar curvature and Ricci curvature are defined in analogous ways in three and higher dimensions. They are particularly important in relativity theory, where they both appear on the side of Einstein's field equations that represents the geometry of spacetime (the other side of which represents the presence of matter and energy). These generalizations of curvature underlie, for instance, the notion that curvature can be a property of a measure.

Another generalization of curvature relies on the ability to compare a curved space with another space that has constant curvature. Often this is done with triangles in the spaces. The notion of a triangle makes senses in metric spaces, and this gives rise to CAT (k) spaces.

14.5 Summary

- All curves with the same tangent vector will have the same normal curvature, which is the same as the curvature of the curve obtained by intersecting the surface with the plane containing T and u. Taking all possible tangent vectors then the maximum and minimum values of the normal curvature at a point are called the principal curvatures, k_1 and k_2, and the directions of the corresponding tangent vectors are called principal directions.
- In contrast to curves, which do not have intrinsic curvature, but do have extrinsic curvature (they only have a curvature given an embedding), surfaces can have intrinsic curvature, independent of an embedding. Gaussian curvature is, however, in fact an intrinsic property of the surface, meaning it does not depend on the particular embedding of the surface; intuitively, this means that ants living on the surface could determine the Gaussian curvature. Formally, Gaussian curvature only depends on the Riemannian metric of the surface. This is Gauss's celebrated Theorema Egregium, which he found while concerned with geographic surveys and map-making.
- An intrinsic definition of the Gaussian curvature at a point P is the following: imagine an ant which is tied to P with a short thread of length r. She runs around P while the thread is completely stretched and measures the length C(r) of one complete trip around P
- The mean curvature is equal to half the sum of the principal curvatures, $(k_1+k_2)/2$. It has the dimension of 1/length. Mean curvature is closely related to the first variation of surface area, in particular a minimal surface such as a soap film, has mean curvature zero and a soap bubble has constant mean curvature. Although an arbitrarily-curved space is very complex to describe, the curvature of a space which is locally isotropic and homogeneous is described by a single Gaussian curvature, as for a surface; mathematically these are strong conditions, but they correspond to reasonable physical assumptions (all points and all directions are indistinguishable). A positive curvature corresponds to the inverse square radius of curvature; an example is a sphere or hypersphere. An example of negatively curved space is hyperbolic geometry. A space or space-time with zero curvature is called flat. For example, Euclidean space is an example of a flat space, and Minkowski space is an example of a flat space-time

14.6 Keywords

Principal directions: Taking all possible tangent vectors then the maximum and minimum values of the normal curvature at a point are called the principal curvatures, k_1 and k_2, and the directions of the corresponding tangent vectors are called principal directions.

Gaussian curvature: An intrinsic definition of the Gaussian curvature at a point P is the following: imagine an ant which is tied to P with a short thread of length r.

Arbitrarily-curved space: An arbitrarily-curved space is very complex to describe, the curvature of a space which is locally isotropic and homogeneous is described by a single Gaussian curvature, as for a surface; mathematically these are strong conditions, but they correspond to reasonable physical assumptions.

14.7 Self Assessment

1. All curves with the same tangent vector will have the same ...., which is the same as the curvature of the curve obtained by intersecting the surface with the plane containing T and u.
2. .... is however in fact an intrinsic property of the surface, meaning it does not depend on the particular embedding of the surface; intuitively, this means that ants living on the surface could determine the Gaussian curvature.
3. Gaussian curvature only depends on the .... of the surface.
4. An .... definition of the Gaussian curvature at a point P is the following: imagine an ant which is tied to P with a short thread of length r
5. A .... corresponds to the inverse square radius of curvature; an example is a sphere or hypersphere.

14.8 Review Question

1. Define Principal curvatures.
2. Discuss the concept of two dimensions.
3. Describe the Curvature of surfaces.
4. Explain the Mean curvature.
5. Explain the Higher dimensions: Curvature of space.

Answers: Self Assessment

1. normal curvature
2. Gaussian curvature
3. Riemannian metric
4. Intrinsic
5. positive curvature

14.9 Further Reading

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati,T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 15: Local Intrinsic Geometry of Surfaces

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define Riemannian Surfaces
- Discuss Lie Derivative
- Explain the Concept of Covariant Differentiation

Introduction

In this unit, we change our point of view, and study intrinsic geometry, in which the starting point is the first fundamental form. Thus, given a parametric surface, we will ignore all information which cannot be recovered from the first fundamental form and its derivatives only.

15.1 Riemannian Surfaces

Definition 1. Let $U \subset \mathbb{R}^2$ be open. A Riemannian metric on U is a smooth function $g : U \to \mathbb{S}_+^{2\times 2}$. A Riemannian surface patch is an open set U equipped with a Riemannian metric.

The tangent space of U at $u \in U$ is $\mathbb{R}^2$. The Riemannian metric g defines an inner-product on each tangent space by:

$$g(Y,Z) = g_{ij}y^i z^j,$$

where y^i and z^j are the components of Y and Z with respect to the standard basis of $\mathbb{R}^2$. We will write $|Y|_g^2 = g(Y, Y),$ and omit the subscript g when it is not ambiguous.

Two Riemannian surface patches (U, g) and $\left(\tilde{U}, \tilde{g}\right)$ are isometric if there is a diffeomorphism $\phi : \tilde{U} \to U$ such that

$$\tilde{g}_{ij} = g_{lm}\phi_i^l\phi_j^m, \qquad ...(1)$$

where $\phi_i^l = \partial u^l / \partial \tilde{u}^i$. In fact, Equation (1) reads:

$$d\phi^* g = \tilde{g},$$

where $d\phi^*$ g is the pull-back of g by the Jacobian of ϕ at $\tilde{u}$. We then say that ϕ is an isometry between (U, g) and $\left(\tilde{U}, \tilde{g}\right)$. As before, we denote by g^{ij} the inverse of the matrix g_{ij}.

We also denote the Riemannian metric:

$$ds^2 = g_{ij}\, du^i\, du^j,$$

and at times refer to it as a line element. The arc length of a curve $\gamma : [a,b] \to U$ is then given by:

$$L_\gamma = \int_a^b \sqrt{g_{ij}\dot{\gamma}^i\dot{\gamma}^j}\, dt.$$

Note that the arc length is simply the integral of $\sqrt{g\left(\dot{\gamma}, \dot{\gamma}\right)}$.

Example: Let $U \subset \mathbb{R}^2$ be open, and let $\text{let}\left(\delta_{ij}\right)$ be the identity matrix, then (U, δ) is a Riemannian surface. The Riemannian metric d will be called the Euclidean metric.

Example: Let $X : U \to \mathbb{R}^3$ be a parametric surface, and let g be the coordinate representation of its first fundamental form, then (U, g) is a Riemannian surface patch. We say that the metric g is induced by the parametric surface X. If $\tilde{X} = X \circ \phi : \tilde{U} \to \mathbb{R}^3$ is a reparametrization of X and $\tilde{g}$ the coordinate representation of its first fundamental form, then $\left(\tilde{U}, \tilde{g}\right)$ is isometric to (U, g).

Example: (The Poincaré Disk). Let $D = \left\{(u, v) : u^2 + v^2 < 1\right\}$ be the unit disk in $\mathbb{R}^2$, and let

$$g_{ij} = \frac{4}{\left(1 - r^2\right)^2}\delta_{ij}$$

where $r = \sqrt{u^2 + v^2}$ is the Euclidean distance to the origin. We can write this line element also as

$$ds^2 = 4\frac{du^2 + dv^2}{\left(1 - u^2 - v^2\right)^2}. \qquad ...(2)$$

The Riemannian surface (D, g) is called the Poincar´e Disk. Let U = {(x, y) : y > 0} be the upper half-plane, and let

$$h_{ij} = \frac{1}{y^2}\delta_{ij}.$$

Then it is not difficult to see that $\left(D, g_{ij}\right)$ and $\left(U, h_{ij}\right)$ are isometric with the isometry given by:

$$\phi:(u,v)\mapsto(x,y)=\left(\frac{2v}{(1+u)^2+v^2},\frac{1-u^2-v^2}{\left(1+u^2\right)+v^2}\right)$$

In fact, a good bookkeeping technique to check this type of identity is to compute the differentials:

$$dx=-4\frac{v(1+u)}{\left((1+u)^2+v^2\right)^2}du+2\frac{(1+u)^2-v^2}{\left((1+u)^2+v^2\right)^2}dv$$

$$dy=-2\frac{(1+u)^2-v^2}{\left((1+u)^2+v^2\right)^2}du+4\frac{v(1+u)}{\left((1+u)^2+v^2\right)^2}dv,$$

substitute into

$$\frac{dx^2+dy^2}{y^2},$$

and then simplify using du dv = dv du to obtain (2). It is not difficult to see that this is equivalent to checking (1).

Definition 2. Let (U, g) be a Riemannian surface. The Christoffel symbols of the second kind of g are defined by:

$$\Gamma^m_{ij}=\frac{1}{2}g^{mn}\left(g_{ni,j}+g_{nj,i}-g_{ij,n}\right). \qquad ...(3)$$

The Gauss curvature of g is defined by:

$$K=\frac{1}{2}g^{ij}\left(\Gamma^m_{ij,m}-\Gamma^m_{ij}\Gamma^m_{nm}-\Gamma^n_{im}\Gamma^m_{nj}\right). \qquad ...(4)$$

If (U, g) is induced by the parametric surface $X:U\to\mathbb{R}^3$, then these definitions agree with those studied earlier.

15.2 Lie Derivative

Here, we study the Lie derivative. We denote the standard basis on $\mathbb{R}^2$ by ∂_1,∂_2. Let f be a smooth function on U, and let $Y=y^i\partial_i\in T_uU$ be a vector at u ∈ U. The directional derivative of f along Y is:

$$\partial_Y f=y^i\partial_i f=y^i f_i. \qquad ...(5)$$

Since $y_i=\partial_Y u^i$ where $\left(u^1,u^2\right)$ are the coordinates on U, we see that Y = Z follows from $\partial_Y=\partial_Z$ as operators. The next proposition shows that the directional derivative of a function is reparametrization invariant.

Proposition 1. Let $\phi:\tilde{U}\to U$ be a diffeomorphism, and let $\tilde{Y}$ be a vector at $\tilde{u}\in\tilde{U}$. Then for any smooth function f on U, we have:

$$\left(\partial_{d\phi(\tilde{Y})}f\right)\circ\phi=\partial_{\tilde{Y}}\left(f\circ\phi\right).$$

Proof. Denoting the coordinates on U by u^j and the coordinates on $\tilde{U}$ by $\tilde{u}^i$, we let $\phi_i^j = \partial u^j / \partial \tilde{u}^i$, and we find, by the chain rule:

$$\partial_{\tilde{Y}}\left(f \circ \phi\right) = \tilde{y}^i \partial_i \left(f \circ \phi\right) = \tilde{y}^i (\partial_j f)\phi_i^j = \left(\partial_{d\phi(\tilde{Y})} f\right) \circ \phi.$$

We define the commutator of two tangent vector fields $Y = y^i \partial_i$ and $Z = z^i \partial_i$, Equation (5):

$$[Y, Z] = \left(y^i \partial_i z^j - z^i \partial d_i y^j\right)\partial_j. \qquad ...(6)$$

Note that

$$\partial_{[Y,Z]} f = \partial_Y \partial_Z f - \partial_Z \partial_Y f. \qquad ...(7)$$

This observation together with Proposition 3.1 are now used to show that the commutator is reparametrization invariant.

***Proposition* 2.** Let $\tilde{Y}$ and $\tilde{Z}$ be vector fields on $\tilde{U}$, and let $\phi : \tilde{U} \to U$ be a diffeomorphism, then

$$d\phi\left(\left[\tilde{Y}, \tilde{Z}\right]\right) = \left[d\phi\left(\tilde{Y}\right), d\phi\left(\tilde{Z}\right)\right].$$

Proof. For any smooth function f on U, we have:

$$\partial_{d\phi\left(\left[\tilde{Y},\tilde{Z}\right]\right)} f = \partial_{\left[\tilde{Y},\tilde{Z}\right]}(f \circ \phi) = \partial_{\tilde{Y}}\partial_{\tilde{Z}}(f \circ \phi) - \partial_{\tilde{Z}}\partial_{\tilde{Y}}(f \circ \phi) \qquad ...(8)$$

$$= \partial_{\tilde{Y}}\left(\partial_{d\phi\left(\tilde{Z}\right)} f\right) \circ \phi - \partial_{\tilde{Z}}\left(\partial_{d\phi\left(\tilde{Y}\right)} f\right) \circ \phi = \partial_{d\phi\left(\tilde{Y}\right)}\partial_{d\phi\left(\tilde{Z}\right)} f - \partial_{d\phi\left(\tilde{Z}\right)}\partial_{d\phi\left(\tilde{Y}\right)} f \;\; = \partial_{\left[d\phi\left(\tilde{Y}\right), d\phi\left(\tilde{Z}\right)\right]} f,$$

and the proposition follows.

We note for future reference that in the proofs of propositions 1 and 2, only the smoothness of the map ϕ is used, and not the fact that it is a diffeomorphism. The operator $Z \mapsto \mathcal{L}_Y Z = [Y, Z]$, also called the Lie derivative, is a differential operator, in the sense that it is linear and satisfies a Leibniz identity: $\mathcal{L}_Y(fZ) = \left(\partial_Y f\right)Z + f\mathcal{L}_Y Z$. However, $\mathcal{L}_Y Z$ depends on the values of Y in a neighborhood of a point as can be seen from the fact that it is not linear over functions in Y , but rather satisfies $\mathcal{L}_{fY} Z = f\mathcal{L}_Y Z - \left(\partial_Z f\right)Y$. Hence the Lie derivative cannot be used as an intrinsic directional derivative of a vector field Z, which should only depend on the direction vector Y at a single point[1].

15.3 Covariant Differentiation

***Definition* 3.** Let (U, g) be a Riemannian metric, and let Z be a vector field on U. The covariant derivative of Z along ∂_i is:

$$\nabla_i Z = \left(\partial_i z^j + \Gamma_{ik}^j z^k\right)\partial_j. \qquad ...(9)$$

[1] Indeed $\partial_Y Z$ as defined does depend only on the value of Y at a single point and satisfies $\partial_{fY} Z = f\partial_Y Z$.

Notes

Let $Y \in T_u U$, the covariant derivative of Z along Y is:

$$\nabla_Y Z = y^i Z_{;i}.$$

We write the components of $\nabla_Y Z$ as:

$$z^j{}_{;i} = z^j{}_{,i} + \Gamma^j_{ik} z^k, \qquad ...(10)$$

so that $\nabla_Y Z = y^i z^j{}_{;i} \partial_j$. Furthermore, note that

$$\nabla_i \partial_j = \Gamma^k_{ij} \partial_k. \qquad ...(11)$$

Our first task is to show that covariant differentiation is reparametrization invariant. However, since the metric g was used in the definition of the covariant derivative, it stands to reason that it would be invariant only under those reparametrization which preserve the metric, i.e., under isometries.

Proposition 3. Let $\phi : (\tilde{U}, \tilde{g}) \to (U, g)$ be an isometry. Let $\tilde{Y} \in T_{\tilde{u}} \tilde{U}$, and let $\tilde{Z}$ be a vector field on $\tilde{U}$. Then

$$d\phi\left(\tilde{\nabla}_{\tilde{Y}} \tilde{Z}\right) = \nabla_{d\phi(\tilde{Y})} d\phi\left(\tilde{Z}\right). \qquad ...(12)$$

Proof. This proof, although tedious, is quite straightforward, and is relegated to the exercises.

Note that on the left hand-side of (12), the covariant derivative $\tilde{\nabla}$ is that obtained from the metric $\tilde{g}$.

Our next observation, which follows almost immediately, gives an interpretation of the covariant derivative when the metric g is induced by a parametric surface X.

Proposition 4. Let the Riemannian metric g be induced by the parametric surface X. Then the image under dX of the covariant derivative $dX(\nabla_i Z)$ is the projection of $\partial_i Z$ onto the tangent space.

Proof. Note that $dX(\partial_i) = X_i$. Thus, if $Z = z^j \partial_j$ then we find:

$$dX(\nabla_i Z) = z^j{}_{;i} X_j = z^j{}_{;i} X_j + \Gamma^j_{ik} z^k X_j = \partial_i \left(z^j X_j\right) - k_{ij} z^j N,$$

which proves the proposition.

We now show that covariant differentiation is in addition well-adapted to the metric g.

Proposition 5. Let (U, g) be a Riemannian surface, and let Y and Z be vector fields on U. Then, we have

$$\partial_{ig}(Y, Z) = g(\nabla_i Y, Z) + g(\nabla_i Y, Z). \qquad ...(13)$$

Proof. We first note that, as in the proof of Theorem 2.29, the definition of the Christoffel symbols (3) :

$$g_{ij,l} = \Gamma^k_{il} g_{kj} + \Gamma^k_{jl} g_{ki}. \qquad ...(14)$$

Notes

Now, setting $Y = y^i\partial_i$ and $Z = z^i\partial_i$, we compute:

$$\partial_{ig}(Y,Z) = \partial_i g_{jk} y^j z^k = \Gamma^m_{ji} g_{km} y^j z^k + \Gamma^m_{ki} g_{mi} y^i z^k + g_{ik} y^j{}_{,i} zk + g_{jk} y^j z^k{}_{,i}$$

$$= g_{jk}\left(y^j{}_{,i} + \Gamma^j_{mi} y^m\right) z^k + g_{ik} y^j \left(z^k{}_{,i} + \Gamma^k_{mi} z^m\right) = g\left(Y_{;i}, Z\right) + g\left(Y, Z_{;i}\right).$$

This completes the proof of (13) and of the proposition.

Definition 4. Let $Y = y^i\partial_i$ be a vector field on the Riemannian surface (U, g). Its divergence is the function:

$$\text{div } Y = \nabla_i y^i = \partial_i y^i + \Gamma^i_{ij} y^j.$$

Note that:

$$\Gamma^i_{ij} = \frac{1}{2} g^{im}\left(g_{mi,j} + g_{mj,i} - g_{ij,m}\right) = \frac{1}{2} g^{im} g_{im,j} = \partial_j \log\sqrt{\det g}.$$

Thus, we see that:

$$\text{div } Y = \frac{1}{\sqrt{\det g}}\partial_i\left(\sqrt{\det g}\; y^i\right) \qquad \text{...(15)}$$

Observe that this implies

$$\int_U \text{div } Y\, dA = \int_U \partial_i\left(\sqrt{\det g}\; y^i\right) du^1\, du^2.$$

Thus, Green's Theorem in the plane implies the following proposition.

Proposition 6. Let Y be a compactly supported vector field on the Riemannian surface (U, g). Then, we have:

$$\int_U \text{div } Y\, dA = 0.$$

Definition 5. If $f : U \to \mathbb{R}$ is a smooth function on the Riemannian surface (U, g), its gradient ∇f is the unique vector field which satisfies $g(\nabla f, Y) = \partial_Y f$.

The Laplacian of f if the divergence of the gradient of f:

$$\nabla f = \text{div } \nabla f.$$

It is easy to see that $\nabla f = g^{ij} f_j \partial_j$, hence

$$\nabla f = \frac{1}{\sqrt{\det g}}\partial_i\left(g^{ij}\sqrt{\det g}\; f_j\right). \qquad \text{...(16)}$$

Thus, in view of Proposition 6, if f is compactly supported, we have:

$$\int_U \nabla f\, dA = 0.$$

15.4 Summary

- Let $U \subset \mathbb{R}^2$ be open. A Riemannian metric on U is a smooth function $g : U \to \mathbb{S}_+^{2\times 2}$. A Riemannian surface patch is an open set U equipped with a Riemannian metric.

  The tangent space of U at $u \in U$ is $\mathbb{R}^2$. The Riemannian metric g defines an inner-product on each tangent space by:

  $$g(Y,Z) = g_{ij}y^i z^j,$$

  where y^i and z^j are the components of Y and Z with respect to the standard basis of $\mathbb{R}^2$. We will write $|Y|_g^2 = g(Y,Y)$, and omit the subscript g when it is not ambiguous.

- Let $U \subset \mathbb{R}^2$ be open, and let $\text{let}(\delta_{ij})$ be the identity matrix, then (U, δ) is a Riemannian surface. The Riemannian metric d will be called the Euclidean metric.
- Let (U, g) be a Riemannian surface. The Christoffel symbols of the second kind of g are defined by:

  $$\Gamma_{ij}^m = \frac{1}{2} g^{mn}\left(g_{ni,j} + g_{nj,i} - g_{ij,n}\right).$$

  The Gauss curvature of g is defined by:

  $$K = \frac{1}{2} g^{ij}\left(\Gamma_{ij,m}^m - \Gamma_{ij}^m \Gamma_{nm}^m - \Gamma_{im}^n \Gamma_{nj}^m\right).$$

  If (U, g) is induced by the parametric surface $X : U \to \mathbb{R}^3$, then these definitions agree.

- Let $\tilde{Y}$ and $\tilde{Z}$ be vector fields on $\tilde{U}$, and let $\phi : \tilde{U} \to U$ be a diffeomorphism, then

  $$d\phi\left(\left[\tilde{Y}, \tilde{Z}\right]\right) = \left[d\phi\left(\tilde{Y}\right), d\phi\left(\tilde{Z}\right)\right].$$

- Let Y be a compactly supported vector field on the Riemannian surface (U, g). Then, we have:

  $$\int_U \text{div } Y \, dA = 0.$$

15.5 Keywords

Riemannian metric: A Riemannian metric on U is a smooth function $g : U \to \mathbb{S}_+^{2\times 2}$. A Riemannian surface patch is an open set U equipped with a Riemannian metric.

Euclidean metric: Let $U \subset \mathbb{R}^2$ be open, and let $\text{let}(\delta_{ij})$ be the identity matrix, then (U, δ) is a Riemannian surface. The Riemannian metric d will be called the Euclidean metric.

15.6 Self Assessment

1. A .... on U is a smooth function $g : U \to \mathbb{S}_+^{2\times 2}$. A Riemannian surface patch is an open set U equipped with a Riemannian metric.

2. Let $U \subset \mathbb{R}^2$ be open, and let $\text{let}(\delta_{ij})$ be the identity matrix, then (U, δ) is a Riemannian surface. The Riemannian metric d will be called the ....

3. Let $\phi : \tilde{U} \to U$ be a diffeomorphism, and let $\tilde{Y}$ be a vector at $\tilde{u} \in \tilde{U}$. Then for any smooth function f on U, we have: ....

4. The .... cannot be used as an intrinsic directional derivative of a vector field Z, which should only depend on the direction vector Y at a single point.

5. Let the Riemannian metric g be induced by the parametric surface X. Then the image under dX of the covariant derivative .... is the projection of $\partial_i Z$ onto the tangent space.

6. The Laplacian of f if the divergence of the gradient of f: ....

15.7 Review Questions

1. Prove that a parametric surface $X : U \to \mathbb{R}^3$ is conformal if and only if its first fundamental form g is conformal to the Euclidean metric δ on U.

2. Two Riemannian metrics g and $\tilde{g}$ on an open set $U \subset \mathbb{R}^2$ are conformal if $\tilde{g} = e^{2\lambda} g$ for some smooth function λ.

Answers: Self Assessment

1.	Riemannian metric	2.	Euclidean metric.
3.	$\left(\partial_{d\phi(\tilde{Y})} f\right) \circ \phi = \partial_{\tilde{Y}}\left(f \circ \phi\right)$.	4.	Lie derivative
5.	$dX(\nabla_i Z)$	6.	∇f = div ∇f.

15.8 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati,T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 16: Geodesics

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define Geodesics
- Explain the Riemann Curvature Tensor
- Discuss the Second Variation of Arc length

Introduction

In particular, we will ignore the Gauss map and the second fundamental form. Thanks to Gauss' Theorema Egregium, we will still be able to take the Gauss curvature into account. In last unit, you have studied about local intrinsic Geometry of Surfaces.

16.1 Geodesics

Definition 1. Let (U, g) be a Riemannian surface, and let $\gamma : I \to U$ be a curve. A vector field along γ is a smooth function $Y : I \to \mathbb{R}^2$. The covariant derivative of $Y = y^i\partial_i$ along γ is the vector field:

$$\nabla_{\dot{\gamma}} Y = \left(\dot{y}^i + \Gamma^i_{jk} y^j \dot{\gamma}^k\right)\partial_i.$$

Note that if Z is any extension of Y, i.e., a any vector field defined on a neighborhood V of the image $\gamma(I)$ of γ in U, then we have:

$$\nabla_{\dot{\gamma}} Y = \nabla_{\dot{\gamma}} Z = \dot{\gamma}^i Z_{;i}.$$

Thus, any result proved concerning the usual covariant differentiation, in particular also for the covariant differentiation along a curve.

Definition 2. A vector field Y along a curve γ is said to be parallel along γ if $\nabla_{\dot{\gamma}} Y = 0$.

Note that if Y and Z are parallel along γ, then g(Y,Z) is constant.

$$\partial_{\dot{\gamma}} g(Y,Z) = g\left(\nabla_{\dot{\gamma}} Y, Z\right) + g\left(Y, \nabla_{\dot{\gamma}} Z\right) = 0.$$

Proposition 1. Let $\gamma : [a,b] \to U$ be a curve into the Riemannian surface (U, g), let $u_0 \in U$, and let $Y_0 \in Tu_0 U$. Then there is a unique vector field Y along γ which is parallel along γ and satisfies $Y(a) = Y_0$.

Proof. The condition that Y is parallel along γ is a pair of linear first-order ordinary differential equations:

$$\dot{y}^i = -\Gamma^i_{jk}(\gamma)\gamma^j y^j.$$

Given initial conditions $\gamma^i(a) = y^i_0$, the existence and uniqueness of a solution on [a, b] follows from the theory of ordinary differential equations.

The proposition together with the comment preceding it shows that parallel translation along a curve γ is an isometry between inner-product spaces $P_\gamma : T_a U \to T_b U$.

Definition 3. A curve γ is a geodesic if its tangent $\dot{\gamma}$ is parallel along γ:

$$\nabla_{\dot{\gamma}} \dot{\gamma} = 0.$$

If γ is a geodesic, then $|\dot{\gamma}|$ is constant and hence, every geodesic is parametrized proportionally to arc length. In particular, if $\beta = \gamma \circ \phi$ is a reparametrization of γ, then β is not a geodesic unless ϕ is a linear map.

Proposition 2. Let (U, g) be a Riemannian surface, let $u_0 \in U$ and let $0 \neq Y_0 \in T_{u0}U$. Then there is and $\varepsilon > 0$, and a unique geodesic $\gamma : (-\varepsilon, \varepsilon) \to U$, such that $\gamma(0) = u_0$, and $\dot{\gamma}(0) = Y_0$.

Proof. We have:

$$\nabla_{\dot{\gamma}} \dot{\gamma} = \left(\ddot{\gamma}^i + \Gamma^i_{jk} \dot{\gamma}^j \dot{\gamma}^k\right) \partial_i.$$

Thus, the condition that γ is a geodesic can written as a pair of non-linear second order ordinary differential equations:

$$\ddot{\gamma}^i = -\Gamma^i_{jk}\left(\gamma(t)\right) \dot{\gamma}^j \dot{\gamma}^k.$$

Given initial conditions $\gamma^i(0) = u^i_0, \dot{\gamma}^i(0) = y^i_0$, there is a unique solution on defined on a small enough interval $(-\varepsilon, \varepsilon)$.

Definition 4. Let $\gamma : [a,b] \to U$ be a curve. We say that γ is length minimizing, or L-minimizing, if:

$$L_\gamma \leq L_\beta$$

for all curves β in U such that $\beta(a) = \gamma(a)$ and $\beta(b) = \gamma(b)$.

Notes

Let $\gamma : [a,b] \to U$ be a curve. A variation of γ is a smooth family of curves $\sigma(t;s) : [a,b] \times (-\varepsilon,\varepsilon) \to I$ such that $\sigma(t;0) = \gamma(t)$ for all $t \in [a,b]$. For convenience, we will denote derivatives with respect to t as usual by a dot, and derivatives with respect to s by a prime. The generator of a variation σ is the vector field Y (t) = $\sigma'(t;\, 0)$ along γ. We say that σ is a fixed-endpoint variation, if $\sigma(a;s) = \gamma(a)$, and $\sigma(b;s) = \gamma(b)$ for all $s \in (-\varepsilon,\varepsilon)$. Note that the generator of a fixed-endpoint variation vanishes at the end points. We say that a variation σ is normal if its generator Y is perpendicular to $\gamma : g(\dot{\gamma}, Y) = 0.$ A curve γ is locally L-minimizing if

$$L_\sigma(s) = \int_a^b \sqrt{g(\dot{\sigma},\dot{\sigma})}\, dt$$

has a local minimum at s = 0 for all fixed-endpoint variations σ. Clearly, an L-minimizing curve is locally L-minimizing.

If γ is locally L-minimizing, then any reparametrization $\beta = \gamma \circ \phi$ of γ is also locally L-minimizing. Indeed, if σ is any fixed-endpoint variation of β, then $\tau(t;s) = \sigma(\phi^{-1}(t);s)$ is a fixed-endpoint variation of γ, and since reparametrization leaves arc length invariant, we see that $L_\tau(s) = L_\sigma(s)$ which implies that L_σ also has a local minimum at s = 0. Thus, local minimizers of the functional L are not necessarily parametrized proportionally to arc length. This helps clarify the following comment: a locally length-minimizing curve is not necessarily a geodesic, but according to the next theorem that is only because it may not be parametrized proportionally to arc length.

Theorem 1. A locally length-minimizing curve has a geodesic reparametrization.

To prove this theorem, we introduce the energy functional:

$$E_\gamma = \frac{1}{2}\int_a^b g(\dot{\gamma},\dot{\gamma})dt$$

We may now speak of energy-minimizing and locally energy-minimizing curves. Our first lemma shows the advantage of using the energy rather than the arc length functional: minimizers of E are parametrized proportionally to arc length.

Lemma 1. A locally energy-minimizing curve is a geodesic.

Proof. Suppose that γ is a locally energy-minimizing curve. We first note that if Y is any vector field along γ which vanishes at the endpoints, then setting $\sigma(t;s) = \gamma(t) + sY(t),$ we see that there is a fixed-endpoint variation of γ whose generator is Y. Since γ is locally energy-minimizing, we have:

$$E'_\sigma(0) = \int_a^b \frac{1}{2}\left(g(\dot{\sigma},\dot{\sigma})\right)'\big|_{s=0}\, dt = 0.$$

We now observe that:

$$\left(\dot{\sigma}^j\right)'\big|_{s=0} = \left(\frac{d}{dt}\sigma^j\right)'\big|_{s=0} = \left(\frac{d}{dt}\sigma^j\right)'\big|_{s=0} = \dot{y}^j.$$

where $Y = y^i\partial_i$ is the generator of the fixed-endpoint variation σ, and:

$$\left(g_{ij}\right)'\big|_{s=0} = g_{ij,k}\left(\sigma^k\right)'\big|_{s=0} = g_{ij,k}\, y^k.$$

Thus, we have:

Notes

$$\frac{1}{2}\left(g(\dot\sigma,\dot\sigma)\right)'|_{s=0}=\frac{1}{2}\left(g_{ij}\dot\sigma^i\dot\sigma^j\right)'|_{s=0}=\frac{1}{2}\left(g_{ij}\right)'|_{s=0}\dot\sigma^i\dot\sigma^j+g_{ij}\dot\sigma^i\left(\dot\sigma^j\right)'|_{s=0}=\frac{1}{2}g_{ij,k}y^k\dot\gamma^i\dot\gamma^j+g_{ij}\dot\gamma^i\dot y^j.$$

Since Y vanishes at the endpoints, we can substitute into $E'_\sigma(0)$, and integrate by parts the second term to get:

$$E'_\sigma(0)=-\int_a^b\left[\frac{d}{dt}\left(g_{ij}\dot\gamma^i\right)-\frac{1}{2}g^{ik,j}\dot\gamma^i\dot\gamma^k\right]y^j.$$

Since:

$$\frac{d}{dt}\left(g_{ij}\dot\gamma^i\right)=g_{ij}\ddot\gamma^i+g_{ij,k}\dot\gamma^i\dot\gamma^k=g_{ij}\ddot\gamma^i+\frac{1}{2}\left(g_{ij,k}+g_{kj,i}\right)\dot\gamma^i\dot\gamma^k,$$

We now see that:

$$E'_\sigma(0)=-\int_a^b\left[g_{ij}\ddot\gamma^i+\frac{1}{2}\left(g_{mj,k}+g_{kj,m}-g_{mk,j}\right)\dot\gamma^m\dot\gamma^k\right]y^j\,dt=-\int_a^b g\left(\nabla_{\dot\gamma}\dot\gamma,Y\right)dt.$$

Since $E'_\sigma(0)=0$ for all vector fields Y along γ which vanish at the endpoints, we conclude that $\nabla_{\dot\gamma}\dot\gamma=0$, and γ is a geodesic.

The Schwartz inequality implies the following inequality between the length and energy functional for a curve γ.

***Lemma* 2.** For any curve γ, we have

$$L_\gamma^2\le 2E_\gamma(b-a),$$

with equality if and only if γ is parametrized proportionally to arc length.

Finally, the last lemma we state to prove Theorem 1, exhibits the relationship between the L and E functionals.

Lemma 3. A locally energy-minimizing curve is locally length-minimizing. Furthermore, if γ is locally length-minimizing and β is a reparametrization of γ by arc length, then β is locally energy-minimizing.

Proof. Suppose that γ is locally energy-minimizing, and let σ be a fixed endpoint variation of γ. For each s, let $\beta_s(t):[a,b]\to U$ be a reparametrization of the curve $t\mapsto\sigma(t;s)$ proportionally to arc length. Let $\tau(t;s)=\beta_s(t)$, then it is not difficult to see, using say the theorem on continuous dependence on parameters for ordinary differential equations, that τ is also smooth. By Lemma 1, γ is a geodesic, hence by Lemma 5, $L_\gamma^2=2E_\gamma(b-a)$. It follows that:

$$L_\sigma^2(0)=L_\gamma^2=2E_\gamma(b-a)=2E_\tau(0)(b-a)\le 2E_\tau(s)(b-a)=L_\tau^2(s)=L_\sigma^2(s).$$

Thus, γ is locally length-minimizing proving the first statement in the lemma.

Now suppose that γ is locally length-minimizing, and let β be a reparametrization of γ by arc length. Then β is also locally length-minimizing, hence for any fixed-endpoint variation σ of β, we have:

$$E_\sigma(0) = E_\beta = \frac{L_\beta^2}{2(b-a)} \le \frac{L_\sigma^2(s)}{2(b-a)} \le E_\sigma(s).$$

Thus, β is locally energy-minimizing.

We note that the same lemma holds if we replace locally energy-minimizing by energy-minimizing. The proof of Theorem 3 can now be easily completed with the help of Lemmas 1 and 3.

Proof of Theorem 1. Let β be a reparametrization of γ by arc length. By Lemma 3, β is locally energy-minimizing. By Lemma 1, β is a geodesic.

16.2 The Riemann Curvature Tensor

Definition 5. Let X, Y, Z, W be vector fields on a Riemannian surface (U, g). The Riemann curvature tensor is given by:

$$R(W, Z, X, Y) = g\left([\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z, W\right).$$

We first prove that R is indeed a tensor, i.e., it is linear over functions. Clearly, R is linear in W, additive in each of the other three variables, and anti-symmetric in X and Y . Thus, it suffices to prove the following lemma.

Lemma 4. Let X, Y, Z, W be vector fields on a Riemannian surface (U, g). Then we have:

$$R(W, Z, fX, Y) = R(W, fZ, X, Y) = fR(W, Z, X, Y).$$

Proof. We have:

$$\nabla_{fX}\nabla_Y Z - \nabla_Y\nabla_{fX}Z - \nabla_{[fX,Y]}Z = f\nabla_X\nabla_Y X - \nabla_Y(f\nabla_X Z) - \nabla_{f[X,Y]-(\partial Yf)X}Z$$

$$= f\nabla_X\nabla_Y Z - \left(\partial_Y f\right)\nabla_X Z - f\nabla_Y\nabla_X Z - f\nabla_{[X,Y]}Z + \left(\partial_Y f\right)\nabla_X Z$$

$$= f\left(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z\right).$$

The first identity follows by taking inner product with W. In order to prove the second identity, note that:

$$\nabla_X\nabla_Y fZ = \nabla_X\left(\left(\partial Yf\right)Z\right) + \nabla_X\left(f\nabla_Y Z\right) = \left(\partial_X\partial_Y f\right)Z + \left(\partial_Y f\right)\left(\nabla_X Z\right) + \left(\partial_X f\right)\left(\nabla_Y Z\right) + f\nabla_X\nabla_Y Z.$$

Interchanging X and Y and subtracting we get:

$$\left[\nabla_X, \nabla_Y\right]fZ = \left(\partial_{[X,Y]}f\right)Z + f[\nabla_X, \nabla_Y]Z.$$

On the other hand, we have also:

$$\nabla_{[X,Y]}fZ = \left(\partial_{[X,Y]}f\right)Z + f\nabla_{[X,Y]}Z.$$

Thus, we conclude:

$$\left[\nabla_X, \nabla_Y\right]fZ - \nabla_{[X,Y]}fZ = f\left([\nabla_X, \nabla_Y]Z - \nabla_{[X,Y]}Z\right).$$

Notes

The second identity now follows by taking inner product with W.

Let

$$R_{ijkl} = R\left(\partial_i, \partial_j, \partial_k, \partial_l\right),$$

be the components of the Riemann tensor. The previous proposition shows that if $X = x^i\partial_i$, $Y = y^i\partial_i, Z = z^i\partial_i, W = w^i\partial_i$, then

$$R(W, Z, X, Y) = w^i z^j x^k y^l R_{ijkl},$$

that is, the value of R(W, Z, X, Y) at a point u depends only on the values of W, Z, X, and Y at u.

Proposition 3. The components R_{ijkl} of the Riemann curvature tensor of any metric g satisfy the following identities:

$$R_{ijkl} = -R_{ijlk} = -R_{jikl} = R_{klij} \qquad ...(1)$$

$$R_{ijkl} + R_{iljk} + R_{iklj} = 0. \qquad ...(2)$$

Proof. We first prove (2). Since $\left[\partial_k, \partial_l\right] = 0$, it suffices to prove

$$\left[\nabla_k, \nabla_l\right]\partial_j + \left[\nabla_j, \nabla_k\right]\partial_l + \left[\nabla_l, \nabla_j\right]\partial_k = 0. \qquad ...(3)$$

Note that the symmetry $\Gamma^m_{lj} = \Gamma^m_{jl}$ imply that $\nabla_l\partial_j = \nabla_j\partial_l$.

Thus, we can write:

$$\left[\nabla_k, \nabla_l\right]\partial_j = \nabla_k\nabla_j\partial_l - \nabla_l\nabla_k\partial_j.$$

Permuting the indices cyclically, and adding, we get (3.19). The first identity in (3.23) is obvious from Definition 5. We now prove the identity:

$$R_{ijkl} = -R_{jikl}.$$

We observe that:

$$\begin{aligned} g\left(\nabla k\nabla_l\partial_j, \partial_i\right) &= \partial_k g\left(\nabla_l\partial_j, \partial_i\right) - \left(\nabla_l\partial_j, \nabla_k\partial_i\right) \\ &= \partial_k\left(\partial_l g_{ij} - g\left(\partial_j, \nabla_l\partial_i\right)\right) - \partial_l g\left(\partial_j, \nabla_k\partial_i\right) + g\left(\partial_j, \nabla_l\nabla_k\partial_i\right) \\ &= \partial_k\partial_l g_{ji} - \partial_k g\left(\partial_j, \nabla_l\partial_i\right) - \partial_l g\left(\partial_j, \nabla_k\partial_i\right) + g\left(\partial_j, \nabla_l\nabla_k\partial_i\right). \end{aligned}$$

It is easy to see that the first term, and the next two taken together, are symmetric in k and l. Thus, interchanging k and l, and subtracting, we get:

$$R_{ijkl} = g\left(\left[\nabla_k, \nabla_l\right]\partial_j, \partial_i\right) = g\left(\partial_j, \left[\nabla_l, \nabla_k\right]\partial_i\right) = -g\left(\left[\nabla_k, \nabla_l\right]\partial_i, \partial_j\right) = -R_{jikl}.$$

The last identity in (1) now follows from the first two and (2). We prove that $B_{ijkl} = R_{ijkl} - R_{klij} = 0$. Note that B_{ijkl} satisfies (1) as well as $B_{ijkl} = -B_{klij}$. Now, in view of the identities already established, we see that:

$$R_{ijkl} = -R_{iljk} - R_{iklj} = -R_{likj} - R_{iklj} = R_{ljik} + R_{lkji} - R_{iklj} = B_{ljik} + R_{klij},$$

Notes

hence, $B_{ijkl} = B_{ljik}$. Using the symmetries of B_{ijkl}, we can rewrite this identity as:

$$B_{ijkl} + B_{iklj} = 0. \qquad ...(4)$$

We now permute the first three indices cyclically:

$$B_{kijl} + B_{kjli} = 0, \qquad ...(5)$$

$$B_{jkil} + B_{jilk} = 0, \qquad ...(6)$$

add (4) to (5) and subtract (6) to get, using the symmetries of B_{ijkl}:

$$B_{ijkl} + B_{iklj} + B_{iklj} + B_{kjli} - B_{kjli} - B_{ijkl} = 2B_{iklj} = 0.$$

This completes the proof of the proposition.

It follows, that all the non-zero components of the Riemann tensor are determined by R_{1212}:

$$R_{1212} = -R_{2112} = R_{2121} = -R_{1221},$$

and all other components are zero. The proposition also implies that for any vectors X, Y, Z, W, the following identities hold:

$$R(W, Z, X, Y) = -R(W, Z, Y, Z) = -R(Z, W, X, Y) = R(X, Y, W, Z), \qquad ...(7)$$

$$R(W, Z, X, Y) + R(W, Y, Z, X) + R(W, X, Y, Z) = 0. \qquad ...(8)$$

Proposition 4. The components R_{ijkl} of the Riemann curvature tensor of any metric g satisfy:

$$g^{mj}R_{imkl} = \Gamma^j_{ik,l} - \Gamma^j_{il,k} + \Gamma^n_{ik}\Gamma^j_{nl} - \Gamma^n_{il}\Gamma^j_{nk}. \qquad ...(9)$$

Furthermore, we have:

$$K = \frac{R_{1212}}{\det(g)}, \qquad ...(10)$$

where K is the Gauss curvature of g.

Proof. Denote the right-hand side of (9) by S^j_{ikl}. We have:

$$\nabla_l \nabla_k \partial_i = \nabla_k \left(\Gamma^j_{ik}\partial_j\right) = \left(\Gamma^j_{ik,l} + \Gamma^n_{ik}\Gamma^j_{nl}\right)\partial_j,$$

or equivalently:

$$\Gamma^j_{ik,l} + \Gamma^n_{ik}\Gamma^j_{nl} = g^{jm}g\left(\nabla_l\nabla_k\partial_i, \partial_m\right).$$

Interchanging k and l and subtracting we get:

$$S^j_{ikl} = g^{jm}g\left(\left[\nabla_l, \nabla_k\right]\partial_i, \partial_m\right) = g^{jm}R_{milk} = g^{jm}R_{imkl}.$$

According (9), we have:

$$K = \frac{1}{2}g^{ik}S^j_{ikj} = \frac{1}{2}g^{ik}g^{il}R_{ijkl}.$$

In view of the comment following Proposition 8, the only non-zero terms in this sum are:

$$K = \frac{1}{2}\left(g^{11}g^{22}R_{1212} + g^{12}g^{21}R_{1221} + g^{21}g^{12}R_{2112} + g^{22}g^{11}R_{2121}\right) = \det\left(g^{-1}\right)R_{1212},$$

which implies (10)

Notes

Corollary 1. The Riemann curvature tensor of any metric g on a surface is given by:

$$R_{ijkl} = K\left(g_{ik}g_{il} - g_{i1}\, g_{jk}\right). \qquad ...(11)$$

Proof. Denote the right-hand side of (11) by S_{ijkl}, and note that it satisfies (1). Thus, the same comment which follows Proposition 8 applies and the only non-zero components of S_{ijkl} are determined by S_{1212}:

$$S_{1212} = -S_{2112} = S_{2121} = -S_{1221}.$$

In view of (11), we have $R_{1212} = S_{1212}$, thus it follows that $R_{ijkl} = S_{ijkl}$

In particular, we conclude that:

$$R(Z,W,X,Y) = K\big(g(W,X)g(Z,Y) - g(W,Y)g(Z,X)\big). \qquad ...(12)$$

16.3 The Second Variation of Arc length

In this section, we study the additional condition $E_\sigma'(0) \geq 0$ necessary for a minimum. This leads to the notion of Jacobi fields and conjugate points.

Proposition 5. Let $\gamma : [a,b] \to U$ be a geodesic parametrized by arc length on the Riemannian surface (U, g), and let σ be a fixed-endpoint variation of γ with generator Y. Then, we have:

$$E_\sigma''(0) = \int_a^b \left(\left|\nabla_{\dot\gamma} Y\right|^2 - K \circ \gamma\left(|Y|^2 - g(\dot\gamma, Y)^2\right)\right)dt, \qquad ...(13)$$

where K is the Gauss curvature of g.

Before we prove this proposition, we offer a second proof of the first variation formula:

$$E_\sigma'(0) = \int_a^b g\left(\nabla_{\dot\gamma}\dot\gamma, Y\right)dt, \qquad ...(14)$$

which is more in spirit with our derivation of the second variation formula. First note that if σ is a fixed-endpoint variation of γ with generator σ′ = Y, and with $\dot\sigma = X$, then [X, Y] = 0. Here Y denotes the vector field σ′ along σ rather than just along γ. Indeed, since X = dσ (d/dt) and Y = dσ(d/ds), it follows, that for any smooth function f on U, we have

$$\partial_{[X,Y]}f = \left[\frac{d}{dt}, \frac{d}{ds}\right] f \circ \sigma = 0.$$

In view of the symmetry $\Gamma^i_{jk} = \Gamma^i_{kj}$, this implies:

$$\nabla_Y X - \nabla_X Y = [X,Y] = 0.$$

We can now calculate:

$$\begin{aligned} E_\sigma'(s) &= \frac{1}{2}\int \partial_Y g(X,X)dt = \int_a^b g\left(\nabla_Y X, X\right)dt = \int_a^b g\left(\nabla X_Y, X\right)dt \\ &= \int_a^b \frac{d}{dt} g(Y,X)dt - \int_a^b g\left(Y, \nabla_X X\right)dt = g(Y,X)\Big|_a^b - \int_a^b g\left(Y, \nabla_X X\right)dt \end{aligned}$$

Notes

Setting s = 0, (14) follows.

Proof of Proposition 5. We compute:

$$
\begin{aligned}
E''_\sigma &= \frac{1}{2}\int_a^b \partial_Y \partial_{Y_g}(X,X)dt = \int_a^b \partial_{Y_g}(\nabla_Y X, X)dt = \int_a^b \partial_{Y_g}(\nabla_X Y, X)dt \\
&= \int_a^b \left(g(\nabla_Y \nabla_X Y, X) + g(\nabla_X Y, \nabla_Y X)\right)dt \\
&= \int_a^b \left(g(\nabla_X \nabla_Y Y, X) + g([\nabla_Y, \nabla_X]Y, X) + g(\nabla_X Y, \nabla_X Y)\right)dt \\
&= \int_a^b \left(\frac{d}{dt} g(\nabla_Y Y, X) - g(\nabla_Y Y, {}_X X) + R(X,Y,Y,X) + g(\nabla_X Y, \nabla_X Y)\right)dt,
\end{aligned}
$$

where as above $X = \dot{\sigma}$, and Y = σ′. Now, the first term integrates to $g(\nabla_Y Y, X)\big|_a^b = 0$, and when we set s = 0, the second term also vanishes since $\nabla_X X = \nabla_{\dot{\gamma}}\dot{\gamma} = 0$. Furthermore, the last term becomes $g(\nabla_{\dot{\gamma}} Y, \nabla_{\dot{\gamma}} Y)$. Hence, we conclude:

$$E''_\sigma(0) = \int_a^b \left(\left|\nabla_{\dot{\gamma}} Y\right|^2 - R(X,Y,X,Y)\right)dt. \qquad \text{...(15)}$$

The proposition now follows from (12).

Thus, $E''_\sigma(0)$ can be viewed as a quadratic form in the generator Y . The corresponding symmetric bilinear form is called the index form of γ:

$$I(Y,Z) = \int_a^b \left(g(\nabla_{\dot{\gamma}} Y, \nabla_{\dot{\gamma}} Z) - K \circ \gamma \left(g(Y,Z) - g(\dot{\gamma}, Y) g(\dot{\gamma}, Z)\right)\right)dt.$$

It is the Hessian of the functional E, and if E has a local minimum, I is positive semi-definite. We will also write I(Y) = I(Y, Y).

Definition 6. Let γ be a geodesic parametrized by arc length on the Riemannian surface (U, g). A vector field Y along γ is called a Jacobi field, if it satisfies the following differential equation:

$$\nabla_{\dot{\gamma}} \nabla_{\dot{\gamma}} Y + K\left(Y - g(\dot{\gamma}, Y)\dot{\gamma}\right) = 0.$$

Two points γ (a) and γ (b) along a geodesic γ are called conjugate along γ if there is a non-zero Jacobi field along γ which vanishes at those two points.

The Jacobi field equation is a linear system of second-order differential equations. Hence given initial data specifying the initial value and initial derivative of Y, a unique solution exists along the entire geodesic γ.

Proposition 6. Let γ be a geodesic on the Riemannian surface (U, g). Then given two vectors $Z_1, Z_2 \in T_{\gamma(a)}U$, there is a unique Jacobi field Y along γ such that Y(a) = Z_1, and $\nabla_{\dot{\gamma}} Y(a) = Z_2$.

In particular, any Jacobi field which is tangent to γ is a linear combination of $\dot{\gamma}$ and $t\dot{\gamma}$. The significance of Jacobi fields is seen in the following two propositions. We say that σ is a variation of γ through geodesics if the curves $t \mapsto \sigma(t;s)$ are geodesics for all s.

Proposition 7. Let γ be a geodesic, and let σ be a variation of γ through geodesics. Then the generator $Y = \sigma'$ of σ is a Jacobi field.

Proof. As before, denote $X = \dot{\sigma}$ and $Y = \sigma'$. We first prove the following identity:

$$[\nabla_Y, \nabla_X]X = -K(Y - g(X,Y)X).$$

Indeed, in the proof of Lemma 7, it was seen that the left-hand side above is a tensor, i.e., is linear over functions, and hence depends only on the values of the vector fields X and Y at one point. Fix that point. If X and Y are linearly dependent, then both sides of the equation above are zero. Otherwise, X and Y are linearly independent, and it suffices to check the inner product of the identity against X and Y. Taking inner product with X, both sides are zero, and equation (11) implies that the inner products with Y are equal. Since $\nabla_X X = 0$, we get:

$$0 = \nabla_Y \nabla_X X = \nabla_X \nabla_Y X + [\nabla_Y, \nabla_X]X = \nabla_X \nabla_X Y - K(Y - g(X,Y)X).$$

Thus, Y is a Jacobi field.

We see that Jacobi fields are infinitesimal generators of variations through geodesics. If there is a non-trivial fixed endpoint variation of γ through geodesics, then the endpoints of γ are conjugate along γ. Unfortunately, the converse is not true but nevertheless, a non-zero Jacobi field which vanishes at the endpoints can be perceived as a non-trivial infinitesimal fixed-endpoint variation of γ through geodesics. This makes the next proposition all the more important.

Proposition 8. Let γ be a geodesic, and let Y be a Jacobi field. Then, for any vector field Z along γ, we have:

$$I(Y,Z) = g\left(\nabla_{\dot{\gamma}} Y, Z\right)\Big|_a^b. \qquad \text{...(16)}$$

In particular, if either Y or Z vanishes at the endpoints, then I(Y,Z) = 0.

Proof. Multiplying the Jacobi equation by Z and integrating, we obtain:

$$0 = \int_a^b \left(g\left(\nabla_{\dot{\gamma}} \nabla_{\dot{\gamma}} Y, Z\right) - K\left(g(Y,Z) - g(\dot{\gamma}, Y)g(\dot{\gamma}, Z)\right)\right) dt$$

$$= \int_a^b \left(\frac{d}{dt} g\left(\nabla_{\dot{\gamma}} Y, Z\right) - g\left(\nabla_{\dot{\gamma}} Y, \nabla_{\dot{\gamma}} Z\right) - K\left(g(Y,Z) - g(\dot{\gamma}, Y)g(\dot{\gamma}, Z)\right)\right) dt$$

Thus, a Jacobi field which vanishes at the endpoints lies in the null space of the index form I acting on vector fields which vanish at the endpoints.

Theorem 2. Let $\gamma : [a,b] \to (U,g)$ be a geodesic parametrized by arc length, and suppose that there is a point $\gamma(c)$ with $a < c < b$ which is conjugate to $\gamma(a)$. Then there is a vector field Z along γ such that $I(Z) < 0$. Consequently, γ is not locally-length minimizing.

Proof. Define:

$$V = \begin{cases} Y & a \le t \le c \\ 0 & c \le t \le b \end{cases}$$

and let W be a vector field supported in a small neighborhood of c which satisfies $W(c) = -\nabla_{\dot{\gamma}} Y(c) \neq 0$. We denote the index form of γ on [a, c] by I_1, and the index form on [c, b] by I_2. Since V is piecewise smooth, we have, in view of (16):

$$I(V,W) = I_1(V,W) + I_2(V,W) = I_1(Y,W) = -\left|\nabla_{\dot\gamma} Y(c)\right|^2 < 0$$

It follows that:

$$I(V + \varepsilon W, V + \varepsilon W) = I(V) + 2\varepsilon I(V,W) + \varepsilon^2 I(W) = 2\varepsilon I(V,W) + \varepsilon^2 I(W)$$

is negative if $\varepsilon > 0$ is small enough. Although $V + \varepsilon W$ is not smooth, there is for any $\delta > 0$ a smooth vector field Z_δ, satisfying $|Y|^2 + \left|\nabla_{\dot\gamma} Z_\delta\right|^2 \le C$ uniformly in $\delta > 0$, which differs from $V + \varepsilon W$ only on $(c - \delta, c + \delta)$. Since the contribution of this interval to both $I(V + \varepsilon W, V + \varepsilon W)$ and $I(Z_\delta, Z_\delta)$ tends to zero with δ, it follows that also $I(Z_\delta, Z_\delta) < 0$ for $\delta > 0$ small enough. Thus, γ is not locally energy-minimizing. Since it is parametrized by arc length, if it was locally length minimizing, it would by Lemma 6 also be locally energy-minimizing. Thus, γ cannot be locally length-minimizing.

A partial converse is also true: the absence of conjugate points along γ guarantees that the index form is positive definite.

Theorem 3. Let $\gamma : [a,b] \to (U,g)$ be a geodesic parametrized by arc length, and suppose that no point $\gamma(t)$, $a < t \le b$, is conjugate to $\gamma(a)$ along γ. Then the index form I is positive definite.

Proof. Let $X = \dot\sigma$, and let Y be a Jacobi field which is perpendicular to X, and vanishes at t = a. Note that the space of such Jacobi fields is 1-dimensional, hence Y is determined up to sign if we also require that $\left|\dot Y(a)\right| = 1$. Since Y is perpendicular to X, it satisfies the equation:

$$\nabla_X \nabla_X Y + KY = 0.$$

Furthermore, since Y never vanishes along γ, the vectors X and Y span $T_{\gamma(t)}U$ for all $t \in (a, b]$. Thus, if Z is any vector field along γ which vanishes at the endpoints, then we can write $Z = fX + hY$ for some functions f and h. Note that $f(a) = f(b) = h(b) = 0$ and $hY(a) = 0$. We then have:

$$I(Z,Z) = I(fX, fX) + 2I(fX, hY) + I(hY, hY).$$

Since $R(X, fX, X, fX) = 0$ and $\nabla_X fX = \dot f X$, it follows from (3.31) that:

$$I(fX, fX) = \int_a^b g\left(\dot f X, \dot f X\right) dt = \int_a^b \dot f^2 dt.$$

Furthermore,

$$\begin{aligned} I(fX, hY) &= \int_a^b g\left(\dot f X, \nabla_X hY\right) dt \\ &= g\left(\dot f X, hY\right)\Big|_a^b - \int_a^b g\left(\nabla_X \dot f X, hY\right) dt = -\int_a^b g\left(\ddot f X, hY\right) dt = 0. \end{aligned}$$

Finally, since $|\nabla_X hY|^2 = g\left(\nabla_X Y, \nabla_X h^2 Y\right) + \dot h^2 |Y|^2$, it follows from Proposition 3.20 that:

$$I(hY, hY) = \int_a^b \dot h^2 |Y|^2 \, dt + I(Y, hY) = \int_a^b h^2 |Y|^2 \, dt.$$

Thus, we conclude that:

Notes

$$I(Z,Z) = \int_a^b \left(\dot{f}^2 + \dot{h}^2 |Y|^2\right) dt \geq 0.$$

If I(Z,Z) = 0, then $\dot{f} = 0$ and $\dot{h}Y = 0$ on [a, b]. Since $Y \neq 0$ on (a, b], we conclude that $\dot{h} = 0$ on (a, b], and in view of h(b) = f(b) = 0, we get that Z = 0. Thus, I is positive definite.

16.4 Summary

- Let (U, g) be a Riemannian surface, and let $\gamma : I \to U$ be a curve. A vector field along γ is a smooth function $Y : I \to \mathbb{R}^2$. The covariant derivative of $Y = y^i \partial_i$ along γ is the vector field:

  $$\nabla_{\dot{\gamma}} Y = \left(\dot{y}^i + \Gamma^i_{jk} y^j \dot{\gamma}^k\right) \partial_i.$$

  Note that if Z is any extension of Y, i.e., a any vector field defined on a neighborhood V of the image $\gamma(I)$ of γ in U, then we have:

  $$\nabla_{\dot{\gamma}} Y = \nabla_{\dot{\gamma}} Z = \dot{\gamma}^i Z_{;i}.$$

- A vector field Y along a curve γ is said to be parallel along γ if $\nabla_{\dot{\gamma}} Y = 0$.

  Note that if Y and Z are parallel along γ, then g(Y,Z) is constant.

  $$\partial_{\dot{\gamma}} g(Y,Z) = g\left(\nabla_{\dot{\gamma}} Y, Z\right) + g\left(Y, \nabla_{\dot{\gamma}} Z\right) = 0.$$

- Let $\gamma : [a,b] \to U$ be a curve into the Riemannian surface (U, g), let $u_0 \in U$, and let $Y_0 \in T_{u_0} U$. Then there is a unique vector field Y along γ which is parallel along γ and satisfies $Y(a) = Y_0$.

- A curve γ is a geodesic if its tangent $\dot{\gamma}$ is parallel along γ:

  $$\nabla_{\dot{\gamma}} \dot{\gamma} = 0.$$

  If γ is a geodesic, then $|\dot{\gamma}|$ is constant and hence, every geodesic is parametrized proportionally to arc length. In particular, if $\beta = \gamma \circ \phi$ is a reparametrization of γ, then β is not a geodesic unless ϕ is a linear map.

- Let (U, g) be a Riemannian surface, let $u_0 \in U$ and let $0 \neq Y_0 \in T_{u0}U$. Then there is and $\varepsilon > 0$, and a unique geodesic $\gamma : (-\varepsilon, \varepsilon) \to U$, such that $\gamma(0) = u_0$, and $\dot{\gamma}(0) = Y_0$.

- Let $\gamma : [a,b] \to U$ be a curve. We say that γ is length minimizing, or L-minimizing, if:

  $$L_\gamma \leq L_\beta$$

  for all curves β in U such that $\beta(a) = \gamma(a)$ and $\beta(b) = \gamma(b)$.

16.5 Keywords

Geodesic: A locally energy-minimizing curve is a geodesic.

Schwartz inequality: The Schwartz inequality implies the following inequality between the length and energy functional for a curve γ.

16.6 Self Assessment

1. If Z is any extension of Y, i.e., a any vector field defined on a neighborhood V of the image $\gamma(I)$ of γ in U, then we have ....
2. A vector field Y along a curve γ is said to be parallel along ....
3. Let (U, g) be a Riemannian surface, let $u_0 \in U$ and let $0 \neq Y_0 \in T_{u0}U$. Then there is and $\varepsilon > 0$, and a unique geodesic ...., such that $\gamma(0) = u_0$, and $\dot{\gamma}(0) = Y_0$.
4. A locally length-minimizing curve has a ....
5. The .... implies the following inequality between the length and energy functional for a curve γ.
6. Let γ be a ...., and let Y be a Jacobi field. Then, for any vector field Z along γ, we have $I(Y, Z) = g\left(\nabla_{\dot{\gamma}} Y, Z\right)\Big|_a^b$.

16.7 Review Questions

1. Let $\tilde{g} = e^{2\lambda} g$ be conformal metrics on U, and let Γ_{ij}^k and $\tilde{\Gamma}_{ij}^k$ be their Christoffel symbols. Prove that:

$$\tilde{\Gamma}_{ij}^k = \Gamma_{ij}^k + \delta_i^k \lambda_j + \delta_j^k \lambda_i + g_{ij} g^{km} \lambda_m$$

2. Let $\tilde{g}$ and g be two conformal metrics on $U, \tilde{g} = e^{2\lambda} g$, and let K and $\tilde{K}$ be their Gauss curvatures. Prove that:

$$\tilde{K} = e^{-2\lambda}(K - \Delta\lambda).$$

Answers: Self Assessment

1.	$\nabla_{\dot{\gamma}} Y = \nabla_{\dot{\gamma}} Z = \dot{\gamma}^i Z_{;i}$.	2.	γ if $\nabla_{\dot{\gamma}} Y = 0$.
3.	$\gamma : (-\varepsilon, \varepsilon) \to U$,	4.	Geodesic reparametrization
5.	Schwartz inequality	6.	Geodesic

16.8 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati,T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 17: Geodesic Curvature and Christoffel Symbols

CONTENTS

Objectives

After studying this unit, you will be able to:

- Explain the Gauss Map and its Derivative dN
- Define the Dupin Indicatrix
- Describe the theorema Egregium of Gauss, the Equations of Codazzi-Mainardi, and Bonnet's Theorem
- Define Lines of Curvature, Geodesic Torsion, Asymptotic Lines

Notes

Introduction

In this unit, we focus exclusively on the study of Geoderic Curvature. In this unit, we will go through the properties of the curvature of curves on a surface. The study of the normal and of the tangential components of the curvature will lead to the normal curvature and to the geodesic curvature. We will study the normal curvature, and this will lead us to principal curvatures, principal directions, the Gaussian curvature, and the mean curvature. In turn, the desire to express the geodesic curvature in terms of the first fundamental form alone will lead to the Christoffel symbols. The study of the variation of the normal at a point will lead to the Gauss map and its derivative, and to the Weingarten equations.

17.1 Geodesic Curvature and the Christoffel Symbols

We showed that the tangential part of the curvature of a curve C on a surface is of the form $k_g \vec{n}_g$.

We now show that k_n can be computed only in terms of the first fundamental form of X, a result first proved by Ossian Bonnet circa 1848.

The computation is a bit involved, and it will lead us to the Christoffel symbols, introduced in 1869.

Since $\vec{n}_g$ is in the tangent space $T_p(X)$, and since (X_u, X_v) is a basis of $T_p(X)$, we can write

$$k_g \vec{n}_g = AX_u + BX_v,$$

form some A, B $\in \mathbb{R}$.

However,

$$k\vec{n} = k_N N + k_g \vec{n}_g,$$

and since N is normal to the tangent space,

$N \cdot X_u = N \cdot X_v = 0$, and by dotting

$$k_g \vec{n}_g = AX_u + BX_v$$

with X_u and X_v, since $E = X_u \cdot X_u$, $F = X_u \cdot X_v$, and $G = X_v \cdot X_v$, we get the equations:

$$k\vec{n} \cdot X_u = EA + FB,$$

$$k\vec{n} \cdot X_v = FA + GB.$$

On the other hand,

$$k\vec{n} = X'' = X_u u'' + X_v v'' + X_{uu}(u')^2 + 2X_{uv}u'v' + X_{vv}(v')^2.$$

Dotting with X_u and X_v, we get

$$k\vec{n} \cdot X_u = Eu'' + Fv'' + (X_{uu} \cdot X_u)(u')^2 + 2(X_{uv} \cdot X_u)u'v' + (X_{vv} \cdot X_u)(v')^2,$$

$$k\vec{n} \cdot Xv = Fu'' + Gv'' + (X_{uu} \cdot X_v)(u')^2 + 2(X_{uv} \cdot X_v)u'v' + (X_{vv} \cdot X_v)(v')^2.$$

Notes

At this point, it is useful to introduce the Christoffel symbols (of the first kind) [α β; γ], defined such that

$$[\alpha\ \beta;\ \gamma] = X_{\alpha\beta} \cdot X_{\gamma},$$

where $\alpha, \beta, \gamma \in \{u, v\}$. It is also more convenient to let $u = u_1$ and $v = u_2$, and to denote $[u_\alpha\ v_\beta;\ u_\gamma]$ as $[\alpha\ \beta;\ \gamma]$.

Doing so, and remembering that

$$k\vec{n} \cdot X_u = EA + FB,$$

$$k\vec{n} \cdot X_v = FA + GB,$$

we have the following equation:

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}\begin{pmatrix} u_1'' \\ u_2'' \end{pmatrix} + \sum_{\substack{\alpha=1,2 \\ \beta=1,2}} \begin{pmatrix} [\alpha\beta;1]u_\alpha' u_\beta' \\ [\alpha\beta;2]u_\alpha' u_\beta' \end{pmatrix}.$$

However, since the first fundamental form is positive definite,

$EG - F^2 > 0$, and we have

$$\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} = (EG - F^2)^{-1}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}.$$

Thus, we get

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} u_1'' \\ u_2'' \end{pmatrix} + \sum_{\substack{\alpha=1,2 \\ \beta=1,2}} (EG - F^2)^{-1}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}\begin{pmatrix} [\alpha\beta;1]u_\alpha' u_\beta' \\ [\alpha\beta;2]u_\alpha' u_\beta' \end{pmatrix}.$$

It is natural to introduce the Christoffel symbols (of the second kind) Γ_{ij}^{k}, defined such that

$$\begin{pmatrix} \Gamma_{ij}^1 \\ \Gamma_{ij}^2 \end{pmatrix} = (EG - F^2)^{-1}\begin{pmatrix} G & -F \\ -F & E \end{pmatrix}\begin{pmatrix} [ij;\ 1] \\ [ij;\ 2] \end{pmatrix}.$$

Finally, we get

$$A = u_1'' + \sum_{\substack{i=1,2 \\ j=1,2}} G_{ij}^1 u_i' u_j',$$

$$B = u_2'' + \sum_{\substack{i=1,2 \\ j=1,2}} \Gamma_{ij}^2 u_i' u_j',$$

Lemma 1. Given a surface X and a curve C on X, for any point p on C, the tangential part of the curvature at p is given by

$$k_g \vec{n}_g = \left(u_1'' + \sum_{\substack{i=1,2 \\ j=1,2}} \Gamma_{ij}^1 u_i' u_j' \right) X_u + \left(u_2'' + \sum_{\substack{i=1,2 \\ j=1,2}} \Gamma_{ij}^2 u_i' u_j' \right) X_v,$$

Notes

where the Christoffel symbols Γ_{ij}^{k} are defined such that

$$\begin{pmatrix}\Gamma_{ij}^{1}\\ \Gamma_{ij}^{2}\end{pmatrix} = \begin{pmatrix}E & F\\ F & G\end{pmatrix}^{-1}\begin{pmatrix}[ij;1]\\ [ij;2]\end{pmatrix},$$

and the Christoffel symbols [i j; k] are defined such that

$$[i\ j;\ k] = X_{ij} \cdot X_k.$$

Note that

$$[i\ j;\ k] = [j\ i;\ k] = X_{ij} \cdot X_k.$$

Looking at the formulae

$$[\alpha\ \beta;\ \gamma] = X_{\alpha\beta} \cdot X_\gamma$$

for the Christoffel symbols [α β; γ], it does not seem that these symbols only depend on the first fundamental form, but in fact they do!

After some calculations, we have the following formulae showing that the Christoffel symbols only depend on the first fundamental form:

$$[1\,1;\,1] = \frac{1}{2}Eu, \qquad [1\,1;\,2] = F_u - \frac{1}{2}E_v,$$

$$[1\,2;\,1] = \frac{1}{2}E_v, \qquad [1\,2;\,2] = \frac{1}{2}G_u,$$

$$[2\,1;\,1] = \frac{1}{2}E_v, \qquad [2\,1;\,2] = \frac{1}{2}G_u,$$

$$[2\,2;\,1] = Fv - \frac{1}{2}G_u, \qquad [2\,2;\,2] = \frac{1}{2}G_u.$$

Another way to compute the Christoffel symbols [α β; γ], is to proceed as follows. For this computation, it is more convenient to assume that $u = u_1$ and $v = u_2$, and that the first fundamental form is expressed by the matrix

$$\begin{pmatrix}g_{11} & g_{12}\\ g_{21} & g_{22}\end{pmatrix} = \begin{pmatrix}E & F\\ F & G\end{pmatrix},$$

where $g_{\alpha\beta} = X_\alpha \cdot X_\beta$. Let

$$g_{\alpha\beta|\gamma} = \frac{\partial g_{\alpha\beta}}{\partial u_\gamma}.$$

Then, we have

$$g_{\alpha\beta|\gamma} = \frac{\partial g_{\alpha\beta}}{\partial u_\gamma} = X_{\alpha\gamma} \cdot X_\beta + X_\alpha \cdot X_{\beta\gamma} = [\alpha\gamma;\beta] + [\beta\gamma;\alpha].$$

From this, we also have

$$g_{\beta\gamma|\alpha} = [\alpha\ \beta;\ \gamma] + [\alpha\ \gamma;\ \beta],$$

Notes

and

$$g_{\alpha\gamma|\beta} = [\alpha\ \beta;\ \gamma] + [\beta\ \gamma;\ \alpha].$$

From all this, we get

$$2[\alpha\ \beta;\ \gamma] = g_{\alpha\gamma|\beta} + g_{\beta\gamma|\alpha} - g_{\alpha\beta|\gamma}.$$

As before, the Christoffel symbols $[\alpha\ \beta;\ \gamma]$ and $\Gamma^{\gamma}_{\alpha\beta}$ are related via the Riemannian metric by the equations

$$\Gamma^{\gamma}_{\alpha\beta} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}^{-1} [\alpha\beta;\gamma].$$

This seemingly bizarre approach has the advantage to generalize to Riemannian manifolds.

17.2 Principal Curvatures, Gaussian Curvature, Mean Curvature

We will now study how the normal curvature at a point varies when a unit tangent vector varies.

In general, we will see that the normal curvature has a maximum value k_1 and a minimum value k_2, and that the corresponding directions are orthogonal. This was shown by Euler in 1760.

The quantity $K = k_1k_2$ called the Gaussian curvature and the quantity $H = (k_1 + k_2)/2$ called the mean curvature, play a very important role in the theory of surfaces.

We will compute H and K in terms of the first and the second fundamental form. We also classify points on a surface according to the value and sign of the Gaussian curvature.

Recall that given a surface X and some point p on X, the vectors X_u, X_v form a basis of the tangent space $T_p(X)$.

Given a unit vector $\vec{t} = X_u x + X_v y$, the normal curvature is given by

$$k_N(\vec{t}) = Lx^2 + 2Mxy + Ny^2,$$

since $Ex^2 + 2Fxy + Gy^2 = 1$.

Usually, (X_u,X_v) is not an orthonormal frame, and it is useful to replace the frame (X_u,X_v) with an orthonormal frame.

One verifies easily that the frame $(\vec{e}_1,\vec{e}_2)$ defined such that

$$\vec{e}_1 = \frac{X_u}{\sqrt{E}},\ \vec{e}_2 = \frac{EX_v - FX_u}{\sqrt{E(EG - F^2)}}.$$

is indeed an orthonormal frame.

With respect to this frame, every unit vector can be written as $\vec{t} = \cos\theta\,\vec{e}_1 + \sin\theta\,\vec{e}_2$, and expressing $(\vec{e}_1,\vec{e}_2)$ in terms of X_u and X_v, we have

$$\vec{t} = \left(\frac{w\cos\theta - F\sin\theta}{w\sqrt{E}}\right)X_u + \frac{\sqrt{E}\sin\theta}{w}X_v,$$

where $w = \sqrt{EG - F^2}$.

Notes

We can now compute $K_N(\vec{t})$, and we get

$$k_N(\vec{t}) = L\left(\frac{w\cos\theta - F\sin\theta}{w\sqrt{E}}\right)^2 + 2M\left(\frac{(w\cos\theta - F\sin\theta)\sin\theta}{w^2}\right) + N\frac{E\sin^2\theta}{w^2}.$$

We leave as an exercise to show that the above expression can be written as

$$k_N(\vec{t}) = H + A\cos 2\theta + B\sin 2\theta,$$

where

$$H = \frac{GL - 2FM + EN}{2(EG - F^2)},$$

$$A = \frac{L(EG - 2F^2) + 2EFM - E^2N}{2E(EG - F^2)}$$

$$B = \frac{EM - FL}{E\sqrt{EG - F^2}}.$$

Letting $C = \sqrt{A^2 + B^2}$, unless A = B = 0, the function

$$f(\theta) = H + A\cos 2\theta + B\sin 2\theta$$

has a maximum $k_1 = H + C$ for the angles θ_0 and $\theta_0 + \pi$, and a minimum $k_2 = H - C$ for the angles $\theta_0 + \frac{\pi}{2}$ and $\theta_0 + \frac{3\pi}{2}$, where $\cos 2\theta_0 = \frac{A}{C}$ and $\sin 2\theta_0 = \frac{B}{C}$.

The curvatures k_1 and k_2 play a major role in surface theory.

Definition 1: Given a surface X, for any point p on X, letting A, B, H be defined as above, and $C = \sqrt{A^2 + B^2}$, unless A = B = 0, the normal curvature k_N at p takes a maximum value k_1 and a minimum value k_2 called principal curvatures at p, where $k_1 = H + C$ and $k_2 = H - C$. The directions of the corresponding unit vectors are called the principal directions at p.

The average $H = \frac{k_1 + k_2}{2}$ of the principal curvatures is called the mean curvature, and the product $K = k_1k_2$ of the principal curvatures is called the total curvature, or Gaussian curvature.

Observe that the principal directions θ_0 and $\theta_0 + \frac{\pi}{2}$ corresponding k_1 and k_2 are orthogonal.

Notes $K = k_1k_2 = (H - C)(H + C) = H^2 - C^2 = H^2 - (A^2 + B^2).$

After some laborious calculations, we get the following (famous) formulae for the mean curvature and the Gaussian curvature:

$$H = \frac{GL - 2FM + EN}{2(EG - F^2)},$$

$$K = \frac{LN - M^2}{EG - F^2}.$$

We showed that the normal curvature k_N can be expressed as

$$k_N(\theta) = H + A\cos 2\theta + B \sin 2\theta$$

over the orthonormal frame $(\vec{e}_1, \vec{e}_2)$.

We also showed that the angle θ_0 such that $\cos 2\theta_0 = \frac{A}{C}$ and $\sin 2\theta_0 = \frac{B}{C}$, plays a special role.

Indeed, it determines one of the principal directions.

If we rotate the basis $(\vec{e}_1, \vec{e}_2)$ and pick a frame $(\vec{f}_1, \vec{f}_2)$ corresponding to the principal directions, we obtain a particularly nice formula for k_N. Indeed, since $A = C \cos 2\theta_0$ and $B = C \sin 2\theta_0$, letting $\varphi = \theta - \theta_0$, we get

$$k_N(\theta) = k_1 \cos^2 \varphi + k_2 \sin^2 \varphi.$$

Thus, for any unit vector $\vec{t}$ expressed as

$$\vec{t} = \cos \varphi\, \vec{f}_1 + \sin \varphi\, \vec{f}_2$$

with respect to an orthonormal frame corresponding to the principal directions, the normal curvature $k_N(\varphi)$ is given by

Euler's formula (1760):

$$kN(\varphi) = k_1 \cos^2 \varphi + k_2 \sin^2 \varphi.$$

Recalling that $EG - F^2$ is always strictly positive, we can classify the points on the surface depending on the value of the Gaussian curvature K, and on the values of the principal curvatures k_1 and k_2 (or H).

Definition 2: Given a surface X, a point p on X belongs to one of the following categories:

(1) Elliptic if $LN - M^2 > 0$, or equivalently $K > 0$.

(2) Hyperbolic if $LN - M^2 < 0$, or equivalently $K < 0$.

(3) Parabolic if $LN - M^2 = 0$ and $L^2 + M^2 + N^2 > 0$, or equivalently $K = k_1k_2 = 0$ but either $k_1 \neq 0$ or $k_2 \neq 0$.

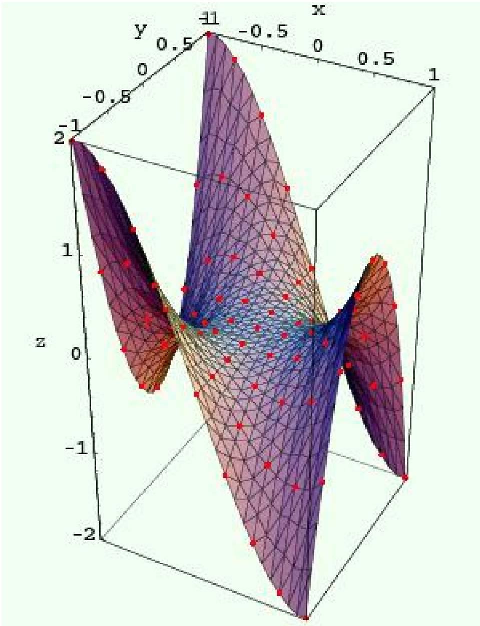
(4) Planar if $L = M = N = 0$, or equivalently $k_1 = k_2 = 0$.

Furthermore, a point p is an umbilical point (or umbilic) if $K > 0$ and $k_1 = k_2$.

- At an elliptic point, both principal curvatures are non-null and have the same sign. For example, most points on an ellipsoid are elliptic.
- At a hyperbolic point, the principal curvatures have opposite signs. For example, all points on the catenoid are hyperbolic.
- At a parabolic point, one of the two principal curvatures is zero, but not both. This is equivalent to $K = 0$ and $H \neq 0$. Points on a cylinder are parabolic.
- At a planar point, $k_1 = k_2 = 0$. This is equivalent to $K = H = 0$. Points on a plane are all planar points! On a monkey saddle, there is a planar point. The principal directions at that point are undefined.

Notes

Figure 17.1: A Monkey Saddle



For an umbilical point, we have $k_1 = k_2 \neq 0$.

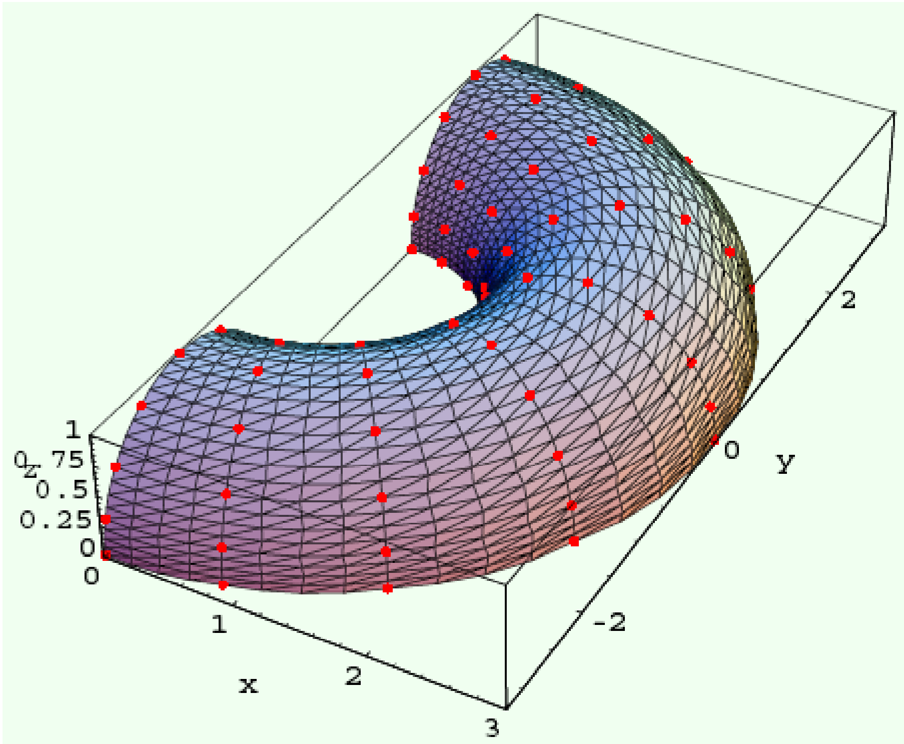
This can only happen when $H - C = H + C$, which implies that $C = 0$, and since $C = \sqrt{A^2 + B^2}$, we have $A = B = 0$.

Thus, for an umbilical point, $K = H^2$. In this case, the function k_N is constant, and the principal directions are undefined. All points on a sphere are umbilics. A general ellipsoid (a, b, c pairwise distinct) has four umbilics. It can be shown that a connected surface consisting only of umbilical points is contained in a sphere. It can also be shown that a connected surface consisting only of planar points is contained in a plane.

A surface can contain at the same time elliptic points, parabolic points, and hyperbolic points. This is the case of a torus.

- The parabolic points are on two circles also contained in two tangent planes to the torus (the two horizontal planes touching the top and the bottom of the torus on the following picture).
- The elliptic points are on the outside part of the torus (with normal facing outward), delimited by the two parabolic circles.
- The hyperbolic points are on the inside part of the torus (with normal facing inward).

Figure 17.2: Portion of Torus



The normal curvature

$$k_N(X_u x + X_v y) = Lx^2 + 2Mxy + Ny^2$$

will vanish for some tangent vector $(x, y) \neq (0, 0)$ iff $M^2 \geq LN \geq 0$.

Since

$$K = \frac{LN - M^2}{EG - F^2},$$

this can only happen if $K \leq 0$.

If $L = N = 0$, then there are two directions corresponding to X_u and X_v for which the normal curvature is zero.

If $L \neq 0$ or $N \neq 0$, say $L \neq 0$ (the other case being similar), then the equation $L\left(\frac{x}{y}\right)^2 + 2M\frac{x}{y} + N = 0$

has two distinct roots iff $K < 0$.

The directions corresponding to the vectors $X_u x + X_v y$ associated with these roots are called the asymptotic directions at p. These are the directions for which the normal curvature is null at p. There are surfaces of constant Gaussian curvature. For example, a cylinder or a cone is a surface of Gaussian curvature $K = 0$. A sphere of radius R has positive constant Gaussian curvature $K = \frac{1}{R^2}$.

Notes

Perhaps surprisingly, there are other surfaces of constant positive curvature besides the sphere. There are surfaces of constant negative curvature, say K = –1. A famous one is the pseudosphere, also known as Beltrami's pseudosphere. This is the surface of revolution obtained by rotating a curve known as a tractrix around its asymptote. One possible parameterization is given by:

$$x = \frac{2\cos v}{e^u + e^{-u}},$$

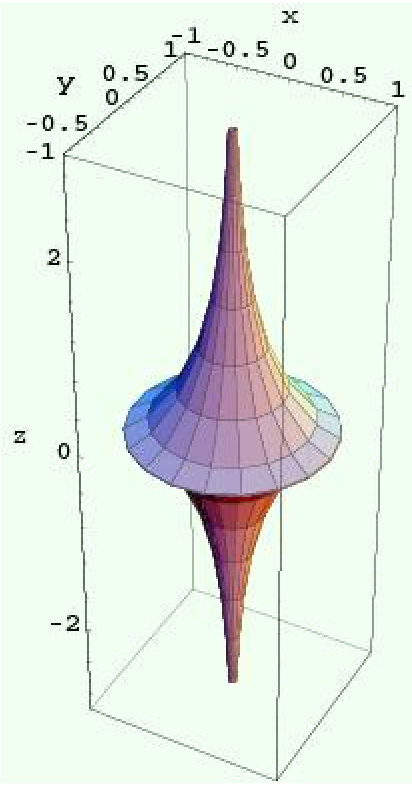
$$y = \frac{2\sin v}{e^u + e^{-u}},$$

$$z = u - \frac{e^u - e^{-u}}{e^u + e^{-u}},$$

over]0, 2π[× ℝ.

The pseudosphere has a circle of singular points (for u = 0). The figure below shows a portion of pseudosphere.

Figure 17.3: A Pseudosphere



Again, perhaps surprisingly, there are other surfaces of constant negative curvature.

The Gaussian curvature at a point (x, y, x) of an ellipsoid of equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

has the beautiful expression

$$K = \frac{p^4}{a^2b^2c^2},$$

where p is the distance from the origin (0, 0, 0) to the tangent plane at the point (x, y, z).

There are also surfaces for which H = 0. Such surfaces are called minimal surfaces, and they show up in physics quite a bit. It can be verified that both the helicoid and the catenoid are minimal surfaces. The Enneper surface is also a minimal surface. We will see shortly how the classification of points on a surface can be explained in terms of the Dupin indicatrix.

The idea is to dip the surface in water, and to watch the shorelines formed in the water by the surface in a small region around a chosen point, as we move the surface up and down very gently. But first, we introduce the Gauss map, i.e. we study the variations of the normal N_p as the point p varies on the surface.

17.3 The Gauss Map and its Derivative dN

Given a surface $X : \Omega \to E^3$, given any point p = X(u, v) on X, we have defined the normal Np at p (or really $N_{(u,v)}$ at (u, v)) as the unit vector

$$N_p = \frac{X_u \times X_v}{\|X_u \times X_v\|}$$

Gauss realized that the assignment $p \mapsto N_p$ of the unit normal Np to the point p on the surface X could be viewed as a map from the trace of the surface X to the unit sphere S^2. If N_p is a unit vector of coordinates (x, y, z), we have $x^2 + y^2 + z^2 = 1$, and N_p corresponds to the point N(p) = (x, y, z) on the unit sphere. This is the so-called Gauss map of X, denoted as $N : X \to S^2$.

The derivative dN_p of the Gauss map at p measures the variation of the normal near p, i.e., how the surface "curves" near p. The Jacobian matrix of dN_p in the basis (X_u, X_v) can be expressed simply in terms of the matrices associated with the first and the second fundamental forms (which are quadratic forms).

Furthermore, the eigenvalues of dN_p are precisely $-k_1$ and $-k_2$, where k_1 and k_2 are the principal curvatures at p, and the eigenvectors define the principal directions (when they are well defined).

In view of the negative sign in $-k_1$ and $-k_2$, it is desirable to consider the linear map $S_p = -dN_p$, often called the shape operator.

Then, it is easily shown that the second fundamental form $II_p(\vec{t})$ can be expressed as

$$II_p(\vec{t}) = \langle S_p(\vec{t}), \vec{t}\rangle_p,$$

where $\langle -, - \rangle$ is the inner product associated with the first fundamental form.

Thus, the Gaussian curvature is equal to the determinant of S_p, and also to the determinant of dN_p, since $(-k_1)(-k_2) = k_1k_2$. We will see in a later section that the Gaussian curvature actually only depends of the first fundamental form, which is far from obvious right now! Actually, if X is not injective, there are problems, because the assignment $p \mapsto N_p$ could be multivalued.

We can either assume that X is injective, or consider the map from Ω to S^2 defined such that

$$(u, v) \mapsto N_{(u,v)}.$$

Then, we have a map from Ω to S^2, where (u, v) is mapped to the point N(u, v) on S^2 associated with N(u,v). This map is denoted as $N : \Omega \to S^2$. It is interesting to study the derivative dN of the Gauss map $N : \to S^2$ (or $N : X \to S^2$). As we shall see, the second fundamental form can be defined in terms of dN. For every $(u, v) \in \Omega$, the map $dN_{(u,v)}$ is a linear map $dN_{(u,v)} : \mathbb{R}^2 \to \mathbb{R}^2$.

It can be viewed as a linear map from the tangent space $T_{(u,v)}(X)$ at X(u, v) (which is isomorphic to $\mathbb{R}^2$) to the tangent space to the sphere at N(u, v) (also isomorphic to $\mathbb{R}^2$).

Recall that $dN_{(u,v)}$ is defined as follows: For every $(x, y) \in \mathbb{R}^2$,

$$dN_{(u,v)}(x, y) = N_u x + N_v y.$$

Thus, we need to compute Nu and Nv. Since N is a unit vector, $N \cdot N = 1$, and by taking derivatives, we have $Nu \cdot N = 0$ and $N_v \cdot N = 0$.

Consequently, N_u and N_v are in the tangent space at (u, v), and we can write

$$N_u = aX_u + cX_v,$$

$$N_v = bX_u + dX_v.$$

Lemma 2. Given a surface X, for any point p = X(u, v) on X, the derivative $dN_{(u,v)}$ of the Gauss map expressed in the basis (X_u, X_v) is given by the equation

$$dN_{(u,v)}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix},$$

where the Jacobian matrix $J(dN_{(u,v)})$ of $dN_{(u,v)}$ is given by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = -\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1}\begin{pmatrix} L & M \\ M & N \end{pmatrix}$$

$$= \frac{1}{EG - F^2}\begin{pmatrix} MF - LG & NF - MG \\ LF - ME & MF - NE \end{pmatrix}.$$

The equations

$$J(dN_{(u,v)}) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$= \frac{1}{EG - F^2}\begin{pmatrix} MF - LG & NF - MG \\ LF - ME & MF - NE \end{pmatrix}.$$

are know as the Weingarten equations (in matrix form).

If we recall the expressions for the Gaussian curvature and for the mean curvature

$$H = \frac{GL - 2FM + EN}{2(EG - F^2)},$$

$$K = \frac{LN - M^2}{EG - F^2},$$

we note that the trace a + d of the Jacobian matrix $J(dN_{(u,v)})$ of $dN_{(u,v)}$ is –2H, and that its determinant is precisely K.

This is recorded in the following lemma that also shows that the eigenvectors of $J(dN_{(u,v)})$ correspond to the principal directions:

Lemma 3. Given a surface X, for any point p = X(u, v) on X, the eigenvalues of the Jacobian matrix $J(dN_{(u,v)})$ of the derivative $dN_{(u,v)}$ of the Gauss map are $-k_1$, $-k_2$, where k_1 and k_2 are the principal curvatures at p, and the eigenvectors of $dN_{(u,v)}$ correspond to the principal directions (when they are defined). The Gaussian curvature K is the determinant of the Jacobian matrix of $dN_{(u,v)}$, and the mean curvature H is equal to $-\frac{1}{2}$ trace $J(dN_{(u,v)})$.

The fact that $Nu = -kX_u$ when k is one of the principal curvatures and when Xu corresponds to the corresponding principal direction (and similarly $N_v = -kX_v$ for the other principal curvature) is known as the formula of Olinde Rodrigues (1815).

The somewhat irritating negative signs arising in the eigenvalues $-k_1$ and $-k_2$ of $dN_{(u,v)}$ can be eliminated if we consider the linear map $S_{(u,v)} = -dN_{(u,v)}$ instead of $dN_{(u,v)}$.

The map $S_{(u,v)}$ is called the shape operator at p, and the map $dN_{(u,v)}$ is sometimes called the Weingarten operator.

The following lemma shows that the second fundamental form arises from the shape operator, and that the shape operator is self-adjoint with respect to the inner product $\langle -,- \rangle$ associated with the first fundamental form:

Lemma 4. Given a surface X, for any point p = X(u, v) on X, the second fundamental form of X at p is given by the formula

$$II_{(u,v)}(\vec{t}) = \left\langle S_{(u,v)}(\vec{t}), \vec{t} \right\rangle,$$

for every $\vec{t} \in \mathbb{R}^2$. The map $S(u,v) = -dN_{(u,v)}$ is self-adjoint, that is,

$$\left\langle S_{(u,v)}(\vec{x}), \vec{y} \right\rangle = \left\langle \vec{x}, S_{(u,v)}(\vec{y}) \right\rangle,$$

for all $\vec{x}, \vec{y} \in \mathbb{R}^2$.

Thus, in some sense, the shape operator contains all the information about curvature.

Remark: The fact that the first fundamental form I is positive definite and that S(u,v) is self-adjoint with respect to I can be used to give a fancier proof of the fact that S(u,v) has two real eigenvalues, that the eigenvectors are orthonormal, and that the eigenvalues correspond to the maximum and the minimum of I on the unit circle.

17.4 The Dupin Indicatrix

The second fundamental form shows up again when we study the deviation of a surface from its tangent plane in the neighborhood of the point of tangency.

A way to study this deviation is to imagine that we dip the surface in water, and watch the shorelines formed in the water by the surface in a small region around a chosen point, as we move the surface up and down very gently.

The resulting curve is known as the Dupin indicatrix (1813).

Formally, consider the tangent plane $T_{(u_0,v_0)}(X)$ at some point $p = X(u_0, v_0)$, and consider the perpendicular distance $\rho(u, v)$ from the tangent plane to a point on the surface defined by $\rho(u, v)$.

Notes

This perpendicular distance can be expressed as

$$\rho(u, v) = (X(u, v) - X(u_0, v_0)) \cdot N_{(u_0,v_0)}.$$

However, since X is at least C^3-continuous, by Taylor's formula, in a neighborhood of (u_0, v_0), we can write

$$X(u, v) = X(u_0, v_0) + X_u(u - u_0) + X_v(v - v_0) + \frac{1}{2}(X_{uu}(u - u_0)^2 + 2X_{uv}(u - u_0)(v - v_0) + X_{vv}(v - v_0)^2) +$$
$$((u - u_0)^2 + (v - v_0)^2)h_1(u, v),$$

where $\lim_{(u,v) \to (u_0,v_0)} h_1(u, v) = 0$.

However, recall that X_u and X_v are really evaluated at (u_0, v_0) (and so are X_{uu}, $X_{u,v}$, and X_{vv}), and so, they are orthogonal to $N_{(u_0,v_0)}$.

From this, dotting with $N_{(u_0,v_0)}$, we get

$$\rho(u, v) = (L(u - u_0)^2 + 2M(u - u_0)(v - v_0) + N(v - v_0)^2) + ((u - u_0)^2 + (v - v_0)^2)h(u, v),$$

where $\lim_{(u,v) \to (u_0,v_0)} h(u, v) = 0$.

Therefore, we get another interpretation of the second fundamental form as a way of measuring the deviation from the tangent plane.

For $\in$ small enough, and in a neighborhood of (u_0, v_0) small enough, the set of points X(u, v) on the surface such that $\rho(u, v) = \pm\frac{1}{2}\in^2$ will look like portions of the curves of equation

$$\frac{1}{2}(L(u - u_0)^2 + 2M(u - u_0)(v - v_0) + N(v - v_0)^2) = \pm\frac{1}{2}\in^2.$$

Letting $u - u_0 = \in x$ and $v - v_0 = \in y$, these curves are defined by the equations

$$Lx^2 + 2Mxy + Ny^2 = \pm 1.$$

These curves are called the Dupin indicatrix.

It is more convenient to switch to an orthonormal basis where $\vec{e}_1$ and $\vec{e}_2$ are eigenvectors of the Gauss map $dN_{(u_0,v_0)}$.

If so, it is immediately seen that

$$Lx^2 + 2Mxy + Ny^2 = k_1x^2 + k_2y^2,$$

where k_1 and k_2 are the principal curvatures. Thus, the equation of the Dupin indicatrix is

$$k_1x^2 + k_2y^2 = \pm 1.$$

There are several cases, depending on the sign of $k_1k_2 = K$, i.e., depending on the sign of $LN - M^2$.

(1) If $LN - M^2 > 0$, then k_1 and k_2 have the same sign. This is the case of an elliptic point.

If $k_1 \neq k_2$, and $k_1 > 0$ and $k_2 > 0$, we get the ellipse of equation

$$\frac{x^2}{\sqrt{\frac{1}{k_1}}} + \frac{y^2}{\sqrt{\frac{1}{k_2}}} = 1,$$

and if $k_1 < 0$ and $k_2 < 0$, we get the ellipse of equation

$$\frac{x^2}{\sqrt{-\frac{1}{k_1}}} + \frac{y^2}{\sqrt{-\frac{1}{k_2}}} = 1.$$

When $k_1 = k_2$, i.e. an umbilical point, the Dupin indicatrix is a circle.

(2) If $LN - M^2 = 0$ and $L^2 + M^2 + N^2 > 0$, then $k_1 = 0$ or $k_2 = 0$, but not both.

This is the case of a parabolic point.

In this case, the Dupin indicatrix degenerates to two parallel lines, since the equation is either

$$k_1x^2 = \pm 1$$

or

$$k_2y^2 = \pm 1.$$

(3) If $LN - M^2 < 0$ then k_1 and k_2 have different signs. This is the case of a hyperbolic point.

In this case, the Dupin indicatrix consists of the two hyperbolae of equations

$$\frac{x^2}{\sqrt{\frac{1}{k_1}}} - \frac{y^2}{\sqrt{\frac{1}{k_2}}} = 1,$$

if $k_1 > 0$ and $k_2 < 0$, or of equation

$$-\frac{x^2}{\sqrt{-\frac{1}{k_1}}} + \frac{y^2}{\sqrt{\frac{1}{k_2}}} = 1,$$

if $k_1 < 0$ and $k_2 > 0$.

These hyperbolae share the same asymptotes, which are the asymptotic directions as defined, and are given by the equation

$$Lx^2 + 2Mxy + Ny^2 = 0.$$

Therefore, analyzing the shape of the Dupin indicatrix leads us to rediscover the classification of points on a surface in terms of the principal curvatures.

It also lends some intuition to the meaning of the words elliptic, hyperbolic, and parabolic (the last one being a bit misleading).

The analysis of $\rho(u, v)$ also shows that in the elliptic case, in a small neighborhood of $X(u, v)$, all points of X are on the same side of the tangent plane. This is like being on the top of a round hill. In the hyperbolic case, in a small neighborhood of $X(u, v)$, there are points of X on both sides of the tangent plane. This is a saddle point, or a valley (or mountain pass).

17.5 Clairaut's Theorem

Clairaut's theorem, published in 1743 by Alexis Claude Clairaut in his *Théorie de la figure de la terre, tirée des principes de l'hydrostatique*, synthesized physical and geodetic evidence that the Earth is an oblate rotational ellipsoid. It is a general mathematical law applying to spheroids of revolution. It was initially used to relate the gravity at any point on the Earth's surface to the position of that point, allowing the ellipticity of the Earth to be calculated from measurements of gravity at different latitudes.

Formula

Clairaut's formula for the acceleration of gravity g on the surface of a spheroid at latitude ϕ, was:

$$g = G\left[1 + \left(\frac{5}{2}m - f\right)\sin^2\phi\right],$$

where G is the value of the acceleration of gravity at the equator, m the ratio of the centrifugal force to gravity at the equator, and f the flattening of a meridian section of the earth, defined as:

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

(where a = semimajor axis, b = semiminor axis).

Clairaut derived the formula under the assumption that the body was composed of concentric coaxial spheroidal layers of constant density. This work was subsequently pursued by Laplace, who relaxed the initial assumption that surfaces of equal density were spheroids. Stokes showed in 1849 that the theorem applied to any law of density so long as the external surface is a spheroid of equilibrium.

The above expression for g has been supplanted by the Somigliana equation:

$$g = G\left[\frac{1 + k\sin^2\phi}{\sqrt{1 - e^2\sin^2\phi}}\right],$$

where, for the Earth, $G = 9.7803267714\ ms^{-2}$; $k = 0.00193185138639$; $e^2 = 0.00669437999013$.

17.6 Gauss–Bonnet theorem

The Gauss–Bonnet theorem or Gauss–Bonnet formula in differential geometry is an important statement about surfaces which connects their geometry (in the sense of curvature) to their topology (in the sense of the Euler characteristic). It is named after Carl Friedrich Gauss who was aware of a version of the theorem but never published it, and Pierre Ossian Bonnet who published a special case in 1848.

17.6.1 Statement of the Theorem

Suppose M is a compact two-dimensional Riemannian manifold with boundary ∂M. Let K be the Gaussian curvature of M, and let k_g be the geodesic curvature of ∂M. Then

$$\int_M K\,dA + \int_{\partial M} k_g\,ds = 2\pi_x(M),$$

where dA is the element of area of the surface, and ds is the line element along the boundary of M. Here, $\chi(M)$ is the Euler characteristic of M.

If the boundary ∂M is piecewise smooth, then we interpret the integral $\int_{\partial M} k_g ds$ as the sum of the corresponding integrals along the smooth portions of the boundary, plus the sum of the angles by which the smooth portions turn at the corners of the boundary.

Notes

17.6.2 Interpretation and Significance

The theorem applies in particular to compact surfaces without boundary, in which case the integral

$$\int_{\partial M} k_g ds$$

can be omitted. It states that the total Gaussian curvature of such a closed surface is equal to 2π times the Euler characteristic of the surface. Note that for orientable compact surfaces without boundary, the Euler characteristic equals 2 – 2g, where g is the genus of the surface: Any orientable compact surface without boundary is topologically equivalent to a sphere with some handles attached, and g counts the number of handles.

If one bends and deforms the surface M, its Euler characteristic, being a topological invariant, will not change, while the curvatures at some points will. The theorem states, somewhat surprisingly, that the total integral of all curvatures will remain the same, no matter how the deforming is done. So for instance if you have a sphere with a "dent", then its total curvature is 4π (the Euler characteristic of a sphere being 2), no matter how big or deep the dent.

Compactness of the surface is of crucial importance. Consider for instance the open unit disc, a non-compact Riemann surface without boundary, with curvature 0 and with Euler characteristic 1: the Gauss–Bonnet formula does not work. It holds true, however, for the compact closed unit disc, which also has Euler characteristic 1, because of the added boundary integral with value 2π.

As an application, a torus has Euler characteristic 0, so its total curvature must also be zero. If the torus carries the ordinary Riemannian metric from its embedding in R^3, then the inside has negative Gaussian curvature, the outside has positive Gaussian curvature, and the total curvature is indeed 0. It is also possible to construct a torus by identifying opposite sides of a square, in which case the Riemannian metric on the torus is flat and has constant curvature 0, again resulting in total curvature 0. It is not possible to specify a Riemannian metric on the torus with everywhere positive or everywhere negative Gaussian curvature.

The theorem also has interesting consequences for triangles. Suppose M is some 2-dimensional Riemannian manifold (not necessarily compact), and we specify a "triangle" on M formed by three geodesics. Then we can apply Gauss–Bonnet to the surface T formed by the inside of that triangle and the piecewise boundary given by the triangle itself. The geodesic curvature of geodesics being zero, and the Euler characteristic of T being 1, the theorem then states that the sum of the turning angles of the geodesic triangle is equal to 2π minus the total curvature within the triangle. Since the turning angle at a corner is equal to ð minus the interior angle, we can rephrase this as follows:

The sum of interior angles of a geodesic triangle is equal to ð plus the total curvature enclosed by the triangle.

In the case of the plane (where the Gaussian curvature is 0 and geodesics are straight lines), we recover the familiar formula for the sum of angles in an ordinary triangle. On the standard sphere, where the curvature is everywhere 1, we see that the angle sum of geodesic triangles is always bigger than ð.

17.6.3 Special Cases

A number of earlier results in spherical geometry and hyperbolic geometry over the preceding centuries were subsumed as special cases of Gauss–Bonnet.

Notes

1. Triangles

In spherical trigonometry and hyperbolic trigonometry, the area of a triangle is proportional to the amount by which its interior angles fail to add up to 180°, or equivalently by the (inverse) amount by which its exterior angles fail to add up to 360°.

The area of a spherical triangle is proportional to its excess, by Girard's theorem – the amount by which its interior angles add up to more than 180°, which is equal to the amount by which its exterior angles add up to less than 360°.

The area of a hyperbolic triangle conversely is proportional to its *defect,* as established by Johann Heinrich Lambert.

2. Polyhedra

Descartes' theorem on total angular defect of a polyhedron is the polyhedral analog: it states that the sum of the defect at all the vertices of a polyhedron which is homeomorphic to the sphere is 4π. More generally, if the polyhedron has Euler characteristic $\chi = 2 - 2g$ (where g is the genus, meaning "number of holes"), then the sum of the defect is $2\pi\chi$. This is the special case of Gauss–Bonnet, where the curvature is concentrated at discrete points (the vertices).

Thinking of curvature as a measure, rather than as a function, Descartes' theorem is Gauss–Bonnet where the curvature is a discrete measure, and Gauss–Bonnet for measures generalizes both Gauss–Bonnet for smooth manifolds and Descartes' theorem.

17.6.4 Combinatorial Analog

There are several combinatorial analogs of the Gauss–Bonnet theorem. We state the following one. Let M be a finite 2-dimensional pseudo-manifold. Let $\chi(v)$ denote the number of triangles containing the vertex v. Then

$$\sum_{v \in \operatorname{int} M}(6-\chi(x)) + \sum_{v \in \partial M}(4-\chi(v)) = 6\chi(M),$$

where the first sum ranges over the vertices in the interior of M, the second sum is over the boundary vertices, and $\chi(M)$ is the Euler characteristic of M.

More specifically, if M is a closed digital 2-dimensional manifold, The genus

$$g = 1 + (M_5 + 2M_6 - M_3)/8,$$

where M_i indicates the set of surface-points each of which has i adjacent points on the surface. See digital topology

17.6.5 Generalizations

Generalizations of the Gauss–Bonnet theorem to n-dimensional Riemannian manifolds were found in the 1940s, by Allendoerfer, Weil and Chern–Weil homomorphism. The Riemann–Roch theorem can also be considered as a generalization of Gauss–Bonnet.

An extremely far-reaching generalization of all the above-mentioned theorems is the Atiyah–Singer index theorem.

Notes

17.7 The Theorema Egregium of Gauss, the Equations of Codazzi-Mainardi, and Bonnet's Theorem

Here, we expressed the geodesic curvature in terms of the Christoffel symbols, and we also showed that these symbols only depend on E, F, G, i.e., on the first fundamental form and we expressed N_u and N_v in terms of the coefficients of the first and the second fundamental form.

At first glance, given any six functions E, F, G, L, M, N which are at least C^3-continuous on some open subset U of $\mathbb{R}^2$, and where E, F > 0 and EG – F^2 > 0, it is plausible that there is a surface X defined on some open subset Ω of U, and having $Ex^2 + 2Fxy + Gy^2$ as its first fundamental form, and $Lx^2 + 2Mxy + Ny^2$ as its second fundamental form.

However, this is false! The problem is that for a surface X, the functions E, F, G, L, M, N are not independent.

In this section, we investigate the relations that exist among these functions. We will see that there are three compatibility equations. The first one gives the Gaussian curvature in terms of the first fundamental form only. This is the famous Theorema Egregium of Gauss (1827).

The other two equations express $M_u – L_v$ and $N_u – M_v$ in terms of L, M, N and the Christoffel symbols. These equations are due to Codazzi (1867) and Mainardi (1856).

Remarkably, these compatibility equations are just what it takes to insure the existence of a surface (at least locally) with $Ex^2 + 2Fxy + Gy^2$ as its first fundamental form, and $Lx^2 + 2Mxy + Ny^2$ as its second fundamental form, an important theorem shown by Ossian Bonnet (1867).

Recall that

$$\begin{aligned} X'' &= X_u u_1'' + X_v u_2'' + X_{uu}(u_1')^2 + 2X_{uv}u_1'u_2' + X_{vv}(u_2')2, \\ &= (L(u_1')^2 + 2Mu_1'u_2' + N(u_2')^2)N + k_g \vec{n}_g, \end{aligned}$$

and since

$$k_g \vec{n}_g = \left(u_1'' + \sum_{\substack{i=1,2\\ j=1,2}} \Gamma_{ij}^1 u_i' u_j' \right) X_u + \left(u_2'' + \sum_{\substack{i=1,2\\ j=1,2}} \Gamma_{ij}^2 u_i' u_j' \right) X_v,$$

we get the equations (due to Gauss):

$$\begin{aligned} X_{uu} &= \Gamma_{11}^1 X_u + \Gamma_{11}^2 X_v + LN, \\ X_{uv} &= \Gamma_{12}^1 X_u + \Gamma_{12}^2 X_v + MN, \\ X_{vu} &= \Gamma_{12}^1 X_u + \Gamma_{21}^2 X_v + MN, \\ X_{vv} &= \Gamma_{22}^1 X_u + \Gamma_{22}^2 X_v + NN, \end{aligned}$$

where the Christoffel symbols Γ_{ij}^k are defined such that

$$\begin{pmatrix} \Gamma_{ij}^1 \\ \Gamma_{ij}^2 \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1} \begin{pmatrix} [ij\,;1] \\ [ij\,;2] \end{pmatrix},$$

and where

$[1\,1;1] = \frac{1}{2}E_u$, $[1\,1;2] = F_u - \frac{1}{2}E_v$,

$[1\,2;1] = \frac{1}{2}E_v$, $[1\,2;2] = \frac{1}{2}G_u$,

$[2\,1;1] = \frac{1}{2}E_v$, $[2\,1;2] = \frac{1}{2}G_u$,

$[2\,2;1] = F_v - \frac{1}{2}G_u$, $[2\,2;2] = \frac{1}{2}G_v$.

Also, recall that we have the Weingarten equations

$$\begin{pmatrix} N_u \\ N_v \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} X_u \\ X_v \end{pmatrix}$$

From the Gauss equations and the Weingarten equations

$$X_{uu} = \Gamma_{11}^{1}X_u + \Gamma_{11}^{2}X_v + LN,$$

$$X_{uv} = \Gamma_{12}^{1}X_u + \Gamma_{12}^{2}X_v + MN,$$

$$X_{vu} = \Gamma_{21}^{1}X_u + \Gamma_{21}^{2}X_v + MN,$$

$$X_{vv} = \Gamma_{21}^{1}X_u + \Gamma_{21}^{2}X_v + NN,$$

$$N_u = aX_u + cX_v,$$

$$N_v = bX_u + dX_v,$$

We see that the partial derivatives of X_u, X_v and N can be expressed in terms of the coefficient E, F, G, L, M, N and their partial derivatives.

Thus, a way to obtain relations among these coefficients is to write the equations expressing the commutation of partials, i.e.,

$$(X_{uu})_v - (X_{uv})_u = 0,$$

$$(X_{vv})_u - (X_{vu})_v = 0,$$

$$N_{uv} - N_{vu} = 0.$$

Using the Gauss equations and the Weingarten equations, we obtain relations of the form

$$A_1X_u + B_1X_v + C_1N = 0,$$

$$A_2X_u + B_2X_v + C_2N = 0,$$

$$A_3X_u + B_3X_v + C_3N = 0,$$

where A_i, B_i, and C_i are functions of E, F, G, L,M,N and their partial derivatives, for i = 1, 2, 3.

However, since the vectors X_u, X_v, and N are linearly independent, we obtain the nine equations

$$A_i = 0,\ B_i = 0,\ C_i = 0, \text{ for } i = 1, 2, 3.$$

Although this is very tedious, it can be shown that these equations are equivalent to just three equations.

Notes

Due to its importance, we state the Theorema Egregium of Gauss.

Theorem 5: Given a surface X and a point p = X(u, v) on X, the Gaussian curvature K at (u, v) can be expressed as a function of E, F,G and their partial derivatives. In fact

$$(EG - F^2)^2K = \begin{vmatrix} C & F_v - \frac{1}{2}G_u & \frac{1}{2}G_v \\ \frac{1}{2}E_u & E & F \\ F_u - \frac{1}{2}E_v & F & G \end{vmatrix} - \begin{vmatrix} 0 & \frac{1}{2}E_v & \frac{1}{2}G_u \\ \frac{1}{2}E_v & E & F \\ \frac{1}{2}G_u & F & G \end{vmatrix}$$

where

$$C = \frac{1}{2}(-E_{vv} + 2F_{vv} - G_{uu}).$$

Proof. Way of proving theorem is to start from the formula

$$K = \frac{LN - M^2}{EG - F^2}$$

and to go back to the expressions of L, M, N using D, D′, D″ as determinants:

$$L = \frac{D}{\sqrt{EG - F^2}},\ M = \frac{D'}{\sqrt{EG - F^2}},\ N = \frac{D''}{\sqrt{EG - F^2}},$$

where

$$D = (X_u, X_v, X_{uu}),$$
$$D' = (X_u, X_v, X_{uv}),$$
$$D'' = (X_u, X_v, X_{vv}).$$

Then, we can write

$$(EG - F^2)^2K = (X_u, X_v, X_{uu})\,(X_u, X_v, X_{vv}) - (X_u, X_v, X_{uv})^2,$$

and compute these determinants by multiplying them out. One will eventually get the expression given in the theorem!

It can be shown that the other two equations, known as the Codazzi-Mainardi equations, are the equations:

$$Mu - Lv = \Gamma_{11}^2 N - (\Gamma_{12}^2 - \Gamma_{11}^1)M - \Gamma_{12}^1 L,$$

$$Nu - Mv = \Gamma_{12}^2 N - (\Gamma_{22}^2 - \Gamma_{12}^1)M - \Gamma_{22}^1 L.$$

We conclude this section with an important theorem of Ossian Bonnet. First, we show that the first and the second fundamental forms determine a surface up to rigid motion. More precisely, we have the following lemma:

Lemma 6. Let $X : \Omega \to E^3$ and $Y : \Omega \to E^3$ be two surfaces over a connected open set Ω. If X and Y have the same coefficients E, F, G, L, M, N over Ω, then there is a rigid motion mapping $X(\Omega)$ onto $Y(\Omega)$.

The above lemma can be shown using a standard theorem about ordinary differential equations. Finally, we state Bonnet's theorem.

Theorem 7: Let E, F, G, L, M, N be any C3-continuous functions on some open set $U \subset \mathbb{R}^2$, and such that $E > 0$, $G > 0$, and $EG - F^2 > 0$. If these functions satisfy the Gauss formula (of the Theorema Egregium) and the Codazzi-Mainardi equations, then for every $(u, v) \in U$, there is an open set $\Omega \subseteq U$ such that $(u, v) \in \Omega$, and a surface $X : \Omega \to E^3$ such that X is a diffeomorphism, and E, F, G are the coefficients of the first fundamental form of X, and L, M, N are the coefficients of the second fundamental form of X. Furthermore, if Ω is connected, then $X(\Omega)$ is unique up to a rigid motion.

17.8 Lines of Curvature, Geodesic Torsion, Asymptotic Lines

Given a surface X, certain curves on the surface play a special role, for example, the curves corresponding to the directions in which the curvature is maximum or minimum.

Definition 3: Given a surface X, a line of curvature is a curve $C : t \mapsto X(u(t), v(t))$ on X defined on some open interval I, and having the property that for every $t \in I$, the tangent vector $C'(t)$ is collinear with one of the principal directions at $X(u(t), v(t))$.

Notes we are assuming that no point on a line of curvature is either a planar point or an umbilical point, since principal directions are undefined as such points.

The differential equation defining lines of curvature can be found as follows:

Remember from lemma that the principal directions are the eigenvectors of $dN_{(u,v)}$.

Therefore, we can find the differential equation defining the lines of curvature by eliminating k from the two equations from the proof of lemma:

$$\frac{MF - LG}{EG - F^2}u' + \frac{NF - MG}{EG - F^2}v' = -ku',$$

$$\frac{LF - ME}{EG - F^2}u' + \frac{MF - NE}{EG - F^2}v' = -kv.$$

It is not hard to show that the resulting equation can be written as

$$\det\begin{pmatrix} (v')^2 & -u'u' & (u')^2 \\ E & F & G \\ L & M & N \end{pmatrix} = 0.$$

From the above equation, we see that the u-lines and the v-lines are the lines of curvatures iff F = M = 0.

Generally, this differential equation does not have closed-form solutions.

There is another notion which is useful in understanding lines of curvature, the geodesic torsion.

Let $C : S \mapsto X(u(s), v(s))$ be a curve on X assumed to be parameterized by arc length, and let $X(u(0), v(0))$ be a point on the surface X, and assume that this point is neither a planar point nor an umbilic, so that the principal directions are defined.

We can define the orthonormal frame $(\vec{e}_1, \vec{e}_2, N)$, known as the Darboux frame, where $\vec{e}_1$ and $\vec{e}_2$ are unit vectors corresponding to the principal directions, N is the normal to the surface at X(u(0), v(0)), and $N = \vec{e}_1 \times \vec{e}_2$.

It is interesting to study the quantity $\dfrac{dN_{(u,v)}}{ds}(0)$.

If $\vec{t} = C'(0)$ is the unit tangent vector at X(u(0), v(0)), we have another orthonormal frame considered in previous Section, namely $(\vec{t}, \vec{n}_g, N)$, where $\vec{n}_g = N \times \vec{t}$, and if φ is the angle between $\vec{e}_1$ and $\vec{t}$ we have

$$\vec{t} = \cos\varphi\, \vec{e}_1 + \sin\varphi\, \vec{e}_2,$$
$$\vec{n}_g = -\sin\varphi\, \vec{e}_1 + \cos\varphi\, \vec{e}_2.$$

Lemma 8. Given a curve $C : s \mapsto X(u(s), v(s))$ parameterized by arc length on a surface X, we have

$$\frac{dN_{(u,v)}}{ds}(0) = -k_N \vec{t} + T_g \vec{n}_g,$$

where k_N is the normal curvature, and where the geodesic torsion T_g is given by

$$T_g = (k_1 - k_2) \sin\varphi \cos\varphi.$$

From the formula

$$T_g = (k_1 - k_2) \sin\varphi \cos\varphi,$$

since φ is the angle between the tangent vector to the curve C and a principal direction, it is clear that the lines of curvatures are characterized by the fact that $T_g = 0$.

One will also observe that orthogonal curves have opposite geodesic torsions (same absolute value and opposite signs).

If $\vec{n}$ is the principal normal, T is the torsion of C at X(u(0), v(0)), and θ is the angle between N and $\vec{n}$ so that $\cos\theta = N \cdot \vec{n}$, we claim that

$$T_g = T - \frac{d\theta}{ds},$$

which is often known as Bonnet's formula.

Lemma 9. Given a curve $C : s \mapsto X(u(s), v(s))$ parameterized by arc length on a surface X, the geodesic torsion T_g is given by

$$T_g = T - \frac{d\theta}{ds} = (k_1 - k_2) \sin\varphi \cos\varphi,$$

where T is the torsion of C at X(u(0), v(0)), and is the angle between N and the principal normal $\vec{n}$ to C at s = 0.

Notes The geodesic torsion only depends on the tangent of curves C. Also, for a curve for which θ = 0, we have $T_g = T$.

Such a curve is also characterized by the fact that the geodesic curvature k_g is null.

As we will see shortly, such curves are called geodesics, which explains the name geodesic torsion for T_g.

Lemma 10: can be used to give a quick proof of a beautiful theorem of Dupin (1813).

Dupin's theorem has to do with families of surfaces forming a triply orthogonal system.

Given some open subset U of E^3, three families F_1, F_2, F_3 of surfaces form a triply orthogonal system for U, if for every point $p \in U$, there is a unique surface from each family F_i passing through p, where i = 1, 2, 3, and any two of these surfaces intersect orthogonally along their curve of intersection.

Theorem 11: The surfaces of a triply orthogonal system intersect each other along lines of curvature.

A nice application of theorem 11 is that it is possible to find the lines of curvature on an ellipsoid.

We now turn briefly to asymptotic lines. Recall that asymptotic directions are only defined at points where $K < 0$, and at such points, they correspond to the directions for which the normal curvature k_N is null.

Definition 4: Given a surface X, an asymptotic line is a curve $C : t \mapsto X(u(t), v(t))$ on X defined on some open interval I where $K < 0$, and having the property that for every $t \in I$, the tangent vector C0(t) is collinear with one of the asymptotic directions at X(u(t), v(t)).

The differential equation defining asymptotic lines is easily found since it expresses the fact that the normal curvature is null:

$$L(u')^2 + 2M(u'v') + N(v')^2 = 0.$$

Such an equation generally does not have closed-form solutions.

Notes The u-lines and the v-lines are asymptotic lines iff $L = N = 0$ (and $F \neq 0$).

Perseverant readers are welcome to compute E, F, G, L, M, N for the Enneper surface:

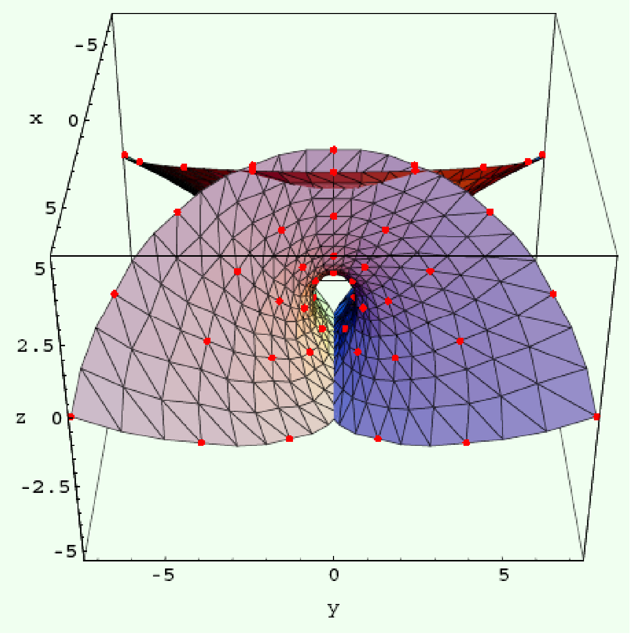
$$x = u - \frac{u^3}{3} + uv^2$$

$$y = v - \frac{u^3}{3} + u^2v$$

$$z = u^2 - v^2.$$

Then, they will be able to find closed-form solutions for the lines of curvatures and the asymptotic lines.

Figure 17.4: The Enneper Surface



Parabolic lines are defined by the equation

$$LN - M^2 = 0,$$

where $L^2 + M^2 + N^2 > 0$.

In general, the locus of parabolic points consists of several curves and points.

We now turn briefly to geodesics.

17.9 Summary

- We now show that k_n can be computed only in terms of the first fundamental form of X, a result first proved by Ossian Bonnet circa 1848.

  The computation is a bit involved, and it will lead us to the Christoffel symbols, introduced in 1869.

  Since $\vec{n}_g$ is in the tangent space $T_p(X)$, and since (X_u, X_v) is a basis of $T_p(X)$, we can write

  $$k_g \vec{n}_g = AX_u + BX_v,$$

  form some $A, B \in \mathbb{R}$.

  However,

  $$k\vec{n} = k_N N + k_g \vec{n}_g,$$

  and since N is normal to the tangent space,

  $N \cdot X_u = N \cdot X_v = 0$, and by dotting

- In general, we will see that the normal curvature has a maximum value k_1 and a minimum value k_2, and that the corresponding directions are orthogonal. This was shown by Euler in 1760.

The quantity $K = k_1k_2$ called the Gaussian curvature and the quantity $H = (k_1 + k_2)/2$ called the mean curvature, play a very important role in the theory of surfaces.

We will compute H and K in terms of the first and the second fundamental form. We also classify points on a surface according to the value and sign of the Gaussian curvature.

Recall that given a surface X and some point p on X, the vectors X_u, X_v form a basis of the tangent space $T_p(X)$.

Given a unit vector $\vec{t} = X_u x + X_v y$, the normal curvature is given by

$$k_N(\vec{t}) = Lx^2 + 2Mxy + Ny^2,$$

since $Ex^2 + 2Fxy + Gy^2 = 1$.

- Given a surface X, for any point p on X, letting A, B, H be defined as above, and $C = \sqrt{A^2 + B^2}$, unless $A = B = 0$, the normal curvature k_N at p takes a maximum value k_1 and a minimum value k_2 called principal curvatures at p, where $k_1 = H + C$ and $k_2 = H - C$. The directions of the corresponding unit vectors are called the principal directions at p.
- It can be shown that a connected surface consisting only of umbilical points is contained in a sphere.

It can also be shown that a connected surface consisting only of planar points is contained in a plane.

A surface can contain at the same time elliptic points, parabolic points, and hyperbolic points. This is the case of a torus.

The parabolic points are on two circles also contained in two tangent planes to the torus (the two horizontal planes touching the top and the bottom of the torus on the following picture).

The elliptic points are on the outside part of the torus (with normal facing outward), delimited by the two parabolic circles.

The hyperbolic points are on the inside part of the torus (with normal facing inward).

17.10 Keywords

Christoffel symbols: The computation is a bit involved, and it will lead us to the Christoffel symbols, introduced in 1869.

Gaussian curvature: The quantity $K = k_1k_2$ called the Gaussian curvature and the quantity $H = (k_1 + k_2)/2$ called the mean curvature, play a very important role in the theory of surfaces.

Elliptic: At an elliptic point, both principal curvatures are non-null and have the same sign. For example, most points on an ellipsoid are elliptic.

Hyperbolic: At a hyperbolic point, the principal curvatures have opposite signs. For example, all points on the catenoid are hyperbolic.

Jacobian matrix: The Jacobian matrix of dN_p in the basis (X_u, X_v) can be expressed simply in terms of the matrices associated with the first and the second fundamental forms (which are quadratic forms).

17.11 Self Assessment

1. The computation is a bit involved, and it will lead us to the .... , introduced in 1869.
2. The quantity $K = k_1k_2$ called the .... and the quantity $H = (k_1 + k_2)/2$ called the mean curvature, play a very important role in the theory of surfaces.
3. At a parabolic point, one of the two principal curvatures is ...., but not both. This is equivalent to $K = 0$ and $H\ ^{1}\ 0$. Points on a cylinder are parabolic.
4. At a planar point, $k_1 = k_2 = 0$. This is equivalent to $K = H = 0$. Points on a plane are all planar points! On a monkey saddle, there is a planar point. The principal directions at that point are ....
5. The derivative dN_p of the .... at p measures the variation of the normal near p, i.e., how the surface "curves" near p.
6. The .... of dN_p in the basis (X_u, X_v) can be expressed simply in terms of the matrices associated with the first and the second fundamental forms (which are quadratic forms).

17.12 Review Questions

1. Explain the Gauss Map and its Derivative dN.
2. Define the Dupin Indicatrix.
3. Describe the theorema Egregium of Gauss, the Equations of Codazzi-Mainardi, and Bonnet's Theorem.
4. Define Lines of Curvature, Geodesic Torsion, Asymptotic Lines.

Answers: Self Assessment

1.	Christoffel symbols	2.	Gaussian curvature
3.	zero	4.	undefined.
5.	Gauss map	6.	Jacobian matrix

17.13 Further Readings

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati,T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.

Unit 18: Joachimsthal's Notations

Notes

CONTENTS

Objectives

After studying this unit, you will be able to:

- Define Geodesic Lines, Local Gauss-Bonnet Theorem
- Discuss Covariant Derivative, Parallel Transport, Geodesics Revisited
- Describe Joachimsthal's Notations

Introduction

In this unit you will go through, Bonnet's theorem about the existence of a surface patch with prescribed first and second fundamental form. This will require a discussion of the Theorema Egregium and of the Codazzi-Mainardi compatibility equations. We will take a Joachimsthal's Notations

18.1 Geodesic Lines, Local Gauss-Bonnet Theorem

Geodesics play a very important role in surface theory and in dynamics. One of the main reasons why geodesics are so important is that they generalize to curved surfaces the notion of "shortest path" between two points in the plane.

More precisely, given a surface X, given any two points $p = X(u_0, v_0)$ and $q = X(u_1, v_1)$ on X, let us look at all the regular curves C on X defined on some open interval I such that $p = C(t_0)$ and $q = C(t_1)$ for some $t_0, t_1 \in I$.

It can be shown that in order for such a curve C to minimize the length $l_C(pq)$ of the curve segment from p to q, we must have $k_g(t) = 0$ along $[t_0, t_1]$, where $k_g(t)$ is the geodesic curvature at $X(u(t), v(t))$.

In other words, the principal normal $\vec{n}$ must be parallel to the normal N to the surface along the curve segment from p to q.

If C is parameterized by arc length, this means that the acceleration must be normal to the surface.

It is then natural to define geodesics as those curves such that $k_g = 0$ everywhere on their domain of definition.

Actually, there is another way of defining geodesics in terms of vector fields and covariant derivatives, but for simplicity, we stick to the definition in terms of the geodesic curvature.

Definition 1. Given a surface $X : \Omega \to E^3$, a geodesic line, or geodesic, is a regular curve $C : I \to E^3$ on X, such that $k_g(t) = 0$ for all $t \in I$.

> *Notes* By regular curve, we mean that $\dot{C}(t) \neq 0$ for all $t \in I$, i.e., C is really a curve, and not a single point.

Physically, a particle constrained to stay on the surface and not acted on by any force, once set in motion with some non-null initial velocity (tangent to the surface), will follow a geodesic (assuming no friction).

Since $k_g = 0$ if the principal normal $\vec{n}$ to C at t is parallel to the normal N to the surface at X(u(t), v(t)), and since the principal normal $\vec{n}$ is a linear combination of the tangent vector $\dot{C}(t)$ and the acceleration vector $\ddot{C}(t)$, the normal N to the surface at t belongs to the osculating plane.

Since the tangential part of the curvature at a point is given by

$$k_g \vec{n}_g = \left(u_1'' + \sum_{\substack{i=1,2 \\ j=1,2}} \Gamma_{ij}^1 u_i' u_j' \right) X_u + \left(u_2'' + \sum_{\substack{i=1,2 \\ j=1,2}} \Gamma_{ij}^2 u_i' u_j' \right) X_v,$$

the differential equations for geodesics are

$$u_1'' + \sum_{\substack{i=1,2 \\ j=1,2}} \Gamma_{ij}^1 u_i' u_j' = 0,$$

$$u_2'' + \sum_{\substack{i=1,2 \\ j=1,2}} \Gamma_{ij}^2 u_i' u_j' = 0,$$

or more explicitly (letting $u = u_1$ and $v = u_2$),

$$u'' + \Gamma_{11}^1 (u')^2 + 2\Gamma_{12}^1 u'v' + \Gamma_{22}^1 (v')^2 = 0,$$
$$v'' + \Gamma_{11}^2 (u')^2 + 2\Gamma_{12}^2 u'v' + \Gamma_{22}^2 (v')^2 = 0.$$

In general, it is impossible to find closed-form solutions for these equations.

Nevertheless, from the theory of ordinary differential equations, the following lemma showing the local existence of geodesics can be shown :

Lemma 1: Given a surface X, for every point p = X(u, v) on X, for every non-null tangent vector $\vec{v} \in T_{(u,v)}(X)$ at p, there is some $\in > 0$ and a unique curve $\gamma :] -\in, \in [\to E^3$ on the surface X, such that γ is a geodesic, $\gamma(0) = p$, and $\gamma'(0) = \vec{v}$.

To emphasize that the geodesic γ depends on the initial direction $\vec{v}$, we often write $\gamma(t, \vec{v})$ instead of (t).

The geodesics on a sphere are the great circles (the plane sections by planes containing the center of the sphere).

More generally, in the case of a surface of revolution (a surface generated by a plane curve rotating around an axis in the plane containing the curve and not meeting the curve), the differential equations for geodesics can be used to study the geodesics.

Example: The meridians are geodesics (meridians are the plane sections by planes through the axis of rotation: they are obtained by rotating the original curve generating the surface).

Also, the parallel circles such that at every point p, the tangent to the meridian through p is parallel to the axis of rotation, is a geodesic.

It should be noted that geodesics can be self-intersecting or closed. A deeper study of geodesics requires a study of vector fields on surfaces and would lead us too far.

Technically, what is needed is the exponential map, which we now discuss briefly.

The idea behind the exponential map is to parameterize locally the surface X in terms of a map from the tangent space to the surface, this map being defined in terms of short geodesics.

More precisely, for every point p = X(u, v) on the surface, there is some open disk $B_\in$ of center (0, 0) in $\mathbb{R}^2$ (recall that the tangent plane $T_p(X)$ at p is isomorphic to $\mathbb{R}^2$), and an injective map

$$\exp_p : B_\in \to X(\Omega),$$

such that for every $\vec{v} \in B_\in$ with $\vec{v} \neq \vec{0}$,

$$\exp_p(\vec{v}) = \gamma(1, \vec{v}),$$

where $\gamma(t, \vec{v})$ is the unique geodesic segment such that $\gamma(0, \vec{v}) = p$ and $\gamma'(0, \vec{v}) = \vec{v}$. Furthermore, for $B_\in$ small enough, $\exp_p$ is a diffeomorphism. It turns out that $\exp_p(\vec{v})$ is the point q obtained by "laying off" a length equal to $\|\vec{v}\|$ along the unique geodesic that passes through p in the direction $\vec{v}$.

Lemma 2: Given a surface $X : \Omega \to E^3$, for every $\vec{v} \neq \vec{0}$ in $\mathbb{R}^2$, if

$$\gamma(-, \vec{v}) :] -\in, \in [\to E^3$$

is a geodesic on the surface X, then for every $\lambda > 0$, the curve

$$\gamma(-, \lambda\vec{v}) :] -\in/\lambda, \in/\lambda [\to E^3$$

Notes

is also a geodesic, and

$$\gamma(t, \lambda\vec{v}) = \gamma(\lambda t, \vec{v}).$$

From lemma 2, for $\vec{v} \neq \vec{0}$, if $\gamma(1,\vec{v})$ is defined, then

$$\gamma\left(\|\vec{v}\|, \frac{\vec{v}}{\|\vec{v}\|}\right) = \gamma(1,\vec{v}).$$

This leads to the definition of the exponential map.

Definition 2: Given a surface $X : \Omega \to E^3$ and a point $p = X(u, v)$ on X, the exponential map expp is the map

$$\exp_p: U \to X(\Omega)$$

defined such that

$$\exp_p(\vec{v}) = \gamma\left(\|\vec{v}\|, \frac{\vec{v}}{\|\vec{v}\|}\right) = \gamma(1,\vec{v}),$$

where $\gamma(0,\vec{v}) = p$ and U is the open subset of $\mathbb{R}^2(= T_p(X))$ such that for every $\vec{v} \neq \vec{0}$, $\gamma\left(\|\vec{v}\|, \frac{\vec{v}}{\|\vec{v}\|}\right)$ is defined. We let $\exp_p(\vec{0}) = p$. It is immediately seen that U is star-like. One should realize that in general, U is a proper subset of Ω.

Example: In the case of a sphere, the exponential map is defined everywhere. However, given a point p on a sphere, if we remove its antipodal point –p, then $\exp_p(\vec{v})$ is undefined for points on the circle of radius π.

Nevertheless, expp is always well-defined in a small open disk.

Lemma 3: Given a surface $X : \Omega \to E^3$, for every point $p = X(u, v)$ on X, there is some $\in > 0$, some open disk $B_\in$ of center (0, 0), and some open subset V of $X(\Omega)$ with $p \in V$, such that the exponential map $\exp_p : B\in \to V$ is well defined and is a diffeomorphism.

A neighborhood of p on X of the form $\exp_p(B\in)$ is called a normal neighborhood of p.

The exponential map can be used to define special local coordinate systems on normal neighborhoods, by picking special coordinates systems on the tangent plane.

In particular, we can use polar coordinates (ρ, θ) on $\mathbb{R}^2$. In this case, $0 < \theta < 2\pi$. Thus, the closed half-line corresponding to $\theta = 0$ is omitted, and so is its image under $\exp_p$. It is easily seen that in such a coordinate system, E = 1 and F = 0, and the ds^2 is of the form

$$ds^2 = dr^2 + G\, d\theta^2.$$

The image under $\exp_p$ of a line through the origin in $\mathbb{R}^2$ is called a geodesic line, and the image of a circle centered in the origin is called a geodesic circle. Since F = 0, these lines are orthogonal.

Notes

It can also be shown that the Gaussian curvature is expressed as follows:

$$K = -\frac{1}{\sqrt{G}}\frac{\partial^2(\sqrt{G})}{\partial\rho^2}.$$

Polar coordinates can be used to prove the following lemma showing that geodesics locally minimize arc length:

However, globally, geodesics generally do not minimize arc length.

For instance, on a sphere, given any two non-antipodal points p, q, since there is a unique great circle passing through p and q, there are two geodesic arcs joining p and q, but only one of them has minimal length.

Lemma 4: Given a surface $X : \Omega \to E^3$, for every point $p = X(u, v)$ on X, there is some $\in > 0$ and some open disk $B_\in$ of center (0, 0) such that for every $q \in \exp_p(B_\in)$, for every geodesic $\gamma :]-\eta, \eta[\to E^3$ in $\exp_p(B_\in)$ such that $\gamma(0) = p$ and $\gamma(t_1) = q$, for every regular curve $\alpha : [0, t_1] \to E^3$ on X such that $\alpha(0) = p$ and $\alpha(t_1) = q$, then

$$l_\gamma(pq) \le l_\alpha(pq),$$

where $l_\alpha(pq)$ denotes the length of the curve segment α from p to q (and similarly for γ). Furthermore, $l_\gamma(pq) = l_\alpha(pq)$ if the trace of γ is equal to the trace of α between p and q.

As we already noted, lemma 4 is false globally, since a geodesic, if extended too much, may not be the shortest path between two points (example of the sphere).

However, the following lemma shows that a shortest path must be a geodesic segment:

Lemma 5: Given a surface $X : \Omega \to E^3$, let $\alpha : I \to E^3$ be a regular curve on X parameterized by arc length. For any two points $p = \alpha(t_0)$ and $q = \alpha(t_1)$ on α, assume that the length $l_\alpha(pq)$ of the curve segment from p to q is minimal among all regular curves on X passing through p and q. Then, α is a geodesic.

At this point, in order to go further into the theory of surfaces, in particular closed surfaces, it is necessary to introduce differentiable manifolds and more topological tools.

Nevertheless, we can't resist to state one of the "gems" of the differential geometry of surfaces, the local Gauss-Bonnet theorem.

The local Gauss-Bonnet theorem deals with regions on a surface homeomorphic to a closed disk, whose boundary is a closed piecewise regular curve α without self-intersection.

Such a curve has a finite number of points where the tangent has a discontinuity.

If there are n such discontinuities $p_1, \ldots, p_n$, let θ_i be the exterior angle between the two tangents at p_i.

More precisely, if $\alpha(t_i) = p_i$, and the two tangents at p_i are defined by the vectors

$$\lim_{t - t_i,\, t < t_i} \alpha'(t) = \alpha'_-(t_i) \neq \vec{0},$$

and

$$\lim_{t - t_i,\, t < t_i} \alpha'(t) = \alpha'_+(t_i) \neq \vec{0},$$

Notes

the angle θ_i is defined as follows:

Let θ_i be the angle between $\alpha'_-(t_i')$ and $\alpha' + (t_i)$ such that $0 < |\theta_i| \le \pi$, its sign being determined as follows:

If p_i is not a cusp, which means that $|\theta_i| \ne \pi$, we give i the sign of the determinant

$$(\alpha'_-(t_i), \alpha'_+(t_i), N_{p_i}).$$

If p_i is a cusp, which means that $|\theta_i| = \pi$, it is easy to see that there is some $\in > 0$ such that the determinant

$$(\alpha'(t_i - \eta), \alpha'(t_i + \eta), N_{p_i})$$

does not change sign for $\eta \in] -\in, \in [$, and we give θ_i the sign of this determinant.

Let us call a region defined as above a simple region.

In order to state a simpler version of the theorem, let us also assume that the curve segments between consecutive points p_i are geodesic lines.

We will call such a curve a geodesic polygon. Then, the local Gauss-Bonnet theorem can be stated as follows:

Theorem 6: Given a surface $X : \Omega \to E^3$, assuming that X is injective, F = 0, and that Ω is an open disk, for every simple region R of $X(\Omega)$ bounded by a geodesic polygon with n vertices $p_1, \ldots, p_n$, letting $\theta_1, \ldots, \theta_n$ be the exterior angles of the geodesic polygon, we have

$$\iint_R K\, dA + \sum_{i=1}^{n} \theta_i = 2\pi.$$

Some clarification regarding the meaning of the integral $\iint_R K\, dA$ is in order.

Firstly, it can be shown that the element of area dA on a surface X is given by

$$dA = \|X_u \times X_v\|\, dudv = \sqrt{EG - F^2}\ dudv.$$

Secondly, if we recall from lemma that

$$\begin{pmatrix} N_u \\ N_v \end{pmatrix} = -\begin{pmatrix} L & M \\ M & N \end{pmatrix}\begin{pmatrix} E & F \\ F & G \end{pmatrix}^{-1}\begin{pmatrix} X_u \\ X_v \end{pmatrix},$$

it is easily verified that

$$N_u \times N_v = \frac{LN - M^2}{EG - F^2} X_u \times X_v = K(X_u \times X_v).$$

Thus,

$$\begin{aligned} \iint_R KdA &= \iint_R K\|X_u \times X_v\|dudv \\ &= \iint_R \|N_u \times N_v\|dudv, \end{aligned}$$

the latter integral representing the area of the spherical image of R under the Gauss map.

This is the interpretation of the integral $\iint_R KdA$ that Gauss himself gave.

If the geodesic polygon is a triangle, and if A,B,C are the interior angles, so that $A = \pi - \theta_1$, $B = \pi - \theta 2$, $C = \pi - 3$, the Gauss-Bonnet theorem reduces to what is known as the Gauss formula:

$$\iint_R K dA = A + B + C - \pi.$$

The above formula shows that if K > 0 on R, then $\iint_R K dA$ is the excess of the sum of the angles of the geodesic triangle over π.

If K < 0 on R, then $\iint_R K dA$ is the efficiency of the sum of the angles of the geodesic triangle over π.

And finally, if K = 0, then $A + B + C = \pi$, which we know from the plane!

For the global version of the Gauss-Bonnet theorem, we need the topological notion of the Euler-Poincare characteristic, but this is beyond the scope of this course.

18.2 Covariant Derivative, Parallel Transport, Geodesics Revisited

Another way to approach geodesics is in terms of covariant derivatives.

The notion of covariant derivative is a key concept of Riemannian geometry, and thus, it is worth discussing anyway.

Let $X : \Omega \to E^3$ be a surface. Given any open subset, U, of X, a vector field on U is a function, w, that assigns to every point, $p \in U$, some tangent vector $w(p) \in T_pX$ to X at p.

A vector field, w, on U is differentiable at p if, when expressed as $w = aX_u + bX_v$ in the basis (X_u, X_v) (of T_pX), the functions a and b are differentiable at p.

A vector field, w, is differentiable on U when it is differentiable at every point $p \in U$.

Definition 3: Let, w, be a differentiable vector field on some open subset, U, of a surface X. For every $y \in T_pX$, consider a curve, $\alpha :]-\in, \in[\to U$, on X, with $\alpha(0) = p$ and $a'(0) = y$, and let $w(t) = (w \circ \alpha)(t)$ be the restriction of the vector field w to the curve α. The normal projection of dw/dt(0) onto the plane T_pX, denoted

$$\frac{Dw}{dt}(0), \quad \text{or} \quad D_{\alpha}'0w(p), \quad \text{or} \quad D_y w(p),$$

is called the covariant derivative of w at p relative to y.

The definition of Dw/dt(0) seems to depend on the curve α, but in fact, it only depends on y and the first fundamental form of X.

Indeed, if $\alpha(t) = X(u(t), v(t))$, from

$$w(t) = a(u(t), v(t))X_u + b(u(t), v(t))X_v,$$

we get

$$\frac{dw}{dt} = a(X_{uu}\dot{u} + X_{uv}\dot{v}) + b(X_{vu}\dot{u} + X_{vv}\dot{v}) + \dot{a}Xu + \dot{b}Xv.$$

However, we obtained earlier the following formulae (due to Gauss) for X_{uu}, X_{uv}, X_{vu}, and X_{vv}:

$$X_{uu} = \Gamma^1_{11}X_u + G^2_{11}X_v + LN,$$

Notes

$$Xuv = \Gamma^1_{12}X_u + G^2_{12}X_v + MN,$$

$$Xvu = \Gamma^1_{21}X_u + G^2_{21}X_v + MN,$$

$$Xvv = \Gamma^1_{22}X_u + G^2_{22}X_v + NN.$$

Now, Dw/dt is the tangential component of dw/dt, thus, by dropping the normal components, we get

$$\frac{Dw}{dt} = (\dot{a} + \Gamma^1_{11}a\dot{u} + \Gamma^1_{12}a\dot{v} + \Gamma^1_{21}b\dot{u} + \Gamma^1_{22}b\dot{v})X_u + (b + \Gamma^2_{11}a\dot{u} + \Gamma^2_{12}a\dot{v} + \Gamma^2_{21}b\dot{u} + G^2_{22}b\dot{v})X_v$$

Thus, the covariant derivative only depends on $y = (\dot{u}, \dot{v})$, and the Christoffel symbols, but we know that those only depends on the first fundamental form of X.

Definition 3: Let $\alpha : I \to X$ be a regular curve on a surface X. A vector field along α is a map, w, that assigns to every $t \in I$ a vector $w(t) \in T_{\alpha(t)}X$ in the tangent plane to X at $\alpha(t)$. Such a vector field is differentiable if the components a, b of $w = aX_u + bX_v$ over the basis (X_u, X_v) are differentiable. The expression Dw/dt(t) defined in the above equation is called the covariant derivative of w at t.

Definition 4: extends immediately to piecewise regular curves on a surface.

Definition 5: Let $\alpha : I \to X$ be a regular curve on a surface X. A vector field along α is parallel if $Dw/dt = 0$ for all $t \in I$.

Thus, a vector field along a curve on a surface is parallel if its derivative is normal to the surface.

For example, if C is a great circle on the sphere S^2 parametrized by arc length, the vector field of tangent vectors C′(s) along C is a parallel vector field.

We get the following alternate definition of a geodesic.

Definition 6: Let $\alpha : I \to X$ be a nonconstant regular curve on a surface X. Then, α is a geodesic if the field of its tangent vectors, $\dot{\alpha}(t)$, is parallel along α, that is

$$\frac{D\dot{\alpha}}{dt}(t) = 0$$

for all $t \in I$.

If we let $\alpha(t) = X(u(t), v(t))$, from the equation

$$\frac{Dw}{dt} = (\dot{a} + \Gamma^1_{11}a\dot{u} + \Gamma^1_{12}a\dot{v} + \Gamma^1_{21}b\dot{u} + \Gamma^1_{22}b\dot{v})X_u + (\dot{b} + \Gamma^2_{11}a\dot{u} + \Gamma^2_{12}a\dot{v} + \Gamma^2_{21}b\dot{u} + \Gamma^2_{22}b\dot{v})X_v$$

with $a = \dot{u}$ and $b = \dot{v}$, we get the equations

$$\ddot{u} + \Gamma^1_{11}(\dot{u})^2 + \Gamma^1_{12}\dot{u}\dot{v} + \Gamma^1_{21}\dot{u}\dot{v} + \Gamma^1_{22}(\dot{v})^2 = 0$$

$$\ddot{v} + \Gamma^2_{11}(\dot{u})^2 + \Gamma^2_{12}\dot{u}\dot{v} + \Gamma^2_{21}\dot{u}\dot{v} + \Gamma^2_{22}(\dot{v})^2 = 0,$$

which are indeed the equations of geodesics found earlier, since $\Gamma^1_{12} = \Gamma^1_{21}$ and $\Gamma^2_{12} = \Gamma^2_{21}$ (except that α is not necessarily parametrized by arc length).

Lemma 7: Let $\alpha : I \to X$ be a regular curve on a surface X, and let v and w be two parallel vector fields along α. Then, the inner product $\langle v(t), w(t) \rangle$ is constant along α (where $\langle -, - \rangle$ is the inner

Notes

product associated with the first fundamental form, i.e., the Riemannian metric). In particular, $\|v\|$ and $\|w\|$ are constant and the angle between v(t) and w(t) is also constant.

The vector field v(t) is parallel if dv/dt is normal to the tangent plane to the surface X at α(t), and so

$$\langle v'(t), w(t)\rangle = 0$$

for all $t \in I$. Similarly, since w(t) is parallel, we have

$$\langle v(t), w'(t)\rangle = 0$$

for all $t \in I$. Then,

$$\langle v(t), w(t)\rangle' = \langle v'(t), w(t)\rangle + \langle v(t), w'(t)\rangle = 0$$

for all $t \in I$. which means that $\langle v(t), w(t)\rangle$ is constant along α.

As a consequence of corollary 14.12.5, if $\alpha : I \to X$ is a nonconstant geodesic on X, then $\|\dot{\alpha}\| = c$ for some constant $c > 0$.

Thus, we may reparametrize α w.r.t. the arc length s = ct, and we note that the parameter t of a geodesic is proportional to the arc length of α.

Lemma 8: Let $\alpha : I \to X$ be a regular curve on a surface X, and for any $t_0 \in I$, let $w_0 \in T_{\alpha(t_0)}X$. Then, there is a unique parallel vector field, w(t), along α, so that $w(t_0) = w_0$.

Lemma is an immediate consequence of standard results on ODE's. This lemma yields the notion of parallel transport.

Definition 7: Let $\alpha : I \to X$ be a regular curve on a surface X, and for any $t_0 \in I$, let $w_0 \in T_{\alpha(t_0)}X$. Let w be the parallel vector field along α, so that $w(t_0) = w_0$, given by Lemma 8. Then, for any $t \in I$, the vector, w(t), is called the parallel transport of w_0 along α at t.

It is easily checked that the parallel transport does not depend on the parametrization of α. If X is an open subset of the plane, then the parallel transport of w_0 at t is indeed a vector w(t) parallel to w_0 (in fact, equal to w_0).

However, on a curved surface, the parallel transport may be somewhat counter-intuitive.

If two surfaces X and Y are tangent along a curve, $\alpha : I \to X$, and if $w_0 \in T_{\alpha(t_0)}X = T_{a(t_0)}Y$ is a tangent vector to both X and Y at t_0, then the parallel transport of w_0 along α is the same, whether it is relative to X or relative to Y .

This is because Dw/dt is the same for both surfaces, and by uniqueness of the parallel transport, the assertion follows.

This property can be used to figure out the parallel transport of a vector w_0 when Y is locally isometric to the plane.

In order to generalize the notion of covariant derivative, geodesic, and curvature, to manifolds more general than surfaces, the notion of connection is needed.

If M is a manifold, we can consider the space, X(M), of smooth vector fields, X, on M. They are smooth maps that assign to every point $p \in M$ some vector X(p) in the tangent space T_pM to M at p.

We can also consider the set $C^\infty(M)$ of smooth functions $f : M \to \mathbb{R}$ on M.

Notes

Then, an affine connection, D, on M is a differentiable map,

$$D : X(M) \times X(M) \rightarrow X(M),$$

denoted DXY (or rXY), satisfying the following properties:

(1) $D_{fX+gY} Z = fD_X Z + gD_Y Z;$

(2) $D_X(\lambda Y + \mu Z) = \lambda D_X Y + \mu D_X Z;$

(3) $D_X(fY) = fD_X Y + X(f)Y$,

for all λ, $\mu \in \mathbb{R}$, all X, Y, Z $\in$ X(M), and all f, g $\in C^\infty(M)$, where X(f) denotes the directional derivative of f in the direction X.

Thus, an affine connection is $C^\infty(M)$-linear in X, $\mathbb{R}$-linear in Y, and satisfies a "Leibnitz" type of law in Y.

For any chart $\varphi : U \rightarrow \mathbb{R}^m$, denoting the coordinate functions by $x_1, \ldots, x_m$, if X is given locally by

$$X(p) = \sum_{i=1}^{m} a_i(p) \frac{\partial}{\partial x_i},$$

then

$$X(f)(p) = \sum_{i=1}^{m} a_i(p) \frac{\partial (f \circ \varphi^{-1})}{\partial x_i}.$$

It can be checked that X(f) does not depend on the choice of chart.

The intuition behind a connection is that $D_X Y$ is the directional derivative of Y in the direction X.

The notion of covariant derivative can be introduced via the following lemma:

Lemma 10: Let M be a smooth manifold and assume that D is an affine connection on M. Then, there is a unique map, D, associating with every vector field V along a curve a : I $\rightarrow$ M on M another vector field, DV/dt, along c, so that:

(1) $\frac{D}{dt}(\lambda V + \mu W) = \lambda \frac{DV}{dt} + \mu \frac{DW}{dt}.$

(2) $\frac{D}{dt}(fV) = \frac{df}{dt} V + f \frac{DV}{dt}.$

(3) If V is induced by a vector field Y $\in$ X(M), in the sense that V (t) = Y (α(t)), then

$$\frac{DV}{dt} = D_{\alpha'(t)} Y.$$

Then, in local coordinates, DV/dt can be expressed in terms of the Christoffel symbols, pretty much as in the case of surfaces.

Parallel vector fields, parallel transport, geodesics, are defined as before.

Affine connections are uniquely induced by Riemmanian metrics, a fundamental result of Levi-Civita.

In fact, such connections are compatible with the metric, which means that for any smooth curve α on M and any two parallel vector fields X, Y along ", the inner product $\langle X, Y \rangle$ i is constant.

Such connections are also symmetric, which means that

$$D_X Y - D_Y X = [X, Y],$$

where [X, Y] is the Lie bracket of vector fields.

18.3 Joachimsthal Theorem and Notation

If the curve of intersection of two surfaces is a line of curvature on both, the surfaces cut at a constant angle. Conversely, if two surfaces cut at a constant angle, and the curve of intersection is a line of curvature on one of them, it is a line of curvature on the other also

Ferdinand Joachimsthal (1818-1861) was a German mathematician and educator famous for the high quality of his lectures and the books he wrote. The notations named after him and discussed below serve one of the examples where the language of mathematics is especially auspicious for derivation and memorization of properties of mathematical objects. Joachimsthal's notations have had extended influence beyond the study of second order equations and conic sections, compare for example the work of F. Morley.

A general second degree equation

$$Ax^2 + 2Bxy + Cy^2 + 2Fx + 2Gy + H = 0 \qquad ...(1)$$

represents a plane conic, or a conic section, i.e., the intersection of a circular two-sided cone with a plane. The equations for ellipses, parabolas, and hyperbolas all can be written in this form. These curves are said to be non-degenerate conics. Non-degenerate conics are obtained when the plane cutting a cone does not pass through its vertex. If the plane does go through the cone's vertex, the intersection may be either two crossing straight lines, a single straight line and even a point. These point sets are said to be degenerate conics. In the following, we shall be only concerned with a non-degenerate case.

The left-hand side in (1) will be conveniently denoted as s:

$$s = Ax^2 + 2Bxy + Cy^2 + 2Fx + 2Gy + H \qquad ...(2)$$

so that the second degree equation (1) acquires a very short form:

$$s = 0. \qquad ...(3)$$

A point $P(x_1, y_1)$ may or may not lie on the conic defined by (1) or (3). If it does, we get an identity by substituting $x = x_1$ and $y = y_1$ into (1):

$$Ax_1^2 + 2Bx_1y_1 + Cy_1^2 + 2Fx_1 + 2Gy_1 + H = 0, \qquad ...(4)$$

which has a convenient Joachimsthal's equivalent

$$s_{11} = 0. \qquad ...(5)$$

For another point $P(x_2, y_2)$ we similarly define s_{22} and, in general, for points $P(x_i, y_i)$ or $P(x_j, y_j)$ we define s_{ii} and s_{jj}, where, for example,

$$s_{ii} = Ax_i^2 + 2Bx_iy_i + Cy_i^2 + 2Fx_i + 2Gy_i + H. \qquad ...(6)$$

Thus, $s_{ii} = 0$ means that $P(x_i, y_i)$ lies on the conic (3), $s_{ii} \neq 0$ that it does not.

There is also a mixed notation. For two points $P(x_i, y_i)$ and $P(x_j, y_j)$, we define

$$s_{ij} = Ax_ix_j + B(x_iy_j + x_jy_i) + Cy_iy_j + F(x_i + x_j) + G(y_i + y_j) + H. \qquad ...(7)$$

Clearly for $P(x_i, y_i) = P(x_j, y_j)$, (7) reduces to (6). An important observation is that s_{ij} is symmetrical in its indices:

$$s_{ij} = s_{ji}. \qquad ...(8)$$

The last of Joachimsthal's conventions brings the first whiff of an indication as to how useful the notations may be. In s_{ij} both $P(x_i, y_i)$ and $P(x_j, y_j)$ are quite generic. The indices are only needed to distinguish between two points. But if we omit the indices from one of them, the points will be as distinct as before. One additional convention accommodates this case: for points $P(x_i, y_i)$ and $P(x, y)$ we write (7) with one index only,

$$s_i = Ax_ix + B(x_iy + xy_i) + Cy_iy + F(x_i + x) + G(y_i + y) + H. \qquad ...(9)$$

The curious thing about (9) is that, although s_{ij} was probably perceived as a number, s_i appears to dependent on "variable" x and y and thus is mostly perceived as a function of these variables. As a function of x and y, (9) is linear, i.e. of first degree, so that $s_i = 0$ is an equation of a straight line. What straight line is it? How does it relate to the conic (1)? The beauty of Joachimsthal's notations is that the relation between s = 0 and $s_i = 0$ is quite transparent.

Theorem 11

Let point $P(x_i, y_i)$ lie on the conic s = 0. In other words, assume that $s_{ii} = 0$. Then $s_i = 0$ is an equation of the line tangent to s = 0 at $P(x_i, y_i)$.

Proof

Any point P(x, y) on the line through two distinct points $P(x_1, y_1)$ and $P(x_2, y_2)$ is a linear combination of the two points:

$$P(x, y) = t \cdot P(x_1, x_1) + (1 - t) \cdot P(x_2, x_2), \qquad ...(10)$$

which is just a parametric equation of the straight line. Substitute (10) into (2). The exercise may be a little tedious but is quite straightforward. The result is a quadratic expression in t:

$$s(t) = t^2 \cdot (s_{11} + s_{22} - 2s_{12}) + 2t \cdot (s_{12} - s_{22}) + s_{22}. \qquad ...(11)$$

Line (10) and conic (1) will have 0, 1, or 2 common points depending on the number of roots of the quadratic equation s(t) = 0, which is determined by the value of the discriminant

$$D= (s_{12} - s_{22})^2 - (s_{11} + s_{22} - 2s_{12}) \cdot s_{22} = s_{12}^{\ 2} - s_{11} \cdot s_{22}. \qquad ...(12)$$

The line is tangent to the conic if the quadratic equation has two equal roots, i.e. when D = 0, or

$$s_{12}^{\ 2} = s_{11} \cdot s_{22}. \qquad ...(13)$$

This is an interesting identity valid for any line tangent to the conic, with $P(x_1, y_1)$ and $P(x_2, y_2)$ chosen arbitrarily on the line. We can use this arbitrariness to our advantage. Indeed, what could be more natural in these circumstances than picking up the point of tangency. Let's $P(x_1, y_1)$ be such a point. This in particular means that the point lies on the conic so that, according to (5), $s_{11} = 0$. But then (12) implies $s_{12} = 0$. Let's say this again: If $P(x_1, y_1)$ is the point of tangency of a conic and a line through another arbitrary point $P(x_2, y_2)$ then

$$s_{12} = 0. \qquad ...(14)$$

Now, since this is true for any point $P(x_2, y_2)$ on the tangent at $P(x_1, y_1)$ we may as well drop the index. The conclusion just drops into our lap: the tangent to a conic s = 0 at point $P(x_1, y_1)$ on the conic is given by

$$s_1 = 0, \qquad ...(15)$$

which proves the theorem.

Tangent Pair

Notes

If two points $P(x_1, y_1)$ and $P(x_2, y_2)$ are such that the line joining them is tangent to a conic s = 0, then as in (13), $s_{12}^2 = s_{11} \cdot s_{22}$. Now, fixing $P(x_1, y_1)$ outside the conic and making $P(x_2, y_2)$ an arbitrary point on the tangent from $P(x_1, y_1)$, we can remove the second index:

$$s_1^2 = s_{11} \cdot s. \quad ...(16)$$

The latter is a quadratic equation which may be factorized into the product of two linear equations each representing a tangent to the conic through $P(x_1, y_1)$.

Example:

Let $s = x^2 + 4y^2 - 25$, so that s = 0 is an ellipse $x^2 + 4y^2 = 25$. What are the tangents from P(0, 0) to the ellipse? Let's see that there are none. First,

$s_{11} = -25$ and $s_1 = -25$.

So that (16) becomes

$s = -25$, or$x^2 + 4y^2 = 0$.

Obviously the equation has no real roots (besides x = y = 0), nor linear factors. We conclude that there are no tangents from (0, 0) to the ellipse. Naturally. But let's now take a different point, say P(5, 5/2). In this case,

$s_{11} = 25$ and $s_1 = 5x + 10y - 25$.

(16) then becomes

$(5x + 10y - 25)^2 = 25 \cdot (x^2 + 4y^2 - 25)$.

First, let's simplify this to

$(x + 2y - 5)^2 = x^2 + 4y^2 - 25$.

Second, let's multiply out and simplify by collecting the like terms:

$2xy - 5x - 10y + 25 = 0$,

which is factorized into

$(x - 5) \cdot (2y - 5) = 0$.

Conclusion: here are two tangents from (5, 5/2) to the ellipse: x = 5 and y = 5/2.

Poles and Polars With Respect To a Conic

Let $P(x_1, y_1)$ be a point outside a conic s = 0 and $P(x_2, y_2)$ and $P(x_3, y_3)$ be the points where the tangents from $P(x_1, y_1)$ meet the conic.

Then the tangents have the equations (15)

$$s_2 = 0 \text{ and } s_3 = 0 \quad ...(17)$$

and also meet at $P(x_1, y_1)$:

$$s_{21} = 0 \text{ and } s_{31} = 0. \quad ...(18)$$

Because of the symmetry of the notations, we have

$$s_{12} = 0 \text{ and } s_{13} = 0, \quad ...(19)$$

which says that points $P(x_2, y_2)$ and $P(x_3, y_3)$ lie on the straight line

$$s_1 = 0. \quad ...(20)$$

The latter is uniquely determined by $P(x_1, y_1)$, which, too, can be retrieved from (20). We define $s_1 = 0$ as the polar of $P(x_1, y_1)$ with respect to the conic $s = 0$. $P(x_1, y_1)$ is said to be the pole of its polar. Obviously, for a point on the conic, the polar is exactly the tangent at this point.

Thus we see that the pole/polar definitions generalize naturally from the circle to other non-degenerate conics. We now prove La Hire's

Theorem 12

If point $P(x_1, y_1)$ lies on the polar of $P(x_2, y_2)$ with respect to a conic $s = 0$, then $P(x_2, y_2)$ lies on the polar of $P(x_1, y_1)$ with respect to the same conic.

Proof

Indeed, $P(x_1, y_1)$ lies on the polar $s_2 = 0$ if and only if $s_{21} = 0$. Because of the symmetry of the notations, this is the same as $s_{12} = 0$, which says that $P(x_2, y_2)$ lies on $s_1 = 0$.

18.4 Tissot's Theorem

At any point on a reference globe there are an infinite number of paired orthogonal directions. When transformed to map they may or may not remain orthogonal

Tissot's theorem states that regardless of the type of transformation, at each point on a sphere there is at least one pair of orthogonal directions that will remain orthogonal when transformed Referred to as principle directions; a and b and it is not important what directions actually are

Tissot's theory of distortions states that

A circle on the datum surface with a centre P and a radius ds may be assumed to be a plane figure within its infinitely small area. This area will remain infinitely small and plane on the projection surface. Generally the circle will be portrayed as a ellipse.

This ellipse is called *Tissot's Indicatrix* as it indicates the characteristics of a projection in the direct environment of a point.

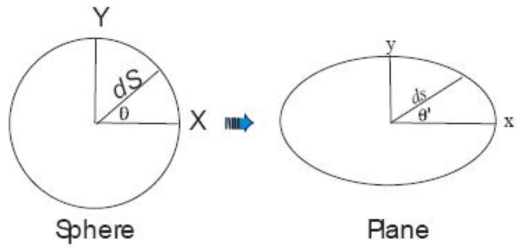
The axes of Tissot's Indicatrix correspond to the two principal directions and the maximum and minimum particular scales, *a* and *b*, at any point, occur in these directions.

Proof That The Projected Circle Is An Ellipse

Notes In figure, the X axis is directed east-west; Y axis is directed north-south.

Remember that capital letters denote elements on the generating globe, and small letters elements on the projection.

Figure 18.1: Plane and Sphere



$$\left.\begin{array}{l} dY = dS \,.\, \text{Sin}\,\theta = \sqrt{G} \,.\, d\lambda \\ dX = dS \,.\, \cos\theta = \sqrt{E} \,.\, d\phi \end{array}\right\} \text{on the generating globe}$$

$$\left.\begin{array}{l} dy = ds \,.\, \text{Sin}\,\theta' = \sqrt{g} \,.\, d\lambda \\ dx = ds \,.\, \cos\theta' = \sqrt{e} \,.\, d\phi \end{array}\right\} \text{on the projection}$$

$$d\phi = \frac{1}{\sqrt{E}} dS \,.\cos\theta$$

$$d\lambda = \frac{1}{\sqrt{G}} dS \,.\sin\theta$$

$$\therefore dy = \sqrt{g} \,.\, \frac{1}{\sqrt{G}} dS \,.\sin\theta$$

$$\text{and} : dx = \sqrt{e} \,.\, \frac{1}{\sqrt{E}} dS \,.\cos\theta$$

$$\frac{dx^2}{E/e} = dS^2 \,.\, \cos^2\theta; \;\; \frac{dy^2}{G/g} = dS^2 \,.\, \sin^2\theta$$

$$\frac{dx^2}{E/e} + \frac{dy^2}{G/g} = (\sin^2 q + \cos^2\theta)dS^2$$

$$\frac{dx^2}{E/e} + \frac{dy^2}{G/g} = dS^2$$

If dS = 1 then the elementary circle on the globe has a radius of 1 (remember that capital letters denote elements on the generating globe, and small letters elements on the projection.)

$$\frac{dx^2}{E/e} + \frac{dy^2}{G/g} = 1$$

This is an equation of an ellipse.

Analysis of Deformation Characteristics using Tissot's Indicatrix

If we call the semi-major and semi-minor axes of the ellipse a, and b, then these are the directions of maximum and minimum distortion i.e. the principal directions. a and b are also thus called the *principal scale factor*

$$\frac{x^2}{b} + \frac{y^2}{a} = 1$$

For convenience we will consider the plane x and y axes to be in the principal directions.

Length Distortion

$$\mu_x = ds \cos\theta' \text{ on the plane}$$

$$\mu_X = dS \cos\theta = 1 \text{ on globe}$$

(There is no distortion on the globe)

Remember from previous section: $m = \frac{ds}{dS}$, so:

$$a = \left(\frac{ds}{dS}\right) = \frac{\mu \cos\theta'}{\cos\theta}$$

$$b = \left(\frac{ds}{dS}\right)_y = \frac{\mu \sin\theta'}{\sin\theta}$$

or

$$a \cos\theta = \mu \cos\theta'$$

$$b \sin\theta = \mu \sin\theta'$$

$$a^2 \cos^2\theta = \mu^2 \cos^2\theta'$$

$$b^2 \sin^2\theta = \mu^2 \sin^2\theta'$$

$$a^2 \cos^2\theta + b^2 \sin^2\theta = \mu^2(\cos^2\theta' + \sin^2\theta')$$

$$\mu^2 = a^2 \cos^2\theta + b^2 \sin^2\theta$$

This formula expresses the length distortion in any direction as a function of the original direction θ, and the principal scale factors, a and b. The angle θ indicates the direction of the parallel with respect to the x axis. The direction of the meridian with respect to the x axis is thus

$$\theta + 90^o = \theta + \pi/2 = \beta$$

The scale distortions along the parallels and meridians (note: not necessarily equal to the maximum and minimum distortions along a and b) are thus:

$$\mu_\lambda^2 = a^2 \sin^2\beta + b^2 \cos^2\beta$$

$$\mu_\phi^2 = a^2 \sin^2\alpha + b^2 \cos^2\alpha = a^2 \cos^2\beta + b^2 \sin^2\beta$$

$$\mu_\lambda^2 + m_\phi^2 = a^2 + b^2$$

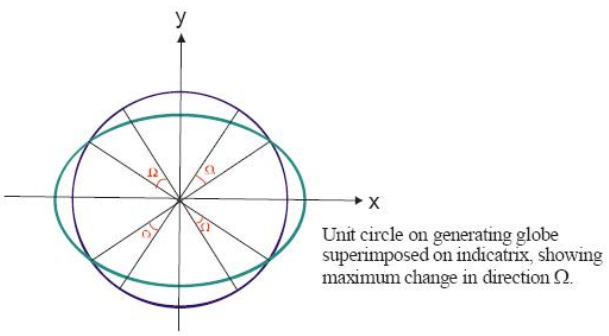
This is known as the First Theorem of Appolonius:

The sum of the squares of the two conjugate diameters of an ellipse is constant.

Angular Distortion 2[redacted]

Without derivation: $2\Omega = 2 \text{ arc} \sin\frac{a-b}{a-b}$, where 2Ω is the maximum angular distortion. The maximum angular deformation occurs in each of the four quadrants.

Figure 18.2: Angular Distortion



If $2\Omega = 0$ then no angular distortion occurs and the projection is called conformal. The property of a conformal projection is that $a = b$ and Tissot's Indicatrix is a circle with equal scale distortion in all directions. This is consistent with the previously derived conditions for conformality, namely that $\mu_\phi = \mu_\lambda$. and $\theta' = \frac{\pi}{2}$. The area is not preserved and the projected circle increases in size as one moves away from the line of zero distortion.

Areal Distortion (s): Second Theorem of Appolonius.

This is found by dividing the projected area by the area of the circle on the globe (radius =1)

$$\sigma = \frac{\pi ab}{\pi R^2} = ab$$

When looking at equal area projections earlier. It was found that:

$$\sigma = \mu_\phi \mu_\lambda \sin\theta', \text{ thus } \mu_\phi \mu_\lambda \sin\theta' = ab.$$

This is called the Second Theorem of Appolonius. When $ab = 1$ then the projection is equal-area or equivalent.

> *Notes* Conformality and equivalence are exclusive: $ab = 1$ and $a = b$ cannot occur at the same time.

18.5 Summary

- Geodesics play a very important role in surface theory and in dynamics.

  One of the main reasons why geodesics are so important is that they generalize to curved surfaces the notion of "shortest path" between two points in the plane.

  More precisely, given a surface X, given any two points $p = X(u_0, v_0)$ and $q = X(u_1, v_1)$ on X, let us look at all the regular curves C on X defined on some open interval I such that $p = C(t_0)$ and $q = C(t_1)$ for some $t_0, t_1 \in I$.

  It can be shown that in order for such a curve C to minimize the length $l_C(pq)$ of the curve segment from p to q, we must have $k_g(t) = 0$ along $[t_0, t_1]$, where $k_g(t)$ is the geodesic curvature at $X(u(t), v(t))$.

- Given a surface $X : \Omega \to E^3$, let $\alpha : I \to E^3$ be a regular curve on X parameterized by arc length. For any two points $p = \alpha(t_0)$ and $q = \alpha(t_1)$ on α, assume that the length $l_\alpha(pq)$ of the curve segment from p to q is minimal among all regular curves on X passing through p and q. Then, α is a geodesic.

  At this point, in order to go further into the theory of surfaces, in particular closed surfaces, it is necessary to introduce differentiable manifolds and more topological tools.

  Nevertheless, we can't resist to state one of the "gems" of the differential geometry of surfaces, the local Gauss-Bonnet theorem.

  The local Gauss-Bonnet theorem deals with regions on a surface homeomorphic to a closed disk, whose boundary is a closed piecewise regular curve α without self-intersection.

  Such a curve has a finite number of points where the tangent has a discontinuity.

18.6 Keywords

Exponential map: The exponential map can be used to define special local coordinate systems on normal neighborhoods, by picking special coordinates systems on the tangent plane.

Polar coordinates can be used to prove the following lemma showing that geodesics locally minimize arc length.

Gauss-Bonnet theorem: The local Gauss-Bonnet theorem deals with regions on a surface homeomorphic to a closed disk, whose boundary is a closed piecewise regular curve α without self-intersection.

18.7 Self Assessment

1. The principal normal $\vec{n}$ must be .... to the normal N to the surface along the curve segment from p to q.
2. Given a surface .... a geodesic line, or geodesic, is a regular curve $C : I \to E^3$ on X, such that $k_g(t) = 0$ for all $t \in I$.
3. Given a surface X, for every point $p = X(u, v)$ on X, for every non-null tangent vector $\vec{v} \in T_{(u,v)}(X)$ at p, there is some .... and a unique curve $\gamma :]-\in, \in[\to E^3$ on the surface X, such that γ is a geodesic, $\gamma(0) = p$, and $\gamma'(0) = \vec{v}$.
4. In this case, ... Thus, the closed half-line corresponding to $\theta = 0$ is omitted, and so is its image under $\exp_p$.
5. The local .... deals with regions on a surface homeomorphic to a closed disk, whose boundary is a closed piecewise regular curve α without self-intersection.
6. Let $X : \Omega \to E^3$ be a surface. Given any open subset, U, of X, a vector field on U is a function, w, that assigns to every point, $p \in U$, some tangent vector ....

18.8 Review Questions

1. Define Geodesic Lines, Local Gauss-Bonnet Theorem.
2. Discuss Covariant Derivative, Parallel Transport, Geodesics Revisited.
3. Describe Joachimsthal's Notations.

Answers: Self Assessment

1.	parallel	2.	$X : \Omega \to E^3$,
3.	$\in > 0$	4.	$0 < \theta < 2\pi$.
5.	Gauss-Bonnet theorem	6.	$w(p) \in T_pX$ to X at p.

18.9 Further Readings

Notes

Books

Ahelfors, D.V. : Complex Analysis

Conway, J.B. : Function of one complex variable

Pati, T. : Functions of complex variable

Shanti Narain : Theory of function of a complex Variable

Tichmarsh, E.C. : The theory of functions

H.S. Kasana : Complex Variables theory and applications

P.K. Banerji : Complex Analysis

Serge Lang : Complex Analysis

H. Lass : Vector & Tensor Analysis

Shanti Narayan : Tensor Analysis

C.E. Weatherburn : Differential Geometry

T.J. Wilemore : Introduction to Differential Geometry

Bansi Lal : Differential Geometry.