SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Riemannian geometry in digital image processing with an application in modeling the cells in the lens of an eye and automating the quantification of a protein

av

Nanna Zhou Hagström

2015 - No 3

Riemannian geometry in digital image processing with an application in modeling the cells in the lens of an eye and automating the quantification of a protein

Nanna Zhou Hagström

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Rikard Bøgvad

2015

RIEMANNIAN GEOMETRY IN DIGITAL IMAGE PROCESSING
WITH AN APPLICATION IN MODELING THE CELLS IN THE LENS OF AN EYE AND AUTOMATING THE QUANTIFICATION OF A PROTEIN

NANNA ZHOU HAGSTRÖM

Abstract. The main objective of this report is understanding mathematics applied in digital imaging processing. We concentrate ourselves on Riemannian structures and study the Riemannian metric on color spaces and image processing of shape. Finally we present an application in modeling the cells in the lens of an eye and automating the quantification of a protein.

Sammanfattning. Syftet med denna rapport är att förstå underliggande matematiken i digital bildbehandling. Vårt fokus ligger på Riemanngeometri. Rapporten presenterar hur Riemanngeometri är tillämpad i färgrum och digitalbildbehandling. Vi presenterar också en tillämpning i modellering av cellerna i ett ögas lins och automatisering av mätningen av en viss protein i cellerna.

Résumé. Le but de ce rapport est de comprendre les mathématiques derrière le traitement d'image numérique. Nous nous sommes concentrés sur les variétés riemanniennes et les tenseurs métriques dans ces variétés appliqués à l'espace des couleurs et au traitement d'images. Nous présenterons aussi une application de ceci dans la modélisation des cellules d'un cristallin et dans la quantification d'une protéine dans ces cellules.

Contents

1. Introduction

Digital image processing is the use of computer algorithms to create, process, communicate, and display digital images. In general it refers to processing of a two dimensional picture by a digital computer. In a broader context it implies digital processing of any two-dimensional data. A variety of rich mathematical topics makes the topic interesting and demanding. Among mathematical subjects appearing in digital image analysis and processing we can find Fourier transform, complex analysis, dynamical system, nonlinear filtering, mathematical morphology, partial differential equations, random fields, and Riemannian geometry, to name a few, in the areas of image perception, sampling and quantization, transformations, for image representation, filtering and restoration, reconstruction from projections, for image data compression and so on. For an overview we refer to [8].

The idea for this report steamed from a research project I participated in. The project was initiated by Professor Carolina Wählby at CBA, Uppsala University affiliated to Science for Life Laboratory. The problem I had been assigned was to create a program that would count the epithelial cells in the lens and compute the intensity of the protein caspase-3 in microscopy images provided to the CBA by the Department of Ophthalmology of Uppsala University. The original purpose was to carry out a two-week internship in my physics program at Université Pierre et Marie Curie, Paris. Without any knowledge of either how microscopy works in medical science and clinical practice or digital image processing or much of underlying mathematics or many experiences of Matlab coding I started a broad program for improving myself, in particular, a better understanding of underlying mathematics.

The focus will be on computation of geodesic distances on Riemannian manifolds for image segmentation, shortest distance and shortest paths, and on geometric transformations of local structure tensor. As pointed out in [17], the notion of Riemannian manifold allows to define a local metric (a symmetric positive tensor field) that encodes the information about the problem one wishes to solve. This takes into account a local isotropic cost (whether some point should be avoided or not) and a local anisotropy (which direction should be preferred). Using this local tensor field, the geodesic distance is used to solve many problems of practical interest such as segmentation using geodesic balls and Voronoi regions, sampling points at regular geodesic distance or meshing a domain with geodesic Delaunay triangles. The shortest path for this Riemannian distance, the so-called geodesics, are also important because they follow salient curvilinear structures in the domain.

Riemannian geometry was a generalization of Gauss theory of surfaces. Riemann introduced the curvature tensor, the sectional curvature and derived the conformal form of the metric of constant curvature. The theory belongs to Differential Geometry. Riemann's construction of the Riemannian manifold consisted first in building the foundation of the smooth manifold. He then established on that foundation the concept of a Riemannian metric. Today it is not too hard to give a correct definition of smooth manifold based on modern general topology and differential calculus. However it took long time. In 1927, Élie Cartan published a textbook on Riemannian manifolds [4] which was the only book on Riemannian geometry up to the 1960's. However, Cartan preferred not to define manifolds precisely. Then many books started to appear. Since the aim of the current report is to giving the author's understanding of some subjects appearing in a practical

problems we are not going to present everything by the style of definition-theorem-proof. We will try to explain why abstraction is needed and how theory can be applied. In this report we also study Riemannian matrices on color spaces and some other issues in image processing. Geometry of color-matching or perception seems to be a fascinating research area since many works in the geometric structure of color are still going on, e.g. [12] and the references therein.

Having decided on doing computation on Riemannian manifolds we meet an immediate difficult task; how to explain and define Riemannian manifolds. It does not seem to be completely possible to do so without speaking of topological and (smooth) manifolds. So we spend some time on these abstract notions and motivates why it is needed by examples.

In §2 we discuss some issues on Euclidean spaces and $\mathbb{R}^n$ for the future use. Then we introduce in §3 the notions of (abstract) manifold and discuss the need of such manifolds in applications. In §4 we collect some basic concepts such as smooth maps, tangent space, tangent bundles, covariant derivatives, connections, curvature and torsion on manifolds. §5 is about Riemannian manifold and metric where we discuss topics like geodesics, curvatures and calculation on moving frames especially as a preparation for §6 where we study geodesic distance together with some examples from image analysis and processing and we do some tensor calculations which appearing in color space of image processing. Finally we present how our project is carried out and concluded by some comments on further possible direction of research. Matlab codes are included in the Appendix with permission from the research team I was involved with.

Acknowledgments. I would like to thank Professor Carolina Wählby, who introduced me to this fascinating research area where mathematics, computer science, physics and medical science meet. She guided me in research topics and helped me with everything, from understanding material to coding with Matlab, through out the project work. Her inspiration and enthusiasm encourage me to overcome difficulties and the time shortage. I would also thank The Physics Department at Université Pierre et Marie Curie, Sorbonne Universités who approved my practice in Uppsala. I would also like to thank PhD candidate Nooshin Talebizadeh and Professor Per Söderberg from the Gullstrand laboratory of Ophthalmology at Akademiska Sjukhus of Uppsala University. Many thanks go to the research team at CBA for invaluable discussions and seminars. I am very grateful to Professor Rikard Bøgvad at Stockholm University for taking care of me for doing mathematics in distance in order to finish this report.

2. **Euclidean spaces and $\mathbb{R}^n$**

The best way to approach the subjects of differential geometry is perhaps doing calculus on manifolds in $\mathbb{R}^n$ as done in e.g. [13, 20] since we are all familiar with the set $\mathbb{R}^n$ and know vector analysis in $\mathbb{R}^3$ and the geometry in a plane and a solid space. After that it will be easier to understand the abstract definition of a manifold. That is perhaps a reason why Cartan avoided using a clear concept of a manifold, rather use examples and considerations in his book on Riemannian manifolds. Since a manifold is considered locally to be like $\mathbb{R}^n$, we discuss different views of this space. A big portion of the text in this section is based on [3].

2.1. **Different views of the space $\mathbb{R}^n$**

The space $\mathbb{R}^n$ is the set of all ordered n-tuples $(x^1, x^2, ..., x^n)$, often denoted x, of real numbers. In other words, it is an n-fold Cartesian product $\underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n}$. In this report we use the topology on $\mathbb{R}^n$ as a metric space with the metric defined by

$$d(x, y) = \left(\sum_{i=1}^{n} (x^i - y^i)^2 \right)^{1/2}.$$

The neighborhoods are open balls with radius $\delta > 0$ and centered at $a \in \mathbb{R}^n$

$$B_\delta(a) = \{x \in \mathbb{R}^n : d(x, a) < \delta\}$$

or open cubes of sides 2δ and centered at a

$$C_\delta(a) = \{x \in \mathbb{R}^n : |x^i - a^i| < \delta, i = 1, ..., n\}$$

In fact, the latter is an open "ball" if we choose to use $d_\infty(x, y) = \max_{1 \le i \le n} |x^i - y^i]$ as another metric on $\mathbb{R}^n$ and these two metrics are equivalent.

The space $\mathbb{R}^n$ will be used in several ways, as a metric space with the topology defined by the metric, or simply a topological space, or sometimes denotes an n-dimensional vector space, and sometimes it is identified with a Euclidean space.

From linear algebra we learned many theorems. Among them is the isomorphism theorem that says *any two vector spaces over* $\mathbb{R}$ *with the same dimension* n *are isomorphic.* However, the isomorphism depends on the choices of bases in the two spaces. In general there is no natural or canonical isomorphism independent of these choices. Nevertheless there does exist one such example of vector space over $\mathbb{R}$. For the vector space of the n-tuple over $\mathbb{R}$ with component wise addition and multiplication by scalar simply denoted as $\mathbb{R}^n$ the basis $e_1 = (1, 0, .., .0)$,..., $e_n = (0, ..., 0, 1)$ are a natural basis, we often call them standard basis in the textbooks.

Sometimes we may mean more by the notation $\mathbb{R}^n$. An abstract vector space over $\mathbb{R}$ is called *Euclidean* if it is equipped with a (positive) inner product, In general there is no natural way to choose such an inner product, but in the case of $\mathbb{R}^n$ we have the natural (standard) inner product

$$(x, y) = \sum_{i=1}^{n} x_i y_i.$$

Often we can see the use of dot for this inner product on $\mathbb{R}^n$, $x \cdot y$. Using this inner product we can characterize geometric concepts such as orthogonality of two vectors. Apparently $(e_i, e_j) = \delta_{ij}$. Thus $\mathbb{R}^n$ as a Euclidean space has a built-in orthonormal basis and inner product. For an abstract vector space even if Euclidean, there is no such preferred basis.

The metric on $\mathbb{R}^n$ defined at the beginning can be defined using the inner product on $\mathbb{R}^n$. We denote $||x||$, the norm of the vector x, by $||x]] = (x, x)^{1/2}$. Then we have

$$d(x, y) = ||x - y||.$$

We use this notation even when we consider $\mathbb{R}^n$ as a metric space without using structure of vector space. In particular, $||x|| = d(x, 0)$, the distance from 0 to x. Note that the x in the left hand side is a vector while in the right hand side it is a point in $\mathbb{R}^n$. This is a clear example to show how the space $\mathbb{R}^n$ can be interpreted in a mixed way.

2.2. More about $\mathbb{R}^n$ as a Euclidean space

The space $\mathbb{R}^n$ plays an important role in linear algebra, e.g., when we study linear transformations from a vector space to another vector space we can use matrix representations which is just like the computations in $\mathbb{R}^n$. It also play an important role as a model for n-dimensional Euclidean space $\mathbb{E}^n$ in the sense of Euclidean geometry.

We are often taught to identify Euclidean spaces with $\mathbb{R}^n$. However it is not a complete picture which is perhaps the obstacle for many of us in understanding the concept of abstract manifolds and the role of coordinates. Next, we'll discuss what more is involved. The identification of $\mathbb{R}^n$ and $\mathbb{E}^n$ dates back to Fermat and Descartes and it led in part to the discovery of non-Euclidean geometries and thus to manifolds. A very careful axiomatic definition of Euclidean space is given by Hilbert [1].

The chronological order of our mathematical training is that we started with definitions and proving theorems in Euclidean plane $\mathbb{E}^2$ without coordinates. Later we introduced coordinates using the notions of length and perpendicularity in choosing two mutually perpendicular number axes which are used to define a one-to-one mapping of $\mathbb{E}^2$ onto $\mathbb{R}^2$ by $p \mapsto (x(p), y(p))$, the coordinates of $p \in \mathbb{R}^2$. This mapping is isometry, preserving distances of points of $\mathbb{E}^2$ and their images in $\mathbb{R}^2$. Finally we obtain further correspondences of essential geometric elements such as lines of $\mathbb{E}^2$ with subsets of $\mathbb{R}^2$ consisting of the solutions of linear equations. Hence we carry each geometric object to a corresponding one in $\mathbb{R}^2$. It is the existence of such coordinate mappings which make the identification of $\mathbb{E}^2$ and $\mathbb{R}^2$ possible. However, there is no natural, geometrically determined way to identify the two spaces. In this sense, we can say that $\mathbb{R}^2$ may be identified with $\mathbb{E}^2$ plus a coordinate system. This being said in this way we still need to define in $\mathbb{R}^2$ the notion of line, angle of lines, and other Euclidean geometric attributes before considering $\mathbb{R}^2$ as a Euclidean space.

Sometimes we do not wish to make the identification, that is use the analytic geometry approach to the study of geometry so to speak. Let's look at an example. Having identified $\mathbb{E}^2$ with $\mathbb{R}^2$ and the lines with the solutions of the linear equations, for example $\ell = \{(x, y) : y = mx+b\}$ we define the slope m and the y-intercept b. This does not give us a geometric meaning in itself because it depends on the choice of the coordinates. Now consider two such lines ℓ_1 and ℓ_2 with slopes m_1 and m_2, respectively, depicted in Figure 1. Here the

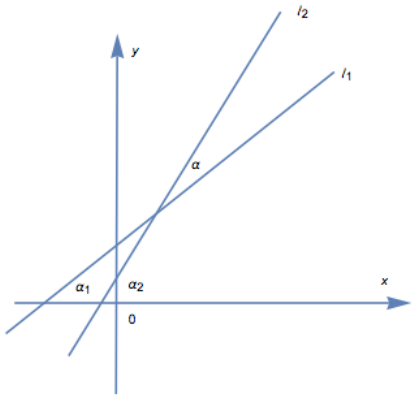


FIGURE 1.

angle between line ℓ_i and x-axis is α_i, $i = 1, 2$ and the angle between the two lines is α.

By Euclidean geometry $\alpha = \alpha_2 - \alpha_1$. We know that $m_i = \tan\alpha_i$. Then we obtain, using a little trigonometry,

$$\tan\alpha = \frac{\tan\alpha_2 - \tan\alpha_1}{1 + \tan\alpha_1 \tan\alpha_2} = \frac{m_2 - m_1}{1 + m_1 m_2}$$

So the quantity $(m_2 - m_1)/(1 + m_1 m_2)$ has a geometric meaning. Basically it describes the angle between the two lines, a concept independent of coordinate choices.

This illustrates the difficulty of doing geometry by working on coordinates alone. It is clear that we need to develop both coordinate methods and coordinate-free methods. Hence mathematicians often look for ways of study manifolds and their geometry which do not involve coordinates, but will use coordinates as e.g. computational tools when necessary.

In conclusion, we usually refer to $\mathbb{R}^n$ as Euclidean space and make the identification. This is particularly true when we are interested in questions involving topology.

3. **Abstract manifolds**

In this section we will follow Cartan at the beginning and give some examples to show what are not manifolds. And later we give the definitions of topological and smooth manifolds. We are not going to repeat the knowledge on vector-valued several variable functions, e.g. [13], and vector analysis at the elementary level. For general topology, we refer to [14].

As we have seen, the metric space $\mathbb{R}^n$ serves as a topological model for Euclidean space $\mathbb{E}^n$, for finite-dimensional vector spaces over $\mathbb{R}$ or $\mathbb{C}$, it is natural for us to study spaces which are locally like $\mathbb{R}^n$.

A map is smooth if it admits derivatives of any order. Roughly speaking an n dimensional *smooth manifold* is a topological space which is everywhere locally smoothly equivalent to $\mathbb{E}^n$. These local equivalences are called charts or coordinate systems, the essential condition being that they overlap, two charts are related by a smooth diffeomorphism, that is, a bijection which is smooth, and so is its inverse. So a loop curve is not a manifold, neither is a surface (say in $\mathbb{E}^3$) with corners or edges. However, a circle and a 2-space which may be defined to be all points of $\mathbb{E}^2$ respectively $\mathbb{E}^3$ at unit distance from a fixed point 0, are manifolds.

However locally being like $\mathbb{E}^n$ is not enough. There are two technical points which make the correct notion of manifold difficult. It is not so difficult to define a smooth manifold as a set covered by charts, which are smoothly related to one another where their domains overlap. But this won't always work. The first problem is that such a manifold can be too large, for example the so-called long-line (see e.g. [27]) which is locally as $\mathbb{R}$ but it is pathological at infinity. The second problem is that it might fail to be separated, i.e. not Hausdorff. A commonly used example is the line with two origins. This space is created by replacing the origin of the real line with two points, an open neighborhood of either of which includes all nonzero numbers in some open interval centered at zero. This space is not Hausdorff because the two origins cannot be separated, [28]. This leads to the following definition of the topological manifolds which can be found in any modern textbooks on differential geometry.

3.1. **Definitions of smooth manifolds**

Definition. A *manifold* M of dimension n, or *n-manifold* is a topological space with the following properties:

(i) M is Hausdorff, i.e. distinct points have disjoint neighborhoods.
(ii) M is locally Euclidean of dimension n, i.e. each point $p \in M$ has a neighborhood U which is homeomorphic to an open subset $U' \subset \mathbb{R}^n$, with n fixed.
(iii) M has a countable basis of open sets.

When M is locally Euclidean of dimension n we say that M has dimension n. When $\dim M = 0$ then M is a countable space with the discrete topology. It is clear by definition, that if $\dim M = 1$ then M is locally homeomorphic to an open interval, if $\dim M = 2$ M is locally homeomorphic to an open disc, and in general an n-manifold is locally homeomorphic to an n-open ball in $\mathbb{R}^n$.

Note that if one is not familiar with topological spaces, just think that M is a subset of $\mathbb{R}^N$ for a large N. An open subset M of $\mathbb{R}^n$ with the subspace topology is an n-manifold. The properties (i) and (iii) are from the topology M equipped (which are satisfied for any subspace of a space which possesses them. We see that (ii) holds with $U = U' = M$ and with the homeomorphism of U to U' being the identity map.

Note also that an n-manifold is not necessarily globally equivalent to $\mathbb{E}^n$, that is not globally homeomorphic to $\mathbb{E}^n$. The following example serves as a counter example.

Example. (Circles S^1 and the 2-spheres S^2). Circles S^1 and the 2-spheres S^2 can be defined to be all points of $\mathbb{E}^2$, or of $\mathbb{E}^3$, respectively, which are at distance from a fixed point 0. (The objects traditionally called "circles" in 2-space, or "surfaces" in 3-space.)

Proof. Since S^1 and S^2 are to be taken with subspace topology so (i) and (iii) are obvious. Now we show that they are locally Euclidean. Introduce coordinate axes with 0 as origin in corresponding ambient Euclidean space. Consider the case S^2. Identify $\mathbb{R}^3$ and $\mathbb{E}^3$. Then S^2 becomes a unit sphere centered at the origin. For any point $p \in S^2$ we have a tangent plane and a unit normal vector N_p. There will be a coordinate axis which is not perpendicular to N_p and some neighborhood U of p on S^2. We project U in a continuous and one-to-one way onto an open set U' of the coordinate plane perpendicular to that axis. See Figure 2 to the left, where N_p is not perpendicular to the x_2-axis. So for $q \in U$, the projection is given explicitly by $\varphi(q) = (x^1(q), 0, x^3(q))$, where $(x^1(q), x^2(q), x^3(q))$ are the coordinates of q in $\mathbb{E}^3$. In a similar way we can prove the local Euclidean property of S^1. Note that S^2 and $\mathbb{R}^2$ cannot be homeomorphic since S^2 is compact but $\mathbb{R}^2$ is not. □

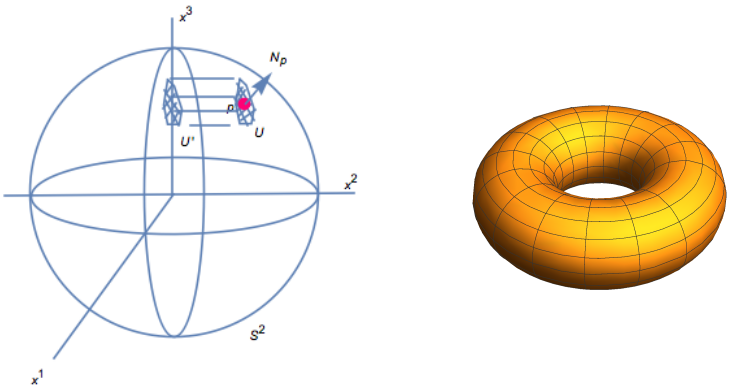


FIGURE 2.

Example. (Tori.) A torus, T^2 is a surface of revolution obtained by moving a circle around an axis which does not intersect it. This figure can be analyzed analytically. It is the image of the map $f : [0, 2\pi) \times [0, 2\pi) \to \mathbb{R}^3$ defined by

$$f(s,t) = ((b + a\cos s)\cos t, (b + a\cos s)\sin t, a\sin s)$$

For $b = 2, a = 1$ the surface is shown in Figure 2 to the right.
We have to prove that it is locally Euclidean. As in the previous example, we consider the normal vector N_p at $p \in T^2$. There will be at least one coordinate axis to which it is not perpendicular, say x^3. Then some neighborhood U of p projects homeomorphically onto a neighborhood U' in the x^1x^2-plane. Since we use the relative topology derived from $\mathbb{E}^3$ the T^2 is necessarily Hausdorff and has a countable basis of open sets. So it satisfies all three conditions in the definition of a topological manifold. So T^2 is a manifold.

There are several observations from these examples. First some subspaces M of $\mathbb{E}^n$ are easily seen to be 2-manifolds; they are surfaces which are "smooth", i.e. there are no corners or edges, so they have at each point $p \in M$ a (unit) normal vector N_p and tangent plane $T_p(M)$, which varies continuously as we move from point to point. It is this smoothness that we use to prove the locally Euclidean property by projection of a neighborhood of p onto a plane as done in the above two examples. Since we use the subspace topology the other two properties are evident. It is also obvious that this method will not always work. The surface of a cube is a 2-manifold which is homeomorphic to S^2, but it has no tangent plane on normal vector at the corners and edges.

The second thing we observe is that the n-sphere S^n is an n-manifold with similar argument for S^2. However the closed n-disc D is not a manifold by definition. This is an example of *manifolds with boundary* (S^{n-1} is the boundary of D^n). The formal definition is as follows

Definition. (Manifold with boundary). A Hausdorff space M is called an *n-manifold with boundary* ($n \geq 1$) if each point in M has a neighborhood homeomorphic to an open set in the half space

$$\mathbb{R}^n_+ = \{(x_1, ..., x_n) \in \mathbb{R}^n : x_n \geq 0\}.$$

We mention two more examples of manifolds with boundary, hemispherical cap (including the equator) and a right circular cylinder (including the circles at the end). They can be used to construct the manifolds 2-sphere S^2 and torus T^2 by pasting two discs (or hemispheres) together so as to form the equator, and T^2 formed by pasting the two end-circles of a cylinder together. In fact new surfaces can be formed by fastening together manifolds with boundary along their boundaries, i.e. by identifying points of various boundary components by a homeomorphism, assuming the necessary condition that such components are homeomorphic. We can even go further and paste any number of cylinders onto a sphere S^2 with "holes" that is, with circular discs removed. This gives variety of Pretzel-like surfaces. In summary, to generate new 2-manifolds from old ones we may cut out two disks, leaving a manifold M with boundary ∂M is the disjoint union of two circles, and then paste on a cylinder or "handle" so that each end-circle is identified with one of the boundary circles of M. For the torus T^2 we can also construct from a square by pasting the outsides to a cylinder then to a torus.

Let U be an open set of the manifold M and φ is a homeomorphism of U to an open subset of $\mathbb{R}^n$. The pair (U, φ) is called a *coordinate neighborhood* or *chart*: To $q \in U$ we assign the n coordinates $x^1(q), x^2(q), ..., x^n(q)$ of its image $\varphi(q)$ in $\mathbb{R}^n$, where each $x^i(q)$ is a real-valued function on U, the ith coordinate function. If q lies also in the second coordinate neighborhood (V, ψ), then it has coordinates $y^1(q), ..., y^n(q)$ in this

neighborhood. Since φ and ψ are homeomorphisms, this defines a homeomorphism

$$\psi \circ \varphi^{-1} : \varphi(U \cap V) \to \psi(U \cap V)$$

the domain and range being the two open subsets of $\mathbb{R}^n$ which correspond to a point in $U \cap V$ by the two coordinate maps φ, ψ, respectively. In coordinates, $\psi \circ \varphi^{-1}$ is given by continuous functions $y^i = h^i(x^1, ..., x^n)$, $i = 1, ..., n$. This gives the y-coordinates of each $q \in U \cap V$ in terms of its x-coordinates. Similarly $\varphi \circ \psi^{-1}$ gives the inverse mapping which express the x-coordinates as functions of the y-coordinates $x^i = g^i(y^1, ..., y^n)$, $i = 1, ..., n$.

Note that the fact $\varphi \circ \psi^{-1}$ and $\psi \circ \varphi^{-1}$ are homeomorphisms and are inverse to each other is equivalent to the continuity of $h^i(x)$ and $g^j(y)$, $i, j = 1, ..., n$ together with the identities

$$h^i(g^1(y), ..., g^n(y)) \equiv y^i, \ i = 1, ..., n$$

and

$$g^j(h^1(x), ..., h^n(x)) \equiv x^j, \ j = 1, ..., n.$$

Therefore every point of a topological manifold M lies in a very large collection of coordinate neighborhoods, but whenever two neighborhoods overlap we have the formulas just given for change of coordinates. The basic idea leading to smooth manifolds is to try to select a family or subcollection of neighborhoods so that the change of coordinates is always given by differentiable functions.

Definition. We say that (U, φ) and (V, ψ) are C^∞-*compatible* if non-emptiness of $U \cap V$ implies that the functions $h^i(x)$ and $g^j(y)$ giving the change of coordinates are C^∞; this is equivalent to requiring $\varphi \circ \psi^{-1}$ and $\psi \circ \varphi^{-1}$ be *diffeomorphisms* of the open subsets $\varphi(U \cap V)$ and $\psi(U \cap V)$ of $\mathbb{R}^n$.

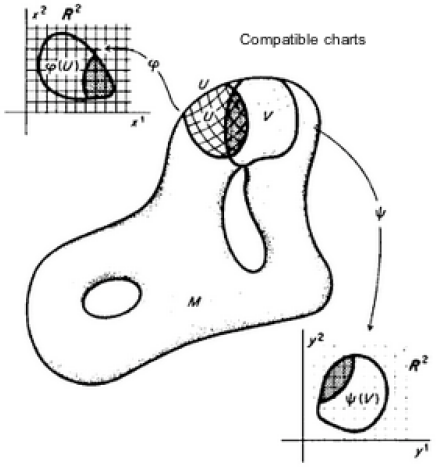


FIGURE 3. Illustration of compatible charts

Definition. A *differentiable* or C^∞ (or *smooth*) structure on a topological manifold M is a family $\mathscr{U} = \{(U_\alpha, \varphi_\alpha) : \alpha \in J\}$ of coordinate neighborhoods, called an *atlas*, such that

(i) the U_α cover M,
(ii) for any α, β the neighborhoods $(U_\alpha, \varphi_\alpha)$ and (U_β, φ_β) are C^∞-compatible,
(iii) any coordinate neighborhood (V, ψ) compatible with every $(U_\alpha, \varphi_\alpha) \in \mathscr{U}$ is itself in $\mathscr{U}$.

A C^∞ *manifold* or *smooth* manifold is a topological manifold together with a C^∞-differentiable structure.

Here we give some examples of smooth manifolds and revisit some examples of topological manifolds.

Example. (0-dimensional manifolds). As shown a topological manifold M of dimension 0 is a countable discrete space. For each point $p \in M$, the only neighborhood of p that is homeomorphic to an open subset of $\mathbb{R}^0$ is $\{p\}$ itself, and there is exactly one coordinate map $\varphi : \{p\} \in \mathbb{R}^0$. Hence the set of all charts on M trivially satisfies the smooth compatibility condition, and each 0-dimensional manifold has a unique smooth structure. So it is a smooth manifold. □

The following theorem is useful for checking if a manifold is smooth.

Theorem 3.1. *Let M be a Hausdorff space with a countable basis of open sets. If (V_β, ψ_β) is a covering of M by C^∞-compactible coordinate neighborhoods, then there is a unique C^∞ structure on M counting these coordinate neighborhoods.*

This theorem shows that (i) and (ii) in the definition of smooth manifolds are the key properties defining a C^∞-structure. Hence we only have to check the compactibility of a covering by neighborhoods.

Example. (The Euclidean plane). As we commented earlier the Euclidean plane $\mathbb{E}^2$ becomes a metric space once we have chosen a unit of length. It is Hausdorff and has a countable basis of open sets. The homeomorphsim $\psi : \mathbb{E}^2 \to \mathbb{R}^2$ can be determined when a choice of an origin and mutually perpendicular coordinate axes is made. Hence we can cover $\mathbb{E}^2$ with a single chart (V, ψ) with $V = \mathbb{E}^2$ and $\psi(V) = \mathbb{R}^2$. This shows that $\mathbb{E}^2$ is a topological manifold and moreover (V, ψ) defines a smooth structure on $\mathbb{E}^2$ by Theorem 3.1. Hence the Euclidean plane is a smooth manifold. □

In particular, the space $\mathbb{R}^2$ as a Euclidean space is determined by the atlas consisting of the single chart $(\mathbb{R}^2, \mathrm{Id}_{\mathbb{R}^2})$. This is called standard smooth structure on $\mathbb{R}^2$ and the resulting coordinate map is called standard coordinates.

Note that there are many other charts on $\mathbb{E}^2$ which are C^∞ compatible with the standard chart. (see e.g. [3]). Similarly we can show that the n-dimensional Euclidean space is a smooth manifold.

Example. (Finite-dimensional vector spaces). Let V be a finite-dimensional real vector space. Any norm on V determines a topology, which is independent of the choice of norm. With this topology V is a topological n-manifold and has a natural smooth structure defined as follows: For any ordered basis $(E_1, ..., E_n)$ of V we define a basis isomorphism $E : \mathbb{R}^n \to V$ by

$$E(x) = \sum_{i=1}^{n} x^i E_i$$

This map is a homeomorphism, so (V, E^{-1}) is a chart. By change of basis we know that if $(\tilde{E}_1, ..., \tilde{E}_n)$ is any other basis and $\tilde{E}(x) = \sum_j x^j \tilde{E}_j$ is the corresponding isomorphism, then there is some invertible matrix (A_i^j) such that $E_i = \sum_j A_i^j \tilde{E}_j$ for each i. Then the transition map between the two charts is given by $\tilde{E}^{-1} \circ E(x) = \tilde{x}$ where $\tilde{x} = (\tilde{x}^1, ..., \tilde{x}^n)$ is determined by

$$\sum_{j=1}^{n} \tilde{x}^j \tilde{E}_j = \sum_{i=1}^{n} x^i E_i = \sum_{i,j=1}^{n} x^i A_i^j \tilde{E}_j.$$

Thus $\tilde{x}^j = \sum_i A_i^j x^i$. Hence, the map sending x to $\tilde{x}$ is an invertible linear map and hence a diffeomorphism. Therefore any two such charts are smoothly compatible. The collection of all such charts defines a smooth structure called the standard smooth structure on V. $\square$

We will use the Einstein summation convention: $E(x) = x^i E_i$ as an abbreviation for $E(x) = \sum_{i=1}^n x^i E_i$. So, $\sum_{i,j=1}^n x^i A_i^j \tilde{E}_j$ will be shortened to $x^i A_i^j \tilde{E}_j$.

Example. (Graph of smooth functions.). Let $U \subseteq \mathbb{R}^n$ be an open subset and $f : U \to \mathbb{R}^k$ be a smooth function. The graph of f is the subset of $\mathbb{R}^n \times \mathbb{R}^k$ defined by

$$\Gamma(f) := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^k : x \in U, y = f(x)\},$$

with the subspace topology. Let $\pi : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ be the projection onto the first factor, and let $\varphi : \Gamma(f) \to U$ be the restriction of π to $\Gamma(f)$:

$$\varphi(x, y) = x, \ (x, y) \in \Gamma(f).$$

Now φ is the restriction of a continuous map, and so it is continuous. Since it has a continuous inverse given by $\varphi^{-1}(x) = (x, f(x))$ it is a homeomorphism. Hence the graph is a topological manifold. Since $\Gamma(f)$ is covered by the single graph coordinate chart φ, we can give a canonical smooth structure on $\Gamma(f)$ by declaring the graph coordinate chart $(\Gamma(f), \varphi)$ to be a smooth chart. $\square$

Example. (Sphere S^2) We have shown that the n-sphere $S^2 \subset \mathbb{R}^3$ is a topological n-manifold. Now we give a smooth structure on S^2. Let

$$U_i^+ = \{(x^1, x^2 x^3) \in \mathbb{R}^3 : x^i > 0\}, U_i^- = \{(x^1, x^2 x^3) \in \mathbb{R}^3 : x^i < 0\}, i = 1, 2, 3.$$

Let D^2 be a unit disk in $\mathbb{R}^2$. Assume that $f : D^2 \to \mathbb{R}$ be the continuous function

$$f(u) = \sqrt{1 - ||u||^2}.$$

Then for $i = 1, 2, 3$ it is easy to check that $U_i^+ \cap S^2$ is respectively the graphs

$$x^1 = f(x^2, x^3), x^2 = f(x^1, x^3), x^3 = f(x^1, x^2),$$

Similarly, $U_i^- \cap S^2$ is the graph of the functions

$$x^1 = -f(x^2, x^3), x^2 = -f(x^1, x^3), x^3 = -f(x^1, x^2),$$

Thus, each subset $U_i^{\pm} \cap S^2$ is locally Euclidean of dimension 2, and the maps $\varphi_i^{\pm} : U_i^{\pm} \cap S^2 \to D^2$ given by

$$\varphi_1^{\pm}(x^1, x^2, x^3) = (x^2, x^3), \varphi_2^{\pm}(x^1, x^2, x^3) = (x^1, x^3), \varphi_3^{\pm}(x^1, x^2, x^3) = (x^1, x^2),$$

are graph coordinates for S^2. Since each point of S^2 is in the domain of at least one of these 6 charts, S^2 is a topological manifold as we already proved. Now we prove that the collection of graph coordinate charts $\{(U_i^{\pm}, \varphi_i^{\pm})\}$ is a smooth atlas. To this end we compute the transition map $\varphi_i^{\pm} \circ (\varphi_j^{\pm})^{-1}$. For $j = i$ we have we have

$$\varphi_i^+ \circ (\varphi_i)^{-1} = \varphi_i^- \circ (\varphi_i^+)^{-1} = \mathrm{Id}_{D^2}.$$

For distinct i and j, for example $\varphi_1^+ \circ (\varphi_2^-)^{-1}$ is given on $U_1^+ \cap U_2^-$ by compositing $(\varphi_2^-)^{-1}$ and φ_1^+ as follows:

$$\begin{aligned}(\varphi_2^-)^{-1} &: (x^1, x^3) \to (x^1, -\sqrt{1 - (x^1)^2 - (x^3)^2}, x^3) \\ \varphi_1^+ &: (x^1, -\sqrt{1 - (x^1)^2 - (x^3)^2}, x^3) \to (-\sqrt{1 - (x^1)^2 - (x^3)^2}, x^3)\end{aligned}$$

Now using (u^1, u^2) as U_2^--coordinates and (v^1, v^2) as U_1^+-coordinates instead of (x^1, x^3) and (x^2, x^3) yields

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

Clearly the v^1, v^2 are C^∞-functions of u^1, u^2 because the square root term is never zero on the open unit disk $\{(u^1, u^2) : (u^1)^2 + (u^2)^2 < 1\}$. Similarly, $\varphi_2^- \circ (\varphi_1^+)^{-1}$ is C^∞ on the open disk $\{(v^1, v^2) : (v^1)^2 + (v^2)^2 < 1\}$. Hence the chart (U_1^+, φ_1^+) and the chart (U_2^-, φ_2^-) are C^∞-compatible. We can do exactly the same computation for other charts. Thus this covering of S^2 by six charts determines a C^∞ structure. So the 2-sphere is a smooth manifold. □

Note that the similar C^∞-structure can be put on any n-sphere in $\mathbb{R}^{n+1}$ so that we can conclude that n-spheres are smooth manifolds.

An easier proof is to use the stereographic projections to show the local Euclidean property. We can cover S^2 by two open subsets

$$U_+ = S^2 \setminus \{(0, 0, -1)\}, U_- = S^2 \setminus \{(0, 0, 1)\}$$

and define two charts (φ_+, U_+) and (φ_-, U_-) by the stereographic projections

$$\varphi_\pm(x^1, x^2, x^3) = \frac{1}{1 \pm x^3}(x^1, x^2).$$

Then $\varphi_\pm$ are continuous, invertible and the inverse is

$$\varphi_\pm^{-1}(y^1, y^2) = \frac{1}{1 + (y^1)^2 + (y^2)^2}(2y^1, 2y^2, \pm(1 - (y^1)^2 - (y^2)^2)),$$

which is also continuous. Now we prove that the two charts are compatible, that is to show that $\varphi_+ \circ \varphi_-^{-1}$ is a diffeomorphism of $\mathbb{R}^2 \setminus \{0\}$, since $\varphi_-(U_+ \cap U_-) = \mathbb{R}^2 \setminus \{0\}$. This follows by

$$\begin{aligned}
&\varphi_+ \circ \varphi_-^{-1}(y^1, y^2) \\
=&\varphi_+ \left(\frac{1}{1 + (y^1)^2 + (y^2)^2}(2y^1, 2y^2, -1 + (y^1)^2 + (y^2)^2) \right) \\
=&\frac{1}{(y^1)^2 + (y^2)^2}(y^1, y^2)
\end{aligned}$$

which is a diffeomorphism of $\mathbb{R}^2 \setminus \{0\}$.

Although life can exist outside $\mathbb{R}^n$ the nice thing about abstract manifolds is that they can be considered as a subset of sufficiently large dimensional flat space. This is the famous imbedding theorem due to Whitney.

Theorem 3.2 (Whitney's Imbedding Theorem, [11])**.** *Any smooth n-manifold may be embedded differentiably into $\mathbb{R}^{2n+1}$.*

3.2. **Why abstract manifolds?**

Since we can embed a smooth manifold in $\mathbb{R}^N$ (with sufficiently large N) by Whitney's Theorem we can ask why we need abstract manifolds. To answer the question we consider some simple examples studying sets of geometric objects.

Example. (*The real projective plane*) The set of all straight lines through origin of $\mathbb{R}^3$, is denoted by $\mathbb{RP}^2$ and called the *real projective plane*.

An intuitive approach will be thinking of the sphere $S^2 \subset \mathbb{R}^3$ centered at the origin and associate to a line the points where it meets the sphere. The problem we then will

immediately meet is that there are two such points, so we need to keep only half of the sphere. So we restrict ourselves to the northern hemisphere. Then there are still two intersection points of horizontal lines with the hemisphere, on the equator. Now if we cut off half of that equator we would have a mess. This piece of a sphere is not a nice surface now, at the equatorial points where the missing half of the equator meets the half still in place. Moreover the construction is not equivariant, we have given some hemisphere higher priority. The original set of lines is acted on by the group of linear maps in an elementary way, but the chopped up sphere is not. Thus we shall find a way out of this mess.

A natural question arises here. Why do we bother with such a set of geometric objects if we do not dream of working on lines through the origin? The motivating example is making a color. It involves mixing the three basic colors in correct proportions. This is represented by a line through the origin in $\mathbb{R}^3$. Color mixing is of vital importance in many applications, e.g. for car makers, printers, graphic artists, in particular in image processing and computer vision with which this report is related. A naive consideration would be that the coefficients must be positive so we may look at only the positive octant of S^2. But it turns out that we really need to work in $\mathbb{RP}^2$, even if only in a part of it.

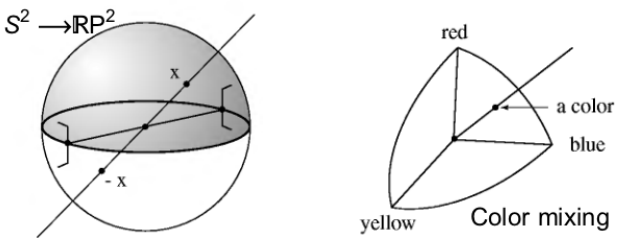


FIGURE 4. Left: Real projective plane; Right: Color mixing model

To make the point more precise, we exemplify by considering the CIE XYZ color space. This color space is also termed as CIE 1931color space, created by the International Commission on Illumination in 1931. See [29].

In color matching experiments negative values or weight factors R, G, B are allowed. Some matchable colors cannot be generated by the Standard Primaries [1]R, G, B. Other light sources are necessary, especially spectral pure sources (mono-chromats). To avoid negative RGB numbers, the CIE had introduced a new coordinate system XYZ. The RGB system is essentially defined by three non-orthogonal base vectors in XYZ. They are related by a linear transformation. Another view is possible by introducing imaginary primaries or synthetical primaries X, Y, Z which are purely mathematical to replace the actual red, green and blue (RGB) primaries for simplifying color calculations. All real colors can be matched using positive proportion of three imaginary primaries. The values of X,Y and Z specify the color stimulus. They are known as the CIE 1931 tristimulus values.

One special feature of this color system is that the luminance is defined by Y only. Roughly speaking the Y tristimulus value represents the lightness of a sample. In the CIE XYZ system, the curve for the Y tristimulus value is equal to the curve of the human eye's response to the total power of a light source. To describe visual attributes of colors in

[1]Primary colors are sets of colors that can be combined to make a useful range of colors. For human applications, three primary colors are usually used, since human color vision is trichromatic.

terms of hue and chroma, the CIE XYZ tristimulus values are used to formulate a new set of chromaticity coordinates that are denoted by xyz. The chromaticity coordinates xyz are obtained by taking the ratio of the tristimulus values to their sum $X+Y+Z$ as given by the equations:

$$\begin{aligned} x &= \frac{X}{X+Y+Z} \\ y &= \frac{Y}{X+Y+Z} \\ z &= \frac{Z}{X+Y+Z} \\ 1 &= x+y+z. \end{aligned}$$

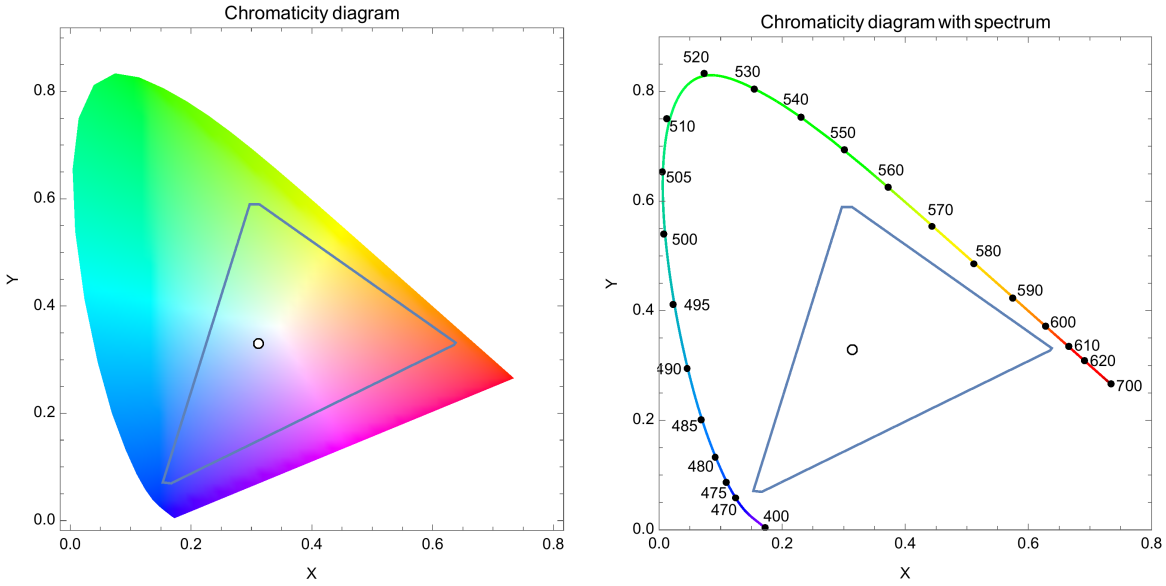


FIGURE 5. Chromaticity diagram

Mathematically, x and y are formulated by the projective transformation of the tristimulus values into two-dimensional plane. The resulting color space specified by x, y and Y is known as the CIE xyY color space. The third dimension is indicated by the tristimulus Y. The scale for Y extends from the white spot in a line perpendicular to the plane formed by x and y using a scale between 0 and 100. A plot of y against x is called a chromaticity diagram Figure 5. The chromaticity diagram is the spectrum locus with horseshoe shape. The colors of the chromaticity diagram occupy a region of the real projective plane.

The chromaticity diagram can be used to visualize distribution of an image's pixels as well as a color space. This is an important step in image processing. Figure 6 shows chromaticity diagram of night views of Paris and Shanghai, respectively. We shall come back to color spaces in §6.

Now we prove that $\mathbb{RP}^2$ is a smooth manifold. To this end we need a little theorems on quotient space/topology. As usual denote by $\sim$ an equivalence relation on a topological space X, $[x] = \{y \in X : y \sim x\}$ the equivalence class of x, $X/_\sim$ the set of equivalent classes. Let $\pi : X \to X/_\sim$ be the natural mapping (projection) taking each $x \in X$ to its

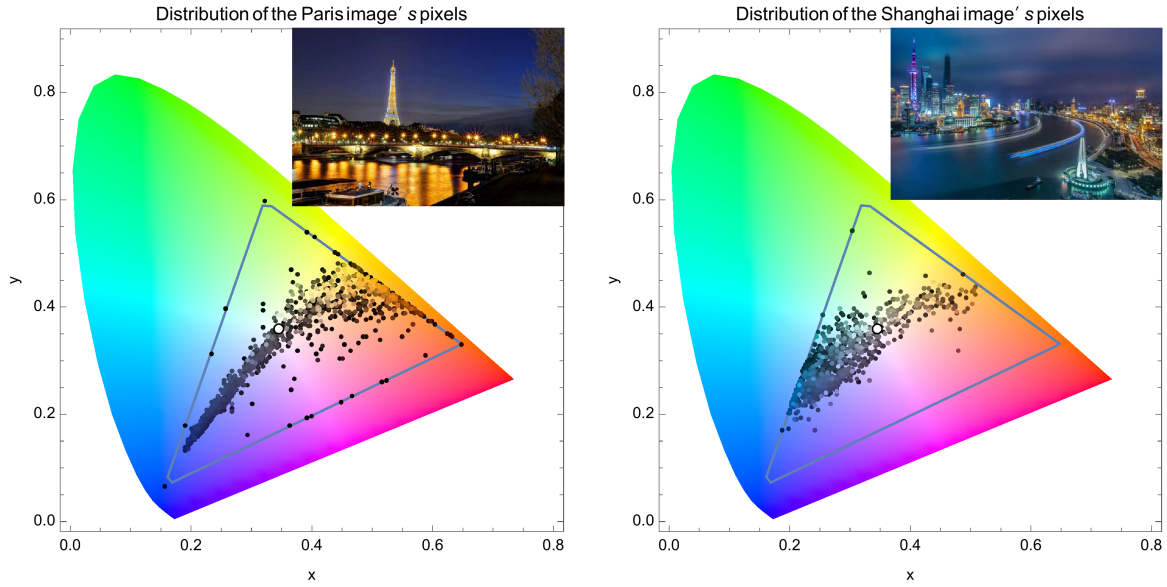


FIGURE 6. Illustration of usage of chromaticity diagram in image processing

equivalent class $[x]$, i.e. $\pi(x) = [x]$. With these notations we define the standard quotient topology on $X/_\sim$ as follows: $U \subset X/_\sim$ is an open subset if $\pi^{-1}(U)$ is open. Then the projection π is continuous.

Now let $\pi : x \mapsto [x]$ denote the natural map of $\mathbb{R}^3 \setminus \{0\}$ onto $\mathbb{RP}^2$ and let S^2 be the unit sphere. The restriction of π to S^2 is one-to-one, for each $p \in \mathbb{RP}^2$ there are precisely two elements $\pm x \in S^2$ with $\pi(x) = p$. Thus we have a model for $\mathbb{RP}^2$ as the set of all pairs of antipodal points in S^2. Further, we equip $\mathbb{RP}^2$ as a Hausdorff topological space as follows. A set $M \subset \mathbb{RP}^2$ is said to be open if and only if its pre-image $\pi^{-1}(A)$ is open in $\mathbb{R}^3$, or equivalently, if $\pi^{-1}(A) \cap S^2$ is open n S^2. We say that $\mathbb{RP}^2$ has the *quotient topology* relative to $\mathbb{R}^3 \setminus \{0\}$. It can be proved that $\mathbb{RP}^2$ is Hausdorff and has countable basis of open sets (see e.g. [26]).

Let $U_i = \{[x] : x_i \neq 0\} \subset \mathbb{RP}^2$, $i = 1, 2, 3$. Clearly it is, for each i, open since $\pi^{-1}(U_i) = \{x : x_i \neq 0\}$ is open in $\mathbb{R}^2$. Let $\varphi_i : \mathbb{R}^2 \to \mathbb{RP}^2$ be the map defined by

$$\varphi_1(u) = [(1, u_2, u_3)], \varphi_2(u) = [(u_1, 1, u_2)], \varphi_3(u) = [(u_1, u_2, 1)], \text{ for } u \in \mathbb{R}^2.$$

They are continuous since they are composed by π and a continuous map $\mathbb{R}^2 \to \mathbb{R}^3$. Furthermore, φ_i's are bijection of $\mathbb{R}^2$ onto U_i, and $\mathbb{RP}^2 = U_1 \cup U_2 \cup U_3$. It remains to show that $\{(\varphi_i, U_i)\}$ defines a smooth structure on $\mathbb{RP}^2$. We have to check the following.:
(1) φ_i is continuous $U_i \to \mathbb{R}^2$, e.g.

$$\sigma_1^{-1}(p) = \left(\frac{x_2}{x_1}, \frac{x_3}{x_1}\right)$$

where $p = \pi([x])$. Since the components in the right hand side are continuous functions on $\mathbb{R}^3 \setminus \{x_1 = 0\}$, $\varphi_1^{-1} \circ \pi$ is continuous.
(2) The overlap between φ_i and φ_j satisfies e.g.

$$\varphi_1^{-1} \circ \varphi_2(u) = \left(\frac{1}{u_1}, \frac{u_2}{u_1}\right),$$

which is smooth map from $\mathbb{R}^2 \setminus \{u : u_i = 0\} \to \mathbb{R}^2$. □

We have seen that there is a homeomorphism

$$\mathbb{RP}^2 \simeq S^2/\{\text{antipodal points}\} = S^2/_\sim.$$

There are other homeomorphisms. Consider the closed upper hemisphere $\mathbb{R}^3_+ = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1, z \geq 0\}$, as defined earlier and the closed unit disk $D^2 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\} \subset \mathbb{R}^2$. These two spaces are homomorphic to each other as shown before via the continuous map

$$f : \mathbb{R}^3_+ \to D^2, f(x, y, z) = (x, y)$$

and its inverse

$$g : D^2 :\to \mathbb{R}^3_+, g(x, y) = (x, y, \sqrt{1 - x^2 - y^2})$$

On $\mathbb{R}^3_+$ define an equivalence relation $\sim$ by identifying the antipodal points on the equator:

$$(x, y, 0) \sim (-x, -y, 0),\ x^2 + y^2 = 1.$$

On D^2 define an equivalence relation $\sim$ by identifying the antipodal points on the boundary circle:

$$(x, y) \sim (-x, -y), x^2 + y^2 = 1.$$

Then f and g induce homeomorphisms

$$\tilde{f} : \mathbb{R}^3_+/_\sim \ \to D^2/_\sim,\ \tilde{g} : D^2/_\sim \ \to \mathbb{R}^3_+/_\sim.$$

Hence we have a sequence of homeomorphisms:

$$\mathbb{RP}^2 \tilde{\to} S^2/_\sim \tilde{\to} \mathbb{R}^3_+/_\sim \tilde{\to} D^2/_\sim$$

that identify the real projective plane as the quotient of the closed disk with the antipodal points on its boundary identified. In general we can show that projective spaces $\mathbb{RP}^n$ are smooth manifold.

Example. (*The set of positions of a rigid body in* $\mathbb{E}^3$). A rigid body has six parameters: three for the location of the center of gravity and three to say how it has been rotated around that center. We can try to avoid working in a six dimensional space, because the center of gravity lives in a three dimensional Euclidean space, but what is the set of rotations, as a three dimensional object? How can we study geometry on it? We would like a general framework, in which the motions of the rigid body will be geometrically meaningful curves. When studying mechanics we struggle for the complicated formulas for Euler angles. Moreover there are positions for which those angles are not well defined using latitude and longitude to describe the sphere. We will refer to the set of rotations in $\mathbb{R}^3$ around a fixed point as $SO(3)$, the *special orthogonal group*. To define Euler angles, an axis is chosen, but $SO(3)$ should look the same near any of its points. Hamilton made this homogeneity of $SO(3)$ manifest by applying the quaternions he discovered. Recall that the *quaternions* are $\mathbb{H} = \mathbb{R} \oplus \mathbb{R}^3$ with multiplication

$$(x_0, x) \cdot (y_0, y) = (x_0 y_0 - \langle x, y \rangle, x_0 y + y_0 x + x \times y).$$

where $\times$ is the cross product and $\langle\ ,\ \rangle$ is the inner product on $\mathbb{R}^3$. If $X = (x_0, x)$ then denote $\overline{X} := (x_0, -x)$. Identify $\mathbb{R}^3 = 0 \oplus \mathbb{R}^3$ with imaginary quaternions that is $x_0 = 0$. Unit length quaternions $Y = (y_0, y)$ act on imaginary quaternions $X = (0, x)$ by

$$X \mapsto YX\overline{Y}.$$

This brings the unit sphere $S^3 \subset \mathbb{H}$ to rotate $\mathbb{R}^3$. Just as for $\mathbb{RP}^2$, although the set of unit length quaternions form a three dimensional sphere S^3, there are two unit length

quaternions $\pm Y$ giving the same rotation. So $S^3/\sim$, where $\sim$ is identifying antipodal points, is $SO(3)$, as the above construction gives all rotations.

Now we take another approach to show that $SO(3)$ and $\mathbb{RP}^3$ are the same smooth manifolds. Since the underlying manifold does not admit a global coordinate system, we have no neat (easy) parametrized matrices of $SO(3)$ unlike those of $SO(2)$ consisting of matrices of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

which is homeomorphic to the circle.

Let's show that $SO(3)$ is homeomorphic to the 3-dimensional real projective space $\mathbb{RP}^3$. Remember that the real projective space $\mathbb{RP}^3$ is the quotient space of $\mathbb{R}^4 \setminus \{0\}$ by the equivalence relation

$$x \sim y \Leftrightarrow y = tx \text{ for some nonzero real number } t, \text{ and } x, y \in \mathbb{R}^4 \setminus \{0\}$$

Denote the equivalence class of a point $(a^0, a^1, a^2, a^3) \in \mathbb{R}^4 \setminus \{0\}$ by $[a^0, a^1, a^2, a^3]$, called *homogeneous coordinates* on $\mathbb{RP}^3$. A possible homeomorphism F is given by

$$[a^0, a^1, a^2, a^3] \mapsto$$
$$\frac{1}{\Delta}\begin{pmatrix} (a^0)^2+(a^1)^2-(a^2)^2-(a^3)^2 & 2(a^1a^2-a^0a^3) & 2(a^1a^3+a^0a^2) \\ 2(a^1a^2+a^0a^3) & (a^0)^2-(a^1)^2+(a^2)^2-(a^3)^2 & 2(a^2a^3-a^0a^1) \\ 2(a^1a^3-a^0a^2) & 2(a^2a^3+a^0a^1) & (a^0)^2-(a^1)^2-(a^2)^2+(a^3)^2 \end{pmatrix}$$

which is an orthogonal matrix with determinant 1 by a straightforward but tedious calculation. To show that this is a homeomorphism we need to give the inverse mapping. Since there is no global coordinate systems it is not immediate how to find an inverse. Assume that $SO(3)$ matrix is given by

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix}$$

Consider the following mapping G_1 from $SO(3)$ to $\mathbb{RP}^3$:

$$R \mapsto [1 + r_{11} + r_{22} + r_{33}, r_{32} - r_{23}, r_{13} - r_{31}, r_{21} - r_{12}]$$

It can be easily checked that

$$G_1 \circ F([a^0, a^1, a^2, a^3]) = \frac{4a^0}{\Delta}[a^0, a^1, a^2, a^3]$$

if $1 + r_{11} + r_{22} + r_{33} \neq 0$, equivalently $a^0 \neq 0$. So G_1 is an inverse to F (since the homogeneous coordinates of the projective space are only defined up to an overall nonzero factor). It is apparent now that the map is not defined on all of $SO(3)$ and it is not onto, because the plane $a^0 = 0$ is not in the image.

Similarly, we define G_2 by

$$R \mapsto [r_{32} - r_{23}, 1 + r_{11} - r_{22} - r_{33}, r_{12} + r_{21}, r_{13} + r_{31}]$$

if $1 + r_{11} - r_{22} - r_{33} \neq 0$, i.e. $a^1 \neq 0$, and G_3

$$R \mapsto [r_{13} - r_{31}, r_{12} + r_{21}, 1 - r_{11} + r_{22} - r_{33}, r_{23} + r_{32}]$$

if $1 - r_{11} + r_{22} - r_{33} \neq 0$, i.e. $a^2 \neq 0$, and finally, G_4

$$R \mapsto [r_{21} - r_{12}, r_{13} + r_{31}, r_{23} + r_{32}, 1 - r_{11} - r_{22} + r_{33}]$$

if $1 - r_{11} - r_{22} + r_{33} \neq 0$, i.e. $a^3 \neq 0$. It can be verified that

$$G_2 \circ F([a^0, a^1, a^2, a^3]) = \frac{4a^1}{\Delta}[a^0, a^1, a^2, a^3],$$
$$G_3 \circ F([a^0, a^1, a^2, a^3]) = \frac{4a^2}{\Delta}[a^0, a^1, a^2, a^3],$$
$$G_4 \circ F([a^0, a^1, a^2, a^3]) = \frac{4a^3}{\Delta}[a^0, a^1, a^2, a^3].$$

These four maps are the inverse of F on the respective subsets. These four maps agree on the regions where they overlap and together cover all of $\mathbb{RP}^3$. Moreover they invert the original map from $\mathbb{RP}^3$ to $SO(3)$. Therefore, the two manifolds are homeomorphic.

The rigid body has a very important application in robotics. The set describing the limb postures and locations of a robot is typically described by an abstract manifold. In order to avoid the robot's movement abrupt the space of its states has to be a smooth manifold. In a similar manner, in statistical mechanics, we have to work with the set made up by the positions of a large collection of particles. Because of collisions, this set is worse (not much worse) than a manifold, which has a corner.

Example. (*Double pendulum*) The space of configuration of a mechanical system form a manifold. The double pendulum is a very simple example. The configuration space is a two dimensional torus T^2, a surface like a doughnut . However, we have to really think of it as an abstract manifold, not as embedded in $\mathbb{R}^3$.

4. **Smooth maps, connections**

In this section collect some basic concepts such as smooth maps, tangent space, tangent bundles, covariant derivatives, connections, curvature and torsion on manifolds.

4.1. **Smooth maps on a manifold**

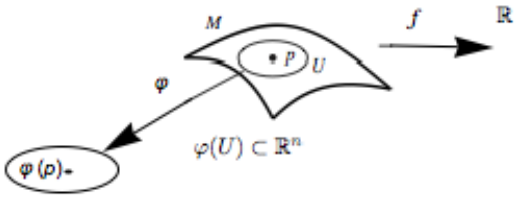
Using coordinate charts, one can transfer the notion of smooth maps from $\mathbb{R}^n$ to manifolds. By the C^∞ compatibility of charts in an atlas, the smoothness of a map turns out to be independent of the choice of charts and is therefore well defined. We give various criteria for the smoothness of a map as well as examples of smooth maps.

Next we transfer the notion of partial derivatives from $\mathbb{R}^n$ to a coordinate chart on a manifold. Partial derivatives relative to coordinate charts allow us to generalize the inverse function theorem to manifolds. Using the inverse function theorem, we formulate a criterion for a set of smooth functions to serve as local coordinates near a point.

4.2. **Smooth functions on a manifold**

Let M be a smooth n-manifold. A function $f : M \to \mathbb{R}$ is said to be C^∞ or smooth at a point $p \in M$ if there is a chart (U, φ) about p such that the function defined on the open subset $\varphi(U) \subset \mathbb{R}^n$, $f \circ \varphi^{-1}$, is C^∞ at $\varphi(p)$. The function is said to be C^∞ on M if it is C^∞ at every point of M. This is illustrated in Figure 7.

Among the C^∞ functions on M are the coordinate functions $(x^1(q), x^2(q), ..., x^n(q))$ of a coordinate neighborhood (U, φ). Note that the definition of smoothness of a function at a point is independent of the chart (U, φ), for if $f \circ \varphi^{-1}$ is C^∞ at $\varphi(p)$ and (V, ψ) is any



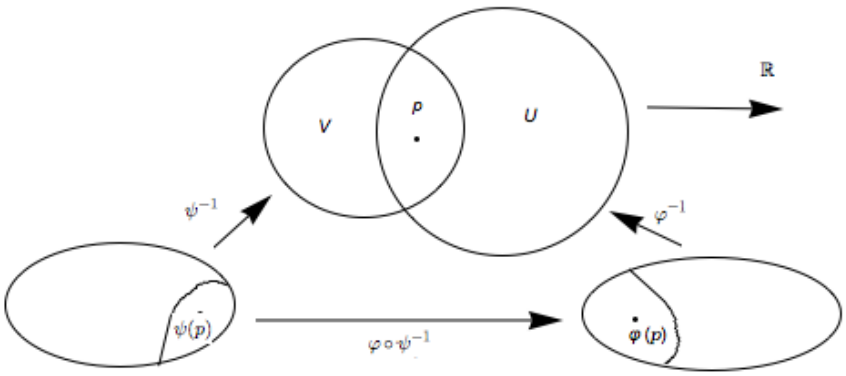
Checking a function f being a C^∞ at p by pulling back to $\mathbb{R}^n$

FIGURE 7.

other chart about $p \in M$ then on $\psi(U \cap V)$,

$$f \circ \psi^{-1} = (f \circ \varphi^{-1}) \circ (\varphi \circ \psi^{-1}),$$

which is C^∞ at $\psi(p)$ as shown in Figure 8.



Checking a function f being a C^∞ at p through two charts

FIGURE 8.

Note also that the function f is not assumed to be continuous. But if f is C^∞ at $p \in M$, then $f \circ \varphi^{-1} : \varphi(U) \to \mathbb{R}$ is continuous at $\varphi(p)$ since it is a C^∞ function at the point $\varphi(p)$ in an open subset of $\mathbb{R}^n$. Now $f = (f \circ \varphi^{-1}) \circ \varphi$ is a composite of continuous functions so it is continuous.

Now it is not hard to prove that the following statements are equivalent:

- $f : M \to \mathbb{R}$ is C^∞.
- The manifold M has an atlas such that for every chart (U, φ) in the atlas, $f \circ \varphi^{-1} : \varphi(U)(\subset \mathbb{R}^n) \to \mathbb{R}$ is C^∞.
- For every chart (V, ψ) on M, the function $f \circ \psi^{-1} : \psi(V)(\subset \mathbb{R}^n) \to \mathbb{R}$ is C^∞.

- its pullback $(\varphi^{-1})^* f$ by φ^{-1} (defined as the composite function $f \circ \varphi^{-1}$) is C^∞ on the subset $\varphi(U)$ of $\mathbb{R}^n$,

4.3. **Smooth maps between manifolds**

Let M and N be m- and n-manifolds respectively. A continuous map $F : M \to N$ is C^∞ at a point $p \in N$ if there exist charts (V, ψ) about $F(p) \in M$ and (U, φ) about $p \in N$ such that the composition $\psi \circ F \circ \varphi^{-1}$ from the open subset $\varphi(F^{-1}(V) \cap U) \subset \mathbb{R}^n$ to $\mathbb{R}^m$, is C^∞.

Let $F : N \to M$ be continuous. Then the following things are equivalent.

- F is C^∞.
- there are atlases $\mathscr{U}_N$ for N and $\mathscr{U}_M$ for M such that for every chart (U, φ) in $\mathscr{U}_N$ and (V, ψ) in $\mathscr{U}_M$, the map
$$\psi \circ F \circ \varphi^{-1} : \varphi(U \cap F^{-1}(v)) \to \mathbb{R}^m$$
is C^∞.
- For every chart (U, φ) on N and (V, ψ) on M, the map
$$\psi \circ F \circ \varphi^{-1} : \varphi(U \cap F^{-1}(v)) \to \mathbb{R}^m$$
is C^∞.

We can also show that the composition of C^∞ maps is C^∞.

4.4. **Diffeomorphisms**

A *diffeomorphism* of manifolds is a bijective C^∞ map $F : N \to M$ whose inverse F^{-1} is also C^∞.

Clearly, if (U, φ) is a chart on an n manifold M , then the coordinate map $\varphi : U \to \varphi(U) \subset \mathbb{R}^n$ is a diffeomorphism. Assume U to be an open subset of n-manifold M. If $F : U \to F(U) \subset \mathbb{R}^n$ is diffeomorphism onto an open subset of $\mathbb{R}^n$, then (U, F) is a chart in the smooth structure of M. Consequently coordinate maps are diffeomorphisms, and conversely, every diffeomorphism of an open subset of a manifold with an open subset of a Euclidean space can serve as a coordinate map.

Partial derivatives. To do calculus we need to define partial derivative on manifolds. For an n-manifold M let (U, φ) be a chart and $f : U \to \mathbb{R}$ a C^∞ function. Assume $r^1, ..., r^n$ are the standard coordinates on $\mathbb{R}^n$, then $x^i = r^i \circ \varphi$. For $p \in U$, we define the partial derivative $\frac{\partial f}{\partial x^i}$ of f with respect to x^i ar p as

$$\left.\frac{\partial}{\partial x^i}\right|_p f := \frac{\partial f}{\partial x^i}(p) := \frac{\partial (f \circ \varphi^{-1})}{\partial x^i}(\varphi(p)) := \left.\frac{\partial}{\partial r^i}\right|_{\varphi(p)} (f \circ \varphi^{-1}).$$

4.5. **Tangent space and tangent bundles**

A *tangent vector* at $p \in M$ is a linear map $v : C^\infty(M) \to \mathbb{R}$ such that the Leibniz rule holds, that is

$$v(fg) = f(p)v(g) + v(f)g(p)$$

for all $f, g \in C^\infty(M)$. Denote by $T_p(M)$ the vector space of all tangent vectors at p. Since this concept is very important we give some examples.

- Let γ be a (smooth) path from an interval I to M with $\gamma(t) = p$. Define $\gamma'(t) \in M$ by
$$\gamma'(t)f = (f \circ \gamma)'(t)$$
In fact we can prove that all $v \in T_p(M)$ are of the form $\gamma'(t)$ for some path γ.
- Let (U, x) be a chart with coordinates $x^1, ..., x^n$ and $x(p) = q \in \mathbb{R}^n$. If we define $\partial_{i|p} f = \frac{\partial(f \circ x^{-1})}{\partial x^i}\big|_p$, then $\partial_{1|p}, ..., \partial_{n|p}$ is a basis for T_pM.
Let $f \in C^\infty(M)$, the *derivative* $\mathrm{d}f_p : T_p(M) \to \mathbb{R}$ of f at $p \in M$ is defined by
$$\mathrm{d}f_p(v) = vf$$
Clearly, this definition is the same as the usual one when M is an open subset of $\mathbb{R}^n$. It's also clear that $\mathrm{d}f_p$ is a linear map and the Leibniz Rule holds:
$$\mathrm{d}(fg)_p = g(p)\mathrm{d}f_p + f(p)\mathrm{d}g_p.$$

Let M and N be two manifolds and $F : M \to N$ be a smooth map. The *tangent map* $\mathrm{d}F_p : T_p(M) \to T_{F(p)(N)}$ at $p \in M$ is the linear map defined by
$$\mathrm{d}F_p(v)f = v(f \circ F),$$
for $v \in T_p(M)$ and $f \in C^\infty(N)$. Then we have the chain rule: for $F : M \to N$ and $G : N \to Z$ and $p \in M$,
$$\mathrm{d}(G \circ F)_p = \mathrm{d}G_{F(p)} \circ \mathrm{d}F_p$$

The *tangent bundle* of M is the disjoint union of the tangent spaces: $TM = \bigcup_{p \in M} T_p(M)$. There is a natural projection map $\pi : TM \to M$ which for each $p \in M$, sends every vector $X \in T_p(M)$ to p. The pre-images $\pi^{-1}(p) = T_p(M)$ are called *fibers* of the bundle. A *section* of TM is a C^∞ mapping $F : M \to TM$ such that $\pi \circ F = \mathrm{Id}_M$.

4.6. **Vector fields**

A *vector field* is a linear map $X : C^\infty(M) \to C^\infty(M)$ such that
$$X(fg) = f(Xg) + g(Xf).$$
In other words we can view a vector field as a map $X : M \to TM$ with $X(p) \in T_p(M)$ with smoothness: for each $f \in C^\infty(M)$ the function $p \mapsto X(p)f$ is C^∞.

To make the definition more intuitive let us consider the $S^3 \subset \mathbb{R}^4$. There are three mutually perpendicular unit vector fields on S^3 given by
$$X = -x^2\partial_1 + x^1\partial_2 + x^4\partial_3 - x^3\partial_4$$
$$Y = -x^3\partial_1 - x^4\partial_2 + x^1\partial_3 + x^2\partial_4$$
$$X = -x^4\partial_1 + x^3\partial_2 - x^2\partial_3 + x^1\partial_4$$
at the point $x = (x^1, x^2, x^3, x^4)$ of S^2. Since at each point these are mutually orthogonal unit vectors in $\mathbb{R}^4$, they are independent. Moreover they are also tangent to S^3. To see this it is sufficient to convince ourselves that they are orthogonal to the radius vector from the origin 0 to the point x on S^3. What remains is to check if they are C^∞ whose proof is referred to [3], to avoid further technical details.

Now let X, Y be vector fields on M. Define the *Lie bracket* of X, Y $[X, Y] : C^\infty(M) \to C^\infty(M)$ given by
$$[X, Y]f = X(Yf) - Y(Xf).$$
Then, $[X, Y]$ is also a vector field.

4.7. **Connection**

Let $\Gamma(TM)$ be the vector space of all vector fields on M. A *connection* or *covariant derivative* on M ([5]) is a map $\nabla : TM \times \Gamma(TM) \to TM$, written as $\nabla_\xi Y$, with the following properties:

(1) $\nabla_\xi Y$ is in the same tangent space as ξ, and for $\alpha, \beta \in \mathbb{R}$, $p \in M$, $\xi, \eta \in T_p(M)$, $Y \in \Gamma(TM)$,
$$\nabla_{\alpha\xi+\beta\eta} Y = \alpha \nabla_\xi Y + \beta \nabla_\eta Y.$$
(2) For $p \in M$, $\xi \in T_p(M)$, $Y, Y_1, Y_2 \in \Gamma(TM)$, $f \in C^1(M)$ we have
$$\nabla_\xi (Y_1 + Y_2) = \nabla_\xi Y_1 + \nabla_\xi Y_2,$$
$$\nabla_\xi (fY) = (\xi f) Y_{|p} + f(p) \nabla Y$$
(3) ∇ is smooth in the sense if X, Y are C^∞ vector fields so is $\nabla_X Y$.

This should be thought of as a generalization of the directional derivative as shown below. Let $\sigma_p : \mathbb{R}^n \to (\mathbb{R}^n)_p$ be the natural identification of $\mathbb{R}^n$ with the abstract tangent space to $(\mathbb{R}^n)_p$ determined by the chart consisting of the identity map of $\mathbb{R}^n$ that is given by
$$\partial_{j|p} = \sigma_p e_j, \text{ for } j = 1, ..., n.$$
Let Y be a differentiable vector field on $\mathbb{R}^n$ such that $Y = \sum_j \eta^j \partial_j$. Then the standard connection on $\mathbb{R}^n$ is given by
$$\nabla_\xi Y = \sum_{j=1}^n (\xi \eta^j) \partial_j.$$
It is apparent that it is a connection. Geometrically, given $\xi \in (\mathbb{R}^n)_p$, let $\gamma : (-\epsilon, \epsilon) \to \mathbb{R}^n$ be a C^1 path in $\mathbb{R}^n$ with $\gamma(0) = p$, $\gamma'(0) = \xi$. Then we have
$$\nabla_\xi Y = \lim_{t \to 0} \frac{\sigma_p \circ \sigma^{-1}_{\gamma(t)} Y_{|\gamma(t)} - Y_{|p}}{t}$$
To show this we compute the quotient on the right hand side for a smooth function f and choose $\gamma(t) = p + t\xi$ (which satisfies $\gamma(0) = p$ $\gamma'(0) = \xi$)
$$\frac{(\sigma_p \circ \sigma^{-1}_{\gamma(t)} Y_{|\gamma(t)} - Y_{|p}) f}{t} = \frac{\sigma_p(\sigma^{-1}_{\gamma(t)}(Yf(\gamma(t)))) - Yf(p)}{t}$$
$$= \frac{Yf(\gamma(t)) - Yf(p)}{t} = \frac{Yf(p + t\xi) - Yf(p)}{t} \to \sum_{j=1}^n (\xi \eta^j) \partial_j f(p)$$
proving the equality. Hence, the natural identification of the tangent space $(\mathbb{R}^n)_p$ and $(\mathbb{R}^n)_q$ via the map $\sigma_q \circ \sigma_p^{-1}$, for any $p, q \in \mathbb{R}^n$ is that which allows for the natural differentiation of vector fields on $\mathbb{R}^n$. In an abstract smooth manifold, no such natural identification exists a priori.

We can show that if two vector fields agree on all of U, an open set consisting $p \in M$, so do their covariant derivatives $\nabla_\xi Y$. So it is well-defined and independent of the choice of extension of Y to $\bar{Y} \in \Gamma(TM)$. In this way we may effectively calculate ∇ by restricting the vector fields at hand to, for example, the domain of a chart.

Moreover, to calculate $\nabla_\xi Y$, for a given $Y \in \Gamma(TM)$, we need only know Y restricted to a path through $p = \gamma(\xi)$ with velocity vector at p equal to ξ. This can be done as follows: let $x : U \to \mathbb{R}^n$ be a chart about p, and ξ given by $\xi = \sum_j \xi^j \partial_{j|p}$. Then $\nabla_\xi Y = \sum_j \xi^j \partial_{j|p} Y$.

Also we have functions $\eta^j : U \to \mathbb{R}$, $j = 1, ..., n$ such that $Y|U = \sum_j \eta^j \partial_j$. Now there are functions $\Gamma_{jk}l : U \to \mathbb{R}$, $j, k, l = 1, ..., n$ (*Christoffel symbols*) such that, on U, it holds that

$$\nabla_{\partial_k} \partial_j = \sum_l \Gamma^l_{jk} \partial_l.$$

Then we have

$$\nabla_\xi Y = \sum_k \xi^k \nabla_{\partial_{k|p}} Y = \sum_k \xi^k \nabla_{\partial_{k|p}} \left(\sum_j \eta^j \partial_j \right)$$

$$= \sum_k \xi^k \left(\sum_j (\partial_k \eta^j)(p) \partial_{j|p} + \sum_{j,l} \eta^j(p) \Gamma^l_{jk}(p) \partial_{l|p} \right)$$

$$= \sum_l \left(\sum_k \xi^k (\partial_k \eta^l)(p) + \sum_{j,k} \Gamma^l_{jk}(p) \eta^j(p) \xi^k \right) \partial_{l|p}$$

This is

$$\nabla_\xi Y = \sum_l \left(\sum_k \xi^k (\partial_k \eta^l)(p) + \sum_{j,k} \Gamma^l_{jk}(p) \eta^j(p) \xi^k \right) \partial_{l|p}$$

In particular,if $\gamma : (\alpha, \beta) \to M$ is differentiable such that $t_0 \in (\alpha, \beta)$, $\gamma(t_0) = p$, $\gamma'(t_0) = \xi$, we will have

$$\nabla_\xi Y = \sum_l \left((\eta^l \circ \gamma)'(t_0) + \sum_{j,k} \Gamma^l_{jk}(p) \eta^j(p) \xi^k \right) \partial_{l|p}$$

4.8. **Torsion and curvature tensors**

Two concepts that are strongly related to the connection are the torsion T_p and curvature tensor R_p at p of ∇. They are multilinear maps:

$$T_p : T_p(M) \times T_p(M) \to T_p(M)$$
$$R_p : T_p(M) \times T_p(M) \times T_p(M) \to T_p(M)$$

given by,

$$T_p(\xi, \eta) = \nabla_\xi Y - \nabla_\eta X - [X, Y]_{|p}$$
$$R_p(\xi, \eta)\zeta = \nabla_\eta \nabla_X Z - \nabla_\xi \nabla_Y Z - \nabla_{[Y,X]_{|p}} Z,$$

where $X, Y, Z \in \Gamma(TM)$ with $X_{|p} = \xi$, $Y_{|p} = \eta$ and $Z_{|p} = \zeta$. These maps are well-defined, meaning that the do not depend on the choice of vector fields X, Y and Z extending ξ, η and ζ.

Obviously we have the identities:

$$T(\xi, \eta) = -T(\eta, \xi), R(\xi, \eta)\zeta = -R(\eta, \xi)\zeta$$

and if $T = 0$ is given on the whole M, then we get the first Bianchi identity:

$$R(\xi, \eta)\zeta + R(\zeta, \xi)\eta + R(\eta, \zeta)\xi = 0$$

A connection ∇ on TN induces a similar operator on vector fields along a map $\phi : M \to N$. More precisely we can show that ([5]) there is a unique bilinear map $TM \times \Gamma(\phi^{-1}TN) \to TN$ defined by $(\xi, X) \mapsto \phi^{-1}\nabla_\xi X$ such that, for $\xi \in T_p(M)$, $X \in \Gamma(TM)$, $Y \in \Gamma(\phi^{-1}TN)$ and $f \in C^\infty(M)$,

(1) $\phi^{-1}\nabla_\xi Y \in T_{\phi(p)}(N)$;
(2) $\phi^{-1}\nabla_\xi(fY) = (\xi f)Y_{|\phi(p)} + f(p)\phi^{-1}\nabla_\xi Y$;
(3) $\phi^{-1}\nabla_X Y \in \Gamma(\phi^{-1}TN)$ (smoothness);
(4) If $Z \in \Gamma(TN)$, then $Z \circ \phi \in \Gamma(\phi^{-1}TN)$ and $\phi^{-1}\nabla_\xi(Z \circ \phi) = \nabla_{\mathrm{d}\phi_p(\xi)}Z$.

Here $\phi^{-1}\nabla$ is called the *pull-back* of ∇ by ϕ. The first three properties above state that $\phi^{-1}\nabla$ behaves like ∇, however the last one essentially defines it in a unique way.

5. **Riemannian structure**

5.1. **An informal discussion**

In the geometric framework, one has to separate the topological and differential properties of the manifold from the geometric and metric ones. The first ones determine the local structure of a manifold M by specifying neighboring points and tangent vectors, which allows us to differentiate smooth functions on the manifold. This also allows us to define continuous paths on the manifold and to classify them by the number of loops they are doing around "holes" in the manifold. However, within each of these homotopy classes, there is no tool to choose something like the "straightest path". To obtain such a notion, we need to add a geometric structure, called a connection, which allows to compare neighboring tangent spaces. Indeed, differentiating a path on a manifold gives tangent vectors belonging at each point to a different tangent vector space. In order to compute the second order derivative (the acceleration of the path), we need a way to map the tangent space at a point to the tangent space at any neighboring point. This is the goal of a connection $\nabla_X Y$, which specifies how the vector field $Y(p)$ is derived in the direction of the vector field $X(p)$. Such a connection operator also describes how a vector is transported from a tangent space to another along a given curve and specifies the local parallel transport. However, there is usually no global parallelism. As a matter of facts, transporting the same vector along two different curves arriving at the same point might lead to different vectors: this is easily seen on the sphere where traveling from north pole to the equator, then along the equator for 90 degrees and back to North pole turns any tangent vector by 90 degrees. This defect of global parallelism is the sign of curvature. By looking for curves that remains locally parallel to themselves (i.e. such that $\nabla_{\dot\gamma}\dot\gamma = 0$), one defines the equivalent concept to "straight lines" in the manifold: geodesics. One should notice that there exists many different choices of connections on a given manifold which lead to different geodesics. Geodesics by themselves do not quantify how far away from each other two points are. For that purpose, we need an additional structure: a distance. By restricting to distances which are compatible with the differential structure, we enter into the realm of Riemannian geometry. A Riemannian metric is defined by a continuous collection of scalar products $\langle\cdot|\cdot\rangle_p$ (or equivalently norms $||\cdot||_p$) on each tangent space T_pM at point p of the manifold. Thus, if we consider a curve on the manifold, we can compute at each point its instantaneous speed vector (this operation only involves the differential structure) and its norm to obtain the instantaneous speed (the Riemannian metric is needed for this operation). To compute the length of the curve, this value is integrated as usual along the curve. The distance between two points of a connected Riemannian manifold is the minimum length among the curves joining these points. The curves realizing this minimum are called metric geodesics. The fundamental theorem of Riemannian geometry states that on any Riemannian manifold there is a unique (torsion-free) connection which is compatible with the metric, called the Levi-Civita (or metric) connection. For

that choice of connection, shortest path are geodesics ("straight lines"). In the following, we only consider the Levi-Civita connection. Moreover, we assume that the manifold is geodesically complete, i.e. that all geodesics can be indefinitely extended. This means that the manifold has neither boundary nor any singular point that we can reach in a finite time. As an important consequence, the Hopf-Rinow-De Rham theorem states that there always exists at least one minimizing geodesic between any two points of the manifold (i.e. whose length is the distance between the two points).

5.2. **Riemannian metric and Riemannian manifolds**

A rich and useful geometry arises when each manifold is equipped with an inner product.

Definition. A *Riemannian metric* g on M is an inner product $g_p : T_p(M) \times T_p(M) \to \mathbb{R}$ on each $T_p(M)$ such that, for all vector fields X and Y, the function

$$p \mapsto g_p(X_{|p}, Y_{|p})$$

is smooth. A *Riemannian manifold* is a pair (M, g) with M a manifold and g a metric on M.

Given a smooth map $\phi : M \to N$ and a metric g on N we can pull g back to a section $\phi^* g$ of $TM \times TM$ by

$$(\phi^* g)_p(X, Y) = g_\phi(p)(D\phi_p(X), D\phi_p(Y)).$$

If $D\phi_p$ is invertible this will again be positive definite, so in particular if ϕ is a diffeomorphism. A diffeomorphism $\phi : M \to N$ between two Riemannian manifolds is an *isometry* if $\phi^* g_N = g_M$.

Examples.

- Let $(\ ,\ )$ denote the inner product on $\mathbb{R}^n$. An open set U in $\mathbb{R}^n$ gets a Riemannian metric via $T_p(U) \simeq \mathbb{R}^n$:
$$g_p(v, w) = (v, w).$$
- Let $S^n \subset \mathbb{R}^{n+1}$ be the unit sphere. Then $T_p(S^2) \simeq p^\perp \subset \mathbb{R}^{n+1}$ and so gets a metric from he inner product on $\mathbb{R}^{n+1}$.
- Let $D^n \subset \mathbb{R}^n$ be the open unit ball define a metric by
$$g_z(v, w) = \frac{4(v, w)}{(1 - |z|^2)^2}$$
(D^n, g) is the so-called hyperbolic space.

Riemannian geometry is powerful thanks to the fact that there is a *canonical* choice of connection. Consider the following two properties for a connection ∇ on (M, g):

- ∇ is *metric*: $Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$.
- ∇ is *torsion-free*: $\nabla_X Y - \nabla_Y X = [X, Y]$.

Theorem 5.1 (Levi-Civita)**.** *If M is a Riemannian manifold, then there exists a unique torsion-free metric connection ∇ (hence called the Levi-Civita connection)*

Proof. Assume that g is metric and torsion-free. Then, for any $X, Y, Z \in \Gamma(TM)$,

$$\begin{aligned} g(\nabla_X Y, Z) &= Xg(Y, Z) - g(Y, \nabla_X Z) \\ &= Xg(Y, Z) - g(Y, \nabla_Z X) - g(Y, [X, Z]) \\ &= Xg(Y, Z) - Zg(Y, X) + g(\nabla_Z Y, X) - g(Y, [X, Z]) \\ &= Xg(Y, Z) - Zg(Y, X) + g(\nabla_Y Z, X) + g([Z, Y], X) - g(Y, [X, Z]) \end{aligned}$$

$$
\begin{aligned}
&= Xg(Y,Z) - Zg(Y,X) + Yg(Z,X) - g(Z,\nabla_Y X) + g([Z,Y],X) - g(Y,[X,Z]) \\
&= Xg(Y,Z) - Zg(Y,X) + Yg(Z,X) - g(Z,\nabla_X Y) \\
&\qquad - g(Z,[Y,X]) + g([Z,Y],X) - g(Y,[X,Z])
\end{aligned}
$$

that is

$$2g(\nabla_X Y, Z) = Xg(Y,Z) - Zg(Y,X) + Yg(Z,X) - g(X,[Y,Z]) + g(Y,[Z,X]) + g(Z,[X,Y])$$

This shows the uniqueness of the connection,because the right hand side is determined by the metric. Moreover, it also defines the connection we look for. Indeed we can show that the ∇ defined by this formula is the torsion-free metric connection. □

In application it is quite often we need express the metric and Levi-Civita connection in terms of local coordinates. Let (U, x) be a chart and $\partial_1, ..., \partial_n$ be the corresponding vector fields on U. Define $g_{ij} \in C^\infty(U)$ by $g_{ij} = g(\partial_i, \partial_j)$ and Christoffel symbols $\Gamma_{ij}^k \in C^\infty(U)$ by

$$\nabla_{\partial_i} \partial_j = \sum_k \Gamma_{ij}^k \partial_k.$$

Then we have $[\partial_i, \partial_j] = 0$. From the definition of the LeviCivita connection we obtain

$$
\begin{aligned}
\sum_k \Gamma_{ij}^k g_{kl} = g(\sum_k \Gamma_{ij}^k \partial_k, \partial_l) &= g(\nabla_{\partial_i} \partial_j, \partial_l) \\
&= \frac{1}{2} \left(\partial_i g_{jl} - \partial_l g_{ji} + \partial_j g_{li} \right) \\
&= \frac{1}{2} \left(\frac{\partial g_{jl}}{\partial x_i} + \frac{\partial g_{li}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_l} \right)
\end{aligned}
$$

where we have used the symmetry of $\Gamma_{ij}^k = \Gamma_{ji}^k$. Now taking the inverse of g, whose element is denoted by g^{kl}, then we can solve above equation for Γ_{ij}^k:

$$\Gamma_{ij}^k = \frac{1}{2} \sum_l g^{kl} \left(\frac{\partial g_{jl}}{\partial x_i} + \frac{\partial g_{li}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_l} \right)$$

5.3. **Geodesics**

In the classic Euclidean geometry the straight lines played an important role. It is viewed as shortest paths between two points. These straight lines can be generalized to a distinguished family of paths called *geodesics* on any Riemannian manifold. Geodesics provide powerful tools in the Riemannian geometry.

Let us first introduce a notation related to bilinear form. Let (M, g). be a Riemannian manifold. For $\xi, \eta \in T_p(M)$, write

$$g(\xi, \eta) = \langle \xi, \eta \rangle, \ \sqrt{g(\xi, \xi)} = |\xi|.$$

Now we claim that (M, g) can be turned to a metric space by a proper definition of a metric (or distance function). To this end we first define the length of a path. A piecewise C^1 path $\gamma : [a, b] \to M$ has *length* $L(\gamma)$:

$$L(\gamma) = \int_a^b |\gamma'(t)| dt.$$

Some remarks are in order: (i) The integrand is a function of t alone, so $\frac{d\gamma}{dt} \in T_{\gamma(t)}(M)$ denotes the tangent vector to the curve $\gamma(t)$. (ii) Reparametrization of a path preserves

the length. This can be seen by the parameter change as follows: $\frac{d\gamma}{ds} = \frac{d\gamma}{dt}\frac{dt}{ds}$ with $t = f(s)$, $c \le s \le d$ is a new parametrization. Then

$$\int_c^d \left|\frac{d\gamma}{ds}\right| ds = \int_a^b \left|\frac{d\gamma}{dt}\left(\frac{dt}{ds}\right)\right| \frac{ds}{dt}dt = \int_a^b |\gamma'(t)|dt.$$

(iii) In particular, we note that the arclength along the curve from $\gamma(a)$ to $\gamma(t)$ which may be denoted by $s = L(t)$ gives a new parameter by the formula

$$s = L(t) = \int_a^t |\gamma'(\tau)|d\tau \Rightarrow \frac{ds}{dt} = |\gamma'(t)| \Leftrightarrow \left(\frac{ds}{dt}\right)^2 = g\left(\frac{d\gamma}{dt}, \frac{d\gamma}{dt}\right)$$

In a single coordinate neighborhood (or chart) (U, φ) with standard coordinates $\partial_1, ..., \partial_n$ we have $g(\partial_i, \partial_j) = g_{ij}(x)$ and $\varphi(\gamma) = x = (x^1, ..., x^n)$ and the curve is given by $\varphi(\gamma(t)) = (x^1(t), ..., x^n(t))$ then

$$s = L(t) = \int_a^t \sqrt{g_{ij}(x(\tau))\frac{dx^i}{d\tau}\frac{dx^j}{d\tau}}d\tau$$

This leads to the classical expression for the Riemannian metric in local coordinates

$$ds^2 = \sum_{j,k} g_{jk}dx^j dx^k.$$

On $\mathbb{R}^n$ this becomes

$$ds^2 = |dx|^2.$$

Since connectedness of M implies path-connectedness of M, [3], we can pretty easily check, for $p, q \in M$, that

$$d(p, q) = \inf\{L(\gamma) : \gamma : [a, b] \to M \text{ is a path with } \gamma(a) = p, \gamma(b) = q\}$$

is well-defined. Moreover we can show that d is a metric on M. The essential parts are the definiteness of d and about the topologies. It is obvious that d is symmetric, nonnegative and satisfies the triangle inequality. So to show d is a distance function it remains to prove that d is positive definite, that is, $d(p, q) > 0$ for $p, q \in M$, $p \neq q$. Assume that M is n-dimensional. Given $p \in M$, let $x : U \to \mathbb{R}^n$ be a chart on M, $p \in U$. Then there exists $r > 0$ for which the open ball $B(x(p); r)$ of $\mathbb{R}^n$ centered at $x(p)$ with radius r satisfies

$$B(x(p); r) \subset x(U)$$

which determines the existence of a constant $K > 0$ such that, for all $\xi \in T(x^{-1}(B(x(p); r)))$

$$|\xi| = \sum_j \xi^j \partial_j$$

we have

$$|\xi| \ge K\sqrt{\sum_j (\xi^j)^2}$$

Then on the open ball $B(x(p); r)$, the Riemannian lengths are uniformly bounded below by the corresponding Euclidean lengths. Therefore, for $q \in x^{-1}(B(x(p); r))$ we have

$$d(p, q) \ge K|x(p) - x(q)] > 0$$

For $q \in M \setminus x^{-1}(B(x(p); r))$ we have obviously $d(p, q) \ge Kr$. So if $p \neq q$ it implies that $d(p, q) > 0$. So d turns M to a metric space. Thus we have

Theorem 5.2. *Any Riemannian manifold (M, d) is a metric space. In fact the metric space topology coincides with the original topology on M.*

We refer to [3] for the proof on the topologies.

5.4. **Parallel vector fields and geodesics**

Let $\gamma : I \to M$ be a path. Recall the pull-back connection $\gamma^{-1}\nabla$ on the space $\Gamma(\gamma^{-1}TM)$ of vector fields along γ. This connection gives rise to a differential operator

$$\nabla_t : \Gamma(\gamma^{-1}TM) \to \Gamma(\gamma^{-1}TM)$$

by

$$\nabla_t Y = (\gamma^{-1}\nabla)_{\partial_1} Y$$

where ∂_1 is the coordinate vector field on I.

Because ∇ is metric, we have the following equality

$$\langle X, Y\rangle' := \frac{d}{dt}\langle X, Y\rangle = \langle \nabla_t X, Y\rangle + \langle X, \nabla_t Y\rangle, \quad \text{for } X, Y \in \Gamma(\gamma^{-1}TM)$$

Definition. $X \in \Gamma(\gamma^{-1}TM)$ is *parallel* to the path if $\nabla_t X = 0$.

By the existence and uniqueness theorem for linear ODE, it is easy to prove

Proposition 5.3. *For $\gamma : [a, b] \to M$ and $U_0 \in T_{\gamma(a)}(M)$ there is a unique parallel vector field U along the curve γ with $U(a) = U_0$. If Y_1, Y_2 are parallel vector fields along γ, then all $\langle Y_i, Y_j\rangle$, and in particular, $|Y_i|$ are constant.*

This can be shown by the fact that ∇ is metric which gives

$$\frac{d}{dt}\langle Y_1, Y_2\rangle = \langle \nabla_t Y_1, Y_2\rangle + \langle Y_1, \nabla_t Y_2\rangle = 0$$

proving that $|Y_i|$ is a constant for $\frac{d}{dt}\langle Y_i, Y_i\rangle = 0$.

Definition. $\gamma : I \to M$ is a *geodesic* if γ' is parallel :

$$\nabla_t \gamma' = 0.$$

Before proceeding further we note that the definition of a geodesic is sometimes given by the formula

$$\nabla_{\dot\gamma}\dot\gamma = 0.$$

Here is a motivation for such a restatement. Note that by definition of pull-back connection we have $\nabla_t Y = \gamma^{-1}\nabla_{\partial_1} Y = \nabla_{\gamma'(t)} Y$. This motivates the notation $\nabla_{\gamma'} Y$ for $\nabla_t Y$. In many situations we also use $\dot\gamma$ for γ'.

From the above proposition it is clear that $|\dot\gamma|$ *is a constant*. And it is easy to show that if γ is a geodesic then the curve $t \mapsto \gamma(st)$ is also a geodesic for all $s \in \mathbb{R}$.

By again the same proposition the following proposition follows.

Proposition 5.4. *Let (M, g) be a Riemannian n-manifold, $p \in M$ and $\{v_1, ..., v_n\}$ be an orthonormal basis for the tangent space $T_p(M)$. Let $\gamma : I \to M$ be a smooth curve such that $\gamma(0) = p$ and $X_1, ..., X_n$ be parallel vector fields along γ such that $X_k(0) = v_k$, $k = 1, ..., n$. Then the set $\{X_1(t), ..., X_n(t)\}$ is an orthonormal basis for the tangent space $T_{\gamma(t)}(M)$ for all $t \in I$.*

Geodesics are of great importance not only in Riemannian geometry but also in many applications. The question now is whether or not they exist and unique if they exist. To this end, we consider a chart (U, x) on M such that $p \in U$. Choose $X_i = \partial_i = \frac{\partial}{\partial x_i} \in C^\infty(TU)$. Let J be an open subset of $I = (-\epsilon, \epsilon)$, sent by a path γ to M with $\gamma(0) = p$

and $\dot{\gamma}(0) = v \in T_p(M)$ such that the image $\gamma(J) \subset U$. Then the tangent of the restriction of γ on J can be written as

$$\dot{\gamma}(t) = \sum_{i=1}^{n} \dot{\gamma}_i(t)(X_i)_{\gamma(t)}.$$

Differentiating this equation yields

$$\begin{aligned}
\nabla_{\dot{\gamma}}\dot{\gamma} &= \sum_{j=1}^{n} \nabla_{\dot{\gamma}}(\dot{\gamma}_j(t)(X_j)_{\gamma(t)}) \\
&= \sum_{j=1}^{n} \left(\ddot{\gamma}_j(t)(X_j)_{\gamma(t)} + \sum_{i=1}^{n} \dot{\gamma}_i(t)\dot{\gamma}_j(t)(\nabla_{X_i}X_j)_{\gamma(t)} \right) \\
&= \sum_{k=1}^{n} \left(\ddot{\gamma}_k(t) + \sum_{i,j=1}^{n} \dot{\gamma}_i(t)\dot{\gamma}_j(t)\Gamma_{ij}^{k}(\gamma(t)) \right) (X_k)_{\gamma(t)}.
\end{aligned}$$

By this computation the curve γ is a a geodesic if and only if

$$\ddot{\gamma}_k(t) + \sum_{i,j=1}^{n} \dot{\gamma}_i(t)\dot{\gamma}_j(t)\Gamma_{ij}^{k}(\gamma(t)) = 0, \ k = 1, ..., n$$

Now applying the well-know theorem of Picard-Lindelöf[2] gives that for initial values $q = x(p)$ and $w = (dx)_p(v)$ there exists an open interval $(-\epsilon, \epsilon)$ and a unique solution $(\gamma_1, ..., \gamma_n)$ satisfying the initial conditions $(\gamma_1(0), ..., \gamma_n(0)) = q$ and $(\dot{\gamma}_1(0), ..., \dot{\gamma}_n(0)) = w$.

These arguments and computations lead to the fundamental uniqueness and existence result

Theorem 5.5. *Let (M, g) be a Riemannian manifold. If $p \in M$ and $v \in T_p(M)$ then there is an open interval $(-\epsilon, \epsilon)$ and a unique geodesic $\gamma : I \to M$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = v$.*

We point out that the image of a geodesic under a local isometry is a geodesic. The argument is that the Levi-Civita connection ∇ on a given Riemannian manifold (M, g) is completely determined by the smooth structure on M and the Riemannian metric g. Hence it also applies to the condition $\nabla_{\dot{\gamma}}\dot{\gamma} = 0$ for a given path γ. This condition is called the *geodesic equation*.

Now we consider $M = \mathbb{R}^n$ with its canonical metric. The geodesic equation reduces to $\ddot{\gamma} = 0$ from which we can conclude that the geodesic are straight lines. More examples will be given in §5.6.

5.5. **Curvature tensors and sectional curvature**

Recall that for $X, Y, Z, W \in \Gamma(TM)$ with M being our Riemannian manifold with the Levi-Civita connection, we have

$$R(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z - \nabla_{[Y,X]} Z,$$

[2]The Picard-Lindelöf Theorem: Let U be an open subset of $\mathbb{R} \times \mathbb{R}^n$. Let $f : U \to \mathbb{R}^n$ be a continuous map and $L \in \mathbb{R}_+$ such that $|f(t, x) - f(t, y)| \leq L|x - y|$ for all $t, x), (t, y) \in U$. If $(t_0, x_0) \in U$ and $x_1 \in \mathbb{R}^n$ there exists a unique local solution $x : I \to \mathbb{R}^n$ to the following initial value problem $x''(t) = f(t, x(t))$, $x(t_0) = x_0$, $x'(t_0) = x_1$.

where R is the curvature tensor of ∇. Let us have a close look at this definition geometrically. What is the geometric meaning of it? Clearly, it measures the failure of the manifold to be flat, i.e. locally isometric to Euclidean space. It also gives geometric meaning on itself: If X and Y are commuting linearly vector fields, we can, for $p \in M$ and sufficiently small $t, s \in \mathbb{R}$, create a closed parallelogram at p by (i) flow for time t along X time s along Y, t along $-X$ and s along $-Y$ and (ii) $R(X, Y)$ measures the failure of parallel transport of Z around an infinitesimally small such parallelogram to give back the vector Z.

An important simplification of the Riemann curvature tensor is *sectional curvature*. The reason will become apparent in a short while.

Definition. Let $\sigma \subset T_p(M)$ be a 2-plane with orthonormal basis ξ, η. The *sectional curvature* $\mathcal{K}(\sigma)$ of σ is defined by

$$\mathcal{K}(\sigma) = -\langle R(\xi, \eta)\xi, \eta\rangle$$

Now we are going to show that the following things:

- This definition is independent of the choice of basis of σ.
- $\mathcal{K}$ defines the curvature tensor R.

It is apparent that the sectional curvature is much simpler to compute than the R. Still, it captures the curvature tensor R and it encapsulates the same information.

To prove we collect some properties of R: For $X, Y, Z, W \in \Gamma(TM)$

(1) R is skew-symmetric in terms of the two entries: $R(X, Y)Z + R(Y, X)Z = 0$
(2) The first Bianchi identity:

$$R(X, Y)Z + R(Z, X)Y + R(Y, Z)X = 0$$

since ∇ is torsion-free
(3) $\langle R(X, Y)Z, W\rangle$ is symmetric between the first pair and the last pair of entries:

$$\langle R(X, Y)Z, W\rangle - \langle R(Z, W)X, Y\rangle = 0$$

and
(4) $\langle R(X, Y)Z, W\rangle$ is skew-symmetric between the last two entries:

$$\langle R(X, Y)Z, W\rangle + \langle R(X, Y)W, Z\rangle = 0$$

These can be proved by a straight-forward but lengthy and tedious computation using the properties of ∇. To argue the independence of basis choice we assume there is another orthonormal basis $\tilde{\xi}$, $\tilde{\eta}$ for σ then $\tilde{\xi} = \alpha\xi + \beta\eta$ and $\tilde{\eta} = \gamma\xi + \delta\eta$ with an orthogonal transformation matrix $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$. Using the above properties of R we have

$$\begin{aligned}
\langle R(\tilde{\xi}, \tilde{\eta})\tilde{\xi}, \tilde{\eta}\rangle &= \langle R(\alpha\xi + \beta\eta, \gamma\xi + \delta\eta)(\alpha\xi + \beta\eta), (\gamma\xi + \delta\eta)\rangle \\
&= (\alpha\delta - \beta\gamma)\langle R(\xi, \eta)(\alpha\xi + \beta\eta), (\gamma\xi + \delta\eta)\rangle \\
&= (\alpha\delta - \beta\gamma)^2\langle R(\xi, \eta)\xi, \eta\rangle = \langle R(\xi, \eta)\xi, \eta\rangle.
\end{aligned}$$

This shows the independence of the basis choice.

Next we establish the relation between the curvature tensor R and the sectional curvature $\mathcal{K}$. To compute the sectional curvature let's study the bilinear form $k(\xi, \eta) = \langle R(\xi, \eta)\xi, \eta\rangle$, which is in fact $-\mathcal{K}(\sigma)$. A straightforward calculation gives

$$\langle R(\xi_1, \eta_1)\xi_2, \eta_2\rangle = \frac{1}{6}\frac{\partial^2}{\partial s \partial t}\bigg|_{s=t=0} (k(\xi_1 + s\xi_2, \eta_1 + t\eta_2) - k(\xi_1 + s\eta_2, \eta_1 + t\xi_2))$$

for $\xi_1, \xi_2, \eta_2, \eta_2 \in \Gamma(TM)$.

Note that if the sectional curvature at p is zero then the curvature $R = 0$ at p. Hence *a Riemannian manifold (M, g) is flat if and only if the sectional curvature is identically zero.*

Spaces of constant sectional curvature. In this paragraph we give a short presentation on how to derive constant curvature. First we need a formal definition.

Definition. (M, g) has constant curvature κ if $\mathcal{K}(\sigma) = \kappa$ for all 2-planes σ in TM.

In this case we have

$$R(\xi, \eta)\zeta = \kappa\left(\langle \xi, \zeta\rangle\eta - \langle \eta, \zeta\rangle\xi\right)$$

$\mathcal{K}$ is a function on the set (in fact manifold) $G_2(TM)$ of all 2-planes in all tangent spaces $T_p(M)$ of M. A diffeomorphism $\Phi : M \to M$ induces $\mathrm{d}\Phi : TM \to TM$ which is a linear isomorphism on each tangent space and so gives a map $\hat{\Phi} : G_2(TM) \to G_2(TM)$. Suppose now that Φ is an *isometry*:

$$\langle \mathrm{d}\Phi_p(\xi), \mathrm{d}\Phi_p(\eta)\rangle = \langle \xi, \eta\rangle \quad \text{for } all \xi, \eta \in T_p(M), p \in M.$$

Since an isometry preserves the metric, it will preserve anything built out of the metric such as the Levi-Civita connection and its curvature. In particular, we have

$$\mathcal{K} \circ \hat{\Phi} = \mathcal{K}.$$

Then it can be proved that (e.g. [3]) the group of all isometries acts *transitively* on $G_2(TM)$ so that $\mathcal{K}$ is constant. We have already proved that $\mathbb{R}^n$ has $\mathcal{K} = 0$, a constant curvature. In a short while we prove that the following two manifolds have constant curvature: the n-sphere $S^n(\rho)$, centered at the origin with radius ρ, has $\mathcal{K} = 1$, the hyperbolic space $\mathbb{R}^n_+$ has $\mathcal{K} = -1$. Moreover it can be proved that all complete, simply-connected Riemann manifolds of constant curvature $\mathcal{K} = 0$, 1, or -1, (and in this order) is isometric to one of these three examples, see again e.g. [3]. Now we turn to computation of the curvatures of the sphere and the hyperbolic spaces.

- The sphere $S^n(\rho)$, $n \geq 2$. There are several ways to compute the curvature of the sphere, perhaps it is easier (at least shorter) to use the differential form (the second fundamental form). However we have to introduce new concepts and theorems. To avoid more new theory in this report we choose to use coordinates on $S^n(\rho)$ given by stereographic projection as used earlier in this report and do the calculation explicitly. Recall that the stereographic projection fixes the "north pole" of the sphere and its equatorial n dimensional hyperplane

  $$\mathbb{R}^n = \{z \in \mathbb{R}^{n+1} : z^{n+1} = 0\}$$

  in $\mathbb{R}^{n+1}$. Then with every point $y \in S^n(\rho) = (y^1, ..., y^{n+1})$ we associate $x = (x^1, ..., x^n) \in \mathbb{R}^n$ to the point of intersection of the line in $\mathbb{R}^{n+1}$ determined by the "north pole" and y with the equatorial hyperplane of $\mathbb{R}^{n+1}$. The line equation for the projection is

$$(y^1, ..., y^{n+1}) = (0, ..., 0, \rho) + t((x^1, ..., x^n, 0) - (0, ..., 0, \rho)) = (tx^1, ..., tx^n, \rho - t\rho)$$

  Illustrated in Figure 9

  From this we have

$$y^j = \frac{x^j(\rho - y^{n+1})}{\rho}, j = 1, ..., n$$

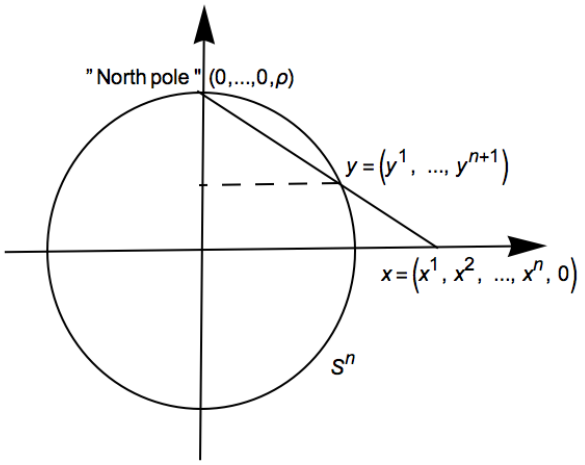


FIGURE 9. Stereographic projection simplified in 1 dimension

This, together with the sphere equation gives

$$y^{n+1} = \frac{|x|^2 - \rho^2}{|x|^2 + \rho^2}\rho, (y^1, ..., y^n) = \frac{2\rho^2 x}{\rho^2 + |x|^2}$$

where $|x|$ is the Euclidean length in $\mathbb{R}^n$. Next we should compute the metric which is given by

$$g_{ij} = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle$$

Remember that the Riemannian metric can be expressed by $ds^2 = \sum_{ij} g_{ij} dy^i dy^j$. We shall use dot for inner product in $\mathbb{R}^n$ and let $\tilde{y} = (y^1, ..., y^n)$. So we compute directly dy^j now:

$$d\tilde{y} = 2\rho^2 \frac{(\rho^2 + |x|^2)dx - 2x(x \cdot dx)}{(\rho^2 + |x|^2)^2} \Rightarrow |d\tilde{y}|^2 = 4r^4 \frac{(\rho^2 + |x|^2)^2 |dx|^2 - 4\rho^2 (x \cdot dx)^2}{(\rho^2 + |x|^2)^4}$$

$$dy^{n+1} = 4\rho^3 \frac{x \cdot dx}{(\rho^2 + |x|^2)^2}, \Rightarrow (dy^{n+1})^2 = 16\rho^6 \frac{(x \cdot dx)^2}{(\rho^2 + |x|^2)^4}$$

All these together yield

$$ds^2 = \sum_{k=1}^{n+1} (dy^k)^2 = |d\tilde{y}|^2 + |dy^{n+1}|^2 = 4\rho^4 \frac{|dx|^2}{(\rho^2 + |x|^2)^2}$$

Hence we obtain

$$g_{ij} = \frac{4\rho^4 \delta_{ij}}{(\rho^2 + |x|^2)^2},$$

and the Christoffel symbol

$$\Gamma^l_{jk} = \frac{1}{2} \sum_i g^{li} (\partial_j g_{ik} + \partial_k g_{ji} - \partial_i g_{jk}) = \frac{-2}{\rho^2 + |x|^2} (\delta_{lk} x^j + \delta_{jl} x^k - \delta_{jk} x^l).$$

Finally we compute the (sectional) curvature $\mathcal{K} = \langle R(\partial_i, \partial_j)\partial_i, \partial_j\rangle$. To this end, we first compute the symbol R^l_{kij} defined by

$$R(\partial_i, \partial_j)\partial_k := \sum_s R^s_{kij}\partial_s$$

Using the fact that $[\partial_i, \partial_j] = 0$ we obtain

$$R(\partial_i, \partial_j)\partial_k = \nabla_{\partial_i}\nabla_{\partial_j}\partial_k - \nabla_{\partial_j}\nabla_{\partial_i}\partial_k = \sum_s \left(\nabla_{\partial_i}(\Gamma^s_{jk}\partial_s) - \nabla_{\partial_j}(\Gamma^s_{ik}\partial_s)\right)$$

$$= \sum_s \left(\partial_i(\Gamma^s_{jk})\partial_s + \sum_r \Gamma^s_{jk}\Gamma^r_{is}\partial_r - \partial_j(\Gamma^s_{ik})\partial_s - \sum_r \Gamma^s_{ik}\Gamma^r_{js}\partial_r\right)$$

$$= \sum_s \left(\partial_i(\Gamma^s_{jk}) - \partial_j(\Gamma^s_{ik}) + \sum_r \left(\Gamma^r_{jk}\Gamma^s_{ir} - \Gamma^r_{ik}\Gamma^s_{jr}\right)\right)\partial_s$$

Therefore,

$$R^s_{kij} = \partial_i(\Gamma^s_{jk}) - \partial_j(\Gamma^s_{ik}) + \sum_r \left(\Gamma^r_{jk}\Gamma^s_{ir} - \Gamma^r_{ik}\Gamma^s_{jr}\right).$$

Similarly we may define the components R_{ijkl} of the Riemannian curvature tensor by the equation

$$R_{ijkl} = \langle R(\partial_i, \partial_i)\partial_k, \partial_l\rangle = \sum_s g_{sl}R^s_{kij}$$

So the sectional curvature for the sphere can be obtained by a lengthy calculation:

$$\mathcal{K}((\partial_1, \partial_2)) = \frac{-R_{1212}}{g_{11}g_{22}} = -\sum_{s=1}^n g_{s2}\left(\partial_1\Gamma^s_{21} - \partial_2\Gamma^s_{11} + \sum_{r=1}^n \left(\Gamma^r_{21}\Gamma^s_{1r} - \Gamma^r_{11}\Gamma^s_{2r}\right)\right)/(g_{11}g_{22})$$

$$= -\left(\partial_1\Gamma^2_{21} - \partial_2\Gamma^2_{11} + \sum_{r=1}^n \left(\Gamma^r_{21}\Gamma^2_{1r} - \Gamma^r_{11}\Gamma^2_{2r}\right)\right)/g_{11}$$

$$= -\left(\partial_1\left(\frac{-2x^1}{\rho^2 + |x|^2}\right) - \partial_2\left(\frac{2x^2}{\rho^2 + |x|^2}\right) + \frac{4(x^3)^2 + \cdots + 4(x^n)^2}{(\rho^2 + |x|^2)^2}\right)/g_{11}$$

$$= -\left(-2\left(\frac{\rho^2 + |x|^2 - 2(x^1)^2}{(\rho^2 + |x|^2)^2}\right) - 2\left(\frac{\rho^2 + |x|^2 - 2(x^2)^2}{(\rho^2 + |x|^2)^2}\right) + \frac{4(x^3)^2 + \cdots 4(x^n)^2}{(\rho^2 + |x|^2)^2}\right)/g_{11}$$

$$= -\left(\frac{-4\rho^2 - 4|x|^2 + 4(x^1)^2 + 4(x^2)^2 + \cdots + (x^n)^2}{(\rho^2 + |x|^2)^2}\right)/g_{11}$$

$$= \frac{4\rho^2}{(\rho^2 + |x|^2)^2}/\left(\frac{4\rho^4}{(\rho^2 + |x|^2)^2}\right) = \frac{1}{\rho^2}$$

Hence the curvature of the sphere is $1/\rho^2$. □

- Hyperbolic space. Consider the upper-half space of $\mathbb{R}^n$ denoted by $\mathbb{R}^n_+$. We identify this space with a unit ball D^n in $\mathbb{R}^n$: Let $e_n = (0, ..., 0, 1)$ be the n-th standard basis vector in $\mathbb{R}^n$. Then $y = x + (\frac{1}{2} - 2x^n)e_n$ takes $x \in \mathbb{R}^n_+$ to the half-space $\{y : y^n < 1/2\}$. Then we map this space to D^n by

$$z = e_n + (y - e_n)/|y - e_n|^2.$$

This gives a diffeomorphism of $\mathbb{R}^n_+$ to D^n. We now define the Riemannian metric $ds^2 := \frac{4|dz|^2}{(1-|z|^2)^2}$ on D^n. Next we have to derive the induced Riemannian metric on $\mathbb{R}^n_+$. So some computations are necessary:

$$dz = \frac{dy}{|y-e_n]^2} - \frac{2((y-e_n)\cdot dy)(y-e_n)}{|y-e_n|^4} \Rightarrow |dz|^2 = \frac{|dy|^2}{|y-e_n|^4}$$

$$1-|z|^2 = 1-\left(1+2\frac{e_n\cdot(y-e_n)}{|y-e_n]^2} + \frac{1}{|y-e_n]^2}\right) = -\frac{1-2y^n}{|y-e_n]^2}$$

These yield

$$ds^2 = \frac{4|dz|^2}{(1-|z|^2)^2} = \frac{4|dy|^2}{(1-2y^n)^2} = \frac{4|dx|^2}{(2x^n)^2} = \frac{|dx|^2}{(x^n)^2}$$

Next we compute the sectional curvature. We work in the half space. From the formula for ds^2 we have $g_{ij} = \frac{\delta_{ij}}{(x^n)^2}$. Then

$$\Gamma^l_{jk} = \frac{1}{2}\sum_i g^{li}(\partial_j g_{ik} + \partial_k g_{ji} - \partial_i g_{jk}) = -\frac{\delta_{jn}\delta_{lk} + \delta_{kn}\delta_{jl} - \delta_{ln}\delta_{jk}}{x^n}$$

Then the sectional curvature is

$$\begin{aligned}
&\mathcal{K}((\partial_1,\partial_2)) = \frac{-R_{1212}}{g_{11}g_{22}} \\
&= -\sum_{s=1}^n g_{s2}\left(\partial_1\Gamma^s_{21} - \partial_2\Gamma^s_{11} + \sum_{r=1}^n(\Gamma^r_{21}\Gamma^s_{1r} - \Gamma^r_{11}\Gamma^s_{2r})\right)/(g_{11}g_{22}) \\
&= -\left(\partial_1\Gamma^2_{21} - \partial_2\Gamma^2_{11} + \sum_{r=1}^n\left(\Gamma^r_{21}\Gamma^2_{1r} - \Gamma^r_{11}\Gamma^2_{2r}\right)\right)/g_{11} \\
&= -\left(\frac{1/(x^n)^2}{1/(x^n)^2}\right) = -1.
\end{aligned}$$

We see that $\mathbb{R}^n_+$, with this induced metric, has a constant curvature -1. It is a hyperbolic space.

5.6. **First integral and Geodesic equation**

The aim of this section is to take a different approach to geodesics. Our main inspiration is from optimal control theory where many things can be (re)cast into optimal control problems. The theory is strongly related to variational calculus. So we are going to derive geodesic equation by the variational calculus. Consider the curve $x = \gamma(t)$ where $\gamma : I \to M$ is smooth and $I = [0,T]$ Let g be the Riemannian metric, then the magnitude of the velocity of a point moving along the curve γ is $v_\gamma(t) = \sqrt{\sum_{ij} g_{ij}(\gamma(t))\dot\gamma^i(t)\dot\gamma^j(t)}$. Following the literature in variational calculus we use the following notations:

$$x = (x^1,...,x^n),\ \dot x = (\dot x^1,...\dot x^n)$$

$$L(x,\dot x) = \sqrt{\sum_{ij} g_{ij}(x)\dot x^i\dot x^j}$$

Then the length of the curve γ is

$$\ell(\gamma) = \int_0^T v_\gamma(t)dt = \int_0^T L(\gamma(t), \dot{\gamma}(t))dt.$$

Note that this integral is also called *the First Fundamental Form*. Using standard technique to get the first variation, we have for small positive $|\epsilon|$

$$\tilde{\ell}(\epsilon) = \ell(\gamma + \epsilon\beta) = \int_0^T L(\gamma(t) + \epsilon\beta(t), \dot{\gamma}(t) + \epsilon\dot{\beta}(t))dt$$

and

$$\tilde{\ell}'(\epsilon) = \frac{d\tilde{\ell}}{d\epsilon}$$

where $\beta : I \to M$ such that $\beta(0) = \beta(T) = 0$. A necessary condition for reaching a minimum is

$$\tilde{\ell}'(0) = 0.$$

That is, by integration by part,

$$0 = \int_0^T \left(\frac{\partial L}{\partial x^i}\beta^i + \frac{\partial L}{\partial \dot{x}^i}\dot{\beta}^i\right)dt = \int_0^T \left(\frac{\partial L}{\partial x^i}\beta^i - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i}\beta^i\right)dt = \int_0^T \left(\frac{\partial L}{\partial x^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i}\right)\beta^i dt.$$

Since this must hold for any arbitrary β_i we get

$$\text{(EL)} \qquad \frac{\partial L}{\partial x^i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} = 0, \qquad \text{for } i = 1, ..., n.$$

This is the well-know Euler-Lagrange equation. In fact it yields the geodesic equation. To see this we have to compute. The procedure is straightforward. First note that $d\ell = Ldt$, So we have the following

$$\begin{aligned}
\frac{d}{dt}\frac{\partial L}{\partial \dot{x}^i} &= \frac{d}{dt}\left(\frac{1}{2L}\left(\sum_{kj} g_{kj}\frac{\partial}{\partial \dot{x}^i}(\dot{x}^k\dot{x}^j)\right)\right) \\
&= \frac{d}{Ldt}\left(\frac{1}{2}\left(\sum_j g_{ij}\dot{x}^j + \sum_k g_{ki}\dot{x}^k)\right)\right) \\
&= \frac{d}{d\ell}\left(\sum_j g_{ij}\dot{x}^j\right) = \frac{d}{d\ell}\left(\sum_j g_{ij}\frac{dx^j}{d\ell}\right)\left(\frac{d\ell}{dt}\right) = \frac{d}{d\ell}\left(\sum_k g_{ik}\frac{dx^k}{d\ell}\right)\left(\frac{d\ell}{dt}\right) \\
&= \left(\sum_k g_{ik}\frac{d^2x^k}{d\ell^2} + \sum_{jk}\frac{\partial g_{ik}}{\partial x^j}\frac{dx^j}{d\ell}\frac{dx^k}{d\ell}\right)\left(\frac{d\ell}{dt}\right) \\
&= \left(\sum_k g_{ik}\frac{d^2x^k}{d\ell^2} + \frac{1}{2}\sum_{jk}\frac{\partial g_{ik}}{\partial x^j}\frac{dx^k}{d\ell}\frac{dx^j}{d\ell} + \frac{1}{2}\sum_{jk}\frac{\partial g_{ij}}{\partial x^k}\frac{dx^k}{d\ell}\frac{dx^j}{d\ell}\right)\left(\frac{d\ell}{dt}\right)
\end{aligned}$$

where we used the symmetry of g. Now we turn to the first term in (EL)

$$\frac{\partial L}{\partial x^i} = \frac{1}{2L}\sum_{kj}\frac{\partial g_{kj}}{\partial x^i}\frac{dx^k}{dt}\frac{x^j}{dt} = \frac{1}{2}\left(\sum_{kj}\frac{\partial g_{kj}}{\partial x^i}\frac{dx^k}{d\ell}\frac{x^j}{d\ell}\right)\left(\frac{d\ell}{dt}\right)$$

Puting above two expressions into (EL) gives

$$\frac{1}{2}\left(\sum_{kj}\frac{\partial g_{kj}}{\partial x^i}\frac{dx^k}{d\ell}\frac{x^j}{d\ell}\right)\left(\frac{d\ell}{dt}\right)=\left(\sum_{k}g_{ik}\frac{d^2x^k}{d\ell^2}+\frac{1}{2}\sum_{jk}\frac{\partial g_{ik}}{\partial x^j}\frac{dx^k}{d\ell}\frac{dx^j}{d\ell}+\frac{1}{2}\sum_{jk}\frac{\partial g_{ij}}{\partial x^k}\frac{dx^k}{d\ell}\frac{dx^j}{d\ell}\right)\left(\frac{d\ell}{dt}\right)$$

or equivalently,

$$\frac{1}{2}\left(\sum_{kj}\frac{\partial g_{kj}}{\partial x^i}\frac{dx^k}{d\ell}\frac{x^j}{d\ell}\right)=\sum_{k}g_{ik}\frac{d^2x^k}{d\ell^2}+\frac{1}{2}\sum_{jk}\frac{\partial g_{ik}}{\partial x^j}\frac{dx^k}{d\ell}\frac{dx^j}{d\ell}+\frac{1}{2}\sum_{jk}\frac{\partial g_{ij}}{\partial x^k}\frac{dx^k}{d\ell}\frac{dx^j}{d\ell}$$

Rearranging and reindexing k to m in the terms of double sum in the last equation yield

$$\sum_{k}g_{ik}\frac{d^2x^k}{d\ell^2}=\frac{1}{2}\sum_{mj}\left(\frac{\partial g_{mj}}{\partial x^i}-\frac{\partial g_{im}}{\partial x^j}-\frac{\partial g_{ij}}{\partial x^m}\right)\frac{dx^m}{d\ell}\frac{dx^j}{d\ell}$$

After apply the inverse of g on the both side of the previous equation and making use of the Christoffel symbol we finally reach the geodesic equation

$$\frac{d^2x^k}{d\ell^2}+\sum_{mj}\Gamma^k_{mj}\frac{dx^m}{d\ell}\frac{dx^j}{d\ell}=0$$

It is our belief that this approach will have some potential, who's motivation will become apparent when determining our solution to the Brachistochrone problem on page 37. It might be possible to approach a geodesic (i.e. the shortest distance) by control theoretic methods. It is a further research topic.

When we did this calculation we found a reference [18] where the similar analysis was done. At the same time when we would like to solve geodesic equations for some manifolds we were lucky to see it in this unexpected paper. The main idea is to use this in color space, however we have not got so far due to the time and limit of this report.

First integrals. Now we restrict in the 2-dimensional manifolds with diagonal metric. Then the geodesic equations are

$$\ddot{x}^1+\frac{\partial_1 g_{11}}{2g_{11}}(\dot{x}^1)^2+\frac{\partial_2 g_{11}}{g_{11}}\dot{x}^1\dot{x}^2-\frac{\partial_1 g_{22}}{2g_{11}}(\dot{x}^2)^2=0$$
$$\ddot{x}^2+\frac{\partial_2 g_{11}}{2g_{22}}(\dot{x}^1)^2-\frac{\partial_1 g_{22}}{g_{22}}\dot{x}^1\dot{x}^2+\frac{\partial_2 g_{22}}{2g_{22}}(\dot{x}^2)^2=0$$

Next we assume that g depends only on x^1, which simplifies the equations further:

$$\ddot{x}^1+\frac{\partial_1 g_{11}}{2g_{11}}(\dot{x}^1)^2-\frac{\partial_1 g_{22}}{2g_{11}}(\dot{x}^2)^2=0$$
$$\ddot{x}^2+\frac{\partial_1 g_{22}}{g_{22}}\dot{x}^1\dot{x}^2=0$$

Later we compute some special metric with $g_{11}=1$ so the equations become

$$\ddot{x}^1-\frac{\partial_1 g_{22}}{2g_{11}}(\dot{x}^2)^2=0$$
$$\ddot{x}^2+\frac{\partial_1 g_{22}}{g_{22}}\dot{x}^1\dot{x}^2=0$$

Now subtracting the second equation multiplied by $\dot{x}^1$ from the first equation by multiplied by $\dot{x}^2$ we get

$$\ddot{x}^1\dot{x}^2 - \ddot{x}^2\dot{x}^1 - \frac{\partial_1 g_{22}}{2}(\dot{x}^2)^3 - \frac{\partial_1 g_{22}}{g_{22}}(\dot{x}^1)^2\dot{y} = 0 \Leftrightarrow \frac{d}{dt}\left(\frac{1}{(g_{22}(x^1(t)))^2}\left(\frac{\dot{x}^1}{\dot{x}^2}\right)^2 + \frac{1}{g_{22}(x^1(t))}\right) = 0$$

In other words we obtain a first integral of the simplified geodesic equations:

$$\frac{1}{g_{22}^2}\left(\frac{dx^1}{dx^2}\right)^2 + \frac{1}{g_{22}} = \text{constant}$$

Computation of geodesics of special metric. Now we compute geodesics on a sphere and a torus.

- *Geodesics on a sphere.* In this case we study the unit sphere. Set $x^1 = \theta$, $x^2 = \varphi$ and $g = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix}$. Then the geodesic equations are

$$\ddot{\theta} - \sin\theta\cos\theta(\dot{\varphi})^2 = 0$$
$$\ddot{\varphi} + 2\frac{\cos\theta}{\sin\theta}\dot{\varphi}\dot{\theta} = 0$$

  and its first integral is

$$\frac{1}{\sin^4\theta}\left(\frac{d\theta}{d\varphi}\right)^2 + \frac{1}{\sin^2\theta} = \frac{1}{\sin\theta_0}$$

  where the constant is determined by $\theta_0 = \min\theta$ and φ_0 which are the coordinates of the "furthest north" point on the curve.

  This is the equation of a circle lying on the plane through the origin. To see this we note that such a plane has the equation

$$x \cdot N = 0$$

  where and N is the normal vector of the plane

$$x = \begin{pmatrix} \sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{pmatrix}, \quad N = \begin{pmatrix} \sin(\pi/2-\theta_0)\cos(\varphi_0+\pi) \\ \sin(\pi/2-\theta_0)\sin(\varphi_0+\pi) \\ \cos(\pi/2-\theta_0) \end{pmatrix}$$

  Since N is the normal of the plane the dot product $x \cdot N = 0$ should give he relation between φ and θ:

$$\begin{aligned} 0 &= \sin\theta\cos\varphi\sin(\pi/2-\theta_0)\cos(\varphi_0+\pi) + \sin\theta\sin\varphi\sin(\pi/2-\theta_0)\sin(\varphi_0+\pi) \\ &\quad + \cos\theta\cos(\pi/2-\theta_0) \\ &= -\sin\theta\cos\varphi\cos\theta_0\cos\varphi_0 - \sin\theta\sin\varphi\cos\theta_0\cos\varphi_0 + \cos\theta\sin\theta_0 \\ &= -\sin\theta\cos\theta_0\cos(\varphi-\varphi_0) + \cos\theta\sin\theta_0 \end{aligned}$$

$$\Rightarrow \varphi = \varphi_0 + \arccos(\tan\theta_0\cot\theta)$$

  Differentiating φ yields

$$\frac{d\varphi}{d\theta} = \frac{1}{\sin^2\theta}\frac{\tan\theta_0}{\sqrt{1-(\tan\theta_0\cot\theta)^2}}$$

  that is,

$$\frac{1}{\sin^4\theta}\left(\frac{d\theta}{d\varphi}\right)^2 + \frac{1}{\sin^2\theta} = \frac{1}{\sin\theta_0}$$

agreeing with the first integral derived above. Therefore, the geodesic is on the great circle on a sphere. See also the argument in the section of geodesic.

- *Geodesics on a torus* For a torus considered being embedded into $\mathbb{R}^3$ we use the usual coordinates: $x^1 = u, x^2 = v$ and the metric $g = \begin{pmatrix} 1 & 0 \\ 0 & (a+\cos u)^2 \end{pmatrix}$ with $a > 1$. The geodesic equations are

$$\ddot{u} + (a + \cos u)\sin u(\dot{v})^2 = 0$$

$$\ddot{v} - 2\frac{\sin u}{a + \cos u}\dot{u}\dot{v} = 0$$

Apparently there are some special cases for solutions

(1) If $\dot{v} = 0$ then we have solutions $u = c_1 t + c_2$ $v = c_3$ with c_1, c_2, c_3 real constants. Figure 10 to the left. There is a geodesic on the symmetric side.
(2) If $u = 0$ we have $v = c_1 t + c_2$. Figure 10 below in the middle.
(3) If $u = \pi$ we have $v = c_1 t + c_2$. Figure 10 below to the right.

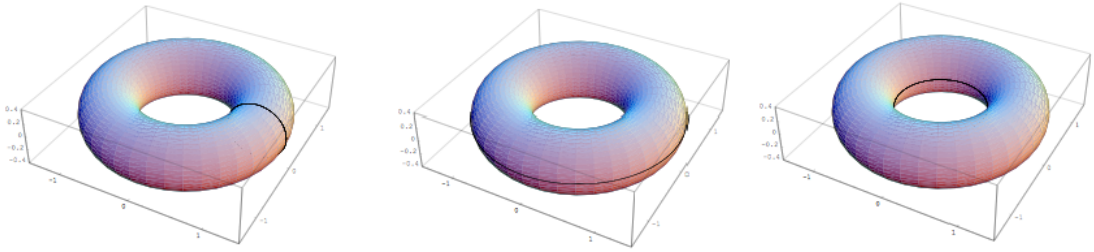


FIGURE 10. Some geodesics on a torus

The first integral for torus is

$$\frac{1}{(a+\cos u)^4}\left(\frac{du}{dv}\right)^2 + \frac{1}{(a+\cos u)^2} = \frac{1}{(a-1)^2}$$

Basically we can get any geodesic from any point on this invariant curve. Due to the limitation of the space we leave the detailed analysis for future case study.

The brachistochrone problem. The Brachistochrone is the curve γ for a ramp along which an object can slide from the rest at a given point (x_1, y_1) to another given point (x_2, y_2) in minimum time. Since the speed of the sliding object is equal to $\sqrt{2gy}$, where g is the gravity of Earth and y is measured vertically downwards from the release point, the differential time it takes the object to traverse the arc $d\ell$ at that speed is $d\ell/\sqrt{2gy}$. The total traveling time T is given by

$$T = \int_{x_1}^{x_2} \frac{d\ell}{\sqrt{2gy}}$$

subject to $\gamma(x_1) = y_1, \gamma(x_2) = y_2$.

Note that the brachistochrone problem is a type of problem of the form: Minimizing

$$k\int_{x_1}^{x_2} y^\alpha \sqrt{1 + (y')^2}dx \quad (k \text{ is a constant})$$

subject to $\gamma(x_1) = y_1, \gamma(x_2) = y_2$.

The special property of these problems is that the Lagrange function V does not explicitly depend on x, which yields $\frac{\partial V}{\partial x} = 0$. So by the chain rule we have

$$\frac{dV}{dx} = y'\frac{\partial V}{\partial y} + y''\frac{\partial V}{\partial y'} + \frac{\partial V}{\partial x} = y'\frac{\partial V}{\partial y} + y''\frac{\partial V}{\partial y'}.$$

Thus

$$\frac{dV}{dx} - (y'\frac{\partial V}{\partial y} + y''\frac{\partial V}{\partial y'}) = 0.$$

Using the Euler-Lagrange equation (EL), we obtain

$$\frac{dV}{dx} - \left(y'\frac{d}{dx}\frac{\partial V}{\partial y'} + y''\frac{\partial V}{\partial y'}\right) = 0.$$

But then the last two terms can be rewritten as

$$\frac{dV}{dx} - \frac{d}{dx}\left(y'\frac{\partial V}{\partial y'}\right) = \frac{d}{dx}\left(V - y'\frac{\partial V}{\partial y}\right) = 0$$

Hence

$$V - y'\frac{\partial V}{\partial y'} = C \tag{BI}$$

for some constant C. This is indeed a first integral.

Using (BI) it is easy to obtain the following differential equation

$$y^\alpha\sqrt{1+(y')^2} - \frac{y^\alpha(y')^2}{\sqrt{1+(y')^2}} = C.$$

It can be simplified to

$$\frac{y^\alpha}{\sqrt{1+(y')^2}} = C \quad \Leftrightarrow \quad \frac{y^{2\alpha}}{1+(y')^2} = \frac{y^{2\alpha}dx^2}{dx^2+dy^2} = C^2. \tag{BIs}$$

So the solution of the brachistochrone problem is the solution of the differential equation

$$\frac{1}{y\sqrt{1+(y')^2}} = C^2 \quad \Leftrightarrow \quad \frac{y^{-1}dx^2}{dx^2+dy^2} = C^2.$$

This is the first integral for shortest curve of the brachistochrone problem. In summary we get, $C^2 = 1/2r$, with the parametric solution (a cycloid)

$$\begin{aligned} x &= r(t-\sin t) \\ y &= r(1-\cos t) \end{aligned} \quad \text{and} \quad \begin{aligned} dx &= r(1-\cos t) \\ dy &= r\sin t \end{aligned}$$

It is worth noting that the brachistochrone problem can be solved by optimal control theory. To this end we rewrite the problem by introducing a control variable $u = y'(t)$. Then the problem is converted to

$$\min \int_{x_1}^{x^2} \frac{\sqrt{1+u^2}}{\sqrt{2gy}}dx$$

subject to $y' = u,\ y(x_1) = y_1, y(x_2) = y_2$.

Now we can use either dynamic programming or Pontryagin's minimum principle to solve this problem. For a historical exposition of relations between the brachistochrone problem and modern control theory we refer to [21] and that Johann Bernoulli almost invented modern control theory about 300 years earlier. This is a reason for our statement made earlier.

Furthermore, we point out that (BIs) brings us some interesting facts, e.g. we can "guess" a solution which in turn determines the corresponding metric. This might have

some practical use. Here are some special cases: (i) when $\alpha = 0$, $x = t$, the parametric solution $y = at + b$, $dx = 1$, $dy = a$ and $C^2 = 1/(1 + a^2)$, So this is a straight line in the Euclidean metric $d\ell^2 = dx^2 + dy^2 = (1 + (y')^2)dx$; (ii) when $\alpha = 1$ the parametric solution is $x = t$, $y = \cosh(t)$, and $dx = 1$, $dy = \sinh(t)$, $C^2 = 1$. So the curve is shortest in the metric of the form $yd\ell$. This is in fact the so-called *Catenary problem.* It considers a chain hanging from two given points from which we want to minimize the total potential energy of the chain; (iii) when $\alpha = -1$, the solution is $x = r\cos(t)$,$y = r\sin(t)$, $dx = -r\sin(t)$, $dy = r\cos(t)$, $C^2 = 1$. So it is a shortest curve in the hyperbolic metric $d\ell/y$; (iv) if $\alpha = 1/2$ we get a solution $y = x^2/4 + 1$. So we can say that for a metric of the form $\sqrt{y}d\ell$ the parabola is the appropriate geodesic. This is an other reason for our earlier statement.

5.7. **Calculations with moving frames**

In this section we review some of previous calculations from the perspective of moving frames due to Cartan (e.g. [4]). *Frame fields* $E_1, ..., E_n$, which are orthonormal vector fields at each point $p \in T_p(M)$ and form a basis of $T_p(M)$. For ξ, η in a real vector space E and α, β in its dual space E^*, the wedge product will be given by

$$(\alpha \wedge \beta)(\xi, \eta) = \alpha(\xi)\beta(\eta) - \alpha(\eta)\beta(\xi).$$

For a smooth manifold M smooth vector fields X Y and a differentiable 1-form ω on M will be given by

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y])$$

Now, let M have a connection ∇ with associated torsion T and curvature tensors R. Then we define the *covariant differential of* 1-form as

$$(\nabla_X\omega)(Y) = X(\omega(Y)) - \omega(\nabla_X Y) \tag{C1}$$

Thus we have

$$d\omega(X, Y) = (\nabla_X\omega)(Y) - (\nabla_Y\omega)(X) - \omega(T(X, Y)).$$

Pick up any tangent vector ξ in TU ((U, φ) is a chart), we can express $\nabla_\xi e_k$ in terms of the basis $\{e_1, ..., e_n\}$:

$$\nabla_\xi e_j = \sum_k \omega_j^k(\xi)e_k.$$

Since $\nabla_\xi e_j$ is linear in ξ, the collection $\{\omega_j^k : j, k = 1, ..., n\}$ forms a matrix of smooth 1-forms, (which can be referred to as the connection 1-forms). Then for any $\xi \in TU$, we have, using (C1)

$$0 = \xi(\omega^l(e_j)) = (\nabla_\xi\omega^l)(e_j) + \omega^l(\nabla_\xi e_j) = (\nabla_\xi\omega^l)(e_j) + \omega_j^l(\xi).$$

Hence,

$$\nabla_\xi\omega^l = \sum_j (\nabla_\xi\omega^l)(e_j)\omega^j = -\sum_j (\omega_j^l(\xi))\omega^j.$$

We computer further, using the previous equalities

$$\begin{aligned}
&d\omega^j(X, Y) - \sum_k (\omega^k \wedge \omega_k^j)(X, Y) \\
=&X(\omega^j(Y)) - Y(\omega^j(X)) - \omega^j([X, Y]) - \sum_k \left(\omega^k(X)\omega_k^j(Y) - \omega^k(Y)\omega_k^j(X)\right) \\
=&X(\omega^j(Y)) + \sum_k \omega^k(Y)\omega_k^j(X) - Y(\omega^j(X)) - \sum_k \omega^k(X)\omega_k^j(Y) - \omega^j([X, Y])
\end{aligned}$$

$$=X(\omega^j(Y)) - (\nabla_X\omega^j)(Y) - Y(\omega^j(X)) + (\nabla_Y\omega^j)(X) - \omega^j([X,Y])$$
$$=\omega^j(\nabla_X Y - \nabla_Y X - [X,Y]) = -\omega^j(T(X,Y)).$$

then we have the following so-called structure form

$$d\omega^j = \sum_k \omega^k \wedge \omega_k^j - \omega^j(T).$$

Here we have to think of T as a 2-form with values in the tangent bundle, and therefore $\omega^j(T)$ is a 2-form on U. Now we calculate the second connection. Consider

$$d\omega_j^k(X,Y) = X(\omega_j^k(Y)) - Y(\omega_j^k(X)) - \omega_j^k([X,Y]).$$

Again we compute the connection:

$$\nabla_Y\nabla_X e_j = \nabla_Y \sum_k \omega_j^k(X)e_k = \sum_k (Y(\omega_j^k(X))e_k + \omega_j^k(X)\nabla_Y e_k)$$

$$=\sum_k \left(Y(\omega_j^k(X)) + \sum_l \omega_j^l(X)\omega_l^k(Y) \right) e_k$$

Using this equality we have

$$\nabla_Y\nabla_X e_j - \nabla_X\nabla_Y e_j - \nabla_{[Y,X]}e_j$$
$$=\sum_k \left(Y(\omega_j^k(X)) - X(\omega_j^k(Y)) + \sum_l \omega_j^l(X)\omega_l^k(Y) - \sum_l \omega_j^l(Y)\omega_l^k(X) - \omega_j^k([Y,X]) \right) e_k$$
$$=\sum_k \left(Y(\omega_j^k(X)) - X(\omega_j^k(Y)) + \sum_l (\omega_j^l \wedge \omega_l^k)(X,Y) - \omega_j^k([Y,X]) \right) e_k$$
$$=\sum_k \left(d\omega_j^k(Y,X) + \sum_l (\omega_j^l \wedge \omega_l^k)(X,Y) - \omega_j^k([Y,X]) \right) e_k$$

Hence, we conclude that if the 2-form Ω_j^k is defined by

$$\Omega_j^k(X,Y) = \omega^k(R(X,Y)e_j)$$

then

$$d\omega_j^k = \sum_l \omega_j^l \wedge \omega_l^k - \Omega_j^k$$

Finally we summarize these results on a Riemannian manifold with the frame $\{e_1, ..., e_n\}$ being orthonormal at every point of U and dual co-frame $\{\omega^1, ..., \omega^n\}$.

Theorem 5.6 (The Cartan structure equations)**.** *The first and the second Cartan structure equations are*

$$d\omega_j = \sum_k \omega^k \wedge \omega_k^j$$

and

$$d\omega_j^k = \sum_l \omega_j^l \wedge \omega_l^k - \Omega_j^k$$

where

$$\omega_j^k(\xi) = \langle \nabla_\xi e_j, e_k \rangle, \quad \omega_j^k = -\omega_k^l,$$
$$\Omega_j^k(X,Y) = \langle R(X,Y)e_j, e_k \rangle, \quad \Omega_j^k = -\Omega_k^j$$

6. **Some applications in color science and image processing**

Now we turn to some applications where Riemannian geometry is extensively used. First we give a very brief introduction on terminologies in image processing. Image Processing treats signals such as photographs, video, or tomographic output. In particular, Computer Graphics consists of image synthesis from some abstract models, while Computer Vision extracts some abstract information: 3-dimensional description (3D) of a scene from video footage of it. From about 2000, analog image processing (by optical devices) gave way to digital processing, and, in particular, digital image editing (for example, processing of images taken by popular digital cameras). Computer graphics (and our brains) deals with vector graphics images, i.e., those represented geometrically by curves, polygons, etc. A *raster graphics image* (or *digital image, bitmap*) in 2D is a representation of a 2D image as a finite set of digital values, called pixels placed on a square grid $\mathbb{Z}^2$ or a hexagonal grid. Typically, the image raster is a square $2^k \times 2^k$ grid with $k = 8, 9$ or 10.

The gray-scale images can be seen as point-weighted binary images. In the gray-scale images, xyi-representation is used, where plane coordinates (x, y) indicate shape, while the weight i (standing for intensity, i.e., brightness) indicates texture. The brightness histogram of a gray-scale image provides the frequency of each brightness value found in that image. If an image has m brightness levels (bins of gray-scale), then there are 2^m different possible intensities. Usually, $m = 8$ and numbers $0, 1, ..., 255$ represent the intensity range from black to white; other typical values are $m = 10, 12, 14, 16$. Humans can differ between around 10 million different colors but between only 30 different gray-levels; so, color has much higher discriminatory power.

For color images, (RGB)-representation is the better known, where space coordinates R, G, B indicate red, green and blue levels; a 3D histogram provides brightness at each point. Among many other 3D color models (spaces) are: (CMY) cube (Cyan, Magenta, Yellow colors), (HSL) cone (Hue-color type given as an angle, Saturation in %, Luminosity in %), and (YUV), (YIQ) used, respectively, in PAL, NTSC television. CIE-approved conversion of (RGB) into luminance (luminosity) of gray-level is 0.299R + 0.587G + 0.114B . The color histogram is a feature vector with components representing either the total number of pixels, or the percentage of pixels of a given color in the image.

Images are often represented by feature vectors, including color histograms, color moments, textures, shape descriptors, etc. Examples of feature spaces are: raw intensity (pixel values), edges (boundaries, contours, surfaces), salient features (corners, line intersections, points of high curvature), and statistical features (moment invariants, centroids). Typical video features are in terms of overlapping frames and motions.

Image Retrieval (similarity search), consists of (as for other data: audio recordings, DNA sequences, text documents, time-series, etc.) finding images whose features have values either mutual similarity, or similarity to a given query or in a given range. There are two methods to compare images directly: intensity-based (color and texture histograms), and geometry-based (shape representations by medial axis, skeletons, etc.). The imprecise term shape is used for the extent (silhouette) of the object, for its local geometry or geometrical pattern (conspicuous geometric details, points, curves, etc.), or for that pattern modulo a similarity transformation group (translations, rotations, and scalings). The imprecise term texture means all that is left after color and shape have been considered, or it is defined via structure and randomness.

Next we describe color distance and color spaces for further study of Riemannian metric in color space and image processing. We pick up two subjects to show how Riemannian

geometry is used: (i) Riemannian approach to compare the CIELAB and CIELUV, based on [15, 16]. (ii) Geodesic distance in shape and surface processing based on [17].

6.1. **Color distance**

The visible spectrum of a typical human eye is about $380 - 760$ nm. It matches the range of wavelengths sustaining photosynthesis. In addition, at those wavelengths opacity often coincides with impenetrability. A light-adapted eye has its maximum sensitivity at ≈ 555 nm (540 THz), in the green region of the optical spectrum. A color space is a 3-parameter description of colors. The need for exactly three parameters comes from the existence of three kinds of receptors (cells on the retina) in the human eye: for short, middle and long wavelengths, corresponding to blue, green, and red. In fact, their respective sensitivity peaks are situated around 570 nm, 543 nm and 442 nm, while wavelength limits of extreme violet and red are about 700 nm and 390 nm, respectively. Some women are tetrachromats, i.e., they have a 4-th type of color receptor. The zebrafish Danio rerio is sensitive to red, green, blue, and ultraviolet light. Color blindness is 10 times more common in males. See [30] for more detailed presentation.

The CIE (International Commission on Illumination) derived $(X\ Y\ Z)$ color space in 1931 from the (RGB)-model and measurements of the human eye. In the CIE $(X\ Y\ Z)$ color space, the values X, Y and Z are also roughly red, green and blue. The basic assumption of Colorimetry ([7]), is that the perceptual color space admits a metric, the true color distance. This metric is expected to be locally Euclidean, i.e., a Riemannian metric. Another assumption is that there is a continuous mapping from the metric space of light stimuli to this metric space. Such a uniform color scale, where equal distances in the color space correspond to equal differences in color, is not obtained yet and existing color distances are various approximations of it. A first step in this direction was given by MacAdam ellipses, i.e., regions on a chromaticity (x, y) diagram which contains all col- ors looking indistinguishable to the average human eye; JND (just-noticeable difference, see below) video quality metric. Here $x = \frac{X}{X+Y+Z}$ and $y = \frac{Y}{X+Y+Z}$ are projective coordinates, and the colors of the chromaticity diagram occupy a region of the real projective plane. See also examples in §3.2.

The CIE (L^*, a^*, b^*) (CIELAB) is an adaptation of CIE 1931 $(X\ Y\ Z)$ color space. It gives a partial linearization of the MacAdam color metric. The parameters L^*, a^*, b^* of the most complete model are derived from L, a, b which are the luminance L, of the color from black $L = 0$ to white $L = 100$, its position a between green $a < 0$ and red $a > 0$ and its position b between green $b < 0$ and yellow $b > 0$.

6.2. **Riemannian color space**

In this section we use Riemannian geometry to study color spaces. The topic belongs to color science and have been used widely in color industry and in image processing. It has a long history dating back to Riemann, Helmholtz, Schrödinger, Stiles in early times. Historically Riemann himself described from his differential geometry that colors and the positions of objects of sense constitute a non-Euclidean manifold. On the basis of Riemannian geometry Helmholtz presented a line element to describe how distance in a color space specifies pairs of color stimuli that give a particular constant perceptual difference in the RGB color space. This quantitative mathematical analysis has progressed further in the color vision by Schrödinger. See [22] for a more detailed overview on color spaces in interaction with Riemannian geometry.

In a color space as a Riemannian space, small color differences can be measured locally as a term of color distance between one point and others in its neighborhood. Such color difference distances represent threshold of color differences. They are described by ellipsoids in three dimensions and ellipses in two dimensions. These representation of color differences can be mapped from one color space to another color space. thanks to the isometry between Riemannian spaces provided that the Gaussian curvatures at the corresponding points must be the same [19]. The isometric property is very important in investigation of the performance of various color spaces and color difference metrics for measuring the perpetual color differences.

The line element is the first fundament form in a color space viewed as a Riemannian space where perceptual or visual color differences are represented by color vectors. The application of the line element is to compute the shortest distance, i.e. the geodesic, between any two points in such a color space from the Riemannian metric.

Now we describe the mathematical form of the line element for the CIE chromaticity color space xy, Y, as a three dimensional Riemannian space. Let p_1 and p_2 be two points in this color space with coordinates (x, y, Y) and $(x + dx, y + dy, Y + dY)$. The distance $d\ell$ between these two points can be expressed by the Riemannian metric, the quadratic form:

$$d\ell^2 = \begin{pmatrix} dx & dy & dY \end{pmatrix} (g_{ij}) \begin{pmatrix} dx & dy & dY \end{pmatrix}^T .$$

As $d\ell$ is constant the Riemannian metric gives an ellipsoid at a color center. In two dimensional case the metric represents the chromaticity difference of any two colors measured along the geodesic of the surface. It gives the chromaticity discrimination ellipses in terms of chromaticity coordinates (x, y) at constant luminance Y. See Figure 5.

Often a surface M is parametrized embedded in Euclidean space $\mathbb{R}^n$ by the mapping

$$u \in E \subset \mathbb{R}^2 \mapsto \varphi(u) \in M$$

Connected with the line element is the geodesic. Some common line elements are

- $d\ell^2 = dr^2 + r^2 d\theta^2$ in plane polar coordinate system (r, θ) with metric $g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}$
- $d\ell^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\varphi^2$ in spherical polar coordinate system (r, θ, φ) with metric $g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2\theta \end{pmatrix}$
- $d\ell^2 = dr^2 + r^2 d\theta^2 + dz^2$ in cylindrical polar coordinate system (r, θ, z) with metric $g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

There are many line element models in color science. We recommend the book [10] for historical development of these models. Here we shall focus on the color metrics proposed by Helmholtz, Luneburg, Stiles and Resnikoff based on the results of [19] which also contains an extensive historical introduction.

Resnikoff in [19] formulated various standard experimental results as axioms delimiting the geometry of the set $\mathscr{P}$ of perceived lights.

Axiom 1. $\mathscr{P}$ is a cone in $\mathbb{R}^3$, that is, if $x \in \mathscr{P}$ and $\alpha > 0$ then $\alpha x \in \mathscr{P}$. In other words, every positive multiple of a perceived light is a perceived light

Axiom 2. $\mathscr{P}$ does not contain any 1 dimensional vector space, i.e. if $x \in \mathscr{P}$ then there is no $y \in \mathscr{P}$ such that $x + y \in \mathscr{P}$. This says that no superposition of perceived lights produces the absence of perceived light.

Axiom 3. (Grassmann, Helmholtz) $\mathscr{P}$ is a convex cone, that is, for any $x, y \in \mathscr{P}$ and $\alpha \in [0, 1]$ we have $\alpha x + (1 - \alpha)y \in \mathscr{P}$.

Axiom 4. Any four perceived lights are linearly dependent, i.e. if $x_k \in \mathscr{P}$, $k = 1, 2, 3, 4$, then there are $\alpha_k \in \mathbb{R}$ not all zero such that $\sum_{k=1}^{4} \alpha_k x_k = 0$. This implies that the vector space $\mathscr{V}$ in which the cone lies is of dimension less than or equal to 3. $\dim \mathscr{V}$ is a characteristic of each observer. Those observers for whom the dimension equal to 3,2,1 and 0 are respectively called *Trichromate, dichromate, monochromate* and *blind*.

These four axioms provide the affine structure of the set $\mathscr{P}$, which is essentially the exposition of Schrödinger.

Next denote by $GL(\mathscr{P})$ the group of orientation preserving linear transformations of $\mathscr{V}$ which preserves the cone $\mathscr{P}$ of perceived colors. The element in $GL(\mathscr{P})$ is called *change of background illumination*.

Axiom 5. $\mathscr{P}$ is locally homogeneous with respect to changes of background illumination. which was shown to be equivalent to

Axiom 5′. $\mathscr{P}$ is (globally) homogeneous with respect to changes of background illumination.

Roughly speaking, Resnikoff [19] showed, using the fact that $GL(\mathscr{P})$ is a Lie-group and the five Axioms, that the only such GL-homogeneous cones $\mathscr{P}$ (i.e., the group of all orientation preserving linear transformations of $\mathbb{R}^3$), carrying $\mathscr{P}$ into itself, acts transitively on $\mathscr{P}$ are either $\mathscr{P}_1 = \mathbb{R}_{>0} \times (\mathbb{R}_{>0} \times \mathbb{R}_{>0})$ or $\mathscr{P}_2 = \mathbb{R}_{>0} \times \mathscr{C}$ where $\mathscr{C} = SL(2, \mathbb{R})/SO(2)$ is the set of 2×2 real symmetric matrices with determinant 1. The first factor $\mathbb{R}_{>0}$ can be identified with variation of brightness and the other with the set of lights of a fixed brightness.

Finally it was postulated:

Axiom 6. The Riemannian metric on $\mathscr{P}$ which measures perceived dissimilarity is a $GL(\mathscr{P})$-invariant metric.

It was shown that Axiom 6 determines the perceptual metric:

$$d\ell^2 = \alpha_1 \left(\frac{dx_1}{x_1}\right)^2 + \alpha_2 \left(\frac{dx_2}{x_2}\right)^2 + \alpha_3 \left(\frac{dx_3}{x_3}\right)^2$$

if $\mathscr{P}_1 = \mathbb{R}_{>0} \times (\mathbb{R}_{>0} \times \mathbb{R}_{>0})$ where $\alpha_i > 0$, $i = 1, 2, 3$. This is the line element proposed by Stiles, which covers a special case where $\alpha_i = 1$ for $i = 1, 2, 3$ which is the Helmholtz line element. The line element defined on $\mathscr{P}_2 = \{(x, u) : x \in \mathbb{R}_{>0}, u \in \mathscr{C}\}$ is

$$d\ell^2 = \alpha_1 \left(\frac{dx_1}{x_1}\right)^2 + \alpha_2 d\ell^2_{\mathscr{C}}$$

where $d\ell^2_{\mathscr{C}} = \frac{|du|^2}{(1-|u|^2)^2}$ is the Poincaré metric on $\mathscr{C}$, $\alpha_1, \alpha_2 > 0$. This is called Resnikoff metric. So $\mathscr{P}_2$ with this metric is not isometric to a Euclidean space.

Now we outline how the metrics above can be derived. For a complete proof see [19]. Let $x \in \mathscr{P}$ be fixed, $T_x(\mathscr{P})$ be the tangent space at $x \in \mathscr{P}$ and G_x be the metric on $T_x(\mathscr{P})$ induced by the given Riemannian metric on $\mathscr{P}$. Due to the fact that $GL(\mathscr{P})$ is transitive on $\mathscr{P}$, the metric is determined everywhere by the metric G_x on $T_x(\mathscr{P})$.

Next we identify x with the coset K in the representation of $\mathscr{P}$ as the homogeneous space $GL(\mathscr{P})/K$. Now apparently $gx = x$ if $g \in K$ and therefore the metric G_x on the

tangent space $T_x(\mathscr{P})$ must be K-invariant: $G_x(dgX) = G_x(X)$ for $g \in K$ and $X \in T_x(\mathscr{P})$, dg is the differential of g. Now we have two cases.

Case 1. $\mathscr{P} = \mathbb{R}_{>0} \times SL(2,\mathbb{R})/SO(2)$. In this case the tangent space $T_x(\mathscr{P}) = \mathbb{R} \oplus T'_x(\mathscr{P})$, where $T'_x(\mathscr{P})$ is the two-dimensional subspace of $T_x(\mathscr{P})$ which is a tangent space to $SL(2,\mathbb{R})/SO(2)$ at $K = SO(2)$. The restriction on the metric to $T'_x(\mathscr{P})$ must be $SO(2)$-invariant, i.e. it is invariant with respect to rotation about the origin in a two-dimensional vector space $T'_x(\mathscr{P})$. Hence it is a multiple of the Euclidean metric. It follows that $G_x = G_{x,1} + G_{x,2}$, where $G_{x,i}$ $(i = 1, 2)$ is the one-dimensional respectively two-dimensional Euclidean metric. This implies that the perceptual metric on $\mathscr{P}$ is unique up to a selection of the units of measure on each of factors of $\mathscr{P} = \mathbb{R}_{>0} \times SL(2,\mathbb{R})/SO(2)$.

Now we try to get an explicit description of this metric. Let $\mathscr{P}$ denote the set of all 2×2 real symmetric positive definite matrices x and $\det(x)$ denote the determinant of x, $\mathrm{tr}(x)$ the trace of x, and $U = \{x \in \mathscr{P} : \det(x) = 1\} \subset \mathscr{P}$. Then we can decompose $x = \sqrt{\det(x)}\left(\frac{x}{\sqrt{\det(x)}}\right)$, this shows that $\mathscr{P} = \mathbb{R}_{>0} \times U$. It is well-know that U is isomorphic to $SL(2,\mathbb{R})/SO(2)$. Further we look at how $\mathbb{R}_{>0} \times SL(2,\mathbb{R}) = GL(2,\mathbb{R})$ acts on $\mathscr{P}$. Take $A \in GL(2,\mathbb{R})$. Then $x \in \mathscr{P}$ is mapped to AxA^t, where A^t is the transpose of A. That is, the group action is a congruent transformation of two matrices. Then

$$d\ell^2 = \mathrm{tr}(x^{-1}dx\ x^{-1}dx)$$

defines a $GL(2,\mathbb{R})$-invariant metric of the perceptual metric on $\mathscr{P}$. To see the $GL(2,\mathbb{R})$-invariance we simply calculate as follows: for $x \in \mathscr{P}$, we have $gx = AxA^T$, and then $x^{-1} = (A^T)^{-1}x^{-1}A^{-1}$, $d(gx) = A(dx)A^T$. Thus

$$\mathrm{tr}((gx)^{-1}d(gx)(gx)^{-1}d(gx)) = \mathrm{tr}((A^T)^{-1}x^{-1}(dx)x^{-1}(dx)A^T) = \mathrm{tr}(x^{-1}dx\ x^{-1}dx)$$

Here in the last equality we used the trace property.

Our final step is to make it explicit. If $\mathscr{P}$ is represented as the set of 2×2 symmetric positive definite matrix $x = \begin{pmatrix} x_1 & x_3 \\ x_3 & x_2 \end{pmatrix}$, then an explicit representation which exhibits the product structure of $\mathscr{P}$ is provided by the new coordinatization (ξ, u) where $u \in SL(2,\mathbb{R})/SO(2)$, as

$$\xi = \sqrt{\det(x)}, u_1 = x_3/x_2, u_2 = \sqrt{\det(x)}/x_2 > 0$$

corresponding to the decomposition of the matrix x as

$$x = \xi \begin{pmatrix} \frac{u_1^2+u_2^2}{u_2} & \frac{u_1}{u_2} \\ \frac{u_1}{u_2} & \frac{1}{u_2} \end{pmatrix}$$

In these coordinates the normalized perceptual metric

$$d\ell^2 = \mathrm{tr}(x^{-1}dx\ x^{-1}dx) = \left(\frac{d\xi}{\xi}\right)^2 + \left(\frac{(du_1)^2 + (du_2)^2}{u_2^2}\right)$$

the general metric is apparently given by

$$d\ell^2 = \alpha_1\left(\frac{d\xi}{\xi}\right)^2 + \alpha_2\left(\frac{(du_1)^2 + (du_2)^2}{u_2^2}\right)$$

where $\alpha_1, \alpha_2 > 0$ are constants. This shows also that in this case perception of the ξ-variable is independent of perception of the variables $u \in SL(2,\mathbb{R})/SO(2)$.

Case 2. $\mathscr{P} = \mathbb{R}_{>0} \times \mathbb{R}_{>0} \times \mathbb{R}_{>0}$. In this case $K = \emptyset$, and K-invariance makes no restriction on the metric. However a $G(\mathscr{P})$-invariant metric must be the sam of metrics on each factor which are $\mathbb{R}_{>0}$-invariant. Since an $\mathbb{R}_{>0}$-invariant metric on $\mathbb{R}_{>0}$ is determined by a positive constant on the tangent space at one point, it is clear that all $\mathbb{R}_{>0}$-invariant metrics on $\mathbb{R}_{>0}$ are proportional. But $d\ell^2 = (dx)^2/x$ is an $\mathbb{R}_{>0}$-invariant metric on $\mathbb{R}_{>0}$. Hence on $\mathscr{P}$ the general $GL(\mathscr{P})$-invariant metric is

$$d\ell^2 = \alpha_1 \left(\frac{dx_1}{x_1}\right)^2 + \alpha_2 \left(\frac{dx_2}{x_2}\right)^2 + \alpha_3 \left(\frac{dx_3}{x_3}\right)^2$$

where all αs are positive constants.

6.3. **Riemannian formulation of color difference formulas**

With respect to color differences, the main objective of color difference formulas is to give quantitative color difference value ΔE that should represent the visual color difference perceived by the human visual system (color difference obtained from the psychophysical experiment). Many color difference formulas like the ΔE^*_{ab} and the ΔE_{00} of the CIELAB space, the ΔE^*_{uv} of the CIELUV space, the ΔE_E of the log compressed OSA-UCS space and so on have been developed to fulfill such an objective. Unfortunately, all these formulas so far developed do not have perfect uniform color spaces. It means they are unable to to measure the visual perception of color differences sufficiently. Theoretically in a perfect uniform color space, the color matching ellipses should become circles.

It has been popular to study various color difference formulas by Riemannian approach. This approach makes it possible to evaluate the performance of various color difference formulas having different color spaces for measuring visual color difference. Here we introduce some of these research work.

We start with a general presentation then do one example to illustrate.

JND ellipsoid. Considering the 2D color space as the Riemannian space, an ellipse whose length is equal to the arc length of a curve between two points is expressed by a line element as introduced before. It is a differential quadratic form. The metric g_{ij} gives intrinsic properties of the color a geometric surface. it represents the chromaticity difference of any two colors measured along the geodesic of the surface. In the study of color vision, MacAdam ellipses refer to the region on a chromaticity diagram which contains all colors which are indistinguishable, to the average human eye, from the color at the center of the ellipse. The contour of the ellipse therefore represents the *just noticeable differences* (commonly JND in the literature) of chromaticity. This is completely determined by the metric g_{ij}. If $d\ell^2 = 1$ then we have an ellipse. The semi-major-axis and the semi-minor-axis are equal to $1/\sqrt{\lambda_1}$ and $1/\sqrt{\lambda_2}$, respectively, where $\lambda_1 \leq \lambda_2$ are the eigenvalues of (g_{ij}). The corresponding orthonormal eigenvectors form the axes in the new coordinate system where which the ellipse is symmetric to. Similarly we have an ellipsoid in 3D.

Riemannian formulation of color difference. Now we demonstrate how to formulate color difference in terms of Riemannian metric. Consider the color difference in the CIELAB color space defined as the Euclidean distance

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$$

The CIBLAB color space defined for moderate to high lightness is given by

$$L^* = 116 \left(\frac{Y}{Y_r}\right)^{\frac{1}{3}} - 16$$

$$a^* = 500\left[\left(\frac{X}{X_r}\right)^{\frac{1}{3}} - \left(\frac{Y}{Y_r}\right)^{\frac{1}{3}}\right]$$
$$b^* = 200\left[\left(\frac{Y}{Y_r}\right)^{\frac{1}{3}} - \left(\frac{Z}{Z_r}\right)^{\frac{1}{3}}\right]$$

where L^*, a^* and b^* correspond to the Lightness, the redness-greenness and the yellowness-blueness scales in the CIELAB color space. Similarly, X, Y, Z and X_r, Y_r, Z_r are the tristimulus values of the color stimuli and white reference, respectively.

The relation between the tristimulus coordinates X, Y, Z and the color coordinates x, y and Y are

$$X = \frac{xY}{y}, Y = Y, Z = \frac{(1-x-y)Y}{y}$$

If the line element distance is meant to measure the infinitesimal color difference at a point in the color space, we will get the differential quadratic form written in matrix form

$$(dE^*_{ab})^2 = \begin{pmatrix} dL^* & da^* & db^* \end{pmatrix} \begin{pmatrix} dL^* \\ da^* \\ db^* \end{pmatrix}$$

We call this color differential form. Clearly we can transform differentials dL^*, da^*, db^* to dX, dY, dZ. This can be done by computing the differential:

$$\begin{pmatrix} dL^* \\ da^* \\ db^* \end{pmatrix} = \begin{pmatrix} \frac{\partial L^*}{\partial X} & \frac{\partial L^*}{\partial Y} & \frac{\partial L^*}{\partial Z} \\ \frac{\partial a^*}{\partial X} & \frac{\partial a^*}{\partial Y} & \frac{\partial a^*}{\partial Z} \\ \frac{\partial b^*}{\partial Z} & \frac{\partial b^*}{\partial Z} & \frac{\partial b^*}{\partial Z} \end{pmatrix} \begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix}$$

$$= \begin{pmatrix} 0 & \frac{116}{3Y_r^{1/3}}\left(\frac{Y}{Y_r}\right)^{-\frac{2}{3}} & 0 \\ \frac{500}{3X_r^{1/3}}\left(\frac{X}{X_r}\right)^{-\frac{2}{3}} & -\frac{500}{3Y_r^{1/3}}\left(\frac{Y}{Y_r}\right)^{-\frac{2}{3}} & 0 \\ 0 & \frac{200}{3Y_r^{1/3}}\left(\frac{Y}{Y_r}\right)^{-\frac{2}{3}} & -\frac{200}{3Z_r^{1/3}}\left(\frac{Z}{YZr}\right)^{-\frac{2}{3}} \end{pmatrix} \begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix}$$

Denote by

$$\frac{\partial(L^*, a^*, b^*)}{\partial(X, Y, Z)} = \begin{pmatrix} \frac{\partial L^*}{\partial X} & \frac{\partial L^*}{\partial Y} & \frac{\partial L^*}{\partial Z} \\ \frac{\partial a^*}{\partial X} & \frac{\partial a^*}{\partial Y} & \frac{\partial a^*}{\partial Z} \\ \frac{\partial b^*}{\partial Z} & \frac{\partial b^*}{\partial Z} & \frac{\partial b^*}{\partial Z} \end{pmatrix}$$

This matrix is called the Jacobi matrix. Thus

$$(dE^*_{ab})^2 = \begin{pmatrix} dX & dY & dZ \end{pmatrix} \left(\frac{\partial(L^*, a^*, b^*)}{\partial(X, Y, Z)}\right)^T \frac{\partial(L^*, a^*, b^*)}{\partial(X, Y, Z)} \begin{pmatrix} dX \\ dY \\ dZ \end{pmatrix}$$

This makes comparison of color difference in different color space.

Basically it is the same calculation for all kinds of color space transformation. However, to make real comparison we need numerical computation.

6.4. **Geodesic distance and geodesic methods for shape and surface processing**

The main focus will be of the shape.

In image processing the manifold is the image domain $M = [0,1]^2$ equipped with a metric derived from the image (for example its gradient). Here is a list of frequently used metric in image processing

- Euclidean space: $\mathbb{R}^n$ and $g_{ij} = \mathrm{id}_n$
- 2D shape: $M \subset \mathbb{R}^2$ and $g_{ij} = \mathrm{id}_2$
- Isotropic metric: $g_{ij} = W(x)\mathrm{Id}_n$, $W(x) > 0$ being some weight function.
- Parametric surface: g_{ij} the first fundamental form.
- Image processing: given an image $\mathcal{I} : [0,1]^2 \to \mathbb{R}$ one can use edge-stopping weight $W(x) = (\epsilon + ||\nabla \mathcal{I}||)^{-1}$. This way, geodesic curve can be used to perform segmentation since they will not cross boundaries of the objects

Geodesic distance and geodesic curves. The local Riemannian metric g_{ij} allows to define a global metric on the space M using shortest paths. This corresponds to the notion of geodesic curves.

Definition. Given a Riemannian manifold (M, g) with $M \subset \mathbb{R}^n$, the *geodesic distance* is defined as

$$\forall (x,y) \in M \times M,\ d_M(x,y) := \min_{\gamma \in \mathcal{P}(x,y)} \ell(\gamma)$$

where $\mathcal{P}(x,y)$ is the set of piecewise smooth curves joining x and y

$$\mathcal{P}(x,y) := \{\gamma : \gamma(0) = x, \gamma(1) = y\}$$

Definition. A *geodesic curve* $\gamma \in \mathcal{P}$ is a smooth curve such that $\ell(\gamma) = d_M(x,y)$

A geodesic curve between two points might not be unique, think for instance about two antipodal points on a sphere. In order to perform the numerical computation of geodesic distances, we fix a set of starting points $\mathcal{S} = (x_k)_k \subset M$ and consider only distance and geodesic curves from this set of points.

Shape matching based on Eccentricity transform. A shape is a planar, connected, bounded and closed set $S \subset \mathbb{R}^2$ with a piecewise smooth boundary ∂S. In practice we consider a discretized version of S which can be represented using an image f_S of n pixels where f_S is the indicator of the shape:

$$f_S(x) = \begin{cases} 1 & \text{for } x \in S \\ 0 & \text{otherwise} \end{cases}$$

The computation of the distance function $U(x) := d_S(x_0, x)$ to some point $x_0 \in S$ can be computed efficiently as the solution of the nonlinear Eikonal equation:

$$\forall x \in S,\ \ ||\mathrm{grad} U|| = 1,\ \text{and } U(0) = 0$$

The Eikonal equation can be solved in $O(n \log(n))$ operations for a grid of n points using the Fast Marching Algorithm which is e.g. implemented in MATLAB.

The eccentricity transform of a shape S, assigns to each point $p \in S$ the shortest geodesic distance to the point of S farthest away from it. The eccentricity of the shape S is defined as

$$\mathrm{ECC}_S(x) := \max_{y \in S} d_S(x,y) = \max_{y \in \partial S} d_S(x,y)$$

This is a continuous and piecewise smooth function. A point y that reaches the global maximum in above equation is called *eccentric*. Denote by $\mathcal{E}(S)$ the set of eccentric points.

As pointed out in [9] $\mathcal{E}(S)$ is included in ∂S. The set of eccentric points allows to define a segmentation of S into eccentric regions

$$S = \cup_{x \in \mathcal{E}(S)} A_x \quad \text{where} \quad A_s = \{y : \mathrm{EEC}_S(y) = d_S(y, x)\}$$

The eccentricity is computed by performing a Fast Marching propagation from each $x \in \partial S$ n order to compute the set of distances $\{d(x, y)\}_{x \in \partial S}$. The set of values $\{\mathrm{ECC}_S(x)\}_{x \in S}$ is invariant under rigid motion and isometric transform of S, which includes bending. It is also nearly-invariant under articulations, see [9].

Another property of ECC_S is that it is robust to salt and pepper noise that might create holes in S, because a hole or a segmentation error of size ε only modifies the eccentricity ECC_S by no more than ε. It is very different from local descriptors such as the curvature or even global ones such as the structure of the skeleton which are not robust to this kind of noise.

In order to match two shapes from two binary images we first create a shape descriptor for each of them and then match these descriptors to obtain a similarity measure. There are many ways to do it here are two of those.

Mono-scale descriptor. The basic building block for this type of shape descriptor is the histogram h_S of the eccentricity transform ECCS of the shape S. m bins are used to estimate the histogram and in numerical applications. The histogram descriptor is then the vector $h_S \in \mathbb{R}^m$ defined by

$$\forall i = 1, ..., m, h_S(i) = \frac{1}{|S|} \# \left\{ x \in S : \frac{i+1}{m} \leq \frac{\mathrm{ECC}_S(x) - \min(\mathrm{ECC}_S)}{\max(\mathrm{ECC}_S) - \min(\mathrm{ECC}_S)} < \frac{i}{m} \right\}$$

Multiscale descriptor. In order to capture more geometric information about a shape S, one computes a nonlinear scale-space of S and extract the histograms of the eccentricity over a scale-space domain.

In order to smooth the shape, we perform the following mean curvature evolution

$$\frac{\partial \gamma_t}{\partial t}(u) = \kappa_t(u) \vec{n}_t(u) \quad \text{with} \quad \gamma_0(u) = \partial S(u)$$

where $\kappa_t(u)$ is the curvature of the curve $\gamma_t(u)$ and $\vec{n}_t(u)$ is the normal vector to the curve. The curve γ_S is thus a smoothed version of the boundary ∂S after a diffusion during time t.

6.5. **On curvature in color Spaces**

In [2] the Cartan structure equations were used to study color, hue and hue curvature. The authors use them to construct a local model for the behavior of the color, which in turn specifies consistency constraints between nearby color measurements. These constraints are then used to replace noisy pixels by examining their spatial context. We'll describe the idea of this approach in this section.

The color space consider here is the HSV color space, where a color image is a mapping $\mathcal{C} : \mathbb{R}^2 \to S^1 \times [0, 1]^2$ where S^1 is the unit circle. The hue component across the image is a mapping $\mathcal{H} : \mathbb{R}^2 \to S^1$. Thus the hue component can be represented as a unit vector over the image plane. So it is called the *hue field*. Attach a frame field $\{\mathcal{H}_T, \mathcal{H}_N\}$ to each point in the image domain. Clearly it represents the hue vector, and also it gives a local coordinate system in which all other vectors can be represented in a natural, object centered view. Next we do differentiation (i.e. compute the covariant derivatives) of $\mathcal{H}_T$ and $\mathcal{H}_N$. These covariant derivatives represent the intial rate of change of the frame in

any given direction $\vec{v}$. The computation can be done by the Cartan structure equation derived earlier. In this particular case, it takes the following simple form

$$\begin{pmatrix} \nabla_{\vec{v}}\mathcal{H}_T \\ \nabla_{\vec{v}}\mathcal{H}_N \end{pmatrix} = \begin{pmatrix} 0 & \omega_1^2(\vec{v}) \\ -\omega_1^2(\vec{v}) & 0 \end{pmatrix} \begin{pmatrix} \mathcal{H}_T \\ \mathcal{H}_N \end{pmatrix}$$

since the ω_i^j is skew-symmetric the non-diagonal entries in the matrix are zero. The coefficient $\omega_1^2(\vec{v})$ is a function of the tangent vector $\vec{v}$. It represents the local behavior of the flow depends on the direction along which it is measured. By the derivation of the Cartan structure equations we know that this is a 1-form and thus linear. This fact makes it possible to represent $\omega_1^2(\vec{v})$ easily in this frame field $\{\mathcal{H}_1, \mathcal{H}_2\}$

$$\omega_1^2(\vec{v}) = \omega_1^2(a\mathcal{H}_1 + b\mathcal{H}_2) = a\omega_1^2(\mathcal{H}_1) + b\omega_1^2(\mathcal{H}_2)$$

Since we have freedom to choose basis we represent it by the $\{\mathcal{H}_T, \mathcal{H}_N\}$. This gives us two scalars:

$$\kappa_T := \omega_1^2(\mathcal{H}_T), \kappa_N := \omega_1^2(\mathcal{H}_N)$$

They are called the *hue's tangential curvature* and *the hue's normal curvature* respectively. They represent the rate of change of hue in the tangential and normal directions respectively. Since differentiation of $\mathcal{H}$ is a 1-form we can write

$$\kappa_T = \text{grad}\mathcal{H} \cdot \mathcal{H}_T = \text{grad}\mathcal{H} \cdot (\cos\mathcal{H}, \sin\mathcal{H})$$
$$\kappa_N = \text{grad}\mathcal{H} \cdot \mathcal{H}_N = \text{grad}\mathcal{H} \cdot (-\sin\mathcal{H}, \cos\mathcal{H})$$

Note that $\mathcal{H}_T$ and $\mathcal{H}_N$ are rigidly coupled, the two curvatures can be rewritten in terms of the hue field ($\mathcal{H}_T$) using the standard operators, curl $\nabla\times$ and divergence ($\nabla\cdot$):

$$\kappa_T = ||\text{curl } \mathcal{H}_T||, \kappa_N = \text{div } \mathcal{H}_T.$$

That means we can solve a PDE for $\mathcal{H}(q)$.

In general we can show: *Given any hue field* $\{\mathcal{H}_1, \mathcal{H}_2\}$ *, its curvature functions* κ_T *and* κ_N *must satisfy*

$$\text{grad } \kappa_T \cdot \mathcal{H}_N - \text{grad } \kappa_N \cdot \mathcal{H}_Y = \kappa_T^2 + \kappa_N^2.$$

This observation has an important implication: Unless the hue function is constant, at lest one of its curvatures must vary, or the two curvatures need to covary in any neighborhood of the color image. Now we give a proof of this proposition.

First note that

$$\kappa_T^2 + \kappa_N^2 = ||\text{grad } \mathcal{H}||^2.$$

Now

$$\begin{matrix} \kappa_T = \text{grad}\mathcal{H} \cdot (\cos\mathcal{H}, \sin\mathcal{H}) \\ \kappa_N = \text{grad}\mathcal{H} \cdot (-\sin\mathcal{H}, \cos\mathcal{H}) \end{matrix} \quad \Leftrightarrow \quad \begin{matrix} \mathcal{H}_{\mathcal{H}_T} = \kappa_T \cos\mathcal{H} - \kappa_N \sin\mathcal{H} \\ \mathcal{H}_{\mathcal{H}_N} = \kappa_T \sin\mathcal{H} + \kappa_N \cos\mathcal{H} \end{matrix}$$

Using the fact that grad $\times$ grad$\mathcal{H} = \mathcal{H}_{\mathcal{H}_T\mathcal{H}_N} - \mathcal{H}_{\mathcal{H}_N\mathcal{H}_T} = 0$. Hence Subtracting the second equation differentiated with respect to $\mathcal{H}_T$ from the fist equation differentiated with respect to $\mathcal{H}_N$ yields

$$\kappa_{T_{\mathcal{H}_N}} \cos\mathcal{H} - \kappa_{N_{\mathcal{H}_N}} \sin\mathcal{H} - \mathcal{H}_{\mathcal{H}_N}^2 = \kappa_{T_{\mathcal{H}_T}} \sin\mathcal{H} + \kappa_{N_{\mathcal{H}_T}} \sin\mathcal{H} + \mathcal{H}_{\mathcal{H}_T}^2 \Leftrightarrow$$

$$\text{grad } \kappa_T \cdot \mathcal{H}_N - \text{grad } \kappa_N \cdot \mathcal{H}_Y = \mathcal{H}_{\mathcal{H}_T}^2 + \mathcal{H}_{\mathcal{H}_N}^2 = \kappa_T^2 + \kappa_N^2.$$

So the observation is proved. □

After the geometry is settled then it will be easier to build up models. In this paper ([2])

the author provided a model for hue coherence. Based on theory for minimal surfaces they derived a model in the neighborhood of q

$$\mathcal{H}(x,y) = \arctan \frac{\kappa_T(q)x + \kappa_N(q)y}{1 + \kappa_T(q)x - \kappa_N(q)y}$$

for further analysis and experiment.

7. Modeling the Cells in the Lens of an Eye and Automating the Quantification of a protein

7.1. Description of the project

During January 2015 I had the opportunity to do a two-week internship at CBA, at the University of Uppsala. The project I had been assigned was to create a program that would count the epithelial cells in the lens and compute the intensity of the protein caspase-3 in microscopy images provided to the CBA by the department of Ophthalmology of Uppsala University. I wrote the code in Matlab.

The code is part of one of the Gullstrand Lab of Ophthalmology's projects. Among other things, they study the effect of UV radiation on the cells of the lens. One way to do this is to study the expression of the protein caspase-3, when it is active it plays an essential role in the nuclear changes in apoptotic cells (programmed cell death). So far they already have models for the spacial distribution of the active caspase-3 in both normal and exposed lenses as well as a model for the time evolution after exposure to the UV radiation.

The experiments had been conducted on 40 Sprague Dawley rats. One of the rat's eye was first exposed to a total dose of $1\text{kJ} \cdot m^{-2}$ (1.1W for 15 min.) of UVR 300nm. The other eye was kept hidden to serve as a comparison. Three samples were taken from each eye. Further information and more detailed explanations can be found in [23, 24] and [25].

To begin, research has been conducted on the spatial distribution of caspase-3 in normal healthy lens epithelium ([23]). It was made clear that there is more active caspase-3 in the anterior pole of the lens than at the lens equator where there is no active caspase-3 (Figure 11). This shows that apoptosis occurs normally in the central zone of the lens and depletes further away. The possible explanation for this is that the central zone is more exposed to light which affects the cells.

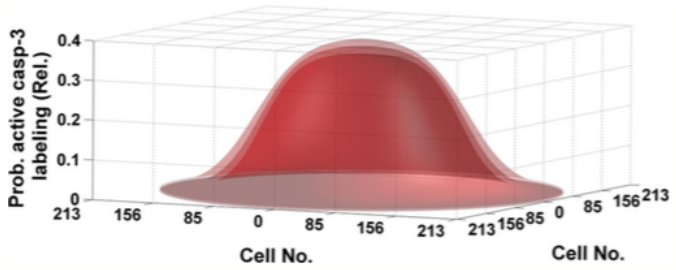


FIGURE 11. Spatial distribution of active caspase-3 activity in the lens epithelium

They have also studied the time evolution of active caspase-3 labelling after in vivo exposure to UVR-300nm ([24]). The rats were randomly divided into 4 post exposure interval groups (0.5, 8, 16 and 24h). The first result was that the caspase-3 expression was higher in the exposed than in the contralateral non exposed eyes (Figure 12).

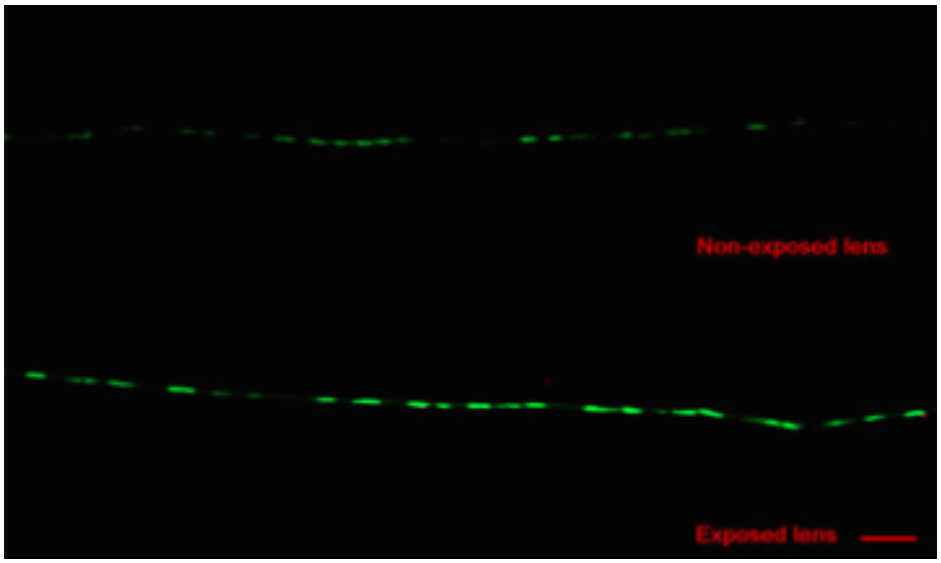


FIGURE 12. Light microscopy of lens epithelium after in vivo exposure to 1kJ$\cdot m^{-2}$ UVR-B in the exposed and contralateral non exposed lens. Scale bar is 25 μm.

The time evolution of the protein expression was also analyzed. The difference between the exposed and non exposed lenses showed a transient maximum in the time interval 8-16 hours. This was later modeled in [25] which we can see in (Figure 13).

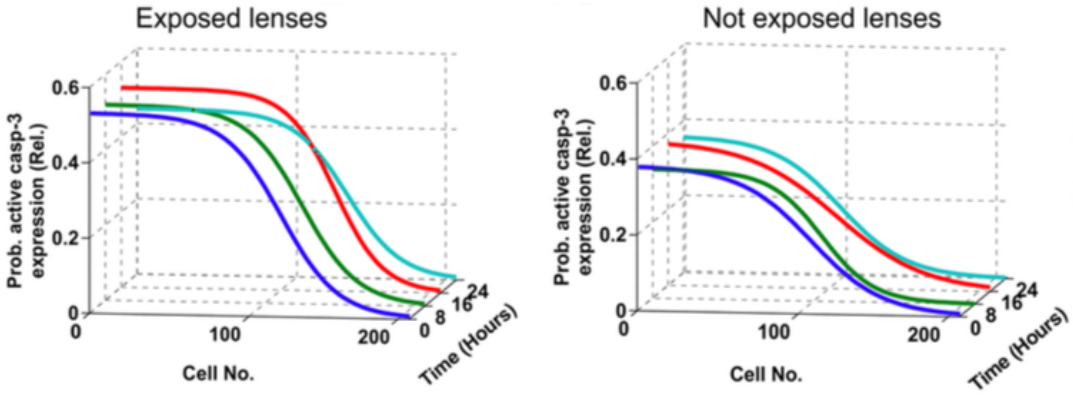


FIGURE 13. Probability of active caspase-3 expression after in vivo exposure to 1kJ$\cdot m^{-2}$ UVR-B in the exposed and non exposed lenses as a function of cell number and time after exposure.

It was concluded that the active caspase-3 expression increases after in vivo exposure to UVR-300nm. It begins 30 min after the exposure, reaches a peak after 8-16 hours and has decreased by 24 hours after the exposure, which is consistent with the last phase of apoptosis (the cell death).

However all the previous research had been done manually, a long and tedious work. In order to accelerate future research they asked the CBA to create a code that would extract all the necessary data from the image.

7.2. **Realization of the project with Matlab**

The cell nuclei were stained with DAPI. The caspase-3 was stained with fluorescein using immunohistochemistry, one of two classical methods to find proteins in cells. (The other being Westernblotting)

I worked with microscopy images taken with a fluorescent microscopy (AxioCamHR – Universal Microscope Axioplan 2 Imaging; Carl Zeiss, Thornwood, NY, USA). They had taken the images with two filters, one blue (matched to the DAPI-stained cells) and one green (matched to the fluorescein-stained caspase-3 - FITC). At least 10 images (size 1300·1030) were needed to reconstruct the whole lens.

The researchers at Gullstrand Lab of Ophthalmology wanted to retrieve specific information from of the images such as the number of cells, their general shape and size, and the expression of caspase-3 in and around the cells. In order to develop such a code I read about the usual methods, especially the one described in [31].

In Matlab if you want to keep the color of the images they can be translated in to RGB-images (as explained before in the RGB color space). You can separate the image into three channels: red, green and blue (all the other colors of the spectrum are in fact a mixture of these three base colors). A large part of this report describes the mathematics behind some of the steps of the code.

When staining the samples the DNA nuclei are colored blue and the caspase-3 is colored green. It is then easy to separate the channels and get a grayscale image of the DNA nuclei (called DAPI-image) and another one of only the protein (FITC-image). Since only the proteins are visible in the FITC image we can get the expression of the protein simply by computing the intensity in each pixel of the image. The DAPI-image is a reference; it is in this image that we find the cell outlines.

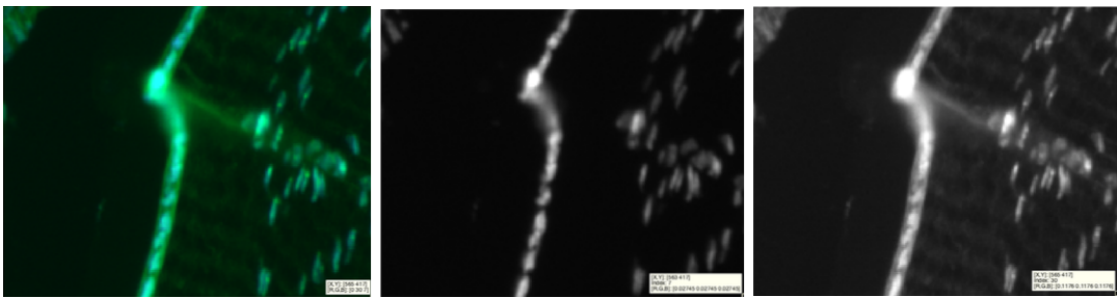


FIGURE 14. Small part of the original image next to the DAPI–image (blue channel) and the FITC–image (green channel)

The idea is to segment the DAPI image to have the location and the shape of the cells. Then this segmentation will be used as a reference for the FITC image in order to compute the intensity in the defined regions.

Extracting the region of interest

The images provided by Gullstrand lab of Ophthalmology have a lot more information than we need. In fact only a small part of the image is relevant, the one surrounding the epithelial cells. The samples and thus the image are not always perfect, sometimes the line that made out by the cells breaks off or cells from another section are trapped in the image. That is why it is better if the scientist himself chooses the relevant part of the picture. A way to quickly extract the relevant part of the image is if, at the lab, they draw a red line (in paint for example) along the cells that are interesting. The idea is to do mathematical approximations as late as possible.

For every sample we then have 2 images, one with the red line and the original untouched one. It is important that the two images are of the same size so that it is really the part that we want. We separate the channels on the image with the red line and keep only the red channel. We create a binary image of it. In the code this image is called the "kakform" (Figure 15). It is the Swedish word for cake pan, or cake mold, the idea is to extract the part that is in white on the "kakform"–image from the DAPI and the FITC images.

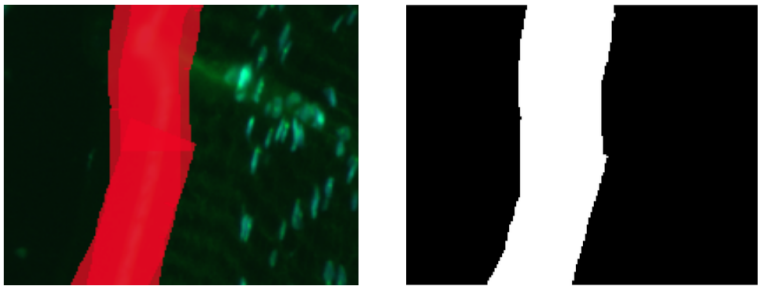


FIGURE 15. The red line along the cells of the image and the black and white "kakform" that will be used repeatedly in the code

Pre-processing and watershed

The first step to segmenting the DAPI image is to smooth it. A simple Gauss filter, that can be manually modified, is enough.

Then we extract the region of interest from the DAPI-image (simply by multiplying the filtered image with the binary "kakform").

The last step is to use a *watershed segmentation* (Figure 17).

The watershed segmentation can be seen as a metaphor. Let's see the image as a landscape, the intensity of each pixel being the height of that point. If a hole is drilled in all minima of the landscape where water is being filled, these minima will become catchment basins. As the water rises, water from neighboring catchment basins will meet. At every point where two catchment basins meet, a dam or a watershed is built. These watersheds are the segmentation of the image.

Since the minimum intensity is 0 we need to invert the image so that the segmentation runs correctly. We also need to settle a threshold so that only the part we are interested in is segmented.

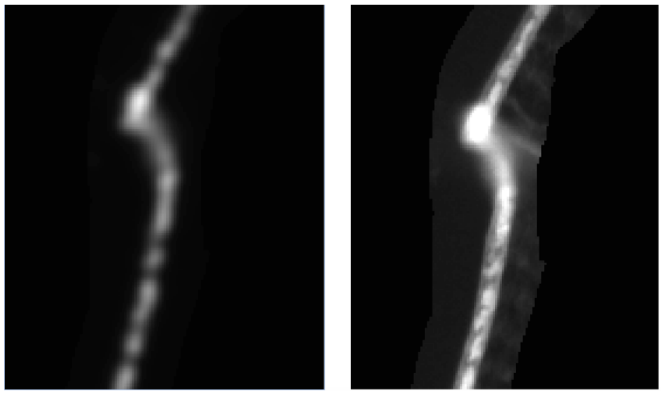


FIGURE 16. The extracted and filtered DAPI-image and the extracted FITC-image

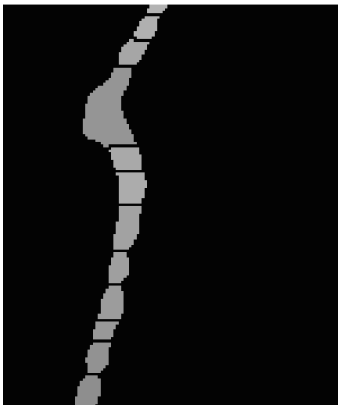


FIGURE 17. Watershed segmentation

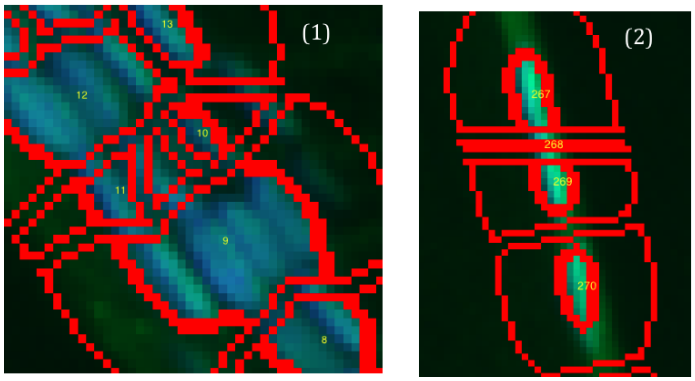


FIGURE 18. Cases of (1) under-segmentation and (2) over-segmentation

When segmenting an image we have to keep an critical eye on the results. Often the image is over- or under-segmented, it depends on the filter used. That is why the Gauss filter used to smooth the DAPI-image can be manually modified, as to find an equilibrium between the over-segmentation and the under-segmentation.

Ordering the cells

Now that we have defined the regions of the cells we want to compute all the data we need from it.

The first step is to order the cells. By default, Matlab orders the regions from left to right. (On a horizontal line this causes no problem, however when the line buckles the cells can be numbered incorrectly.) However, we want to label them along the line. To do this we will call upon a function that computes the geodesic distance (with the metric $\max\{|x_1 - x_2|, |y_1 - y_2|\}$ where the points are numbered $(x_1, y_1), (x_2, y_2)$) along a black and white image, in this case the image with the white line. This function needs a seed location, ideally one of the extremities of the line.

Seed location

One way to choose the seed location is by, on yet another image with the red line, draw a black dot where you want to start counting. By transforming the red channel of that image into a binary image you can easily locate the dot and make it the seed location. However, the images we are working on are very big (around 150MB), so to have 3 images of that size for every sample takes up a lot a space. That is why I had to find another solution.

There are a couple of functions that apply morphological operations on binary images. One of them computes the skeleton of the image, while another one finds the endpoints of that skeleton. Usually the skeleton of a black and white image is not a single line but has many spurious edges. In order to find the real endpoint we must remove these edges (Figure 19). Then it is easy to find the two ends of the line that represent the first and the last cell of the line. To make the analysis of the data as efficient as possible we chose to always start counting the first cell in the bottom left corner.

FIGURE 19. Skeleton and pruned skeleton of the white line

Ordering the cells

When we compute the geodesic distance of the line starting at the left corner we will get a image that gets brighter the further away it gets from the starting point. We can now order the cells the way we want them, this is to say along the line starting in the left corner.

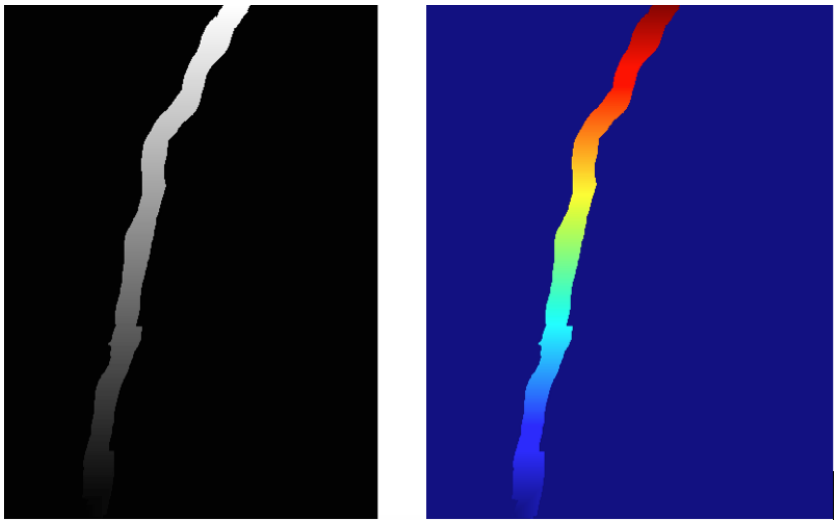


FIGURE 20. The geodesic of the line, in grayscale and in color

For this we use "regionprops" a function that measures different proprieties in the regions of our labelmatrix (the watershed image). By finding the maximum intensity in the geodesic of each corresponding region of the label matrix we can order the regions in an increasing order and thus order the cells along the line as we wanted.

Drawing outlines

The Gullstrand Lab. wanted to know the expression of the protein in the cells, as well as around them. We have already defined the regions of the cells. With a similar method we can define the background corresponding to each cell.

One way to create the background is by doing a watershed segmentation of a black and white version of the label matrix (the segmented image) – see Figure 21 to the left. It is then important to label the background exactly the same way as the label matrix, otherwise the analysis of the data will be faulty.

Now that we have the segmented cells and the corresponding background, it is easy to save the outlines of the cells, as well as the outlines of the area surrounding them – see Figure 21 to the right.

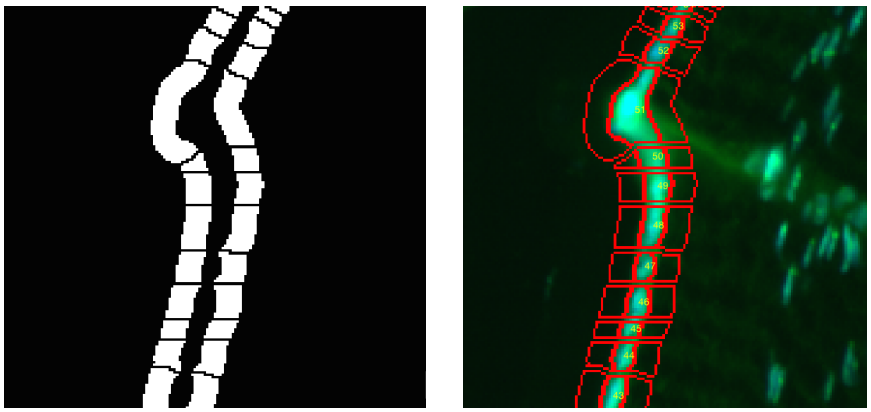
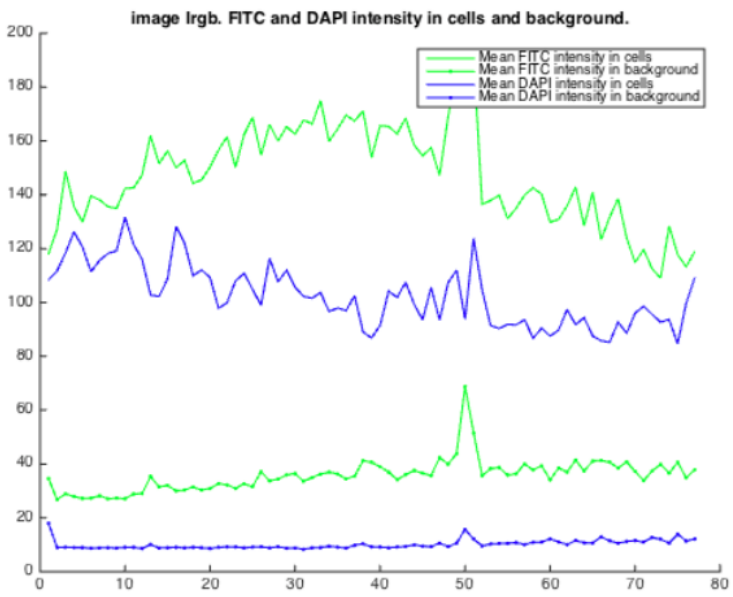


FIGURE 21. The backgrounds (left) and the outlines of the cells and the backgrounds on the original image (right)

Extracting the data

With "regionprops" it is simple to compute the necessary information such as the mean, the maximum and the minimum intensity of the cells and the background in both the DAPI and the FITC images, as well as the area that is being analyzed and the size of the cells - Figure 22. Then these values can be plotted and be sent to a mat-file.



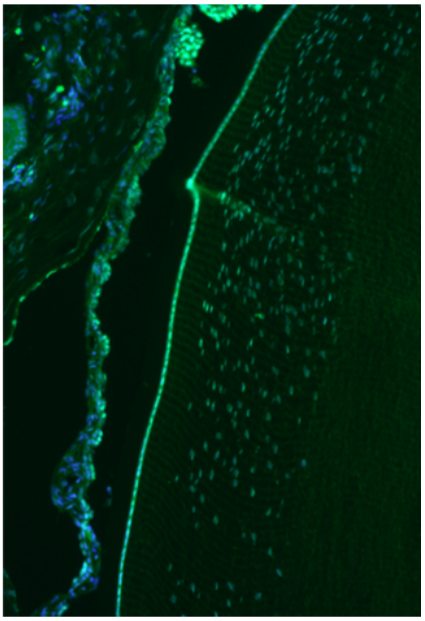


FIGURE 22. Plots of the data extracted out of the image

8. Concluding remarks

In this report we studied Riemannian geometry, in particular, differential calculus on Riemannian manifolds, geodesics and curvature which are related to the applications in color science, image processing etc. In order to understand the mathematics we made an effort to compute some classroom examples and tried to view the underlying mathematical theory from different point of view if possible. Some examples were revisited many times with a belief of deepening the author's understanding.

Since the geodesic distance was our primary driving instrument for this study we made several simplification to study the geodesic equation, which led to the first integral of these geodesics. Then we applied them to sphere and torus. Since many image processing problems are parametric surfaces as the Riemannian space this seems to have some potential for determining the geodesics in these applications by the observation that we could transform one color space to another. The idea is to see if it is possible to transform a color space to some simpler systems where the metric would have some special forms.

We also derived the geodesic equations based on variational calculus. By doing so we had a feeling that other approaches for example optimal control theory, to determine the geodesics would be more efficient. This would too become a further investigation.

To have a real touch on the research areas in color science and image processing we studied some recent research works. It turned out that numerical computation is a very important part of further investigation. Many theoretical sound methods may not feasible in reality. This brings the research to a new dimension.

As for the project carried out at CBA, Uppsala University we have some ideas, purely from the view of mathematical analysis and methodology, that one perhaps could compare some 3D-image with the 2D one we had in the project. Of course this brings out additional difficulties and we are not sure about the partial possibilities.

References

[1] R. Abraham and J.E. Marsden, Foundations of Mechanics, 2ed. Benjamin/Cumming, Reading, Massachusetts, 1978.

[2] O. Ben-Shahar and S.W. Zucker, Hue fields and color curvatures: a perceptual organization approach to color image denoising, In the Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, 2003.

[3] W.M. Boothby, An Introduction to Differential Manifolds and Riemanninan Geometry, 2ed ed. Academic Press, 1986.

[4] É. Cartan, Leçons sur la géométrie des espaces de Riemann, Gauthier-Villares, Paris, 1928.

[5] I. Chavel, Riemannian geometry – a modern introduction, 2nd ed., Cambridge University Press, Cambridge, 2006.

[6] L.I. Cohen, Minimal paths and fast matching methods for image analysis. In Handbook of Mathametical Methods in Computer Vision, N. Paragios and Y. Chen and O. Faugeras Eds, Springer, 2005.

[7] T. Indow, A critical review of Luneburg's model with regard to global structure of visual space, Psychological Review (1991), Vol. 98, No. 3430–45

[8] A.K. Jain, Fundamentals of Digital Image Processing, Prentice Hall, 1989.

[9] W.G.Kropatsch,A.Ion,Y.Haxhimusa,andT.Flanitzer.Theeccentricitytransform(ofadigitalshape). In 13th DGCI, pages 437-448, Hungary, October 2006. Springer.

[10] R.G. Kuehni, Color Space and Its Division: Color Order from Antiquity to the Present, Wiley-Interscience, 2003.

[11] J.M. Lee, Introduction to Smooth Manifolds, 2nd ed, Springer, 2003.

[12] , A.D. Logvinenko, The geometric structure of color, Journal of Vision, vol.15, (2015), 1–9.

[13] J.R. Munkres, Analysis on Manifolds, Advanced Books Classics, Westview Press, 1997.
[14] J.R. Munkres, Topology, 2ed. Pearson, 2000.
[15] D.R. Pant and I. Farup, Evaluating color difference formulae by Riemannian metric, 5th Conference on Colour in Graphics, Images,Joensuu, Finland.
[16] D.R. Pant and I. Farup, Geodesic calculation of color difference formulas and comparison with the munsell color order system, Color Research & Application, Volume 38, pp. 259–266, 2013
[17] G. Peyré and L.I. Cohen, Geodesic Methods for shape and surface processing, Advances in Computational Vision and Medical Image Processing: Methods and Medical Image Processing, Methods and Applications, Springer, 2009.
[18] P. Pokorny ,Geodesics Revisited Chaotic Modeling and Simulation (CMSIM) : 281–298, 2012.
[19] H.L. Resnikoff, Differential geometry and color perception, Journal of Mathematical Biology, 1 (1974), 97–131.
[20] M. Spivak, Calculus on Manifolds, Westview Presse, 1971.
[21] H. Sussmann and J:c: Willems, The brachistochrone problem and modern control theory, In Contemporary trends in non-linear geometric control theory and its applications (proceedings of the conference on "Geometric Control Theory and Applications" held in Mexico City on September 4-6, 2000.
[22] M. Tkalcic, J.Tasic, Colour spaces – perceptual, historical and applicational background, (2003) `http://ldos.fe.uni-lj.si/docs/documents/20030929092037_markot.pdf`
[23] N. Talebizadeh, Z. Yu, M. Kronschläger, F. Halbröök, P. Söderberg, (2014) Specific spatial distribution of caspase-3 in normal lenses. Acta Ophthalmologica
[24] N. Talebizadeh, Z. Yu, M. Kronschläger, F. Halbröök, P. Söderberg, (2014) Time evolution of active caspase-3 labelling after in vivo exposure to UVR-300nm. Acta Ophthalmologica
[25] N. Talebizadeh, Z. Yu, M. Kronschläger, F. Halbröök, P. Söderberg, (2014) Modeling the time evolution of active caspase-3 protein in the rat lens after in vivo exposure to ultraviolet radiation-B. PlosOne
[26] L.W. Tu, An introduction to Manifolds, Springer, 2008.
[27] `http://en.wikipedia.org/wiki/Long_line_%28topology%29`
[28] `http://en.wikipedia.org/wiki/Topological_manifold`
[29] `http://en.wikipedia.org/wiki/Color_space`.
[30] `http://en.wikipedia.org/wiki/Color_vision`.
[31] C. Wählby, I.M. Sintorn, F. Erlandsson, B. Borgefors and E. Bengtsson, Combining intensity, edge, and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. Journal of Microscopy, 215(1):67-76, July 2004.

Appendix – Matlab-code

In this project we use Matlab image processing toolbox to code.

```
%Choose the directory of images we are working on

mydir=uigetdir;
k = dir([mydir,'/','*.tif']);
imagefiles = {k.name}';
number_of_images=length(imagefiles);
for i = 1 : number_of_images;
    [dummy1,dummy2]=strsplit(char(imagefiles(i)),'.tif');
    [dummy1,dummy2]=strsplit(char(dummy1(1)),'_line');
    dummy3(i)=dummy1(1);
end

imagename=unique(dummy3);
number_of_images=length(imagename);

%Separate the channels of the original image.

for i = 1:number_of_images
    current_image=[mydir,'/',char(imagename(i))];
    I=imread([current_image,'.tif']);
    Ig=I(:,:,2);
    Ib=I(:,:,3);
    %Call on the function kakform to get the binary form of the line.
    [Ir_bw]=Kakform(current_image);

    %Create a filter for smoothing the image. It can be modified by the
    %user to improve the segmentation.
    h = fspecial('gaussian', 7 , 3);

    %Segment the DNA image, limited to the region marked by the red line
    [unorderedL]=Watershed_segmentation(h,Ib,Ir_bw);

    %Ordered the segmented cells by the red line (first find seed, then
    %distance from seed)
    %Seed location of the mask to compute the geodesic distance.
    D=Seed_location(current_image,Ir_bw);
    Cells=OrderLabels(unorderedL,D);

    %Find the local background around each cell
    %Specify how many pixels away from the cells the backgrounds should be
    %found in the parameter bg_width, typically about 10 pixels
    bg_width=10;
    Backgrounds=FindBackground(Cells,bg_width);

    %Draw the detected outlines on an input image (color or grayscale)
    [ColorImageWithLines,centroids] = ...
        DrawOutlines(Cells,Backgrounds,I,imagename{i});

```

```
48      %Compute and save the different intensities in and around the cells,
49      %as well as the areas that they are computed on.
50      [FITCCells, FITCBackgrounds, DAPICells, ...
            DAPIBackgrounds]=ExtractData(Cells,Backgrounds,Ig,Ib);
51      save(char(current_image),'imagename','FITCCells', ...
            'FITCBackgrounds', ...
            'DAPICells','DAPIBackgrounds','centroids','ColorImageWithLines');
52  end
53
54
55  %Example of plotting the different values
56  for i = 1:number_of_images
57      current_image=[mydir,'/',char(imagename(i))];
58      load([current_image,'.mat']);
59      figure;
60      hold on;
61      plot(cat(1,FITCCells.MeanIntensity),'g-')
62      plot(cat(1,FITCBackgrounds.MeanIntensity),'g.-')
63      plot(cat(1,DAPICells.MeanIntensity),'b-')
64      plot(cat(1,DAPIBackgrounds.MeanIntensity),'b.-')
65
66      title(['image ',imagename{i},'. FITC and DAPI intensity in cells ...
            and background.'])
67      legend('Mean FITC intensity in cells','Mean FITC intensity in ...
            background', 'Mean DAPI intensity in cells','Mean DAPI intensity ...
            in background')
68  end
```

```
1   function[Ir_bw]=Kakform(imagename);
2   % *Create the "kakform"*
3
4   line=strcat(imagename,'_line','.tif');
5   %Open the image with the red line
6   I_line=imread(line); %title('Image with the line')
7
8   %Extract only the red channel
9   Ir=I_line(:,:,1);
10
11  %Create the binary version of the image
12  Ir_bw=zeros(size(Ir));
13  Ir_bw(Ir>1)=1;
```

```
function[L]=Watershed_segmentation(h,Ib,Ir_bw);
%*Segmentation of the image*

%Erase small noise with an Gauss filter called h, defined in Main
Ifilt=imfilter(Ib,h);
Ifilt(Ifilt<50)=0;

Ib_ext=Ifilt.*uint8(Ir_bw);

%Do a watershed segmentation of the inverse (so that it is on a light
%background) of the image.
Iinv=max(Ib_ext(:))-Ib_ext;
L=watershed(Iinv);

%Create a threshold.
L(Ib_ext<1)=0;
%figure,imshow(L,[]); title('Watershed image')
```

```
function[D]=Seed_location(imagename,Ir_bw);
% *Choosing the seed location*

%Compute the skeleton of the black and white kakform, then the endpoints
%and the branchpoints.
skel= bwmorph(Ir_bw,'skel',Inf);

B = bwmorph(skel, 'branchpoints');
E = bwmorph(skel, 'endpoints');

%Remove the spurious edge of the skeleton.
[y,x] = find(E);
B_loc = find(B);

Dmask = false(size(skel));
for k = 1:numel(x)
    d = bwdistgeodesic(skel,x(k),y(k));
    distanceToBranchPt = min(d(B_loc));
    Dmask(d < distanceToBranchPt) =true;
end
skelD = skel - Dmask;

%Now compute the new endpoints of the skeleton and save them in two array
%vectors x and y.
e=bwmorph(skelD,'endpoints');
[y,x]=find(e);

%Create the mask.
mask=zeros(size(Ir_bw));
mask(y(1,1),x(1,1))=1;
%We choose to always start counting from the first cells in the bottom
%left corner.

%Use bwdistgeodesic to calculate the distance between the centers.
D= bwdistgeodesic(logical(Ir_bw),logical(mask));
D(D==NaN)=0; %Remove unwanted areas.
```

```
function[L]=OrderLabels(unorderedL,D)

%Compute the mean intensity of every region of L
M=regionprops(unorderedL,D,'MaxIntensity');
%M is a n*1 struct where n is the number of cells found in the line.
%Convert into double
D1=cat(1,M.MaxIntensity);

%Create the "order" matrix.
L=zeros(size(unorderedL));

for i=1:length(D1)
    L(unorderedL==i)=D1(i);
end

U=unique(D1);
for i=1:length(U)
    L(L==U(i))=i;
end
```

```
function[Backgrounds]=FindBackground(L,bg_width)

%Create the background by segmenting L.
L_bw=zeros(size(L));
L_bw(L>0)=1; %Create the binary image of the label
map=bwdist(L_bw);%Create a distance map on which to grow backgrounds
map(map>bg_width)=Inf; %Exclude the values that are too high
background_all=watershed(map); %Do a watershed segmentation of the map
background_all(map>bg_width)=0; %Create a threshold

%Make sure background gets the same labelling as the foreground.
BG_numbering=regionprops(background_all,L,'MaxIntensity');
new_numbers=cat(1,BG_numbering.MaxIntensity);

Backgrounds=zeros(size(L));

for i=1:length(new_numbers)
    Backgrounds(background_all==i)=new_numbers(i);
end

Backgrounds=double(Backgrounds)-L;
```

```
function [ColorImageWithLines,centroids] = ...
    DrawOutlines(Cells,Backgrounds,InputImage,imagename)

%Find and combine the edges of the cells and the background : "the ...
    analyzed area around the cells".
normalizedColor = [1,0,0];

Background_edges = bwperim(Backgrounds);
Cell_edges=bwperim(Cells);

all_edges=or(Background_edges,Cell_edges);
if size(InputImage,3)==1
    red = InputImage;
    green = InputImage;
    blue = InputImage;
    red(all_edges) = 255 * normalizedColor(1);
    green(all_edges) = 255 * normalizedColor(2);
    blue(all_edges) = 255 * normalizedColor(3);
    ColorImageWithLines=cat(3,red,green,blue);
end

if size(InputImage,3)==3
    red = InputImage(:,:,1);
    green = InputImage(:,:,2);
    blue = InputImage(:,:,3);
    red(all_edges) = 255 * normalizedColor(1);
    green(all_edges) = 255 * normalizedColor(2);
    blue(all_edges) = 255 * normalizedColor(3);
    ColorImageWithLines=cat(3,red,green,blue);
end

figure;imshow(ColorImageWithLines); hold on;
s  = regionprops(Cells, 'centroid');
centroids = cat(1, s.Centroid);
hold on;
for i=1:length(centroids)
    text(centroids(i,1), centroids(i,2), ...
        num2str(i),'FontSize',14,'Color','yellow')
end

title(['Image ',imagename,'. Outlines of cells and their surrounding ...
    background. Cells numbered based on manually drawn line.'])
```

```
function[FITCCells,FITCBackgrounds,DAPICells,DAPIBackgrounds]=
ExtractData(Cells,Backgrounds,Ig,Ib)

%Compute the mean, max and min intensity in the cells as well as the size
%and the length of the cells.
FITCCells=regionprops(Cells,Ig,'MeanIntensity','MinIntensity','MaxIntensity',
'Area','MajorAxisLength');
FITCBackgrounds=regionprops(Backgrounds,Ig,'MeanIntensity','
MinIntensity','MaxIntensity','Area');
DAPICells=regionprops(Cells,Ib,'MeanIntensity','MinIntensity',
'MaxIntensity','Area','MajorAxisLength');
DAPIBackgrounds=regionprops(Backgrounds,Ib,'MeanIntensity','MinIntensity',
'MaxIntensity');
```