Graduate Texts in Mathematics

D. Bao
S.-S. Chern
Z. Shen

An Introduction to Riemann–Finsler Geometry

Springer

Graduate Texts in Mathematics 200

Springer Science+Business Media, LLC

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces. 2nd ed.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.
61 WHITEHEAD. Elements of Homotopy

(continued after index)

D. Bao
S.-S. Chern
Z. Shen

An Introduction to Riemann–Finsler Geometry

With 20 Illustrations

Springer

D. Bao
Department of Mathematics
University of Houston
University Park
Houston, TX 77204
USA
bao@math.uh.edu

S.-S. Chern
Department of Mathematics
University of California at Berkeley
Berkeley, CA 94720
USA

Z. Shen
Department of Mathematical Sciences
Indiana University–Purdue University Indianapolis
Indianapolis, IN 46202
USA
zshen@math.iupui.edu

Back cover illustration: Courtesy of Mrs. Chern.

Mathematics Subject Classification (1991): 53C60, 58B20

Library of Congress Cataloging-in-Publication Data
Bao, David Dai-Wai.
An introduction to Riemann–Finsler geometry/D. Bao, S.-S. Chern, Z. Shen.
p. cm. — (Graduate texts in mathematics; 200)
Includes bibliographical references and index.
ISBN 978-1-4612-7070-6 ISBN 978-1-4612-1268-3 (eBook)
DOI 10.1007/978-1-4612-1268-3
1. Geometry, Riemannian. 2. Finsler spaces. I. Chern, Shiing-Shen, 1911– II. Shen, Zhongmin, 1963– III. Title. IV. Series.
QA649.B358 2000
516.3'73—dc21 99-049808

Printed on acid-free paper.

Production managed by Frank McGuckin; manufacturing supervised by Erica Bresler.
Camera-ready copy prepared from the authors' AMSTeX and LaTeX files.

9 8 7 6 5 4 3 2 1

ISBN 978-1-4612-7070-6

To Our Teachers
in Life and in Mathematics

Preface

A historical perspective

The subject matter of this book had its genesis in Riemann's 1854 "habilitation" address: "Uber die Hypothesen, welche der Geometrie zu Grunde liegen" (On the Hypotheses, which lie at the Foundations of Geometry). Volume II of Spivak's *Differential Geometry* contains an English translation of this influential lecture, with a commentary by Spivak himself.

Riemann, undoubtedly the greatest mathematician of the 19th century, aimed at introducing the notion of a manifold and its structures. The problem involved great difficulties. But, with hypotheses on the smoothness of the functions in question, the issues can be settled satisfactorily and there is now a complete treatment.

Traditionally, the structure being focused on is the Riemannian metric, which is a quadratic differential form. Put another way, it is a smoothly varying family of inner products, one on each tangent space. The resulting geometry — Riemannian geometry — has undergone tremendous development in this century. Areas in which it has had significant impact include Einstein's theory of general relativity, and global differential geometry.

In the context of Riemann's lecture, this restriction to a quadratic differential form constitutes only a special case. Nevertheless, Riemann saw the great merit of this special case, so much so that he introduced for it the curvature tensor and the notion of sectional curvature. Such was done through a Taylor expansion of the Riemannian metric.

The Riemann curvature tensor plays a major role in a fundamental problem. Namely: how does one decide, in principle, whether two given Riemannian structures differ only by a coordinate transformation? This was solved in 1870, independently by Christoffel and Lipschitz, using different methods and *without* the benefit of tensor calculus. It was almost 50 years later, in 1917, that Levi-Civita introduced his notion of parallelism (equivalent to a connection), thereby giving the solution a simple geometrical interpretation.

Riemann saw the difference between the quadratic case and the general case. However, the latter had no choice but to lay dormant when he remarked that "The study of the metric which is the fourth root of a quartic differential form is quite time-consuming and does not throw new light to the problem." Happily, interest in the general case was revived in 1918 by Paul Finsler's thesis, written under the direction of Carathéodory. For this reason, we refer to the general case as Riemann–Finsler geometry, or Finsler geometry for short.

Finsler geometry is closely related to the calculus of variations. See §1.0. As such its deeper study went back at least to Jacobi and Adolf Kneser. In his Paris address in 1900, Hilbert formulated 23 unsolved problems. The last one was devoted to the geometry of the calculus of variations. It is the only problem for which he did not formulate a specific question/conjecture. Hilbert gave praise to Kneser's book, then new. He provided an account of the invariant integral, and emphasized the importance of the problem of multiple integrals. The Hilbert invariant integral plays an important role in all modern treatments of the subject.

The geometrical data in Finsler geometry consists of a smoothly varying family of Minkowski norms (one on each tangent space), rather than a family of inner products. This family of Minkowski norms is known as a Finsler structure. Just like Riemannian geometry, there is the equivalence problem: how can one decide (in principle) whether two given Finsler structures differ only by a transformation induced from a coordinate change? It is not unreasonable to expect that the solution of the equivalence problem will again involve a connection and its curvature, together with the proper space on which these objects live.

In Riemannian geometry, the connection of choice was that constructed by Levi-Civita, using the Christoffel symbols. It has two remarkable attributes: metric-compatibility and torsion-freeness. Although we now know that in Finsler geometry *proper*, these cannot both be present in the same connection, such was perhaps not common knowledge during the turn of the century. Even after reaching this realization, one still faces the daunting task of writing down viable structural equations for the connection. Furthermore, the Levi-Civita (Christoffel) connection operates on the tangent bundle TM of our underlying manifold M. But the same cannot be said of its Finslerian counterpart.

It was not until 1926 that significant progress was made by Ludwig Berwald (1883–1942), from an analytical perspective. See the poignant and informative obituary by Max Pinl in Scripta Math. **27** (1965), 193–203.

Berwald's work stemmed from the study of systems of differential equations, and was very much rooted in the calculus of variations. He introduced a connection and two curvature tensors, all rightfully bearing his name. See Matsumoto's appendix ("A History of Finsler Geometry") in *Proceedings of the 33rd Symposium on Finsler Geometry* (ed. Okubo), 1998, Lake Yamanaka. (A revised version is scheduled to appear in Tensor.) The Berwald connection is torsion-free, but is (necessarily) not metric-compatible. The Berwald curvature tensors are of two types: an *hh*- one not unlike the Riemann curvature tensor, and an *hv*- one which automatically vanishes in the Riemannian setting. Berwald's constructions have, since their inception, been indispensable to the geometry of path spaces.

Enthusiasts of metric-compatibility were not to be outdone. It is an amusing irony that although Finsler geometry starts with only a norm in any given tangent space, it regains an entire family (!) of inner products, one for each direction in that tangent space. This is why one can still make sense of metric-compatibility in the Finsler setting. In 1934, Elie Cartan introduced a connection that is metric-compatible but has torsion. The Cartan connection remains, to this day, immensely popular with the Matsumoto and the Miron schools of Finsler geometry. Besides the curvature tensors of *hh*- and *hv*- type, there is a third curvature tensor associated with the Cartan connection. It is of *vv*- type. Curiously, this last tensor is numerically identical to the curvature of a canonical (albeit singular) Riemannian metric on each tangent space.

Back in the torsion-free camp, the next progress came in 1948, when the Chern connection was discovered. Its formula differs from that of Berwald's by an $\dot{A}$ term. In natural coordinates on the slit tangent bundle $TM \setminus 0$, the Chern connection coefficients are given by

$$\frac{g^{is}}{2} \left(\frac{\delta g_{sj}}{\delta x^k} - \frac{\delta g_{jk}}{\delta x^s} + \frac{\delta g_{ks}}{\delta x^j} \right) .$$

To get those for the Berwald connection, one simply adds on the tensor $\dot{A}^i{}_{jk}$. More importantly, replacing the operator $\frac{\delta}{\delta x}$ by $\frac{\partial}{\partial x}$ gives the familiar Levi-Civita (Christoffel) connection of Riemannian metrics.

The connections of Berwald and Chern are both torsion-free. They also fail, slightly but expectedly, to be metric-compatible. Of the two, the Chern connection is simpler in form, while the Berwald connection effects a leaner *hh*-curvature for spaces of constant flag curvature. These connections coincide when the underlying Finsler structure is of Landsberg type. They further reduce to a *linear* connection on M, one which operates on TM, when the Finsler structure is of Berwald type.

In the generic Finslerian case, none of the connections we mentioned operates directly on the tangent bundle TM over M. Chern realized in his solution of the equivalence problem that, by pulling back TM so that it sits over the manifold of rays SM rather than M, one provides a natural vector bundle on which these connections may operate. It is within this geometrized setting that the equivalence problem and its solution admit a sound conceptual interpretation.

The layout of the book

The Riemann–Finsler manifolds form a much larger class than the Riemannian manifolds. Correspondingly, the former has a much more extensive literature, connected with the names Synge, Berwald, E. Cartan, Busemann, Rund, and many of our contemporaries. It is *not* the objective of this book to provide a comprehensive survey. Rather, following the general outline of Riemann and Hilbert, our aim is to develop the subject somewhat independently, with Riemannian geometry as a special case. We hope our attempt at least reflects some of the spirits of those two pioneers.

This book is comprised of three parts:

* Finsler Manifolds and Their Curvature: four chapters.
* Calculus of Variations and Comparison Theorems: five chapters.
* Special Finsler Spaces over the Reals: five chapters.

The key points of each chapter are detailed in our table of contents. Given that, we refrain from discussing here the specific topics covered.

There are fourteen chapters with an average of 30 pages each. The chapters are intentionally kept short. It seems that psychologically, one's progress through the Finsler landscape is more easily monitored this way. Every chapter is devoted to (only) one or two major results. This constraint allows us to base each chapter on a *single* theme, thereby rendering the book more teachable.

Regarding classroom use, the students we have in mind are advanced undergraduates or first-year graduate students. They are assumed to have had at least a small amount of tensor analysis, to the extent that they are comfortable with the gymnastics of raising and lowering indices. It would also help if they have had some exposure to manifolds in the abstract, so that pull-backs and push-forwards are familiar operations. Some computational experience with the Gaussian curvature of Riemannian surfaces would provide adequate motivation and intuition. This book contains enough material for roughly three semester courses.

We have adopted a candid style of writing. If something is deemed simple or straightforward, then it really is. If an omitted calculation is long, we say so. Details, annotations, and remarks are provided for the harder or subtler topics. Perhaps these gestures will help encourage the newly initiated to stay the course and not give up too easily.

At the end of every chapter, one finds a list of references. Other than a few books, these consist primarily of research papers mentioned in that chapter. We have chosen to list them there for a reason. It is helpful to be able to tell, at a glance, the research territories and boundaries with which the chapter in question has made contact. We hope this feature helps foster the book's image as an invitation to ongoing research. Incidentally, a master bibliography also appears at the end of the book.

We have compiled 393 exercises. Among those, there are 80 that prompt the reader to fill in some of the steps that we have omitted. Nothing was left out due to laziness on our part. Instead, the omissions are to be thought of as casualties of the editorial process. Their inclusion would either prove to be too distracting, or add unnecessarily to the size of the book. Those 80 problems aside, the remaining 313 exercises explore examples, touch upon new frontiers, and prepare for developments in later chapters.

If the purpose of the reader is to gain a nodding acquaintance of Finsler geometry, then the exercises can be skipped without harm, until some specific ones are referred to later. If the reader plans to do research in Finsler geometry, then practically all the exercises need to be carefully worked out. And, to assist those in the second group, we have provided detailed step-by-step guidance on the more challenging problems. The adventurous reader can always restore as much challenge as he or she wants by blocking out some of our suggestions. We simply want to ensure that no one feels demoralized by any of the exercises.

A good number of *explicit* examples are presented in this book. Those discussed in the sections proper include:

* Minkowski spaces: §1.3A, §14.1.
* Riemannian spaces: §13.3, especially §13.3B, §13.3C.
* Berwald spaces: §10.3, §11.6B.
* Randers spaces: §1.3C, §11.0, §11.6B, §12.6.
* Spaces of scalar curvature: §3.9B.
* Spaces of constant flag curvature: §12.6, §12.7.

Many more can be found among the exercises.

The above examples all involve *y-global* Finsler structures F, with the exception of the Berwald–Rund example treated in §10.3. By y-global, we mean that F is smooth and strongly convex on $TM \setminus 0$. The said example does not meet this stringent criterion, but is nevertheless included because it illustrates some computation well. It also provides excellent motivation for the rest of Chapter 10 and all of Chapter 11.

By no means have we exhausted the realm of interesting examples, y-global or not. For instance, it is with great reluctance that we have omitted Antonelli's Ecological Models, Matsumoto's Slope of a Mountain Metric, and Models of Physiological Optics discussed by Ingarden. The interested reader can consult the book *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology* written by these three authors.

It is true that Finsler geometry has not been nearly as popular as its progeny—Riemannian geometry. One reason is that deceptively simple formulas can quickly give rise to complicated expressions and mind-boggling computations. With the effort of many dedicated practitioners, this situation is slowly being turned around. Nonetheless, some intrinsic aspects of the subject are suggesting bounds on what one can do with mere pencil and paper.

Fortunately, we are in a technological age. Symbolic computations and large-scale computations on the computer are readily accessible. We took the first step in that direction by writing Maple codes for the Finslerian analogue of the Gaussian curvature. Then we implemented those codes on some explicit examples in Chapter 12. We hope this modest attempt represents the start of a trend. This could also be the venue by which a geometry-minded computer scientist helps advance the field significantly.

As we mentioned earlier, this book is not intended to be a comprehensive survey. Furthermore, our choice of topics and examples is guided by an eye towards the global geometry. The picture we paint can possibly be rather idiosyncratic. In spite of that, the material covered here is fundamental enough to be considered essential to all branches of Finsler geometry.

To our colleagues

In earlier versions of the manuscript, our definitions of the *nonlinear connection* and related objects on $TM \setminus 0$ differed from those of our fellow researchers by factors involving the Finsler function F. In this final version, we have decided to match their notations exactly. It is hoped that by removing an unnecessary accent, we have enhanced the book's suitability as a textbook or as a basic desk reference. Here are the specifics:

$$N^i{}_j \;:=\; \gamma^i{}_{jk}\, y^k \;-\; \frac{A^i{}_{jk}}{F}\, \gamma^k{}_{rs}\, y^r\, y^s \;=\; \gamma^i{}_{jk}\, y^k \;-\; C^i{}_{jk}\, \gamma^k{}_{rs}\, y^r\, y^s \,,$$

$$\frac{\delta}{\delta x^j} \;:=\; \frac{\partial}{\partial x^j} \;-\; N^i{}_j\, \frac{\partial}{\partial y^i}\,, \qquad \delta y^i \;:=\; dy^i \;+\; N^i{}_j\, dx^j \,.$$

We have *not* changed our philosophy of working, as much as possible, with objects that are homogeneous of degree zero in y. Our reason for doing so is that they make intrinsic sense on the manifold of rays SM. For instance, we prefer to work with $N^i{}_j/F$ rather than just $N^i{}_j$. But, unlike our earlier notation, the $N^i{}_j$ here is identical to the $N^i{}_j$ used by others.

Next, our convention on the wedge product does *not* contain the normalization factors $\frac{1}{2!}$, $\frac{1}{3!}$, etc. For example, if θ, ζ, and ξ are 1-forms, then:

$$\begin{aligned} \theta\wedge\zeta \;&:=\; \theta\otimes\zeta \;-\; \zeta\otimes\theta \,, \\ \theta\wedge\zeta\wedge\xi \;&:=\; \theta\otimes\zeta\otimes\xi \;-\; \theta\otimes\xi\otimes\zeta \\ &\quad+\zeta\otimes\xi\otimes\theta \;-\; \zeta\otimes\theta\otimes\xi \\ &\quad+\xi\otimes\theta\otimes\zeta \;-\; \xi\otimes\zeta\otimes\theta \,. \end{aligned}$$

Our placement of indices and sign convention on the curvature tensor are adequately illustrated by what we do in the Riemannian case:

$$\gamma^i{}_{jk} \;:=\; \frac{g^{is}}{2}\left(\frac{\partial g_{sj}}{\partial x^k} \;-\; \frac{\partial g_{jk}}{\partial x^s} \;+\; \frac{\partial g_{ks}}{\partial x^j} \right),$$

$$R_j{}^i{}_{kl} \;:=\; \frac{\partial \gamma^i{}_{jl}}{\partial x^k} \;-\; \frac{\partial \gamma^i{}_{jk}}{\partial x^l} \;+\; \gamma^i{}_{hk}\,\gamma^h{}_{jl} \;-\; \gamma^i{}_{hl}\,\gamma^h{}_{jk} \,.$$

Finally, our $G^i := \gamma^i{}_{jk}\, y^j y^k$ is *twice* the G^i of Matsumoto.

Houston, Texas D. Bao
Berkeley, California S.-S. Chern
Indianapolis, Indiana Z. Shen

Acknowledgments

This book project began as an attempt to sort through the literature on Finsler geometry. It was our intention to write a systematic account about that part of the material which is both elementary and indispensable. We want to thank many fellow geometers for their encouragement, for answering our email calls for help, and for steering us towards the pertinent references. Some of these colleagues also helped us by proof-reading parts of the manuscript.

We are grateful to Peter Kabal (Electrical Engineering, McGill University) and Brad Lackey for the roles they played in our 20 pictorial illustrations. Brad spent countless hours researching the various software for drawing pictures. It was he who introduced us to Peter Kabal's very user-friendly TeXdraw. Brad then provided a number of examples, one of which we retained as Figure 2.1. After Brad initiated us into TeXdraw, Peter supplied us with a copy of the manual. Incidentally, Brad also alerted us to a theoretical blunder. Happily, it was something that we were able to fix.

One of us (Bao) would like to thank the University of Houston for a Limited-Grant-In-Aid (LGIA) which partially funded Brad's efforts.

There are several cherished monographs on Finsler geometry that we have not directly referenced in this book, although it is certain that we have benefited from them in many ways. We are referring to the books by Abate and Patrizio [AP], Bejancu [Bej], Miron and Anastasiei [MA].

Our book does not discuss the applications of Finsler geometry to biology, engineering, and physics. For this reason, we are especially thankful for the monographs [Asan] by Asanov, [AB] by Antonelli and Bradbury, [AIM] by Antonelli, Ingarden, and Matsumoto, and [AZas] by Antonelli and Zastawniak. We have also gained much insight from the four expository essays by Antonelli [Ant], Ingarden [Ing], Gardner and Wilkens [GW], and Beil [Bl] in the Seattle proceedings volume [BCS2].

The details for all the references cited here can be found in the master bibliography at the end of the book.

We would like to acknowledge the always enthusiastic support of Ina Lindemann (Mathematics Editor), the TeX-nical expertise of Fred Bartlett (Supervising Developer), Kanitra Fletcher and Yong-Soon Hwang for help with miscellaneous issues, and Frank McGuckin (Production Editor) for his editorial guidance. Finally, we want to thank Valerie Greco, who copy-edited the manuscript, for a splendidly accurate and constructive job.

Houston, Texas	D. Bao
Berkeley, California	S.-S. Chern
Indianapolis, Indiana	Z. Shen

Contents

Chapter 1

Finsler Manifolds and the Fundamentals of Minkowski Norms

1.0 Physical Motivations

Finsler geometry has its genesis in integrals of the form

$$\int_a^b F\Big(x^1, \ldots, x^n ;\, \frac{dx^1}{dt}, \ldots, \frac{dx^n}{dt}\Big)\, dt\ .$$

The function $F(x^1, \ldots, x^n ; y^1, \ldots, y^n)$ is positive unless all the y^i are zero. It is also homogeneous of degree one in y. Let us single out some contexts in which this integral arises.

* In certain physical examples, x stands for position, y for velocity. Then F would have the meaning of speed, and t would play the

role of time. In these cases, the above integral measures distance traveled. However, other interpretations are possible.

* Take optics for instance. Keep in mind that in an anisotropic medium, the speed of light depends on its direction of travel. At each location x, visualize y as an arrow that emanates from x. Now measure the amount of time it takes light to travel from x to the tip of y, and call the result $F(x, y)$. The hypothesized homogeneity allows us to rewrite the displayed integral as $\int_a^b F(x, dx)$. This then represents the total time it takes light to traverse a given (possibly curved) path in this medium. See Ingarden's exposition in [AIM].

* There are many variations on the theme we just described. A particularly interesting one concerns the time it takes to negotiate any given path on a hillside. It was originally mentioned by Finsler to Matsumoto [M1]. The premise here is that one's walking speed depends heavily on the slope of the terrain, and hence on one's direction of travel. See Matsumoto's account in [AIM].

* Mathematical ecology provides more esoteric examples. For instance, x could stand for the state of a coral reef, and y the displacement vector from the state x to a new state. The quantity $F(x, dx)$ represents the energy one needs in order to evolve from the state x to the neighboring state $x + dx$. Hence the integral $\int_a^b F(x, dx)$ is the total energy cost of a given path of evolution. See Antonelli's treatment in [AIM], as well as the book by Antonelli and Bradbury [AB].

For explicit mathematical examples, see §1.3.

1.1 Finsler Structures: Definitions and Conventions

Let M be an n-dimensional C^∞ manifold. Denote by T_xM the tangent space at $x \in M$, and by $TM := \cup_{x\in M} T_xM$ the tangent bundle of M. Each element of TM has the form (x, y), where $x \in M$ and $y \in T_xM$. The natural projection $\pi : TM \to M$ is given by $\pi(x, y) := x$. The dual space of T_xM is T_x^*M, called the cotangent space at x. The union $T^*M := \cup_{x\in M} T_x^*M$ is the cotangent bundle of M.

A (globally defined) **Finsler structure** of M is a function

$$F \,:\, TM \;\to\; [0, \infty)$$

with the following properties:

(i) **Regularity**: F is C^∞ on the entire **slit tangent bundle** $TM \setminus 0$.

(ii) **Positive homogeneity**: $F(\,x\,,\,\lambda\,y\,) = \lambda\,F(x, y)$ for all $\lambda > 0$.

(iii) **Strong convexity**: The $n \times n$ Hessian matrix

$$(g_{ij}) := \left(\left[\frac{1}{2} F^2 \right]_{y^i y^j} \right)$$

is positive-definite at every point of $TM \setminus 0$.

* In some situations, the Finsler structure F satisfies the criterion $F(x, -y) = F(x, y)$. In that case we have **absolute homogeneity** instead: $F(x, \lambda y) = |\lambda| F(x, y)$ for all $\lambda \in \mathbb{R}$. In general, we find this property to be too restrictive, because it would immediately exclude some interesting examples such as Randers spaces (see §1.3C).

* Let us make sense of the y^i in criterion (iii). Fix any basis $\{b_i\}$ for $T_x M$. Out of habit, one can take b_i to be $\frac{\partial}{\partial x^i}$, although this restriction is unnecessary. Express y as $y^i b_i$. The Finsler structure F is then a function of (x^i, y^i), and the partial derivatives of $\frac{1}{2} F^2$ are taken with respect to the y^i. It can be checked that the positive-definiteness stipulated in (iii) is independent of our choice of $\{b_i\}$.

Given a manifold M and a Finsler structure F on TM, the pair (M, F) is known as a Finsler manifold. See §1.3 for explicit examples of some important Finsler manifolds.

Throughout the book, the rules that govern our index gymnastics are as follows:

- Lower case Latin indices (except the alphabet n) run from 1 to n .
- Lower case Greek indices run from 1 to $n - 1$.
- Vector indices are up; covector indices are down.
- Any repeated pair of indices—provided that one is up and the other is down—is automatically summed.
- The lowering and raising of indices are carried out by the g_{ij} defined above, and its matrix inverse g^{ij}.

Let $(x^1, \ldots, x^n) = (x^i) : U \to \mathbb{R}^n$ be a local coordinate system on an open subset $U \subset M$. As usual, $\{\frac{\partial}{\partial x^i}\}$ and $\{dx^i\}$ are, respectively, the induced coordinate bases for $T_x M$ and $T_x^* M$. The said x^i give rise to local coordinates (x^i, y^i) on $\pi^{-1} U \subset TM$ through the mechanism

$$y = y^i \frac{\partial}{\partial x^i} .$$

The y^j are fiberwise global. Whenever possible, let us make no distinction between (x, y) and its coordinate representation (x^i, y^i). Functions F that are defined on TM can be locally expressed as

$$F(x^1, \ldots, x^n ; y^1, \ldots, y^n) .$$

We continue a convention employed in criterion (iii) above; namely, denote by F_{y^i}, $F_{y^i y^j}$, ..., etc. the partial derivative(s) of F with respect to the coordinates y^i. Adopt a similar notation for the partial derivatives with respect to the coordinates x^i.

We close this section with some cautionary remarks about our notation. For the sake of concreteness, we focus our attention on the various objects that the symbol $\frac{\partial}{\partial x^i}$ comes to represent throughout this book.

* When evaluated at the point $x \in M$, $\frac{\partial}{\partial x^i}$ refers to a coordinate vector on M.
* When evaluated at the point $(x, y) \in TM$, the same notation $\frac{\partial}{\partial x^i}$ stands for a coordinate vector on TM. As such, it would be on the same footing as the $\frac{\partial}{\partial y^i}$, which are also coordinate vectors on the tangent bundle TM.
* Later on, we use the restricted projection $\pi : TM \setminus 0 \to M$ to pull the tangent bundle TM back, producing a vector bundle $\pi^* TM$ that sits over $TM \setminus 0$. In that case, when $\frac{\partial}{\partial x^i}$ is evaluated at the point $(x, y) \in TM \setminus 0$, it will take on yet another meaning, namely, as (the value of) a basis section of the bundle $\pi^* TM$.

In short, we are using the same symbol $\frac{\partial}{\partial x^i}$ to denote objects that belong to three different spaces. Furthermore, they do not obey the same transformation law. Indeed, let

$$x^i = x^i(\tilde{x}^1, \ldots, \tilde{x}^n)$$

be a local change of coordinates on M. Correspondingly, the chain rule gives

$$y^i = \frac{\partial x^i}{\partial \tilde{x}^p} \, \tilde{y}^p . \tag{1.1.1}$$

One can apply the chain rule carefully to deduce that:

* As coordinate vector fields on M, or as basis sections of $\pi^* TM$, the $\frac{\partial}{\partial x^i}$ transform like

$$\frac{\partial}{\partial \tilde{x}^p} = \frac{\partial x^i}{\partial \tilde{x}^p} \, \frac{\partial}{\partial x^i} . \tag{1.1.2}$$

* On the other hand, as coordinate vector fields on TM, the $\frac{\partial}{\partial x^i}$ transform like

$$\frac{\partial}{\partial \tilde{x}^p} = \frac{\partial x^i}{\partial \tilde{x}^p} \, \frac{\partial}{\partial x^i} + \frac{\partial^2 x^i}{\partial \tilde{x}^p \partial \tilde{x}^q} \, \tilde{y}^q \, \frac{\partial}{\partial y^i} . \tag{1.1.3}$$

Nevertheless, we have decided that the inherent risks of this practice do not outweigh its virtue, which is simplicity. We feel (or hope!) that it is easier for the reader to ferret out the proper meaning of $\frac{\partial}{\partial x^i}$ from the context of our discussion or computation, than to create a different symbol for each of the three objects described.

Exercises

Exercise 1.1.1: Recall that the definition of the g_{ij} involves a choice of basis for each T_xM. Explain why the positive-definiteness of the matrix (g_{ij}) is a basis-free concept.

Exercise 1.1.2: Derive the induced transformation laws (1.1.1)–(1.1.3).

1.2 Two Basic Properties of Minkowski Norms

The restriction of a Finsler structure F to any specific tangent space T_xM gives what is known as a Minkowski norm on T_xM. Thus a Finsler structure of M may be viewed as a smoothly varying family of Minkowski norms. Generically, this family has rather limited (to be precise, no more than C^1) differentiability along the zero section of the tangent bundle TM. Such regularity issues are dealt with later. Here, let us concern ourselves with certain geometrical aspects of Minkowski norms.

Every n-dimensional vector space is linearly isomorphic to $\mathbb{R}^n$, whose elements y have the form $(y^1, \ldots, y^n)$. Thus there is no loss of generality in confining our discussion to Minkowski norms on $\mathbb{R}^n$.

1.2 A. Euler's Theorem

First, let us dispense with a technical ingredient that manifests itself repeatedly in our arguments. It is known as **Euler's theorem** for homogeneous functions.

Theorem 1.2.1. *Suppose a real-valued function H on $\mathbb{R}^n$ is differentiable away from the origin of $\mathbb{R}^n$. Then the following two statements are equivalent:*

- *H is positively homogeneous of degree r. That is,*

$$H(\lambda y) = \lambda^r H(y) \qquad \text{for all } \lambda > 0 .$$

- *The radial directional derivative of H is r times H. Namely,*

$$\boxed{y^i H_{y^i}(y) = r H(y)} \, .$$

Proof.

* Suppose H satisfies $H(\lambda y) = \lambda^r H(y)$ for all positive λ. Fix y. Differentiating this equation with respect to the parameter λ gives

$$y^i H_{y^i}(\lambda y) = r \lambda^{r-1} H(y) .$$

Setting λ equal to 1 gives the criterion sought.

* Conversely, suppose $y^i H_{y^i}(y) = r\, H(y)$. Fix y and consider the function $H(\lambda y)$ with $\lambda > 0$. By the chain rule, we have

$$\frac{d}{d\lambda} H(\lambda y) = y^i H_{y^i}(\lambda y) = \frac{1}{\lambda} (\lambda y)^i H_{y^i}(\lambda y) .$$

Using our supposition, we see that the last term equals $\frac{1}{\lambda} r\, H(\lambda y)$. Since we have not assumed that H is nonzero away from the origin, we cannot read the above as $\frac{d}{d\lambda} \log H(\lambda y) = \frac{r}{\lambda} = \frac{d}{d\lambda} \log \lambda^r$. Instead, we rewrite it as the ODE

$$\frac{d}{d\lambda} H(\lambda y) - \frac{r}{\lambda} H(\lambda y) = 0 .$$

The integrating factor $\frac{1}{\lambda^r}$ then gives $H(\lambda y) = C\, \lambda^r$, where C is some constant that depends on our fixed y. Setting λ equal to 1 shows that $C = H(y)$. ☐

In particular, if F is positively homogeneous of degree 1, then

$$y^i F_{y^i}(y) = F(y) , \qquad \text{equivalently} \qquad \frac{y^i}{F} F_{y^i} = 1 , \tag{1.2.1}$$

$$y^j F_{y^i y^j}(y) = 0 . \tag{1.2.2}$$

1.2 B. A Fundamental Inequality

The next theorem tells us that positivity and the triangle inequality are actually *consequences* of the defining properties of Minkowski norms. It also calls our attention to a multifaceted fundamental inequality.

Theorem 1.2.2. *Let F be a nonnegative real-valued function on $\mathbb{R}^n$ with the properties:*

* *F is C^∞ on the punctured space $\mathbb{R}^n \setminus 0$.*
* *$F(\lambda y) = \lambda F(y)$ for all $\lambda > 0$.*
* *The $n \times n$ matrix (g_{ij}), where $g_{ij\,(y)} := [\,\frac{1}{2} F^2\,]_{y^i y^j}(y)$, is positive-definite at all $y \neq 0$.*

Then we have the following conclusions:

* *(Positivity)*

$$F(y) > 0 \qquad \text{whenever } y \neq 0 .$$

* *(Triangle inequality)*

$$F(y_1 + y_2) \leqslant F(y_1) + F(y_2) ,$$

where equality holds if and only if $y_2 = \alpha\, y_1$ or $y_1 = \alpha\, y_2$ for some $\alpha \geqslant 0$.

- *(Fundamental inequality)*

(1.2.3) $$\boxed{w^i \, F_{y^i}(y) \leqslant F(w) \quad \textit{at all } y \neq 0} \,,$$

and equality holds if and only if $w = \alpha\, y$ *for some* $\alpha \geqslant 0$.

Remarks:

** The hypotheses of the above theorem define what one means by a **Minkowski norm** on $\mathbb{R}^n$. According to this theorem, there is no need to hypothesize that F be positive at $y \neq 0$; it is necessarily so.

** If the Minkowski norm satisfies $F(-y) = F(y)$, then one has the absolute homogeneity $F(\lambda\, y) = |\lambda|\, F(y)$. The simplest example of an absolutely homogeneous Minkowski norm on $\mathbb{R}^n$ is

$$F(y) := \sqrt{y \bullet y} \,,$$

where $\bullet$ denotes the canonical inner product $u \bullet v := \delta_{ij}\, u^i\, v^j$. This F is called the standard **Euclidean norm** of $\mathbb{R}^n$.

** In view of the first two conclusions of this theorem, every *absolutely* homogeneous Minkowski norm is a norm in the sense of functional analysis.

In preparation for the proof of Theorem 1.2.2, we observe the following:

- One can check that

(1.2.4) $$g_{ij\,(y)} := \left(\frac{1}{2}\, F^2\right)_{y^i y^j} (y) = \Big[F\, F_{y^i y^j} + F_{y^i}\, F_{y^j} \Big](y) \,.$$

The g_{ij} are C^∞ functions on $\mathbb{R}^n \setminus 0$ and, in typical examples (that are not Riemannian), they cannot even be extended continuously to all of $\mathbb{R}^n$.

- Applying the consequences (1.2.1), (1.2.2) of Euler's theorem to the above formula for g_{ij} gives

(1.2.5) $$g_{ij\,(y)}\, y^i\, y^j = F^2(y) \,, \quad \text{equivalently} \quad g_{ij}\, \frac{y^i}{F}\, \frac{y^j}{F} = 1 \,.$$

We now give a proof, adapted from Rund [R], of Theorem 1.2.2.

Proof of the theorem.

(i) Positivity:

Consider (1.2.5); namely, $g_{ij\,(y)}\, y^i\, y^j = F^2(y)$. The hypothesized strong convexity of F says that the left-hand side is positive whenever $y \neq 0$, thus F is strictly positive on $\mathbb{R}^n \setminus 0$.

(ii) The triangle inequality:

At each point $y \in \mathbb{R}^n \setminus 0$, the matrix (g_{ij}) defines an inner product. So we have the Cauchy–Schwarz type inequality

$$(1.2.6) \qquad [g_{ij\,(y)}\, \xi^i \eta^j\,]^2 \;\leqslant\; [g_{ij\,(y)}\, \xi^i \xi^j\,]\,[g_{kl\,(y)}\, \eta^k \eta^l\,] \qquad \forall\, \xi, \eta \in \mathbb{R}^n ,$$

where equality holds if and only if $\xi = (\xi^i)$, $\eta = (\eta^i)$ are collinear. Setting $\eta^i = y^i$ and using (1.2.5), we obtain

$$(1.2.7) \qquad [g_{ij\,(y)}\, \xi^i y^j\,]^2 \;\leqslant\; F^2(y)\,[g_{ij\,(y)}\, \xi^i \xi^j\,] \qquad \forall\, \xi \in \mathbb{R}^n ,$$

where equality holds if and only if ξ and y are collinear. On the other hand, the formula (1.2.4) for g_{ij} leads us to

$$(1.2.8) \quad F_{y^i y^j}(y)\; \xi^i \xi^j \;=\; \frac{1}{F^3(y)} \left\{ F^2(y)\,[g_{ij\,(y)}\, \xi^i \xi^j\,] \;-\; [g_{ij\,(y)}\, y^i \xi^j\,]^2 \right\}$$

which, in conjunction with (1.2.7), gives

$$(1.2.9) \qquad F_{y^i y^j}(y)\; \xi^i \xi^j \;\geqslant\; 0 \qquad \forall\, \xi \in \mathbb{R}^n .$$

Here, equality holds if and only if ξ and y are collinear.

Next we prove that

$$(1.2.10) \qquad 2\,F(y) \;\leqslant\; F(y+\xi) \;+\; F(y-\xi) \qquad \forall\, y, \xi \in \mathbb{R}^n ,$$

and equality holds if and only if $\xi = \lambda\, y$ for some $|\lambda| \leqslant 1$.

Let us begin by analyzing all the linearly dependent cases:

* If $\xi = \lambda\, y$ for some $|\lambda| \leqslant 1$, the (positive) homogeneity of F implies that both sides of (1.2.10) are equal to $2F(y)$.
* If $\xi = \lambda\, y$ with $|\lambda| > 1$, the right-hand side of (1.2.10) reduces to the form $(2+\alpha)F(y)+\beta$, where α, β are positive. Hence the inequality in question is strict as claimed.
* The case of $\xi = 0$ is covered by the above. The only scenario left is when $\xi \neq 0$ but $y = 0$, for which the inequality is strict by the positivity we just established.

Now suppose y, ξ are linearly independent. Consider $F(y+t\,\xi)$, which is a C^∞ function in the real variable t. By the second mean-value theorem, we have

$$(1.2.11) \qquad F(y \pm \xi) \;=\; F(y) \;\pm\; F_{y^i}(y)\; \xi^i \;+\; \frac{1}{2}\, F_{y^i y^j}(y \pm \epsilon\xi)\; \xi^i \xi^j$$

for some $0 < \epsilon < 1$. Since $y \pm \epsilon\xi$ and ξ are linearly independent, (1.2.9) tells us that the quadratic term in (1.2.11) is positive. Thus

$$(1.2.12) \qquad F(y+\xi) \;>\; F(y) \;+\; F_{y^i}(y)\; \xi^i ,$$

$$(1.2.13) \qquad F(y-\xi) \;>\; F(y) \;-\; F_{y^i}(y)\; \xi^i .$$

These add to yield the strict inequality part of (1.2.10).

By setting $y := \frac{1}{2}(y_1 + y_2)$ and $\xi := \frac{1}{2}(y_1 - y_2)$ in (1.2.10), we obtain the triangle inequality stated in the theorem. The fact that (1.2.10) becomes an equality only when $\xi = \lambda y$ for some $|\lambda| \leqslant 1$ now implies that the triangle inequality is strict except when $y_1 = \alpha y_2$ or $y_2 = \alpha y_1$ for some $\alpha \geqslant 0$.

Let us note in passing the following. We have seen that (1.2.10) implies the triangle inequality. The converse is quite straightforward. So the two are actually equivalent.

(iii) The fundamental inequality:

Finally, we ascertain (1.2.3):

$$w^i \, F_{y^i}(y) \leqslant F(w) \qquad \text{at all } \; y \neq 0 \,,$$

where equality is supposed to hold if and only if $w = \alpha y$ for some $\alpha \geqslant 0$. The consequences of Euler's theorem, as described in (1.2.1) and (1.2.2), are used repeatedly without mention. As before, we enumerate all possibilities:

* When $w = \alpha y$ for some $\alpha \geqslant 0$, both sides equal $\alpha F(y)$.
* When w is a negative multiple of y, the inequality is strict because its left-hand side becomes a negative multiple of $F(y)$.
* The case $w \neq 0$ but $y = 0$ is disallowed.
* Lastly, suppose y, w are linearly independent; then so are y and $\xi := y - w$. Inequality (1.2.13) now reads

(1.2.14)
$$F(w) > F(y) - F_{y^i}(y)\,(y^i - w^i) \,,$$

which readily reduces to the strict part of (1.2.3).

We have completely proved Theorem 1.2.2. □

1.2 C. Interpretations of the Fundamental Inequality

In this subsection, let us explore the many different faces of the fundamental inequality (1.2.3).

- At face value, (1.2.3) says that

$$\boxed{w^i \, F_{y^i}(y) \leqslant F(w)} \,.$$

And, the latter becomes an equality if and only if $w = \alpha y$ for some $\alpha \geqslant 0$. Note that the said equality, after cancelling off α (if > 0), is none other than Euler's theorem (1.2.1): $y^i \, F_{y^i}(y) = F(y)$. Hence (1.2.3) may be viewed as *an extension of Euler's theorem*, from an equation to an inequality.

- Adding the equation $F(y) - y^i \, F_{y^i}(y) = 0$ to (1.2.3) gives

(1.2.15)
$$\boxed{F(y) + F_{y^i}(y)\,(w - y)^i \leqslant F(w)} \,,$$

where equality holds only when $w = \alpha y$ with $\alpha \geqslant 0$. Think of y as fixed and w as the independent variable. The left-hand side is then the linear approximation of the value $F(w)$. So, at any fixed $(y, F(y))$ on the graph of F, the tangent hyperplane touches the graph only along the ray $(\alpha y, \alpha F(y))$, $\alpha \geqslant 0$. Everywhere else, the tangent hyperplane lies below the graph of F. This is depicted in Figure 1.1. In this way, (1.2.3) tells us that *the graph of F is a convex cone* with its vertex at the origin of our Minkowski space.

- Since $F(y) > 0$ for $y \neq 0$, we can multiply (1.2.3) by $F(y)$ to get $w^i\, F(y)\, F_{y^i}(y) \leqslant F(y)\, F(w)$. Now, (1.2.4), (1.2.2), and (1.2.1) together give $y^j\, g_{ij} = F F_{y^i}$. Thus (1.2.3) is equivalent to

$$\boxed{g_{ij\,(y)}\; w^i\, y^j \;\leqslant\; F(w)\; F(y)}\;. \tag{1.2.16}$$

  Consider first the case in which F is the norm associated with an inner product on $\mathbb{R}^n$. Here, $F(y) = \sqrt{g_{ij}\, y^i y^j}$, where the g_{ij} are constants. Almost by inspection, we see that the fundamental tensor $g_{ij\,(y)} := (\frac{1}{2}F^2)_{y^i y^j}$ is simply given by the inner product g_{ij}. Also in this case, $F(-y) = F(y)$. These observations allow us to deduce that (1.2.16) is equivalent to $|g_{ij}\, w^i y^j| \leqslant F(w)\, F(y)$, which is the standard Cauchy–Schwarz inequality. So, in the general case, we may view (1.2.16) [equivalently (1.2.3)] as *a generalization of the Cauchy–Schwarz inequality*, from inner products to Minkowski norms. Note however that, when spelled out, (1.2.16) implies that

$$[\, g_{ij\,(y)}\; w^i\, y^j\,]^2 \;\leqslant\; [\, g_{pq\,(w)}\; w^p\, w^q\,]\, [\, g_{rs\,(y)}\; y^r\, y^s\,]\;.$$

  We emphasize that in the first term on the right, it is $g_{pq\,(w)}$ and not $g_{pq\,(y)}$. As such, this last inequality is distinctly different from, and much more subtle than, the Cauchy–Schwarz type inequality (1.2.6) encountered during the proof of Theorem 1.2.2.

- Finally, the fundamental inequality (1.2.3) plays a pivotal role in the proof (Theorem 6.3.1) that short Finslerian geodesics are minimal. Upon this edifice rests the Hopf–Rinow theorem (see §6.6) and the enterprise of cut versus conjugate loci (treated systematically in Chapter 8). As we show, the fundamental inequality comes to the rescue when the Riemannian proof of Theorem 6.3.1 breaks down in the generic Finsler case. The same technique saves the day again in Proposition 9.2.2, when we prove that for (forward geodesically) complete connected Finsler manifolds of nonpositive flag curvature, the exponential map is a covering projection.

We have interpreted the fundamental inequality (1.2.3) in several contexts. In each case, something interesting and important emerges. This is a testimonial to the inequality's depth and significance.

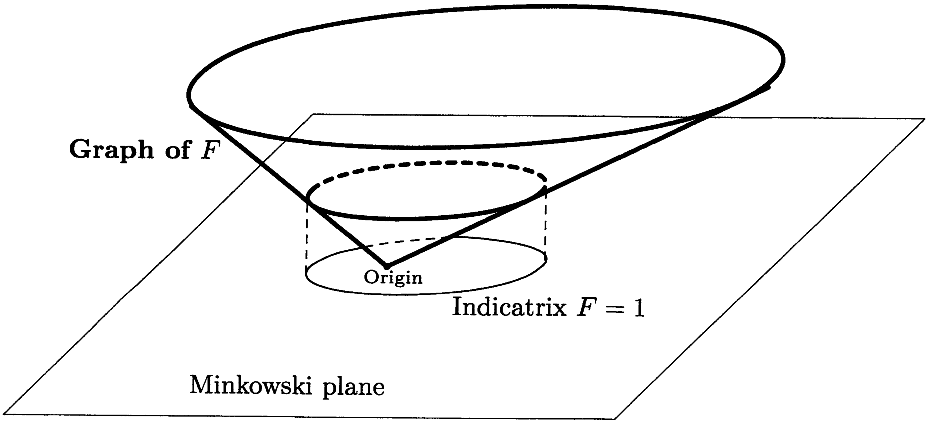

Figure 1.1

The graph of a Minkowski norm is a convex cone with its vertex at the origin. The one shown here "tilts to the right."

Exercises

Exercise 1.2.1: Let F be positively homogeneous of degree 1 on $\mathbb{R}^n$. Use Euler's theorem to show that

(a) $y^i \, F_{y^i} = F$.

(b) $y^j \, F_{y^i y^j} = 0$.

(c) $y^k \, F_{y^i y^j y^k} = - F_{y^i y^j}$.

(d) $y^l \, F_{y^i y^j y^k y^l} = - 2 \, F_{y^i y^j y^k}$.

Here, all formulas are supposed to be evaluated at y.

Exercise 1.2.2: Let F be the standard Euclidean norm on $\mathbb{R}^n$. Show that its g_{ij} is simply the Kronecker delta δ_{ij} .

Exercise 1.2.3: Derive (1.2.5).

Exercise 1.2.4: A Minkowski norm F on $\mathbb{R}^n$ is said to be **Euclidean** if it arises from an inner product $\langle \ , \ \rangle$ through $F(y) = \sqrt{\langle y , y \rangle}$. Prove that the following three criteria are equivalent:

(a) The Minkowski norm F is Euclidean.

(b) The functions g_{ij} defined in (1.2.4) are constant.

(c) The functions $A_{ijk\,(y)} := \frac{F}{4} \, (F^2)_{y^i y^j y^k}$ are all zero.

Exercise 1.2.5: Let F be a Minkowski norm on $\mathbb{R}^n$.

(a) Explain why its Hessian matrix $(F_{y^i y^j})$ is positive *semi*definite.

(b) Prove that its rank is $n - 1$.

(c) Verify that its 1-dimensional null space at any point $y \neq 0$ is spanned by the vector $y^i \frac{\partial}{\partial y^i}$.

Hint: you may want to review the discussions that center around (1.2.9).

Exercise 1.2.6: A domain D in $\mathbb{R}^n$ is said to be **strictly convex** if it contains the interior of every line segment joining any two points of the topological closure $\overline{D}$. Let F be a Minkowski norm. Given any $r > 0$, define the ball $\mathrm{B}^n(r)$ and the sphere $\mathrm{S}^{n-1}(r)$ of radius r (centered at the origin) as follows:

$$\begin{aligned} \mathrm{B}^n(r) &:= \{ y \in \mathbb{R}^n : F(y) < r \} , \\ \mathrm{S}^{n-1}(r) &:= \{ y \in \mathbb{R}^n : F(y) = r \} . \end{aligned}$$

Show that:

(a) Each $\mathrm{B}^n(r)$ is a strictly convex domain with C^∞ boundary $\mathrm{S}^{n-1}(r)$.

(b) Explain what it means to say that strong convexity implies strict convexity.

Exercise 1.2.7: Suppose B is a strictly convex open domain "centered" at the origin, with smooth boundary $S := \partial B$. Define a nonnegative function F on $\mathbb{R}^n$ as follows:

$$F(y) := \frac{1}{t} , \quad \text{where } t > 0 \text{ is such that } ty \in S .$$

(a) Check that $F(y) > 0$ for all $y \neq 0$.

(b) Verify that $F(\lambda y) = \lambda F(y)$ for all positive numbers λ. Also, ascertain that if the domain B satisfies $y \in B \Leftrightarrow -y \in B$, then we have $F(\lambda y) = |\lambda| F(y)$ for all $\lambda \in \mathbb{R}$.

(c) Show that

$$F(y) \quad \text{is} \quad \begin{cases} < 1 & \text{if and only if } \ y \in B \\ = 1 & \text{if and only if } \ y \in S \\ > 1 & \text{if and only if } \ y \notin \overline{B} \ . \end{cases}$$

(d) Prove that F satisfies the triangle inequality.

(e) Then check that

$$F(y + t\xi) + F(y - t\xi) - 2F(y) \geqslant 0 .$$

(f) Explain why F is C^∞ on $\mathbb{R}^n \setminus 0$, but is typically not differentiable at the origin.

(g) Fix $y \neq 0$ and divide the inequality in part (e) by t^2. Then take the limit as $t \to 0^+$. Show that the result is

$$F_{y^i y^j}(y) \, \xi^i \xi^j \geqslant 0 .$$

In other words, the Hessian of F is positive-semidefinite. Hint: expand the terms $F(y \pm t\xi)$ using the second mean-value theorem.

(h) Use formula (1.2.4) and part (g) to help you deduce that the Hessian $g_{ij} := [\frac{1}{2} F^2]_{y^i y^j}$ is typically only positive-semidefinite. Can you exhibit a nonzero ξ such that $g_{ij}\,\xi^i\,\xi^j = 0$? Hint: see Exercise 1.2.9.

The moral here is that:

There are homogeneous functions F with strictly convex unit balls but fail (just barely) to be strongly convex. Hence they do *not* define Minkowski norms.

Exercise 1.2.8:

(a) Let S be some smooth hypersurface in $\mathbb{R}^n$ that is defined by an equation $P(v) = 0$. Suppose we want to find a function F on $\mathbb{R}^n$ that has the constant value 1 on S. Explain why $F(y)$ is characterized by the equation

$$P\Big[\frac{y}{F(y)}\Big] = 0 .$$

Occasionally, such an equation can be solved to give an explicit formula for $F(y)$. The method we have just described is known affectionately as **Okubo's technique**.

(b) As a concrete example, let S be the convex limaçon in $\mathbb{R}^2$. In polar coordinates, it has the description

$$\rho = 3 + \cos\phi , \quad 0 \leqslant \phi \leqslant 2\pi .$$

Sketch S and check that its Cartesian description is

$$(y^1)^2 + (y^2)^2 = 3\sqrt{(y^1)^2 + (y^2)^2} + y^1 .$$

(c) Apply Okubo's technique to show that the function F which has constant value 1 on S is

$$F(y) = \frac{(y^1)^2 + (y^2)^2}{3\sqrt{(y^1)^2 + (y^2)^2} + y^1} .$$

(d) Can you prove that this F has all the defining properties (especially strong convexity) of a Minkowski norm?

Exercise 1.2.9: Let

$$B := \{(y^1, y^2) \in \mathbb{R}^2 : (y^1)^4 + (y^2)^4 < 1\} .$$

(a) Check that B is strictly convex (as defined in Exercise 1.2.6) and has a smooth boundary.

(b) Consider the F defined in Exercise 1.2.7. Use Okubo's technique to deduce that here, it has the explicit formula

$$F(y) = [(y^1)^4 + (y^2)^4]^{1/4} .$$

(c) Calculate the matrix (g_{ij}) and check that it is singular (that is, not invertible) on the y^1 and y^2 axes. As a result, it cannot possibly be positive-definite at these points. Show that these are the only points at which it fails to be positive-definite.

(d) Explain what it means to say that strict convexity does *not* imply strong convexity.

Exercise 1.2.10: In $\mathbb{R}^2$, abbreviate y^1, y^2 as p, q, respectively. Define

$$F(p,q) \ := \ \alpha \ (p^2 + q^2)^{1/2} \ + \ \beta \ (p^4 + q^4)^{1/4} \ ,$$

where $\alpha > 0, \ \ \beta \geqslant 0$ are constants.

(a) Calculate the functions g_{ij} and check that they are homogeneous of degree zero.

(b) Identify all constants α and β for which F defines a Minkowski norm on $\mathbb{R}^2$.

Exercise 1.2.11: Let F be a Minkowski norm on $\mathbb{R}^n$.

(a) Show that if two vectors y and w satisfy $g_{ij\,(y)}\ y^j = g_{ij\,(w)}\ w^j$, then $y = w$.

(b) Decide whether anything can be concluded if those two vectors satisfy the following identity instead: $F_{y^i}(y) = F_{w^i}(w)$.

1.3 Explicit Examples of Finsler Manifolds

1.3 A. Minkowski and Locally Minkowski Spaces

A Finsler manifold (M, F) is said to be **locally Minkowskian** if, at every point $x \in M$, there is a local coordinate system (x^i), with induced tangent space coordinates y^i, such that F has no dependence on the x^i.

In order to construct locally Minkowskian manifolds, one might intuitively begin with a smooth manifold M and try to put the "same" Minkowski norm on each of its tangent spaces. However, a good amount of caution should be exercised, because there are topological obstruction(s) that one must overcome. For example, if M is compact and boundaryless, then having a locally Minkowskian structure will force its Euler characteristic to vanish. See [BC2].

The simplest locally Minkowskian manifolds are of the following type:

* We start with a Minkowski norm F on $\mathbb{R}^n$.
* We change our perspective and regard $\mathbb{R}^n$ as a manifold, albeit a linear one.
* Given any tangent vector v based at $y \in \mathbb{R}^n$, we slide it (without twisting) until it emanates from the origin "o" instead. Then we evaluate F at the tip of this translated vector.

* In terms of a formula, we have

$$F(y\,,v) \;=\; F\left(y\,,\, v^i\, \frac{\partial}{\partial y^i}_{|y} \right) \;:=\; F\left(v^i\, \frac{\partial}{\partial y^i}_{|o} \right) .$$

One is certainly justified in saying that the locally Minkowskian examples we just cited are actually Minkowski norms in trivial disguise! Yet, this does not detract from the fact that such examples are among the most important ones in practice.

For numerical explorations, a particularly instructive family of Minkowski norms is the following. Here, λ can be any nonnegative constant.

$$\boxed{F(v^1\,,v^2) \;:=\; \sqrt{\sqrt{(v^1)^4 \,+\, (v^2)^4} \;+\; \lambda\,[\,(v^1)^2 \,+\, (v^2)^2\,]}\,} .$$

This may be viewed as a **perturbation of the quartic metric**. See also Exercise 1.2.11. Let us demonstrate that the perturbation serves to regularize the singularity in the quartic metric. To this end, we first relabel v^1 as p and v^2 as q in order to avoid clutter. Straightforward computations then give

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} \;=\; \begin{pmatrix} \lambda \,+\, \frac{p^2\,(p^4\,+\,3\,q^4)}{(p^4\,+\,q^4)^{3/2}} & \frac{-2\,p^3\,q^3}{(p^4\,+\,q^4)^{3/2}} \\ \frac{-2\,p^3\,q^3}{(p^4\,+\,q^4)^{3/2}} & \lambda \,+\, \frac{q^2\,(q^4\,+\,3\,p^4)}{(p^4\,+\,q^4)^{3/2}} \end{pmatrix} .$$

Hence

$$\begin{aligned} \det(g_{ij}) \;&=\; \lambda^2 \,+\, \lambda\, \frac{(p^2\,+\,q^2)^3}{(p^4\,+\,q^4)^{3/2}} \,+\, \frac{3\,p^2\,q^2}{p^4\,+\,q^4}\,, \\ \operatorname{trace}(g_{ij}) \;&=\; 2\,\lambda \,+\, \frac{(p^2\,+\,q^2)^3}{(p^4\,+\,q^4)^{3/2}}\,. \end{aligned}$$

Note that

- If $\lambda = 0$, then $\det(g_{ij})$ vanishes on the p and q axes in each tangent plane. In that case, the Finsler function F, whose fundamental tensor is called the **quartic metric**, fails to be a Minkowski norm because strong convexity is violated at some nonzero y.
- If $\lambda > 0$, then both the determinant and the trace of (g_{ij}) are positive away from the origin in each tangent plane. In that case, g_{ij} is positive-definite because both its eigenvalues are positive. The Finsler structure F is then a Minkowski norm. In this sense, the perturbation has regularized the quartic metric.

1.3 B. Riemannian Manifolds

Let M be an n-dimensional C^∞ (smooth) manifold. A smooth Riemannian metric g on M is a family $\{g_x\}_{x\in M}$ of inner products, one for each tangent space T_xM, such that the functions $g_{ij\,(x)} := g_x(\,\frac{\partial}{\partial x^i}\,,\,\frac{\partial}{\partial x^j}\,)$ are C^∞. Since each g_x is an inner product, the matrix (g_{ij}) is positive-definite at every $x \in M$. We can write

$$g = g_{ij\,(x)}\, dx^i \otimes dx^j \ .$$

This g defines a symmetric Finsler structure F on TM by the mechanism

$$F(x,y) := \sqrt{g_x(y,y)} \ .$$

Every Riemannian manifold (M,g) is therefore a Finsler manifold. A Finsler structure F is said to be **Riemannian** if it arises from a Riemannian metric g in the manner we just described. In practice, one ascertains this by showing that the fundamental tensor computed from F via (1.2.4) has no y dependence. As a matter of fact,

$$g_{ij} := \left(\frac{1}{2} F^2 \right)_{y^i y^j} = g_{ij\,(x)} \ .$$

Let us describe some fundamental Riemannian metrics. To this end, let $\mathfrak{s}_\lambda(t)$ be the unique solution to the ODE

$$\mathfrak{s}_\lambda'' + \lambda\, \mathfrak{s}_\lambda = 0 \ , \qquad \text{with initial data} \quad \mathfrak{s}_\lambda(0) = 0 \ , \ \mathfrak{s}_\lambda'(0) = 1 \ .$$

Here, λ is an arbitrary but *fixed* real number. Explicitly, we have

$$\mathfrak{s}_\lambda(t) = \begin{cases} \dfrac{\sin(\sqrt{\lambda}\, t)}{\sqrt{\lambda}} & \text{if } \lambda > 0 \\ t & \text{if } \lambda = 0 \\ \dfrac{\sinh(\sqrt{-\lambda}\, t)}{\sqrt{-\lambda}} & \text{if } \lambda < 0 \ . \end{cases} \tag{1.3.1}$$

Let (x^i) denote the natural coordinates of $\mathbb{R}^n$. At any point $x \in \mathbb{R}^n$, introduce the abbreviations

$$|x| := \sqrt{\delta_{ij}\, x^i x^j} \qquad \text{and} \qquad \psi := \left[\frac{\mathfrak{s}_\lambda(\,|x|\,)}{|x|} \right]^2$$

to avoid clutter. Define

$$\boxed{ g_{ij\,(x)} := (1 - \psi)\, \frac{(\delta_{ik}\, x^k)\, (\delta_{jl}\, x^l)}{|x|^2} + \psi\, \delta_{ij} } \ . \tag{1.3.2}$$

One can verify that:

* These g_{ij} can be extended smoothly to the origin $x = 0$.
* $g := g_{ij\,(x)}\, dx^i \otimes dx^j$ is a Riemannian metric on $\mathbb{R}^n$ if $\lambda \leqslant 0$.
* If $\lambda > 0$, our g is a Riemannian metric on the open ball

$$\left\{ x \in \mathbb{R}^n : |x| < \frac{\pi}{\sqrt{\lambda}} \right\} .$$

As we show in §13.3, these Riemannian metrics have constant sectional curvature λ. They are the **Riemannian space forms.**

1.3 C. Randers Spaces

In 1941, G. Randers [Ra] studied a very interesting class of Finsler manifolds. Let M be an n-dimensional manifold. A **Randers metric** is a Finsler structure F on TM that has the form

$$F(x,y) := \alpha(x,y) + \beta(x,y) , \tag{1.3.3}$$

where

$$\alpha(x,y) := \sqrt{\tilde{a}_{ij\,(x)}\, y^i\, y^j} , \tag{1.3.4}$$

$$\beta(x,y) := \tilde{b}_{i\,(x)}\, y^i . \tag{1.3.5}$$

- * The $\tilde{a}_{ij}$ are the components of a Riemannian metric and the $\tilde{b}_i$ are those of a 1-form. Both objects live on M, and are understood to be fixed throughout the discussion.
- * Due to the presence of the β term, Rander's metrics do *not* satisfy $F(x,-y) = F(x,y)$ when $\tilde{b} \neq 0$. In fact, the Finsler function of a Randers space is absolutely homogeneous if and only if it is Riemannian.

The indices on certain objects are lowered and raised by $(\tilde{a}_{ij})$ and its inverse matrix $(\tilde{a}^{ij})$. Such objects are decorated with a tilde.

Since $\beta(x,y)$ is linear in y, it cannot possibly have a fixed sign. Thus, in order for F to be positive on $TM \setminus 0$, the size of the components $\tilde{b}_i$ must be suitably controlled. It can be shown (see §11.1) that the said positivity holds if and only if

$$\|\tilde{b}\| := \sqrt{\tilde{b}_i\, \tilde{b}^i} < 1 , \tag{1.3.6}$$

where

$$\tilde{b}^i := \tilde{a}^{ij}\, \tilde{b}_j . \tag{1.3.7}$$

We also need to address the issue of strong convexity. The g_{ij} associated with F can be computed according to formula (1.2.4). One finds that

$$g_{ij} = \frac{F}{\alpha}\,(\tilde{a}_{ij} - \tilde{\ell}_i\,\tilde{\ell}_j) + (\tilde{\ell}_i + \tilde{b}_i)\,(\tilde{\ell}_j + \tilde{b}_j) , \tag{1.3.8}$$

where

$$\tilde{\ell}_i := \alpha_{y^i} = \frac{\tilde{a}_{ij}\, y^j}{\alpha} . \tag{1.3.9}$$

Equivalently,

$$g_{ij} = \frac{F}{\alpha}\,\tilde a_{ij} - \frac{\beta}{\alpha}\,\tilde\ell_i\,\tilde\ell_j + \tilde\ell_i\,\tilde b_j + \tilde\ell_j\,\tilde b_i + \tilde b_i\,\tilde b_j \,. \tag{1.3.10}$$

It turns out (see §11.1) that the criterion $\|\tilde b\| < 1$, which guarantees the positivity of F, also ensures strong convexity. And the crux of the argument involves the following computational fact:

$$\det(g_{ij}) = \left(\frac{F}{\alpha}\right)^{n+1} \det(\tilde a_{ij}) \,. \tag{1.3.11}$$

Its derivation can be found in [M2], albeit in the more general context of (α, β) metrics. A direct and expository account of (1.3.11) is given in §11.2.

Let us borrow an explicit example of a Randers metric from [AIM]. Set $M := \mathbb{R}^2 \setminus 0$. At each $x \in M$, the indicatrix is to have the following properties:

* It is an ellipse with eccentricity e, possibly depending on x, in the tangent plane T_xM.
* One of its foci is located at the orgin $y = 0$ of T_xM.
* The directrix (corresponding to the above focus) passes through the deleted point 0 of M, and is perpendicular to the line segment from 0 to x.

Using Okubo's technique (see Exercise 1.2.8), it can be shown that the formula for F is

$$\boxed{F(x^1, x^2; y^1, y^2) = \frac{1}{e}\sqrt{\frac{|y|^2}{|x|^2}} - \frac{x \bullet y}{|x|^2}} \,. \tag{1.3.12}$$

Here, we have introduced some temporary abbreviations

$$|y| := \sqrt{(y^1)^2 + (y^2)^2}$$
$$x \bullet y := x^1 y^1 + x^2 y^2$$

in order to avoid clutter.

1.3 D. Berwald Spaces

Berwald spaces are just a bit more general than Riemannian and locally Minkowskian spaces. They provide examples that are more properly Finslerian, but only slightly so. The most easily described characteristic of a **Berwald space** is that all its tangent spaces are linearly isometric to a common Minkowski space. One might say that the Berwald space in question is modeled on a single Minkowski space. For a precise definition of Berwald spaces, see Chapter 10.

We shall focus on Finsler structures F that are smooth and strongly convex on all of $TM \setminus 0$. Let us refer to these F as **y-global** for emphasis. As we show, it takes some work to explicitly locate a y-global Berwald space that is neither Riemannian nor locally Minkowskian. In fact, according to a rigidity result (see §10.6) of Szabó's, these do not even exist in dimension two. Fortunately, examples of the desired vintage can be found in dimension three or higher. The ones we know had their genesis in a result of Matsumoto [M4], Hashiguchi–Ichijyō [HI], Shibata–Shimada–Azuma–Yasuda [SSAY], and Kikuchi [Ki]. By contrast, y-local Berwald surfaces do exist, and an explicit example of such is analyzed in §10.3.

Let us quote (from §11.6) an example of **a 3-dimensional y-global Berwald space that is neither Riemannian nor locally Minkowskian**. It is given by a Randers metric constructed with the following data:

- The underlying manifold is the Cartesian product
$$M \ := \ S^2 \times S^1 \ .$$
It is compact and boundaryless. As local coordinates, one can use the usual spherical θ, ϕ on S^2, and t for S^1. For concreteness, we measure ϕ from the positive z axis down. Also, t is such that $(\cos t, \sin t, 0)$ parametrizes S^1.
- The Riemannian metric $\tilde{a}$ is the product metric on $S^2 \times S^1$. Here, S^2 and S^1 are given the standard Riemannian metrics that they inherited as submanifolds of Euclidean $\mathbb{R}^3$. Explicitly, one finds that
$$\tilde{a} \ := \ (\sin^2\!\phi \ d\theta \otimes d\theta \ + \ d\phi \otimes d\phi) \ + \ dt \otimes dt \ .$$
This metric is not flat because it has nonzero curvature tensor.
- The 1-form we need is
$$\tilde{b} \ := \ \epsilon \ dt \ ,$$
where ϵ is any (fixed) positive constant less than 1. This $\tilde{b}$ is globally defined on M, even though the coordinate t is not. It is nonvanishing by inspection, and has Riemannian norm $\| \tilde{b} \| = \epsilon < 1$. A straightforward calculation shows that it is parallel with respect to the Levi-Civita (Christoffel) connection of $\tilde{a}$.

We now write down the resulting Randers metric. Let x be any point on M, with coordinates (θ, ϕ, t). Let the arbitrary tangent vector $y \in T_xM$ be expanded as $y^\theta \, \partial_\theta + y^\phi \, \partial_\phi + y^t \, \partial_t$. Then

$$\boxed{F(x,y) \ := \ \sqrt{\sin^2\!\phi \ (y^\theta)^2 \ + \ (y^\phi)^2 \ + \ (y^t)^2} \ + \ \epsilon \ y^t} \ . \tag{1.3.13}$$

Since this F is of Randers type, its fundamental tensor is in principle given by (1.3.8) or (1.3.10), although a direct computation is probably more efficient. The reader is asked to do this calculation in Exercise 1.3.6.

1.3 E. Finsler Spaces of Constant Flag Curvature

An extensive discussion of Finsler spaces with **constant** (flag) **curvature** is given in Chapter 12.

There are non-Riemannian Finsler structures on $\mathbb{R}^2$ with negative constant Gaussian curvature. These are discussed in [Br3]. In §12.6, we construct one (known to Okada [Ok]) using the Yasuda–Shimada theorem [YS] as an inspiration (because we do not prove that theorem in this book). We then directly *verify* that **it has constant negative Gaussian curvature** $K = -\frac{1}{4}$. The explicit formula for the Finsler function is

$$\text{(1.3.14)} \quad \boxed{F(x,y) := \frac{1}{1-\frac{r^2}{4}} \sqrt{y \bullet y} + \frac{r}{\left(1-\frac{r^2}{4}\right)\left(1+\frac{r^2}{4}\right)} \, dr(y)} \, .$$

Here, $r^2 := (x^1)^2 + (x^2)^2$, where $x = (x^1, x^2)$ is any point on the Poincaré disc $M := \{x \in \mathbb{R}^2 : r < 2\}$. And y is an arbitrary vector in the tangent plane T_xM.

This is a very special Randers metric:

* It has constant negative (Finslerian) Gaussian curvature $-\frac{1}{4}$.
* It violates some completeness assumption in Akbar-Zadeh's [AZ] rigidity theorem.
* The Finslerian metric distance from the origin to the rim of the disc is infinite. But that coming back from the rim to the origin has the finite value $\log 2$!
* Its geodesics are, trajectorywise, the same as the geodesics of the Riemannian Poincaré disc.

For these reasons, we would like to view it as the **Finslerian analogue of the Poincaré disc.**

Next, we turn to positive flag curvatures. In two dimensions, we have explicit non-Riemannian examples with constant positive Gaussian curvature $K = 1$, due to Bryant [Br1, Br2]. Here, we focus on a 2-parameter family from [Br2]. Each Finsler structure in this family has $K = 1$ and is projectively flat. In [Br2], it is explained how these are related to some of Funk's earlier works [F1, F2].

Let V be a 3-dimensional real vector space with basis $\{b_1, b_2, b_3\}$. Let ρ, γ be two fixed angles satisfying

$$(*) \qquad |\gamma| \leqslant \rho < \frac{\pi}{2} \, .$$

Define a ρ and γ dependent, complex-valued quadratic form Q on V by

$$Q(u,v) := e^{i\rho} \, u^1 v^1 + e^{i\gamma} \, u^2 v^2 + e^{-i\rho} \, u^3 v^3 \, .$$

In the above exponentials, i means $\sqrt{-1}$. Also, $u = u^i \, b_i$; $v = v^i \, b_i$.

Let S^2 denote the set of rays in V. Equivalently, we are identifying X and X^* in V whenever $X^* = \lambda X$ for some $\lambda > 0$. Each point of S^2 can thus be denoted as an equivalence class $[X]$, with $0 \neq X \in V$. A moment's thought shows that every tangent vector at the point $[X]$ on S^2 is the initial velocity to a curve of the form $[X+tY]$, for some $Y \in V$. Each such curve is half of a great circle on S^2. And it makes sense to denote the said tangent vector by $[X,Y]$. Note that $[X',Y'] = [X,Y]$ if and only if $X' = \lambda X$ and $Y' = \lambda Y + \mu X$, for some $\lambda > 0$ and $\mu \in \mathbb{R}$.

The Finsler function $F : TS^2 \to [0,\infty)$ for **Bryant's family of metrics**, indexed by ρ and γ, is

(1.3.15)

$$\boxed{F([X,Y]) \ := \ \mathrm{Re}\left[\sqrt{\frac{Q(Y,Y)\,Q(X,X) - Q^2(X,Y)}{Q^2(X,X)}} \ - \ i\,\frac{Q(X,Y)}{Q(X,X)} \right]} \ ,$$

where "Re" means taking the real part. The complex square root function is taken to be branched along the negative real axis, and to satisfy $\sqrt{1} = 1$. In other words,

$$\mathrm{Re}\,\sqrt{a+ib} \ := \ + \sqrt{\frac{a \ + \ \sqrt{a^2+b^2}}{2}} \ .$$

It is not difficult to check that when $\rho = 0 = \gamma$, the above F is Riemannian. It is also instructive to work out a manifestly real formula of F for specific choices of ρ and γ. See §12.7 for a sample.

Bryant assures us that his methods in [Br2] give the following:

* The above F is indeed a Finsler structure in the sense of §1.1. Unless $\rho = 0 = \gamma$, this Finsler function is non-Riemannian and is only positively homogeneous.
* Each great semicircle $[X+tY]$ is a geodesic of the Finsler structure. (Incidentally, such curves are not yet parametrized to have constant speed. Nevertheless, the Finslerian length of each great circle is 2π.)
* The Gaussian curvature of the Finsler surface (S^2, F) has the constant positive value 1.

Exercises

Exercise 1.3.1: Show that the Minkowski spaces arising from the Minkowski norms

$$F(y) \ := \ \sqrt{\sqrt{\sum_{i=1}^{n} (y^i)^4} \ + \ \lambda \sum_{i=1}^{n} (y^i)^2}$$

are all nonisometric for different values of $\lambda \geqslant 0$.

Exercise 1.3.2: Recall the g_{ij} defined in (1.3.2). Prove that they can be smoothly extended to the origin $x = 0$ of $\mathbb{R}^n$.

Exercise 1.3.3: This again concerns (1.3.2). Show that

(a) For $\lambda \leqslant 0$, the matrix (g_{ij}) is positive-definite at every $x \in \mathbb{R}^n$.

(b) For $\lambda > 0$, the matrix (g_{ij}) is positive-definite if $|x| < \frac{\pi}{\sqrt{\lambda}}$.

Exercise 1.3.4: Verify formula (1.3.8) for the g_{ij} of Randers metrics.

Exercise 1.3.5: (Borrowed from [AIM].) Let $M := \mathbb{R}^2 \setminus 0$. Suppose at each $x \in M$, we want the indicatrix to have the following properties:

* It is an ellipse with eccentricity e in the tangent plane T_xM.
* One of its foci is located at the origin $y = 0$ of T_xM.
* The directrix (corresponding to the above focus) passes through the deleted point 0 of M, and is perpendicular to the line segment from 0 to x.

(a) Draw a picture of what we have just described.

(b) Use Okubo's technique to derive formula (1.3.12) for F.

(c) For this Randers metric, identify $\tilde{a}_{ij}$ and $\tilde{b}_i$.

(d) Directly compute the fundamental tensor and its determinant.

Exercise 1.3.6: For the 3-dimensional Berwald space given in (1.3.13), compute directly the fundamental tensor and its determinant.

Exercise 1.3.7: Show that the Finsler function F, given in (1.3.15), is well defined on TS^2. In other words, that expression for F remains unchanged upon replacing X by λX and Y by $\lambda Y + \mu X$, where $\lambda > 0$ and $\mu \in \mathbb{R}$.

1.4 The Fundamental Tensor and the Cartan Tensor

Let F be a Minkowski norm on $\mathbb{R}^n$. We have seen the utility of the functions

(1.4.1) $$\boxed{g_{ij} \;:=\; \left(\frac{1}{2}\, F^2\right)_{y^i y^j} \;=\; F\, F_{y^i y^j} \;+\; F_{y^i}\, F_{y^j}}$$

in §1.2. Next define

(1.4.2) $$\boxed{A_{ijk\,(y)} \;:=\; \frac{F}{2}\,\frac{\partial g_{ij}}{\partial y^k} \;=\; \frac{F}{4}\,(F^2)_{y^i y^j y^k}}\,,$$

which are manifestly symmetric in the three indices i, j, k. All these functions are homogeneous of degree zero. In other words, they are invariant under the rescaling $y \mapsto \lambda y$. In a context that we postpone describing until Chapter 2, the g_{ij} and the A_{ijk} are, respectively, the components of two important tensors, called the **fundamental tensor** and the **Cartan tensor**. Incidentally, some authors have chosen to call

$$C_{ijk} \;:=\; \frac{1}{F}\, A_{ijk}$$

the Cartan tensor instead.

Exercise 1.2.1 enumerated some specific consequences of Euler's theorem. Using these, one gets the following identities:

$$\boxed{g_{ij}\,\frac{y^i}{F} \;=\; F_{y^j}}\,, \tag{1.4.3}$$

$$g_{ij}\,\frac{y^i}{F}\,\frac{y^j}{F} \;=\; 1\,, \tag{1.4.4}$$

$$y^i\,\frac{\partial g_{ij}}{\partial y^k} \;=\; 0\,, \quad y^j\,\frac{\partial g_{ij}}{\partial y^k} \;=\; 0\,, \quad y^k\,\frac{\partial g_{ij}}{\partial y^k} \;=\; 0\,. \tag{1.4.5}$$

The last one can be re-expressed as

$$y^i\,A_{ijk\,(y)} \;=\; y^j\,A_{ijk\,(y)} \;=\; y^k\,A_{ijk\,(y)} \;=\; 0\,. \tag{1.4.6}$$

The g_{ij} define a natural Riemannian metric

$$g_{ij\,(y)}\;dy^i \otimes dy^j$$

on the punctured linear manifold $\mathbb{R}^n \setminus 0$. Here, we use $dy^i \otimes dy^j$ instead of $dx^i \otimes dx^j$ because, throughout our discussions of Minkowski norms, the natural coordinates on $\mathbb{R}^n$ have been denoted by y^i and not x^i. Some features of this punctured Riemannian manifold are worth noting. We list them here and refer the details to the exercises at the end of this section.

- Its volume form is *chosen* to be $\sqrt{g}\;dy^1\wedge\cdots\wedge dy^n$, where $\sqrt{g}$ stands for $\sqrt{\det(g_{ij})}$.
- It admits the hypersurfaces $\mathrm{S}(r) := \{y \in \mathbb{R}^n : F(y) = r\}$ as smooth Riemannian submanifolds.
- Each $\mathrm{S}(r)$ is the boundary of a strictly convex domain and, with respect to $g_{ij}\;dy^i \otimes dy^j$, its outward-pointing unit normal [Y] is

$$\hat{n}_{\text{out}} \;:=\; \frac{y^i}{F}\,\frac{\partial}{\partial y^i}\,. \tag{1.4.7}$$

- The volume form of the Riemannian submanifold $\mathrm{S}(r)$ is

$$\sqrt{g}\;\sum_{j=1}^{n}\;(-1)^{j-1}\,\frac{y^j}{F}\;dy^1\wedge\cdots\wedge dy^{j-1}\wedge dy^{j+1}\wedge\cdots\wedge dy^n\,. \tag{1.4.8}$$

The significance of the A_{ijk} lies in the fact that their vanishing characterizes Euclidean norms among Minkowski norms.

Theorem 1.4.1 (Deicke) [D]. *Let F be a Minkowski norm on $\mathbb{R}^n$. The following three statements are equivalent:*

(a) *F is Euclidean. That is, it arises from an inner product.*

(b) $A_{ijk} = 0$ *for all* i, j, k.

(c) $A_k := g^{ij} A_{ijk} = 0$ *for all* k. *Here,* (g^{ij}) *denotes the inverse matrix of* (g_{ij}).

Remarks:

* The equivalence between the first two statements comes directly from (1.4.2). Namely, A_{ijk} is proportional to the vertical derivative of g_{ij}. Thus this derivative vanishes if and only if g_{ij} has no y-dependence (which means that F comes from an inner product). This equivalence between (a) and (b) constitutes the easy part of Theorem 1.4.1, and is used without mention.

* It is clear that (b) implies (c). Thus it remains to prove the converse or, equivalently, that (c) implies (a). Such is a result of Deicke's. However, the proof that we give for (c) $\Rightarrow$ (a) is due to Brickell [B1]. It involves tools that we have yet to learn. For this reason, the proof is postponed until Chapter 14, which studies Minkowski spaces. Needless to say, the implication (c) $\Rightarrow$ (a) is never used before that proof.

Exercises

Exercise 1.4.1: Let F be a Minkowski norm on $\mathbb{R}^n$. Show that under a change of coordinates of the special type $\tilde{y}^p = c^p{}_i \, y^i$, where the $c^p{}_i$ are *constants*, the g_{ij} and A_{ijk} transform like:

$$\begin{aligned} g_{ij} &= c^p{}_i \, c^q{}_j \, \tilde{g}_{pq} \,, \\ A_{ijk} &= c^p{}_i \, c^q{}_j \, c^r{}_k \, \tilde{A}_{pqr} \,. \end{aligned}$$

Exercise 1.4.2: On $\mathrm{S}(r)$, the function F has the constant value r.

(a) Take the directional derivative of the above statement along an arbitrary vector $v^i \frac{\partial}{\partial y^i}$ which is *tangent* to $\mathrm{S}(r)$.

(b) Re-express your answer with the help of the identity (1.4.3).

(c) Interpret what you obtain in order to conclude (1.4.7). Be sure to explain why that $\hat{n}_{\mathrm{out}}$ has unit length.

Exercise 1.4.3: Suppose we have an ambient Riemannian manifold (M, g) with volume form dV, and a submanifold S with outward-pointing unit normal field $\hat{n}_{\mathrm{out}}$. By simply restricting g to vectors tangent to S, one induces a Riemannian metric on S, whose volume form can be obtained by contracting $\hat{n}_{\mathrm{out}}$ into the first slot of dV. Use this procedure to derive formula (1.4.8).

Exercise 1.4.4: Refer to Exercise 1.3.1, where we introduced a family of Minkowski norms, indexed by a parameter λ. At each $y \in \mathbb{R}^n \smallsetminus 0$, define

$$\| A \|_y \ := \ \sup_U \ A_{ijk\,(y)} \ \frac{U^i\, U^j\, U^k}{(\ \sqrt{g_{pq\,(y)}}\ U^p\, U^q\ )^3} \ .$$

Also, set $\| A \| := \sup_{y \neq 0} \ \| A \|_y$. Show that there is a constant c, depending only on the dimension n, such that

$$\| A \| \ \geqslant \ c\,(\lambda - 1)\ .$$

References

[AB] P. L. Antonelli and R. H. Bradbury, *Volterra–Hamilton Models in the Ecology and Evolution of Colonial Organisms*, World Scientific, 1996.

[AIM] P. L. Antonelli, R. S. Ingarden, and M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology*, FTPH 58, Kluwer Academic Publishers, 1993.

[AZ] H. Akbar-Zadeh, *Sur les espaces de Finsler à courbures sectionnelles constantes*, Acad. Roy. Belg. Bull. Cl. Sci. (5) **74** (1988), 281–322.

[B1] F. Brickell, *A new proof of Deicke's theorem on homogeneous functions*, Proc. AMS **16** (1965), 190–191.

[BC2] D. Bao and S.S. Chern, *A note on the Gauss–Bonnet theorem for Finsler spaces*, Ann. Math. **143** (1996), 233–252.

[Br1] R. Bryant, *Finsler structures on the 2-sphere satisfying $K = 1$*, Cont. Math. **196** (1996), 27–42.

[Br2] R. Bryant, *Projectively flat Finsler 2-spheres of constant curvature*, Selecta Mathematica, New Series **3** (1997), 161–203.

[Br3] R. Bryant, *Finsler surfaces with prescribed curvature conditions*, Aisenstadt Lectures, in preparation.

[D] A. Deicke, *Über die Finsler-Raume mit $A_i = 0$*, Arch. Math. **4** (1953), 45–51.

[F1] P. Funk, *Über zweidimensionale Finslersche Räume, insbesondere über solche mit geradlinigen Extremalen und positiver konstanter Krümmung*, Math. Zeitschr. **40** (1936), 86–93.

[F2] P. Funk, *Eine Kennzeichnung der zweidimensionalen elliptischen Geometrie*, Österreichische Akad. der Wiss. Math., Sitzungsberichte Abteilung II **172** (1963), 251–269.

[HI] M. Hashiguchi and Y. Ichijyō, *On some special (α, β)-metrics*, Rep. Fac. Sci. Kagoshima Univ. **8** (1975), 39–46.

[Ki] S. Kikuchi, *On the condition that a space with (α, β)-metric be locally Minkowskian*, Tensor, N.S. **33** (1979), 242–246.

[M1] M. Matsumoto, *A slope of a mountain is a Finsler surface with respect to a time measure*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **29-1** (1989), 17–25.

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

[M4] M. Matsumoto, *On Finsler spaces with Randers' metric and special forms of important tensors*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **14** (1974), 477–498.

[Ok] T. Okada, *On models of projectively flat Finsler spaces of constant negative curvature*, Tensor, N.S. **40** (1983), 117–124.

[R] H. Rund, *The Differential Geometry of Finsler Spaces*, Springer-Verlag, 1959.

[Ra] G. Randers, *On an asymmetric metric in the four-space of general relativity*, Phys. Rev. **59** (1941), 195–199.

[SSAY] C. Shibata, H. Shimada, M. Azuma, and H. Yasuda, *On Finsler spaces with Randers' metric*, Tensor, N.S. **31** (1977), 219–226.

[Y] H. Yasuda, *On the indicatrices of a Finsler space*, Tensor, N.S. **33** (1979), 213–221.

[YS] H. Yasuda and H. Shimada, *On Randers spaces of scalar curvature*, Rep. on Math. Phys. **11** (1977), 347–360.

Chapter 2

The Chern Connection

2.0 Prologue

The Chern connection that we construct is a linear connection that acts on a distinguished vector bundle π^*TM, sitting over the manifold $TM \setminus 0$ or SM. It is *not* a connection on the bundle TM over M. Nevertheless, it serves Finsler geometry in a manner that parallels what the Levi-Civita (Christoffel) connection does for Riemannian geometry. This connection is on equal footing with, but is different from, those due to Cartan, Berwald, and Hashiguchi (to name just a few).

In the exercise portion of §5.2, we use this linear connection to induce nonlinear covariant derivatives on M. These derivatives involve "correction" terms highlighted by certain connection coefficients. Our covariant derivatives are nonlinear in the generic Finsler setting because the said connection coefficients have a dependence on either the direction of differentiation or the vector that is being differentiated. Such connection coefficients reduce to the usual Christoffel symbols when the Finsler structure is Riemannian. In that case, the corresponding covariant derivative on M becomes the familiar one due to Levi-Civita (Christoffel).

2.1 The Vector Bundle π^*TM and Related Objects

Recall the fundamental tensor $g_{ij\,(x,y)}$ that we introduced in (1.2.4) and revisited in (1.4.1). It is defined at all $(x, y) \in TM \setminus 0$, and is invariant under positive rescaling in y. One could imagine the formal object $g_{ij}\, dx^i \otimes dx^j$. That would have behaved exactly like an inner product on the tangent space T_xM, if it were not for the dependence on $y \neq 0$. Happily, this conceptual difficulty can be overcome without too much trouble. We first give a heuristic description of the resolution, followed by a more technical version.

The collection of all (x, y), with $y \neq 0$, constitutes the slit tangent bundle $TM \setminus 0$. Let us view it as a parameter space. Over each point (x, y) in this parameter space, we erect a copy of T_xM. We then form $g_{ij\,(x,y)} dx^i \otimes dx^j$ and declare it the inner product on this T_xM. Note that the vector space T_xM is determined solely by the position parameters x in (x, y). The directional parameters y have no say in this matter. Collectively, the vector spaces we have erected form a vector bundle (with fiber dimension n) over the parameter space $TM \setminus 0$, which is $2n$-dimensional. What we have just described is depicted schematically in Figure 2.1.

There is, however, some redundancy in the above scheme. Consider all points in $TM \setminus 0$ of the form $(x, \lambda y)$, with x, y fixed and λ an arbitrary positive number. Over each such point, we have erected the *same* vector space T_xM. Since $g_{ij\,(x,y)}$ is invariant under the rescaling $y \mapsto \lambda y$, the inner products we assigned to these copies of T_xM are also *identical.* This is the redundancy to which we referred. There is a simple way to restore

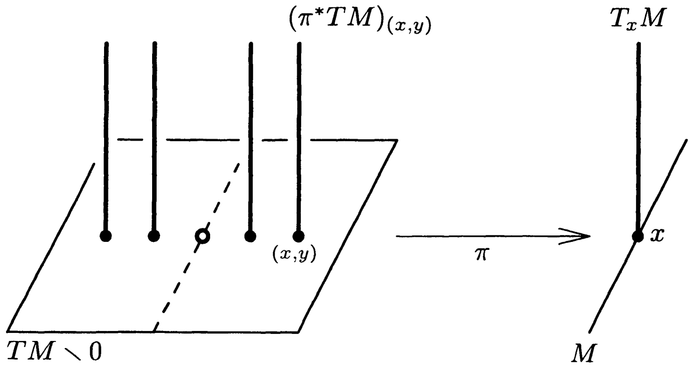

Figure 2.1

The pulled-back tangent bundle π^*TM is a vector bundle over the slit tangent bundle $TM \setminus 0$. The fiber over a typical point (x, y) is a copy of T_xM. The dotted part is the deleted zero section.

economy. First, we can treat the ray $\{(x, \lambda y) : \lambda > 0\}$ as a single point in the projective sphere bundle SM. Then over that point, we erect a single copy of T_xM and endow it with the inner product $g_{ij\,(x,y)} dx^i \otimes dx^j$. The resulting vector bundle has fiber dimension n as before, but now it sits over the $2n - 1$ dimensional sphere bundle SM. In case the Finsler structure is absolutely homogeneous rather than positively homogeneous, we can economize further by replacing SM with the projectivized tangent bundle PTM. The latter treats each line $\{(x, \lambda y) : \lambda \in \mathbb{R}\}$ as a single point.

Now we give the technical equivalent of what has just been said. Over M, we have the tangent bundle TM and the cotangent bundle T^*M. Much of Finsler geometry's tensor calculus can be handled in any one of the following two environments.

- **For analytical and global purposes**, it is often advantageous to work with a compact parameter space. In that case, the **base manifold** of choice is the sphere bundle SM, or PTM when F happens to be absolutely homogeneous (of degree one). Let $p : SM \to M$ be the canonical projection map. A good number of our geometrical objects are sections of the **pulled-back bundle** p^*TM or its dual p^*T^*M, or their tensor products. These bundles sit over SM and not M.
- **For ease of local computations**, it is to our advantage to work on an affine parameter space, where natural coordinates are readily available. In that case, the preferred **base manifold** is the slit tangent bundle $TM \setminus 0$. A good number of our geometrical objects are sections of the **pulled-back bundle** π^*TM or its dual π^*T^*M, or their tensor products. These sit over $TM \setminus 0$ and not M.

There is no consensus among Finsler geometers as to which approach one should take in a *book*. And picking one instead of the other will inevitably render the book less useful to some. Happily, there is a way to retain the essence of both approaches.

> We work on $TM \setminus 0$. But, unlike other authors who prefer the slit tangent bundle, we rigidly use only objects that are invariant under positive rescaling in all our important computations. For example, we use $\frac{1}{F} N^i{}_j$ and $\frac{1}{F} \delta y^i$ instead of $N^i{}_j$ and δy^i by themselves. This way, all calculations can be done in natural coordinates. And, since all the steps are manifestly invariant under positive rescaling in y, one can correctly view them as having been carried out on the sphere bundle SM (or PTM) using homogeneous coordinates.

Local coordinates (x^i) on M produce the basis sections $\{\frac{\partial}{\partial x^i}\}$ and $\{dx^i\}$, respectively, for TM and T^*M. Now, over each point (x, y) on the manifold $TM \setminus 0$, the fiber of π^*TM is the vector space T_xM while that of π^*T^*M is the covector space T^*_xM. Thus, the $\frac{\partial}{\partial x^i}$ and dx^i give rise to sections of the pulled-back bundles, in a rather simple-minded way. These **transplanted sections** are defined locally in x and globally in y. This global nature in y is automatic because once x is fixed, the transplanted sections do not change as we vary y.

In order to keep the notation simple, we also use the symbols $\frac{\partial}{\partial x^i}$ and dx^i to denote the transplanted sections.

We hope this does not cause too much confusion.

There is a **distinguished section** ℓ of π^*TM. It is defined by

$$\boxed{\ell = \ell_{(x,y)} := \frac{y^i}{F(y)} \frac{\partial}{\partial x^i} = \frac{y^i}{F} \frac{\partial}{\partial x^i} =: \ell^i \frac{\partial}{\partial x^i}} \,. \tag{2.1.1}$$

Its natural dual is the **Hilbert form** ω, which is a section of π^*T^*M. We have

$$\boxed{\omega = \omega_{(x,y)} := F_{y^i}(x,y)\, dx^i = F_{y^i}\, dx^i} \,. \tag{2.1.2}$$

Both ℓ and ω are globally defined on the manifold $TM \setminus 0$. The asserted duality means that

$$\omega(\ell) = \frac{y^i}{F} F_{y^i} = 1 \,,$$

which is a consequence of Euler's theorem. See (1.2.1).

The pulled-back vector bundle π^*TM admits a natural Riemannian metric

$$g = g_{ij}\, dx^i \otimes dx^j \,. \tag{2.1.3}$$

This is the **fundamental tensor** that we have alluded to before. It is a symmetric section of $\pi^*T^*M \otimes \pi^*T^*M$. Likewise, there is the **Cartan tensor**

$$A = A_{ijk}\, dx^i \otimes dx^j \otimes dx^k \,, \tag{2.1.4}$$

which is a symmetric section of $\otimes^3 \pi^*T^*M$. In these formulas, we have suppressed the point of evaluation (x,y) in order to avoid clutter. The components g_{ij} and A_{ijk} have already been defined in (1.4.1) and (1.4.2). We reproduce them here for convenience:

$$\boxed{g_{ij} := \left(\frac{1}{2} F^2\right)_{y^i y^j} = F\, F_{y^i y^j} + F_{y^i}\, F_{y^j}} \,, \tag{2.1.5}$$

$$\boxed{A_{ijk} := \frac{F}{2} \frac{\partial g_{ij}}{\partial y^k} = \frac{F}{4} (F^2)_{y^i y^j y^k}} \,. \tag{2.1.6}$$

As we mentioned before, the object

$$C_{ijk} := \frac{1}{F} A_{ijk}$$

is called the Cartan tensor in the literature at large. But we prefer to work with quantities that are homogeneous of degree 0 because they make sense on the (projective) sphere bundle SM.

By another consequence [see (1.2.5)] of Euler's theorem, we find that our distinguished section ℓ has norm 1 with respect to the said Riemannian metric:

$$g(\ell, \ell) = g_{ij} \frac{y^i}{F} \frac{y^j}{F} = 1 . \tag{2.1.7}$$

Exercises

Exercise 2.1.1: Consider the components $\ell^i := \frac{y^i}{F}$ of our distinguished section ℓ. Show that

$$\boxed{\ell_i := g_{ij}\, \ell^j = F_{y^i}} .$$

Thus the Hilbert form is expressible as $\omega = \ell_i\, dx^i$.

Exercise 2.1.2: Let $\sigma : [a, b] \to M$ be a piecewise C^∞ curve. Its integral length is defined as $L(\sigma) := \int_a^b F(\sigma, \frac{d\sigma}{dt})\, dt$. Prove that this length can be re-expressed as the line integral of the Hilbert form along σ. Namely:

$$L(\sigma) = \int_\sigma \omega .$$

The right-hand side is sometimes called **Hilbert's invariant integral.**

Exercise 2.1.3:

(a) Check that A_{ijk} is totally symmetric on all its indices.

(b) Show that $\ell^i A_{ijk} = 0$.

Exercise 2.1.4: Show that $\frac{\partial g^{ij}}{\partial y^k} = -\frac{2}{F} A^{ij}{}_k$.

2.2 Coordinate Bases Versus Special Orthonormal Bases

We have described how the coordinate bases $\{\frac{\partial}{\partial x^i}\}$ and $\{dx^i\}$, respectively, for TM and T^*M, can be transplanted from M to the manifold $TM \setminus 0$. These transplants form basis sections of the bundles π^*TM and π^*T^*M. We have also decided, in the name of keeping things simple, *not* to create new symbols for these sections.

Occasionally, the exposition can benefit from the use of g-orthonormal basis sections. The ones we use are defined as follows:

- For π^*TM, the **special g-orthonormal basis** $\{e_a\}$ must satisfy

$$g(e_a, e_b) = \delta_{ab}$$
$$e_n = \ell , \quad \text{the distinguished section.}$$

- For $\pi^* T^* M$, the special basis $\{\omega^b\}$ is dual to $\{e_a\}$, thus

$$\omega^b(e_a) = \delta^b{}_a$$
$$\omega^n = \omega \,, \qquad \text{the Hilbert form.}$$

Note that:

* The special g-orthonormal bases we have just introduced make sense only on the manifold $TM \setminus 0$; in general they have no analogues on M. This is because the natural Riemannian metric g lives on $\pi^* TM$ and not TM, unless F is Riemannian.

* Furthermore, we have unequivocally specified only the last member in each such basis. The residual "gauge" freedom is equal to the orthogonal group $O(n-1)$.

The bases $\{\frac{\partial}{\partial x^i}\}$ and $\{e_a\}$ can be expressed in terms of each other. The same can be said of $\{dx^i\}$ and $\{\omega^a\}$. That is,

(2.2.1) $$e_a = u_a{}^i \, \frac{\partial}{\partial x^i}$$

(2.2.2) $$\frac{\partial}{\partial x^i} = v^a{}_i \, e_a$$

(2.2.3) $$\omega^a = v^a{}_i \, dx^i$$

(2.2.4) $$dx^i = u_a{}^i \, \omega^a \,.$$

- A basic relationship between $(u_a{}^i)$ and $(v^a{}_i)$ is that they are matrix inverses of each other. Namely,

$$v^a{}_i \, u_b{}^i = \delta^a{}_b \,,$$

$$u_a{}^i \, v^a{}_j = \delta^i{}_j \,.$$

- Since we have already specified that

$$e_n := \ell = \ell_i \, \frac{\partial}{\partial x^i} \,,$$

$$\omega^n := \omega = F_{y^i} \, dx^i = \ell_i \, dx^i \,,$$

we see that $u_n{}^i = \ell^i$ and $v^n{}_i = \ell_i$.

These, together with other identities of interest, are addressed in the exercises below.

Exercises

Exercise 2.2.1: Verify the following statements:

(a) $v^a{}_i \, u_b{}^i = \delta^a{}_b$.

(b) $u_a^{\ i} \, v^a_{\ j} = \delta^i_{\ j}$.
(c) $v^\alpha_{\ i} \, y^i = 0$.
(d) $u_\alpha^{\ i} \, F_{y^i} = 0$.

Exercise 2.2.2: Show that

(a) $u_a^{\ i} \, g_{ij} \, u_b^{\ j} = \delta_{ab}$.
(b) $v^a_{\ i} \, g^{ij} \, v^b_{\ j} = \delta^{ab}$.
(c) $g_{ij} = v^a_{\ i} \, \delta_{ab} \, v^b_{\ j}$.
(d) $g^{ij} = u_a^{\ i} \, \delta^{ab} \, u_b^{\ j}$.

Exercise 2.2.3: Deduce the following identities.

(a) $u_a^{\ i} = \delta_{ab} \, v^b_{\ j} \, g^{ji}$.
(b) $v^a_{\ i} = \delta^{ab} \, u_b^{\ j} \, g_{ji}$.
(c) $v^\alpha_{\ i} \, v^\beta_{\ j} \, \delta_{\alpha\beta} = F \, F_{y^i y^j}$.
(d) $u_\alpha^{\ i} \, u_\beta^{\ j} \, F \, F_{y^i y^j} = \delta_{\alpha\beta}$.

Hint: you may need to use the results of the above exercises.

2.3 The Nonlinear Connection on the Manifold $TM \setminus 0$

The components g_{ij} [see (2.1.5)] of the fundamental tensor are functions on $TM \setminus 0$, and are invariant under positive rescaling in y. We use them to define the **formal Christoffel symbols** of the second kind

$$\boxed{\gamma^i_{\ jk} := g^{is} \, \frac{1}{2} \left(\frac{\partial g_{sj}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^s} + \frac{\partial g_{ks}}{\partial x^j} \right)} \,, \tag{2.3.1}$$

and also the quantities

$$\boxed{N^i_{\ j} := \gamma^i_{\ jk} \, y^k - C^i_{\ jk} \, \gamma^k_{\ rs} \, y^r \, y^s} \,. \tag{2.3.2a}$$

Our preference for objects invariant under $y \mapsto \lambda y$ dictates that we work with

$$\boxed{\frac{N^i_{\ j}}{F} := \gamma^i_{\ jk} \, \ell^k - A^i_{\ jk} \, \gamma^k_{\ rs} \, \ell^r \, \ell^s} \tag{2.3.2b}$$

instead.

The transformation law for $\frac{1}{F} N^i_{\ j}$ is quite elegant. Let

$$x^i = x^i(\tilde{x}^1, \ldots, \tilde{x}^n)$$

be a local change of coordinates on M. Correspondingly, the chain rule gives

$$y^i = \frac{\partial x^i}{\partial \tilde{x}^p} \, \tilde{y}^p \,.$$

It can be shown, albeit with some tedium, that

$$\frac{1}{\tilde{F}}\,\tilde{N}^p{}_q \;=\; \frac{\partial \tilde{x}^p}{\partial x^i}\,\frac{\partial x^j}{\partial \tilde{x}^q}\,\frac{N^i{}_j}{F} \;+\; \frac{\partial \tilde{x}^p}{\partial x^i}\,\frac{\partial^2 x^i}{\partial \tilde{x}^q \partial \tilde{x}^s}\,\tilde{\ell}^s \,.$$

Transformation laws of various geometrical objects are treated systematically elsewhere.

We show momentarily why the above $N^i{}_j$ are known in the trade as the **nonlinear connection** on $TM \setminus 0$.

In order to set the stage for this realization, observe that:

* The tangent bundle of the manifold TM has a local coordinate basis that consists of the $\frac{\partial}{\partial x^j}$ and the $\frac{\partial}{\partial y^j}$. However, under the transformation on TM induced by a coordinate change on M, the vectors $\frac{\partial}{\partial x^j}$ transform in a somewhat complicated manner, as exhibited in (1.1.3). On the other hand, the $\frac{\partial}{\partial y^j}$ do not have this "problem."

* The cotangent bundle of TM has a local coordinate basis $\{dx^i, dy^i\}$. Here, under the said transformation, the dx^i behave simply while the dy^i do not. See Exercise 2.3.1.

The remedy lies in replacing $\frac{\partial}{\partial x^j}$ by

(2.3.3) $$\boxed{\frac{\delta}{\delta x^j} \;:=\; \frac{\partial}{\partial x^j} \;-\; N^i{}_j\,\frac{\partial}{\partial y^i}}\,,$$

and dy^i by

(2.3.4a) $$\boxed{\delta y^i \;:=\; dy^i \;+\; N^i{}_j\, dx^j}\,.$$

As before, we prefer to work with

(2.3.4b) $$\boxed{\frac{\delta y^i}{F} \;=\; \frac{1}{F}\,(\, dy^i \;+\; N^i{}_j\, dx^j\,)}\,,$$

which is invariant under rescaling in y. Note that

(2.3.5)
$$\frac{\delta}{\delta x^j} \quad \xrightarrow[\text{dual}]{\text{natural}} \quad dx^j\,,$$
$$F\,\frac{\partial}{\partial y^i} \quad \xrightarrow[\text{dual}]{\text{natural}} \quad \frac{\delta y^i}{F}\,.$$

- These objects typically only make sense on $TM \setminus 0$. With the exception of the dx^i, the rest are *nonholonomic*. That is, they are neither coordinate vector fields nor coordinate 1-forms.
- They indeed have simple behavior under transformations induced by coordinate changes on M. The verification of this claim though, is quite tedious. It is not pursued in the current book.

It turns out that the manifold $TM \smallsetminus 0$ has a natural Riemannian metric

$$g_{ij}\, dx^i \otimes dx^j \;+\; g_{ij}\, \frac{\delta y^i}{F} \otimes \frac{\delta y^j}{F} \;,$$

known as a **Sasaki** (type) **metric**. With respect to this metric, the **horizontal subspace** spanned by the $\frac{\delta}{\delta x^j}$ is orthogonal to the **vertical subspace** spanned by the $F\frac{\partial}{\partial y^i}$. The manifold $TM \smallsetminus 0$ therefore admits an **Ehresmann connection** through this splitting, and the latter owes its existence directly to the quantities $N^i{}_j$. This is why the $N^i{}_j$ are collectively known as the nonlinear connection.

We have just introduced two new natural (local) bases that are dual to each other:

* $\{ \frac{\delta}{\delta x^i} \,,\, F\frac{\partial}{\partial y^i} \}$ for the tangent bundle of $TM \smallsetminus 0$,
* $\{ dx^i \,,\, \frac{\delta y^i}{F} \}$ for the cotangent bundle of $TM \smallsetminus 0$.

Since $TM \smallsetminus 0$ is a Riemannian manifold with the Sasaki (type) metric, the above bases have orthonormal analogues:

* $\{ \hat{e}_a \,,\, \hat{e}_{n+a} \}$ for $T(TM \smallsetminus 0)$,
* $\{ \omega^a \,,\, \omega^{n+a} \}$ for $T^*(TM \smallsetminus 0)$.

The relationship between the natural bases and the orthonormal ones can be readily written down, thanks to the "**n-beins**" $u_a{}^i$ and $v^a{}_i$ we encountered at the end of §2.2 :

(2.3.6) $$\hat{e}_a \;=\; u_a{}^i\, \frac{\delta}{\delta x^i}$$

(2.3.7) $$\hat{e}_{n+a} \;=\; u_a{}^i\, F\frac{\partial}{\partial y^i}$$

(2.3.8) $$\omega^a \;=\; v^a{}_i\, dx^i$$

(2.3.9) $$\omega^{n+a} \;=\; v^a{}_i\, \frac{\delta y^i}{F} \;.$$

Recall that the n-beins were originally introduced to relate the natural and the g-orthonormal bases of the pulled-back bundle π^*TM, which sits *over* the manifold $TM \smallsetminus 0$. Surprisingly, they also serve as a go-between for the "natural" versus the Sasaki-orthonormal bases on $TM \smallsetminus 0$.

Exercises

Exercise 2.3.1: Consider local coordinate changes on M, say

$$x^i \;=\; x^i(\tilde{x}^1, \ldots, \tilde{x}^n) \;,$$

and their inverses $\tilde{x}^p = \tilde{x}^p(x^1, \ldots, x^n)$. They induce transformations on the manifold TM, as described at the end of §1.1. Show that, as differential forms on TM, the dx^i and dy^i behave as follows:

$$d\tilde{x}^p = \frac{\partial \tilde{x}^p}{\partial x^i}\, dx^i \,,$$

$$d\tilde{y}^p = \frac{\partial \tilde{x}^p}{\partial x^i}\, dy^i + \frac{\partial^2 \tilde{x}^p}{\partial x^i \partial x^j}\, y^j\, dx^i \,.$$

Exercise 2.3.2: Define the quantity

$$G_i := g_{is}\, \gamma^s{}_{jk}\, y^j\, y^k =: \gamma_{ijk}\, y^j\, y^k \,.$$

Show that

$$\frac{1}{2}\, \frac{\partial G_i}{\partial y^j} = \gamma_{ijk}\, y^k \,.$$

Exercise 2.3.3: Let

$$\boxed{G^i := \gamma^i{}_{jk}\, y^j\, y^k} \,.$$

This is *twice* the G^i in [AIM]. By combining the above exercise with the conclusion of Exercise 2.1.4, prove that

$$\boxed{\frac{1}{2}\, \frac{\partial G^i}{\partial y^j} = N^i{}_j} \,.$$

We later show that Finslerian geodesics are curves in M which obey the equation $\ddot{x}^i + G^i = 0$, where in G^i we set $y^i := \dot{x}^i$. Thus, if the geodesic equation is somehow known (say, through a shortcut), the nonlinear connection $N^i{}_j$ can be computed without having to first calculate the Cartan tensor $A^i{}_{jk}$ and the formal Christoffel symbols $\gamma^i{}_{jk}$.

Exercise 2.3.4: Directly rewrite the Sasaki type metric on $TM \setminus 0$ as

$$\delta_{ab}\, \omega^a \otimes \omega^b + \delta_{ab}\, \omega^{n+a} \otimes \omega^{n+b} \,.$$

Exercise 2.3.5:

(a) Recall the fact $\ell_i = F_{y^i}$. Prove that

$$\boxed{\ell_i\, \frac{\delta y^i}{F} = d(\log F) = \omega^{n+n}} \,.$$

(b) A curve in $TM \setminus 0$ is said to be *horizontal* if all its velocity vectors are horizontal, in the sense defined near the end of this section. Explain why F is constant along all horizontal curves. Namely,

$$\boxed{\frac{\delta F}{\delta x^i} = 0 \quad \text{for all } i} \,.$$

2.4 The Chern Connection on π^*TM

The distinguished section

$$\ell := \frac{y^i}{F} \frac{\partial}{\partial x^i} = \ell^i \frac{\partial}{\partial x^i}$$

and the fundamental tensor

$$g := (F F_{y^i y^j} + F_{y^i} F_{y^j}) dx^i \otimes dx^j = g_{ij} dx^i \otimes dx^j$$

are both sections of tensor bundles that sit over the manifold $TM \setminus 0$. As one moves around on $TM \setminus 0$, not only do the components ℓ^i and g_{ij} vary, the basis sections $\frac{\partial}{\partial x^i}$ and dx^i change as well. Thus, when measuring the rate of change $\nabla_v E$ of any tensor field E, along a given direction v at the point p, we must invoke the product rule. For example:

* $\nabla_v \ell = (d\ell^j)(v) \frac{\partial}{\partial x^j} + \ell^j \nabla_v \frac{\partial}{\partial x^j}$.
* $\nabla_v g = (dg_{ij})(v) dx^i \otimes dx^j + g_{ij} (\nabla_v dx^i) \otimes dx^j + g_{ij} dx^i \otimes (\nabla_v dx^j)$.

The terms on the right-hand sides of these formulas split into two groups:

(1) The first group consists of $(d\ell^j)(v) \frac{\partial}{\partial x^j}$ and $(dg_{ij})(v) dx^i \otimes dx^j$. They come from taking the ordinary directional derivative of the components, which are scalars, but leaving the basis sections alone.

(2) In the second group, the components are left untouched, but we have yet to make sense of the quantities $\nabla_v \frac{\partial}{\partial x^j}$ and $\nabla_v dx^i$. Intuitively, this is done by tabulating the values of the basis section $\frac{\partial}{\partial x^j}$ or dx^i as we move away from p in the direction v. These are then compared to its value at p in order to produce the requisite rate of change.

However, before a meaningful comparison can be carried out, the tabulated values must first be transported back to p. In general, on a manifold there is no canonical way to carry out this transport. The best we can hope for is to specify one that does not run afoul of any *a priori* geometrical or topological constraint. These specifications are usually spelled out in the form of so-called structural equations. One then solves these equations to obtain the **connection 1-forms** $\omega_j^{\ i}$, using which the **covariant derivatives** $\nabla_v \frac{\partial}{\partial x^j}$ and $\nabla_v dx^i$ can be explicitly written down:

$$\boxed{\nabla_v \frac{\partial}{\partial x^j} := \omega_j^{\ i}(v) \frac{\partial}{\partial x^i}} , \tag{2.4.1}$$

$$\boxed{\nabla_v dx^i := - \omega_j^{\ i}(v) dx^j} . \tag{2.4.2}$$

See Exercise 2.4.1 for an explanation of the minus sign.

Suppose the structural equations have been proposed, and the connection forms have been solved for. (We carry out these steps momentarily, in the proof of Theorem 2.4.1 below.) Let us substitute (2.4.1), (2.4.2) into our formulas for $\nabla_v \ell$ and $\nabla_v g$, relabel some summation indices, and suppress v. The results read:

(2.4.3) $$\nabla \ell = (d\ell^i + \ell^j \omega_j{}^i) \otimes \frac{\partial}{\partial x^i} ,$$

(2.4.4) $$\nabla g = (dg_{ij} - g_{kj}\, \omega_i{}^k - g_{ik}\, \omega_j{}^k) \otimes dx^i \otimes dx^j .$$

The operator ∇, or the $\omega_j{}^i$ collectively, defines what is called a **linear connection** on π^*TM and its associated tensor products. Each linear connection, for example, the Chern connection that we introduce, is fully characterized by its structural equations. Nevertheless, there is a general set of axioms that *all* linear connections must satisfy. They are:

* $\nabla_v (f E) = (df)(v) E + f \nabla_v E$,
* $\nabla_v (E_1 + E_2) = \nabla_v E_1 + \nabla_v E_2$,
* $\nabla_{\lambda v} E = \lambda \nabla_v E$ for all constants λ ,
* $\nabla_{v_1 + v_2} E = \nabla_{v_1} E + \nabla_{v_2} E$.

Theorem 2.4.1 (Chern) [Ch1]. *Let (M, F) be a Finsler manifold. The pulled-back bundle π^*TM admits a unique linear connection, called the* **Chern connection**. *Its connection forms are characterized by the structural equations:*

* **Torsion freeness:**

(2.4.5) $$d(dx^i) - dx^j \wedge \omega_j{}^i = - dx^j \wedge \omega_j{}^i = 0 .$$

* **Almost g-compatibility:**

(2.4.6) $$\boxed{dg_{ij} - g_{kj}\, \omega_i{}^k - g_{ik}\, \omega_j{}^k = 2 A_{ijs} \frac{\delta y^s}{F}} .$$

In fact:

- *Torsion freeness is equivalent to the absence of dy^k terms in $\omega_j{}^i$; namely,*

(2.4.7) $$\boxed{\omega_j{}^i = \Gamma^i{}_{jk}\, dx^k} ,$$

together with the symmetry

(2.4.8) $$\boxed{\Gamma^i{}_{kj} = \Gamma^i{}_{jk}} .$$

- *Almost metric-compatibility then implies that*

(2.4.9)
$$\boxed{\Gamma^l{}_{jk} = \gamma^l{}_{jk} - g^{li} \left(A_{ijs} \frac{N^s{}_k}{F} - A_{jks} \frac{N^s{}_i}{F} + A_{kis} \frac{N^s{}_j}{F} \right)} .$$

Equivalently,

(2.4.10)
$$\boxed{\Gamma^i_{\ jk} = \frac{g^{is}}{2}\left(\frac{\delta g_{sj}}{\delta x^k} - \frac{\delta g_{jk}}{\delta x^s} + \frac{\delta g_{ks}}{\delta x^j}\right)}\ .$$

Here:

* g is the fiber Riemannian metric of π^*TM. See §2.1. It is not to be confused with the Sasaki (type) metric that we encountered in §2.3.

* δy^s was defined in (2.3.4a), and $\frac{\delta}{\delta x^s}$ in (2.3.3). Namely,

$$\delta y^s := dy^s + N^s_{\ j}\, dx^j\ ,$$
$$\frac{\delta}{\delta x^s} := \frac{\partial}{\partial x^s} - N^i_{\ s}\, \frac{\partial}{\partial y^i}\ .$$

* The $\gamma^l_{\ jk}$, defined in in (2.3.1), are the formal Christoffel symbols of the second kind, associated with the fundamental tensor.

* The $N^i_{\ j}$ were defined in (2.3.2a). They represent the nonlinear Ehresmann connection on the tangent bundle of $TM \setminus 0$.

Remarks. Formula (2.4.10) is the *raison d'être* of the Chern connection. Note its structural similarity to the Christoffel symbols in (2.3.1). However, there are three famous connections in the literature, each with its own merit. We express them in terms of the Chern connection.

- The **Cartan connection** is given by $\omega_j^{\ i} + A^i_{\ jk}\frac{\delta y^k}{F}$. It is metric-compatible but has torsion. One of its merits is that it makes readily accessible a specific Bianchi identity for Landsberg spaces. See [M2] or [BCS1].
- The **Hashiguchi connection** is given by $\omega_j^{\ i} + A^i_{\ jk}\frac{\delta y^k}{F} + \dot{A}^i_{\ jk}dx^k$. Here, $\dot{A} = \nabla_{\hat{\ell}} A$ is the horizontal covariant derivative of the Cartan tensor A along the distinguished (horizontal) direction $\hat{\ell} := \ell^i\frac{\delta}{\delta x^i}$. See Exercise 2.5.5.
- The **Berwald connection** is given by $\omega_j^{\ i} + \dot{A}^i_{\ jk}dx^k$. Like the Chern connection, it has no torsion. The Berwald connection is particularly convenient when dealing with Finsler spaces of constant flag curvature. It is most directly related to the nonlinear connection $N^i_{\ j}$, and most amenable to the study of the geometry of paths. See Exercises 2.3.3, 3.8.3, 3.8.4, 3.8.5, and 3.10.7.

For a systematic treatment of these connections, see [M2] and [AIM]. Practitioners of Finsler geometry may wonder why the **Rund connection** is conspicuously absent from the above. The reason is that it coincides with the Chern connection, as pointed out by Anastasiei [A].

Proof of Theorem 2.4.1.

(A) The consequences of torsion freeness:

A priori, we have $\omega_j{}^i = \Gamma^i{}_{jk}\, dx^k + Z^i{}_{jk}\, dy^k$. Substituting this expression into the torsion freeness criterion, namely, $dx^j \wedge \omega_j{}^i = 0$, we immediately get

$$\Gamma^i{}_{jk}\, dx^j \wedge dx^k \ + \ Z^i{}_{jk}\, dx^j \wedge dy^k \ = \ 0 \ .$$

Thus the $Z^i{}_{jk}$ must vanish. The same applies to the antisymmetric part (in the lower indices j and k) of $\Gamma^i{}_{jk}$. That is,

$$\Gamma^i{}_{kj} \ = \ \Gamma^i{}_{jk} \ .$$

(B) The explicit formula for $\Gamma^i{}_{jk}$:

Let us substitute (2.4.7) and the formula (2.3.4a) for δy^s into the criterion (2.4.6). Equating the coefficients of the dx^k terms, we get

$$\frac{\partial g_{ij}}{\partial x^k} \ = \ g_{sj}\, \Gamma^s{}_{ik} \ + \ g_{is}\, \Gamma^s{}_{jk} \ + \ 2\, A_{ijs}\, \frac{N^s{}_k}{F} \ .$$

In other words,

$$\boxed{\Gamma_{ijk} \ + \ \Gamma_{jik} \ = \ \frac{\partial g_{ij}}{\partial x^k} \ - \ 2\, A_{ijs}\, \frac{N^s{}_k}{F}} \ . \tag{2.4.11}$$

Now we use the so-called **Christoffel's trick**. Namely, apply (2.4.11) to the combination

$$(\Gamma_{rjk} \ + \ \Gamma_{jrk}) \ - \ (\Gamma_{jkr} \ + \ \Gamma_{kjr}) \ + \ (\Gamma_{krj} \ + \ \Gamma_{rkj}) \ ,$$

and impose the symmetry (2.4.8). After much cancellation, this will result in a formula for Γ_{rjk}. Raising the index r will give (2.4.9) as desired.

Finally, using the operators $\frac{\delta}{\delta x}$ defined in (2.3.3), we can re-express the Chern connection coefficients in the elegant form (2.4.10). □

Let us describe the Chern connection for two important families of Finsler spaces:

- *Riemannian manifolds.* These are characterized by F^2 having only an explicit quadratic dependence on y. As a result, y dependence will be absent from the fundamental tensor, which then coincides with the Riemannian metric on the underlying manifold M. See §1.3. Since the Cartan tensor A vanishes in this case, formula (2.4.9) reduces to $\Gamma^i{}_{jk} = \gamma^i{}_{jk}$. Thus, on Riemannian manifolds, the Chern connection coefficients $\Gamma^i{}_{jk}$ are simply the Riemannian metric's Christoffel symbols of the second kind.
- *Locally Minkowski spaces.* These are characterized by F having no x dependence in some privileged coordinate charts. Consequently the fundamental tensor vanishes and so do its formal Christoffel symbols. A quick glance at (2.3.2) shows that the nonlinear connection

${N^i}_j$ is zero too. These reduce formula (2.4.9) to read ${\Gamma^i}_{jk} = 0$. So, for locally Minkowski spaces, the Chern connection coefficients completely vanish in certain natural coordinates. Exercise 2.4.8 assures us that in arbitrary natural coordinates, these connection coefficients can have at most an x (but no y) dependence.

Exercises

Exercise 2.4.1:

(a) Let $E := E^i \frac{\partial}{\partial x^i}$ be an arbitrary section of π^*TM. Show that

$$\nabla E \;=\; (\, dE^i \;+\; E^j\, {\omega_j}^i\,) \otimes \frac{\partial}{\partial x^i} \;.$$

(b) Given any section $\theta := \theta_i\, dx^i$ of π^*T^*M, the quantity $\theta(E)$ is a scalar. Insist that the following **Leibniz rule** holds:

$$\nabla[\,\theta(E)\,] \;=\; (\nabla\theta)(E) \;+\; \theta(\nabla E) \;.$$

Prove that the above can be manipulated to yield the statement

$$\nabla\theta \;=\; (\, d\theta_i \;-\; \theta_j\, {\omega_i}^j\,) \otimes dx^i \;.$$

(c) Explain how you would deduce (2.4.2) from part (b).

Exercise 2.4.2: Recall that torsion freeness forces the ${\omega_j}^i$ to have the structure ${\Gamma^i}_{jk}\, dx^k$. In other words, there are no dy^k terms.

(a) As in the proof of Theorem 2.4.1, substitute the above and (2.3.4a) (the formula for δy^s) into the almost g-compatibility criterion (2.4.6). Show that equating the coefficients of dy terms simply recovers the definition of the Cartan tensor.

(b) Suppose, in the compatibility criterion (2.4.6), we replace the A_{ijs} on the right-hand side by some other functions. Explain why the resulting structural equations have *no* solutions.

(c) A connection is said to be **g-compatible** if

$$dg_{ij} \;-\; g_{kj}\, {\omega_i}^k \;-\; g_{ik}\, {\omega_j}^k \;=\; 0 \;.$$

Prove that the following two statements are equivalent:

- There exists a torsion free g-compatible linear connection on the pulled-back bundle π^*TM.
- The Finsler structure F is Riemannian.

Exercise 2.4.3: The proof of Theorem 2.4.1 spelled out Christoffel's trick, but did not exhibit all the interesting cancellations that eventually led to (2.4.9). Supply the missing details of that computation.

Exercise 2.4.4: If we had discussed the connection ∇ using the special g-orthonormal bases, we would have

$$\boxed{\nabla_v \, e_b \;:=\; \omega_b{}^a(v)\, e_a} \;,$$

$$\boxed{\nabla_v \, \omega^a \;:=\; -\,\omega_b{}^a(v)\, \omega^b} \;.$$

instead of (2.4.1) and (2.4.2). Recall the n-beins $u_a{}^i$ and $v^a{}_i$ of §2.2.

(a) Apply ∇ to both sides of $e_b = u_b{}^k \frac{\partial}{\partial x^k}$. Show that after relabeling a summation index, we have

$$\omega_b{}^a \, e_a \;=\; (\, du_b{}^k \;+\; u_b{}^l \, \omega_l{}^k \,) \, \frac{\partial}{\partial x^k} \;.$$

(b) In that equation, check that expanding out the e_a and equating coefficients of $\frac{\partial}{\partial x^k}$ then leads to

$$\omega_b{}^a \, u_a{}^k \;=\; du_b{}^k \;+\; u_b{}^l \, \omega_l{}^k \;.$$

(c) Invert away that $u_a{}^k$ to obtain

$$\boxed{\omega_b{}^a \;=\; (\, du_b{}^i \,)\, v^a{}_i \;+\; u_b{}^j \, \omega_j{}^i \, v^a{}_i} \;.$$

This expresses the connection forms for the orthonormal basis in terms of those for the natural basis.

(d) Likewise, prove that

$$\boxed{\omega_j{}^i \;=\; (\, dv^a{}_j \,)\, u_a{}^i \;+\; v^b{}_j \, \omega_b{}^a \, u_a{}^i} \;.$$

It expresses the connection forms for the natural basis in terms of those for the orthonormal basis.

Exercise 2.4.5: Use the last formula in the above exercise, together with suitable items from §2.2 and §2.3, to re-express the structural equations of the Chern connection.

(a) Check that torsion freeness (2.4.5) now reads

$$\boxed{d\omega^a \;-\; \omega^b \wedge \omega_b{}^a \;=\; 0} \;.$$

(b) Show that the "almost g-compatibility" criterion (2.4.6) becomes

$$\boxed{\omega_{ab} \;+\; \omega_{ba} \;=\; -\,2\, A_{abc}\, \omega^{n+c}} \;,$$

where ω_{ba} abbreviates $\omega_b{}^c \, \delta_{ca}$.

(c) Explain why

$$\boxed{\omega_{nn} \;=\; 0 \;=\; \omega_n{}^n}$$

in a special g-orthonormal basis. Caution: this is *not* a statement about ω^{n+n}.

Exercise 2.4.6:

(a) Deduce from (2.4.9) that $\boxed{\Gamma^i{}_{jk}\,\ell^j \;=\; \Gamma^i{}_{kj}\,\ell^j \;=\; \frac{N^i{}_k}{F}}$.

(b) Use this to rewrite $\frac{\delta y^i}{F}$ as

$$\boxed{\frac{\delta y^i}{F} \;=\; (\nabla\ell)^i \;+\; \ell^i\, d(\log F)} \,,$$

where $\nabla\ell = d\ell^i + \ell^j\,\omega_j{}^i$.

(c) Show that part (b) leads us to

$$\omega^{n+a} \;=\; \omega^a(\,\nabla\ell\,) \;+\; \delta^a{}_n\; d(\log F)\;.$$

Equivalently,

$$\omega^{n+a}\otimes e_a \;=\; \nabla\ell \;+\; \ell\; d(\log F)\;.$$

(d) Contract the formula in part (b) with ℓ_i . With the help of Exercise 2.3.5, show that $\ell_i\,(\nabla\ell)^i = 0$. In coordinate free notation:

$$\boxed{g(\,\ell\,,\,\nabla\ell\,) \;=\; 0}\;.$$

Exercise 2.4.7:

(a) Check that $\nabla\ell = \omega_n{}^a\, e_a$.

(b) Use part (d) of Exercise 2.4.6 to reduce this to $\nabla\ell = \omega_n{}^\alpha\, e_\alpha$.

(c) Substitute the above into part (c) of Exercise 2.4.6. Show that one obtains

$$\boxed{\omega_n{}^\alpha \;=\; \omega^{n+\alpha}}$$

in a special g-orthonormal basis.

(d) Explain why

$$\boxed{\omega_n{}^n \;\neq\; \omega^{n+n}}\;.$$

Exercise 2.4.8: Consider local coordinate changes on M, say

$$x^i \;=\; x^i(\,\tilde{x}^1\,,\ldots,\,\tilde{x}^n\,)\;,$$

and their inverses $\tilde{x}^p = \tilde{x}^p(\,x^1\,,\ldots,\,x^n\,)$. Denote the Chern connection coefficients in the natural coordinates (x^i, y^i) by $\Gamma^i{}_{jk}$, and their counterparts in the natural coordinates $(\tilde{x}^p, \tilde{y}^p)$ by $\tilde{\Gamma}^p{}_{qr}$.

(a) Imitate the technique in Exercise 2.4.4 to show that

$$\tilde{\Gamma}^p{}_{qr} \;=\; \frac{\partial\tilde{x}^p}{\partial x^i}\,\frac{\partial^2 x^i}{\partial\tilde{x}^q\,\partial\tilde{x}^r} \;+\; \frac{\partial\tilde{x}^p}{\partial x^i}\;\Gamma^i{}_{jk}\;\frac{\partial x^j}{\partial\tilde{x}^q}\,\frac{\partial x^k}{\partial\tilde{x}^r}\;.$$

(b) As explained at the end of this section, the Chern connection coefficients of a locally Minkowski space vanish in certain privileged natural coordinates. Explain why, in arbitrary natural coordinates, they can have at most an x (but no y) dependence.

Exercise 2.4.9:

(a) Derive formula (2.4.11) for the Chern connection coefficients.

(b) Prove that $d \log \sqrt{g} = \frac{1}{2} g^{ij} dg_{ij}$, where $\sqrt{g} := \sqrt{\det(g_{ij})}$.

(c) Show that $\Gamma^i{}_{ki} = \Gamma^i{}_{ik} = \frac{1}{2} g^{ij} \frac{\delta g_{ij}}{\delta x^k} = \frac{\delta}{\delta x^k} \log \sqrt{g}$.

Exercise 2.4.10: Derive formula (2.4.10) in detail.

2.5 Index Gymnastics

Let $T := T^j{}_i \frac{\partial}{\partial x^j} \otimes dx^i$ be an arbitrary smooth local section of $\pi^* TM \otimes \pi^* T^* M$. It is a **tensor field of rank** $\binom{1}{1}$ on the manifold $TM \setminus 0$. Its **covariant differential** is

$$\nabla T := (\nabla T)^j{}_i \frac{\partial}{\partial x^j} \otimes dx^i ,$$

where

(2.5.1) $$(\nabla T)^j{}_i := dT^j{}_i + T^k{}_i \omega_k{}^j - T^j{}_k \omega_i{}^k .$$

2.5 A. The Slash $(\ldots)_{|s}$ *and the Semicolon* $(\ldots)_{;s}$

The $(\nabla T)^j{}_i$ are 1-forms on $TM \setminus 0$. They can therefore be expanded in terms of the natural basis $\{ dx^s , \frac{\delta y^s}{F} \}$:

(2.5.2) $$(\nabla T)^j{}_i = T^j{}_{i|s} dx^s + T^j{}_{i;s} \frac{\delta y^s}{F} .$$

In order to obtain formulas for the coefficients, we evaluate equation (2.5.2) on each individual member of the dual basis $\{ \frac{\delta}{\delta x^s} , F \frac{\partial}{\partial y^s} \}$. We also use the fact that the Chern connection forms for the natural basis have no $\frac{\delta y^s}{F}$ terms, and are given by $\omega_j{}^i = \Gamma^i{}_{js} dx^s$. The results are:

(2.5.3) $$\boxed{T^j{}_{i|s} = \left(\nabla_{\frac{\delta}{\delta x^s}} T \right)^j{}_i = \frac{\delta T^j{}_i}{\delta x^s} + T^k{}_i \Gamma^j{}_{ks} - T^j{}_k \Gamma^k{}_{is}} ,$$

(2.5.4) $$\boxed{T^j{}_{i;s} = \left(\nabla_{F \frac{\partial}{\partial y^s}} T \right)^j{}_i = F \frac{\partial T^j{}_i}{\partial y^s}} .$$

As a reminder:

(2.5.5)
$$\boxed{\frac{\delta T^j{}_i}{\delta x^s} := \frac{\partial T^j{}_i}{\partial x^s} - N^r{}_s \frac{\partial T^j{}_i}{\partial y^r}} ;$$

(2.5.6) $F \dfrac{\partial}{\partial y^s}$ is the homogenized usual partial derivative.

Note that with respect to *the natural basis*:

- The **horizontal covariant derivative** $T^j{}_{i|s}$ is comprised of a horizontal directional derivative $\frac{\delta T^j{}_i}{\delta x^s}$ and correction terms.
- The **vertical covariant derivative** $T^j{}_{i;s}$ consists merely of a homogenized partial derivative. There are no correction terms.

The treatment for tensor fields of higher rank is similar. There will simply be more correction terms because of the additional indices. We ask the reader to provide the details. Instead, let us now illustrate the above formalism with two basic examples.

2.5 B. Covariant Derivatives of the Fundamental Tensor g

Criterion (2.4.6) says that the Chern connection is almost g-compatible:

(2.5.7)
$$(\nabla g)_{ij} = dg_{ij} - g_{kj}\,\omega_i{}^k - g_{ik}\,\omega_j{}^k = 2\,A_{ijs}\,\frac{\delta y^s}{F} .$$

This immediately gives:

(2.5.8)
$$\boxed{g_{ij|s} = 0} .$$

(2.5.9)
$$\boxed{g_{ij;s} = 2\,A_{ijs}} .$$

Next, it can be shown (see Exercise 2.5.2) that

(2.5.10)
$$\boxed{\delta_i{}^j{}_{|s} = 0 \quad \text{and} \quad \delta_i{}^j{}_{;s} = 0} .$$

Thus $(g^{ij}\,g_{jk})_{|s} = 0$ and $(g^{ij}\,g_{jk})_{;s} = 0$. These then yield:

(2.5.11)
$$\boxed{g^{ij}{}_{|s} = 0} .$$

(2.5.12)
$$\boxed{g^{ij}{}_{;s} = -2\,A^{ij}{}_s} .$$

The above statements say that the fundamental tensor (with all possible index configurations) is covariantly constant along horizontal directions. Its vertical derivatives are, as expected, proportional to the Cartan tensor.

2.5 C. Covariant Derivatives of the Distinguished ℓ

Part (b) of Exercise 2.4.6 says that

$$(\nabla \ell)^i \ = \ \frac{\delta y^i}{F} \ - \ \ell^i \, d(\log F) \ .$$

According to part (a) of Exercise 2.3.5, the $d(\log F)$ term can be rewritten as $\ell_s \frac{\delta y^s}{F}$. Thus

$$(\nabla \ell)^i \ = \ (\delta^i{}_s \ - \ \ell^i \, \ell_s)\, \frac{\delta y^s}{F} \ . \tag{2.5.13}$$

Hence

$$\boxed{\ell^i{}_{|s} \ = \ 0} \ , \tag{2.5.14}$$

$$\boxed{\ell^i{}_{;s} \ = \ \delta^i{}_s \ - \ \ell^i \, \ell_s} \ . \tag{2.5.15}$$

These, together with (2.5.8) and (2.5.9), can then be used to deduce that

$$\boxed{\ell_{i|s} \ = \ 0} \ , \tag{2.5.16}$$

$$\boxed{\ell_{i;s} \ = \ g_{is} \ - \ \ell_i \, \ell_s} \ . \tag{2.5.17}$$

Our formulas show that the distinguished section ℓ and the Hilbert form ω are both covariantly constant along horizontal directions. Their vertical derivatives are equal to suitable configurations of the **angular metric**.

Exercises

Exercise 2.5.1: Instead of writing $T := T^j{}_i \, \frac{\partial}{\partial x^j} \otimes dx^i$, let us expand it in terms of a special g-orthonormal basis. Namely: $T := T^b{}_a \ e_b \otimes \omega^a$. Thus $\nabla T = (\nabla T)^b{}_a \, e_b \otimes \omega^a$, and the analogue of (2.5.2) is

$$(\nabla T)^b{}_a \ = \ T^b{}_{a|c} \, \omega^c \ + \ T^b{}_{a;c} \, \omega^{n+c} \ .$$

Show that:

- $T^b{}_{a|c} \ = \ v^b{}_j \ u_a{}^i \ u_c{}^k \ T^j{}_{i|k}$.
- $T^b{}_{a;c} \ = \ v^b{}_j \ u_a{}^i \ u_c{}^k \ T^j{}_{i;k}$.

In particular, the formula for $T^b{}_{a;c}$ is practically as simple as (2.5.4). This simplicity is *unexpected* because, according to part (c) of Exercise 2.4.4, the Chern connection forms $\omega_b{}^a$ for our orthonormal basis do contain ω^{n+c} terms.

Exercise 2.5.2:

(a) To what is $d\delta_i{}^j - \delta_k{}^j \, \omega_i{}^k + \delta_i{}^k \, \omega_k{}^j$ equal?

(b) Explain why $\delta_i{}^j{}_{|s} = 0$ and $\delta_i{}^j{}_{;s} = 0$.

Exercise 2.5.3: Derive (2.5.16) and (2.5.17).

Exercise 2.5.4: The covariant differential of the Cartan tensor A is

$$\nabla A \;=\; (\, dA_{ijk} - A_{ljk}\, \omega_i{}^l - A_{ilk}\, \omega_j{}^l - A_{ijl}\, \omega_k{}^l\,)\; dx^i \otimes dx^j \otimes dx^k \;.$$

The quantity inside the parentheses is $(\nabla A)_{ijk}$. It can be expanded as

$$(\nabla A)_{ijk} \;=\; A_{ijk|s}\; dx^s \;+\; A_{ijk;s}\; \frac{\delta y^s}{F} \;.$$

(a) Explain why $A_{ijk|s}$ and $A_{ijk;s}$ are both symmetric in the first three indices i, j, k.

(b) Show that $A_{ijk|s}\,\ell^k = 0$.

(c) Show that $A_{ijk;s}\,\ell^k = -\,A_{ijs}$.

Exercise 2.5.5: Define the quantities

$$\dot{A}_{ijk} \;:=\; A_{ijk|s}\; \ell^s \;,$$

and set $\dot{A} := \dot{A}_{ijk}\, dx^i \otimes dx^j \otimes dx^k$.

(a) Show that $\dot{A} = \nabla_{\hat{\ell}} A$, where $\hat{\ell} := \ell^i \frac{\delta}{\delta x^i}$.

(b) Check that $\dot{A}_{ijk}\, \ell^k = 0$.

(c) Explain why the quantities $A_{ijk;s}\, \ell^s$ are uninteresting.

Exercise 2.5.6: With the help of (2.5.3) and part (a) of Exercise 2.4.6, show that:

$$\boxed{\ell^r\, \frac{\delta T^j{}_i}{\delta x^s} \;=\; \frac{\delta}{\delta x^s}\left(\ell^r\, T^j{}_i\right) \;+\; T^j{}_i\; \frac{N^r{}_s}{F}} \;.$$

Exercise 2.5.7: Here is one practical use of Exercise 2.5.6. We learn from Exercise 3.3.4 that the expression

$$\ell^j \left(\frac{\delta}{\delta x^k} \frac{N^i{}_j}{F} \;-\; \frac{\delta}{\delta x^j} \frac{N^i{}_k}{F} \right)$$

describes something of paramount importance. It is numerically equal to the predecessor $R^i{}_k$ of the flag curvature. Let us now use some material developed in §2.3 to rewrite that expression into a more computationally friendly form.

(a) Manipulate the first term $\ell^j \frac{\delta}{\delta x^k} \frac{N^i{}_j}{F}$ as follows. Use Exercise 2.5.6 to move ℓ past $\frac{\delta}{\delta x}$. This introduces a "correction" term $(N^i{}_j\, N^j{}_k)/F^2$ which, through Exercise 2.3.3, can be expressed as y derivatives of G^i. Next, (2.3.2b) shows that $\ell^j\,(N^i{}_j/F) = G^i/F^2$. Use part (b) of Exercise 2.3.5 to move the $1/F^2$ outside the $\frac{\delta}{\delta x}$ derivative. Finally, spell out $\frac{\delta}{\delta x}$ with (2.3.3). These maneuvers should produce

$$\frac{1}{F^2}\left[(G^i)_{x^k} - \frac{1}{4}(G^i)_{y^j}(G^j)_{y^k}\right] .$$

(b) Now work on the second term $-\ell^j \frac{\delta}{\delta x^j}\frac{N^i{}_k}{F}$. As in part (a), move the $1/F$ past $\frac{\delta}{\delta x}$. Then spell out $\frac{\delta}{\delta x}$ and use $\ell^j N^s{}_j = G^s/F$. Check that one obtains

$$\frac{1}{F^2}\left[-\frac{1}{2}y^j(G^i)_{y^k x^j} + \frac{1}{2}G^j(G^i)_{y^k y^j}\right] .$$

In short, the expression we stated at the beginning is equal to

$$\frac{1}{F^2}\left[2(\bar{G}^i)_{x^k} - (\bar{G}^i)_{y^j}(\bar{G}^j)_{y^k} - y^j(\bar{G}^i)_{y^k x^j} + 2\bar{G}^j(\bar{G}^i)_{y^k y^j}\right] ,$$

where

$$\bar{G}^s := \frac{G^s}{2} .$$

The utility of this result is shown in §3.9B.

References

[A] M. Anastasiei, *A historical remark on the connections of Chern and Rund*, Cont. Math. **196** (1996), 171–176.

[AIM] P. L. Antonelli, R. S. Ingarden, and M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology*, FTPH 58, Kluwer Academic Publishers, 1993.

[BCS1] D. Bao, S. S. Chern, and Z. Shen, *On the Gauss–Bonnet integrand for 4-dimensional Landsberg spaces*, Cont. Math. **196** (1996), 15–25.

[Ch1] S. S. Chern, *Local equivalence and Euclidean connections in Finsler spaces*, Sci. Rep. Nat. Tsing Hua Univ. Ser. A **5** (1948), 95–121; or Selected Papers, vol. II, 194–212, Springer 1989.

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

Chapter 3

Curvature and Schur's Lemma

3.1 Conventions and the *hh*-, *hv*-, *vv*-curvatures

The curvature 2-forms of the Chern connection are

$$\boxed{\Omega_j^{\ i} := d\omega_j^{\ i} - \omega_j^{\ k} \wedge \omega_k^{\ i}} \,. \tag{3.1.1}$$

Our convention on the wedge product does *not* include any normalization factor. Thus, for example, the wedge product of two 1-forms is

$$\theta \wedge \zeta := \theta \otimes \zeta - \zeta \otimes \theta \ , \tag{3.1.2}$$

without the factor of $\frac{1}{2}$.

Since the $\Omega_j{}^i$ are 2-forms on the manifold $TM\setminus 0$, they can be generically expanded as

$$\Omega_j{}^i := \frac{1}{2} R_j{}^i{}_{kl}\, dx^k \wedge dx^l + P_j{}^i{}_{kl}\, dx^k \wedge \frac{\delta y^l}{F} + \frac{1}{2} Q_j{}^i{}_{kl}\, \frac{\delta y^k}{F} \wedge \frac{\delta y^l}{F} .$$

The objects R, P, Q are respectively the *hh*-, *hv*-, *vv*-**curvature tensors** of the Chern connection. There is no loss in generality in supposing that

$$\boxed{R_j{}^i{}_{lk} = -R_j{}^i{}_{kl}} , \tag{3.1.3}$$

$$Q_j{}^i{}_{lk} = -Q_j{}^i{}_{kl} . \tag{3.1.4}$$

As we soon show, the *vv*-curvature Q actually vanishes for the Chern connection.

Exercises

Exercise 3.1.1: Explain why it does not make sense to put a factor of $\frac{1}{2}$ in front of the term $P_j{}^i{}_{kl}\, dx^k \wedge \frac{\delta y^l}{F}$.

Exercise 3.1.2: Why is (3.1.3) an assumption instead of a consequence?

3.2 First Bianchi Identities from Torsion Freeness

The Chern connection is torsion free: $dx^j \wedge \omega_j{}^i = 0$. Exterior differentiation then gives

$$dx^j \wedge d\omega_j{}^i = 0 .$$

Since the term $dx^j \wedge \omega_j{}^k \wedge \omega_k{}^i$ vanishes by torsion freeness, it can be subtracted from the left-hand side of the above equation without affecting anything. Thus

$$dx^j \wedge \Omega_j{}^i = 0 . \tag{3.2.1}$$

Into this we substitute our expansion for $\Omega_j{}^i$, and obtain:

$$\begin{aligned} 0 = &\ \frac{1}{2} R_j{}^i{}_{kl}\, dx^j \wedge dx^k \wedge dx^l \\ &+ P_j{}^i{}_{kl}\, dx^j \wedge dx^k \wedge \frac{\delta y^l}{F} \\ &+ \frac{1}{2} Q_j{}^i{}_{kl}\, dx^j \wedge \frac{\delta y^k}{F} \wedge \frac{\delta y^l}{F} . \end{aligned} \tag{*}$$

The three terms on the right are of completely different types. Therefore each must vanish. Let us discuss them in turn.

- One important consequence comes from the vanishing of the third term on the right. It says that Q must be symmetric in the indices k and l. Yet (3.1.4) tells us that there is antisymmetry in those two indices. So

$$\boxed{Q_j{}^i{}_{kl} = 0} \; .$$

This simplifies the curvature 2-forms to

$$\boxed{\Omega_j{}^i = \frac{1}{2} R_j{}^i{}_{kl} \, dx^k \wedge dx^l + P_j{}^i{}_{kl} \, dx^k \wedge \frac{\delta y^l}{F}} \; . \tag{3.2.2}$$

- The vanishing of the second term on the right of (*) uncovers a symmetry

$$\boxed{P_k{}^i{}_{jl} = P_j{}^i{}_{kl}} \tag{3.2.3}$$

of P, between its first and third indices.

- Finally, the vanishing of the first term on the right of (*) gives the **first Bianchi identity for R:**

$$\boxed{R_j{}^i{}_{kl} + R_k{}^i{}_{lj} + R_l{}^i{}_{jk} = 0} \; . \tag{3.2.4}$$

Exercises

Exercise 3.2.1: Consider the connection forms $\omega_b{}^a$ for our g-orthonormal basis sections. According to Exercise 2.4.5, in that situation the criterion for vanishing torsion reads $d\omega^a = \omega^b \wedge \omega_b{}^a$.

(a) Show that the analogue of (3.2.1) is

$$\omega^b \wedge \Omega_b{}^a = 0 \, ,$$

where $\Omega_b{}^a := d\omega_b{}^a - \omega_b{}^c \wedge \omega_c{}^a$.

(b) Check that $\Omega_b{}^a$ has the structure

$$\Omega_b{}^a = \frac{1}{2} R_b{}^a{}_{cd} \, \omega^c \wedge \omega^d + P_b{}^a{}_{cd} \, \omega^c \wedge \omega^{n+d} \, .$$

Exercise 3.2.2:

(a) Is the vv-curvature Q always zero for torsion-free connections?

(b) Can it vanish for connections that have torsion?

3.3 Formulas for R and P in Natural Coordinates

The left-hand side of (3.2.2) is $\Omega_j{}^i$. We replace it by the defining expression given in (3.1.1):

$$(3.3.1) \qquad d\omega_j{}^i - \omega_j{}^h \wedge \omega_h{}^i = \frac{1}{2} R_j{}^i{}_{kl} \, dx^k \wedge dx^l + P_j{}^i{}_{kl} \, dx^k \wedge \frac{\delta y^l}{F} .$$

* Note that $d\omega_j{}^i = d\Gamma^i{}_{jl} \wedge dx^l$. Since the differential $d\Gamma^i{}_{jl}$ is a 1-form on $TM \setminus 0$, it can be expanded in terms of dx^k and $\frac{\delta y^k}{F}$. Carrying that out, we get

$$d\omega_j{}^i = \frac{\delta \Gamma^i{}_{jl}}{\delta x^k} \, dx^k \wedge dx^l + F \frac{\partial \Gamma^i{}_{jl}}{\partial y^k} \frac{\delta y^k}{F} \wedge dx^l .$$

Relabeling the second term on the right then gives:

$$d\omega_j{}^i = \frac{\delta \Gamma^i{}_{jl}}{\delta x^k} \, dx^k \wedge dx^l - F \frac{\partial \Gamma^i{}_{jk}}{\partial y^l} \, dx^k \wedge \frac{\delta y^l}{F} .$$

* Also, $-\omega_j{}^h \wedge \omega_h{}^i = \omega_h{}^i \wedge \omega_j{}^h = \Gamma^i{}_{hk} \Gamma^h{}_{jl} \, dx^k \wedge dx^l$.

Substituting these into (3.3.1), we see that the expression

$$\left(\frac{\delta \Gamma^i{}_{jl}}{\delta x^k} + \Gamma^i{}_{hk} \Gamma^h{}_{jl} \right) dx^k \wedge dx^l - F \frac{\partial \Gamma^i{}_{jk}}{\partial y^l} \, dx^k \wedge \frac{\delta y^l}{F}$$

is supposed to be equal to

$$\frac{1}{2} R_j{}^i{}_{kl} \, dx^k \wedge dx^l + P_j{}^i{}_{kl} \, dx^k \wedge \frac{\delta y^l}{F} .$$

Hence the antisymmetric part (which involves a factor of $\frac{1}{2}$) of $(\cdots)$ must equal $\frac{1}{2} R_j{}^i{}_{kl}$. In other words,

$$(3.3.2) \qquad \boxed{ R_j{}^i{}_{kl} = \frac{\delta \Gamma^i{}_{jl}}{\delta x^k} - \frac{\delta \Gamma^i{}_{jk}}{\delta x^l} + \Gamma^i{}_{hk} \Gamma^h{}_{jl} - \Gamma^i{}_{hl} \Gamma^h{}_{jk} } .$$

Also,

$$(3.3.3) \qquad \boxed{ P_j{}^i{}_{kl} = - F \frac{\partial \Gamma^i{}_{jk}}{\partial y^l} } .$$

Note that (3.3.3) implies (3.2.3). As a reminder:

$$\frac{\delta}{\delta x^k} = \frac{\partial}{\partial x^k} - N^i{}_k \frac{\partial}{\partial y^i} .$$

See §2.3.

We conclude by discussing the curvatures R and P of two important classes of Finsler spaces. This is a follow-up of the discussion we gave, at the end of §2.4, of their Chern connections.

- *Riemannian manifolds.* In natural coordinates, the Chern connection coefficients $\Gamma^i{}_{jk}$ of Riemannian manifolds simply equal the underlying Riemannian metric's Christoffel symbols $\gamma^i{}_{jk}$ of the second kind. The latter do not have any y dependence. Therefore $P = 0$ by (3.3.3), and those $\frac{\delta}{\delta x}$ in (3.3.2) reduce to $\frac{\partial}{\partial x}$. In other words,

$$R_j{}^i{}_{kl} = \frac{\partial \gamma^i{}_{jl}}{\partial x^k} - \frac{\partial \gamma^i{}_{jk}}{\partial x^l} + \gamma^i{}_{hk}\gamma^h{}_{jl} - \gamma^i{}_{hl}\gamma^h{}_{jk} .$$

- *Locally Minkowski spaces.* In certain natural coordinates, the Chern connection coefficients $\Gamma^i{}_{jk}$ vanish identically. Thus (3.3.2) and (3.3.3) tell us that both R and P must be zero in these privileged natural coordinates. But curvatures are tensorial objects, so they remain zero in all natural coordinates (and actually, in all bases).

Exercises

Exercise 3.3.1: Review part (b) of Exercise 3.2.1. Show that

(a) $R_b{}^a{}_{cd} = u_b{}^j \, v^a{}_i \, u_c{}^k \, u_d{}^l \, R_j{}^i{}_{kl}$.

(b) $P_b{}^a{}_{cd} = u_b{}^j \, v^a{}_i \, u_c{}^k \, u_d{}^l \, P_j{}^i{}_{kl}$.

Exercise 3.3.2: Explain how Euler's theorem can be used to deduce the statement

$$\boxed{P_j{}^i{}_{kl}\, \ell^l = 0} .$$

Exercise 3.3.3: Let (M, F) be a locally Minkowski space, as defined in §1.3. Show that its $R_j{}^i{}_{kl}$ and $P_j{}^i{}_{kl}$ are both identically zero in *all* natural coordinates.

Exercise 3.3.4: The quantities (see §3.9)

$$R^i{}_k := \ell^j \, R_j{}^i{}_{kl} \, \ell^l$$

are of paramount importance in Finsler geometry. Prove that

$$\boxed{R^i{}_k = \ell^j \left(\frac{\delta}{\delta x^k} \frac{N^i{}_j}{F} - \frac{\delta}{\delta x^j} \frac{N^i{}_k}{F} \right)} .$$

A viable strategy is as follows:

* Start with the formula (3.3.2) for $R_j{}^i{}_{kl}$ in natural coordinates.
* Use Exercise 2.5.6 to move ℓ^j and ℓ^l past the appropriate $\frac{\delta}{\delta x}$.
* Then refer to part (a) of Exercise 2.4.6.

Exercise 3.3.5: Recall from Exercise 2.4.6 that

$$\frac{\delta y^i}{F} = d\ell^i + \ell^j \, \omega_j{}^i + \ell^i \, d(\log F) .$$

Prove that its exterior differential is given by the formula

$$\boxed{d\Big(\frac{\delta y^i}{F}\Big) \;=\; \ell^j\, \Omega_j^{\ i} \;+\; \frac{\delta y^j}{F} \wedge \Big(\omega_j^{\ i} \;-\; \ell_j\, \frac{\delta y^i}{F}\Big)} \; .$$

Exercise 3.3.6: For our orthonormal frame, recall from Exercises 2.4.7, 2.3.5, and 2.4.5 that:

* $\omega^{n+\alpha} = \omega_n^{\ \alpha}$,
* $\omega^{n+n} = d(\log F)$,
* $\omega_n^{\ n} = 0$.

Without doing any computation, explain why one immediately has:

(a) $d\omega^{n+\alpha} = \Omega_n^{\ \alpha} + \omega_n^{\ b} \wedge \omega_b^{\ \alpha} = \Omega_n^{\ \alpha} + \omega^{n+b} \wedge (\omega_b^{\ \alpha} - \ell_b\, \omega^{n+\alpha})$.

(b) $d\omega^{n+n} = 0$.

3.4 First Bianchi Identities from "Almost" g-compatibility

The Chern connection is almost metric-compatible, in the sense that

$$dg_{ij} \;-\; g_{kj}\, \omega_i^{\ k} \;-\; g_{ik}\, \omega_j^{\ k} \;=\; 2\; A_{ijs}\, \frac{\delta y^s}{F} \; .$$

After exterior differentiation and some manipulations, we get

$$\Omega_{ij} \;+\; \Omega_{ji} \;=\; -2\,(\nabla A)_{ijk} \wedge \frac{\delta y^k}{F} \;-\; 2\, A_{ijk} \Big[d\Big(\frac{\delta y^k}{F}\Big) \;+\; \omega_l^{\ k} \wedge \frac{\delta y^l}{F} \Big] \; .$$

The expansion of $(\nabla A)_{ijk}$ was considered in Exercise 2.5.4, and $d(\frac{\delta y^k}{F})$ was computed in Exercise 3.3.5. These turn the above into the following **fundamental identity**:

$$\begin{aligned}
&\Omega_{ij} \;+\; \Omega_{ji} \\
&= \frac{1}{2}\,(\,R_{ijkl} + R_{jikl}\,)\; dx^k \wedge dx^l \;+\; (\,P_{ijkl} + P_{jikl}\,)\; dx^k \wedge \frac{\delta y^l}{F} \\
\text{(3.4.1)}\qquad &= \;-\; (\,A_{iju}\; R^u_{\ kl}\,)\; dx^k \wedge dx^l \\
&\quad -\; 2\,(\,A_{iju}\; P^u_{\ kl} \;+\; A_{ijl|k}\,)\; dx^k \wedge \frac{\delta y^l}{F} \\
&\quad +\; 2\,(\,A_{ijk;l} - A_{ijk}\; \ell_l\,)\; \frac{\delta y^k}{F} \wedge \frac{\delta y^l}{F} \; .
\end{aligned}$$

Here, we have introduced the abbreviations

$$\text{(3.4.2)} \qquad \boxed{R^i_{\ kl} \;:=\; \ell^j\; R_j^{\ i}{}_{kl}}$$

$$\text{(3.4.3)} \qquad \boxed{P^i_{\ kl} \;:=\; \ell^j\; P_j^{\ i}{}_{kl}}$$

in order to reduce clutter.

There is a wealth of information that one can uncover from this fundamental identity. We carry that out systematically below.

3.4 A. Consequences from the $dx^k \wedge dx^l$ Terms

In (3.4.1), the coefficients of the $dx^k \wedge dx^l$ terms tell us that

$$\boxed{R_{ijkl} \;+\; R_{jikl} \;=\; 2\,(\,-\,A_{iju}\;R^u{}_{kl}\,) \;=:\; 2\;B_{ijkl}}\;, \tag{3.4.4}$$

where we have introduced a temporary abbreviation

$$B_{ijkl} \;:=\; -\,A_{iju}\;R^u{}_{kl}\;.$$

It is symmetric in i, j but skew-symmetric in k, l. Thus R_{jikl} is in general *not* skew on its first two indices. Formula (3.4.4), together with the first Bianchi identity for R [namely (3.2.4)] and (3.1.3), can be used to deduce the following:

(3.4.5)

$$\boxed{R_{klji} \;-\; R_{jikl} \;=\; (B_{klji} - B_{jikl}) + (B_{kilj} + B_{ljki}) + (B_{iljk} + B_{jkil})}\;.$$

The procedure is spelled out in Exercise 3.4.2.

In Exercise 3.3.4, we encountered the quantities

$$\boxed{R_{ik} \;:=\; \ell^j\;R_{jikl}\;\ell^l}\;. \tag{3.4.6}$$

Using (3.4.4) and (3.4.5), we see that

$$\ell^l\;R_{lkij}\;\ell^j \;=\; \ell^l\;R_{klji}\;\ell^j \;=\; \ell^j\;R_{jikl}\;\ell^l\;.$$

In other words,

$$\boxed{R_{ki} \;=\; R_{ik}}\;. \tag{3.4.7}$$

This symmetry was *not* apparent from the said exercise.

3.4 B. Consequences from the $dx^k \wedge \frac{1}{F}\delta y^l$ Terms

The coefficients of the $dx^k \wedge \frac{\delta y^l}{F}$ terms in our fundamental identity tell us that

$$\boxed{P_{ijkl} \;+\; P_{jikl} \;=\; -\;2\,A_{iju}\;P^u{}_{kl} \;-\; 2\,A_{ijl|k}}\;. \tag{3.4.8}$$

Let us use this to derive a constitutive relation for P_{jikl} :

- Apply (3.4.8) three times to the combination

  $$(\,P_{ijkl} + P_{jikl}\,) \;-\; (\,P_{jkil} + P_{kjil}\,) \;+\; (\,P_{kijl} + P_{ikjl}\,)\;.$$

  With the help of a temporary abbreviation

  $$E_{ijkl} \;:=\; -\;A_{iju}\;P^u{}_{kl}\;,$$

the result takes the form

$$P_{jikl} = -(A_{ijl|k} - A_{jkl|i} + A_{kil|j}) + (E_{ijkl} - E_{jkil} + E_{kijl}) .$$

- Contract this with ℓ^j and $\ell^j \ell^k$, respectively, and use (2.5.14) (which says that the horizontal covariant derivative of ℓ is zero). One can check that the second contraction gives $P_{jikl}\, \ell^j\, \ell^k = 0$, which then reduces the first contraction to the important statement

(3.4.9)
$$\boxed{P_{ikl} := \ell^j\, P_{jikl} = -\, \dot{A}_{ikl}} .$$

Here,

(3.4.10)
$$\boxed{\dot{A}_{ijk} := A_{ijk|s}\, \ell^s} .$$

- It follows that $E_{ijkl} = A_{ij}{}^u\, \dot{A}_{ukl}$. This updates the above intermediate formula for P_{jikl} to the **constitutive relation**

(3.4.11)
$$\boxed{\begin{aligned} P_{jikl} = &- (A_{ijl|k} - A_{jkl|i} + A_{kil|j}) \\ &+ A_{ij}{}^u\, \dot{A}_{ukl} - A_{jk}{}^u\, \dot{A}_{uil} + A_{ki}{}^u\, \dot{A}_{ujl} . \end{aligned}}$$

Therefore:

The second Chern curvature tensor P is a functional of the Cartan tensor A_{ijk} and its horizontal covariant derivatives $A_{ijk|s}$.

Formula (3.4.9) can be used to re-express (3.4.8) as

(3.4.12)
$$\boxed{A_{ijl|k} = A_{ij}{}^u\, \dot{A}_{ukl} - \frac{1}{2} (P_{ijkl} + P_{jikl})} ,$$

which is a converse to the constitutive relation (3.4.11). In particular, we now see that

(3.4.13)
$$P_{jikl} = 0 \quad \text{if and only if} \quad A_{ijk|l} = 0 .$$

3.4 C. Consequences from the $\frac{1}{F}\delta y^k \wedge \frac{1}{F}\delta y^l$ Terms

Finally, the coefficients of the $\frac{\delta y^k}{F} \wedge \frac{\delta y^l}{F}$ terms in (3.4.1) tell us that

(3.4.14)
$$\boxed{A_{ijk;l} - A_{ijl;k} = A_{ijk}\, \ell_l - A_{ijl}\, \ell_k} .$$

Exercises

Exercise 3.4.1: Supply all the details that are involved in the derivation of our fundamental identity (3.4.1).

Exercise 3.4.2: Derive (3.4.5) as follows.

(a) Cyclicly permute the first, third, and fourth indices of R_{jikl}. We simply get the first Bianchi identity $R_{jikl} + R_{kilj} + R_{lijk} = 0$. Next do the same to R_{iklj}, and likewise to R_{klji} and to R_{ljik}.

(b) Add the four resulting equations together; then use (3.1.3) and (3.4.4). Check that after some appropriate relabeling, one gets formula (3.4.5) as claimed.

Exercise 3.4.3: Derive the constitutive relation (3.4.11) by following the guidelines given in this section.

Exercise 3.4.4:

(a) Explain why $P_{ikl} := \ell^j \, P_{jikl}$ is totally symmetric in all its indices. Also, explain why $P_{ikl}\, \ell^l = 0$.

(b) Recall the statement $P_j{}^i{}_{kl}\, \ell^l = 0$ from Exercise 3.3.2. There, it was derived from first principles. Check that it is also a consequence of the constitutive relation (3.4.11). Does $P_{ikl}\, \ell^l = 0$ follow immediately from the said statement?

(c) Use (3.4.4) to help show that $\ell^j \, \ell^i \, R_{jikl} = 0 = R_{jikl}\, \ell^k \, \ell^l$. In particular, deduce that

$$\boxed{\ell^i \, R_{ik} \;=\; 0 \;=\; R_{ik}\, \ell^k} \,.$$

Exercise 3.4.5:

(a) Contract the fundamental identity (3.4.1) with ℓ^i. With the help of part (c) of Exercise 2.5.4, show that one gets

$$\ell^i \, \Omega_{ij} \;=\; -\, \Omega_{ji}\, \ell^i \;.$$

(b) Explain why the above carries no more information than (3.4.4) and (3.4.8).

Exercise 3.4.6: Consider the following two statements:

(1) The Chern connection coefficients $\Gamma^i{}_{jk}$ have no y-dependence, in which case the Finsler structure is said to be of Berwald type.

(2) $\frac{\partial^2}{\partial y^p \partial y^q}(y^j \, \Gamma^i{}_{jk}) \;=\; 0$.

We surely have (1) $\Rightarrow$ (2), and it would be intuitively appealing to have the converse as well. This is indeed the case. To demonstrate that, adopt the following strategy:

(a) First show that

$$F \, \frac{\partial^2}{\partial y^p \partial y^q}(y^j \, \Gamma^i{}_{jk}) \;=\; -\; P_p{}^i{}_{kq} \;+\; \dot{A}^i{}_{kp;q} \;.$$

Hint: use (3.3.3) and perhaps (3.4.11).

(b) Show that $-P_p{}^i{}_{kq}$ and $\dot{A}^i{}_{kp;q}$ *each* produces a copy of $-\dot{A}_{kpq}$ if we contract them with ℓ_i.

(c) Now explain in detail why (2) $\Rightarrow$ (1).

3.5 Second Bianchi Identities

Exterior differentiation of (3.1.1) gives the **second Bianchi identity**

$$d\Omega_j{}^i \; - \; \omega_j{}^k \wedge \Omega_k{}^i \; + \; \omega_k{}^i \wedge \Omega_j{}^k \; = \; 0 \, . \tag{3.5.1}$$

Into this we substitute the expression (3.2.2) of $\Omega_j{}^i$ in terms of R and P.

- In the computation of $d\Omega_j{}^i$, let us use

$$d\left(\frac{\delta y^i}{F}\right) \; = \; \ell^j \, \Omega_j{}^i \; + \; \frac{\delta y^j}{F} \wedge \left(\omega_j{}^i \; - \; \ell_j \, \frac{\delta y^i}{F}\right)$$

  whenever we encounter $d(\frac{\delta y^i}{F})$. See Exercise 3.3.5.

- The combination

$$dR_j{}^i{}_{kl} - R_t{}^i{}_{kl} \, \omega_j{}^t + R_j{}^t{}_{kl} \, \omega_t{}^i - R_j{}^i{}_{tl} \, \omega_k{}^t - R_j{}^i{}_{kt} \, \omega_l{}^t$$

  will show up. Being a 1-form on $TM \smallsetminus 0$, it can be re-expressed as

$$R_j{}^i{}_{kl|t} \, dx^t \; + \; R_j{}^i{}_{kl;t} \, \frac{\delta y^t}{F} \, .$$

- Likewise, we replace the 1-form

$$dP_j{}^i{}_{kl} - P_t{}^i{}_{kl} \, \omega_j{}^t + P_j{}^t{}_{kl} \, \omega_t{}^i - P_j{}^i{}_{tl} \, \omega_k{}^t - P_j{}^i{}_{kt} \, \omega_l{}^t$$

  with the expression

$$P_j{}^i{}_{kl|t} \, dx^t \; + \; P_j{}^i{}_{kl;t} \, \frac{\delta y^t}{F} \, .$$

The result is

$$\begin{aligned} 0 \; = \; & \frac{1}{2} \, (R_j{}^i{}_{kl|t} \; - \; P_j{}^i{}_{ku} \, R^u{}_{lt}) \, dx^k \wedge dx^l \wedge dx^t \\ & + \frac{1}{2} \, (R_j{}^i{}_{kl;t} \; - \; 2 \, P_j{}^i{}_{kt|l} \; + \; 2 \, P_j{}^i{}_{ku} \, \dot{A}^u{}_{lt}) \, dx^k \wedge dx^l \wedge \frac{\delta y^t}{F} \\ & + \quad (P_j{}^i{}_{kl;t} \; - \; P_j{}^i{}_{kl} \, \ell_t) \, dx^k \wedge \frac{\delta y^l}{F} \wedge \frac{\delta y^t}{F} \, . \end{aligned} \tag{3.5.2}$$

This is a useful restatement of the master second Bianchi identity. It is equivalent to the following three identities:

(3.5.3)

$$\boxed{R_j{}^i{}_{kl|t} + R_j{}^i{}_{lt|k} + R_j{}^i{}_{tk|l} \; = \; P_j{}^i{}_{ku} \, R^u{}_{lt} \; + \; P_j{}^i{}_{lu} \, R^u{}_{tk} \; + \; P_j{}^i{}_{tu} \, R^u{}_{kl}} \; ,$$

$$(3.5.4) \quad \boxed{R_j{}^i{}_{kl;t} = P_j{}^i{}_{kt|l} - P_j{}^i{}_{lt|k} - (P_j{}^i{}_{ku}\, \dot{A}^u{}_{lt} - P_j{}^i{}_{lu}\, \dot{A}^u{}_{kt})} \; .$$

$$(3.5.5) \quad \boxed{P_j{}^i{}_{kl;t} - P_j{}^i{}_{kt;l} = P_j{}^i{}_{kl}\, \ell_t - P_j{}^i{}_{kt}\, \ell_l} \; .$$

It can be shown that (3.5.4) embodies a **constitutive relation** for the first Chern curvature tensor R. Namely, $R_j{}^i{}_{kl}$ is a functional of the tensor $R^i{}_k$ and its first and second vertical covariant derivatives, together with $\dot{A}$ and its first horizontal covariant derivatives. The tensor R_{ik} becomes known in §3.9 as the predecessor of flag curvatures. Explicitly, the said constitutive relation reads:

$$(3.5.6) \quad \boxed{\begin{aligned} R_j{}^i{}_{kl} = \ & \frac{1}{3}\,(R^i{}_{k;l;j} - R^i{}_{l;k;j} + \ell_j\, R^i{}_{k;l} - \ell_j\, R^i{}_{l;k}) \\ & + \frac{2}{3}\,(R^i{}_{k;j}\, \ell_l - R^i{}_{l;j}\, \ell_k + R^i{}_k\, g_{jl} - R^i{}_l\, g_{jk}) \\ & - (\dot{A}^i{}_{jl|k} - \dot{A}^i{}_{jk|l} + \dot{A}^i{}_{uk}\, \dot{A}^u{}_{jl} - \dot{A}^i{}_{ul}\, \dot{A}^u{}_{jk}) \, . \end{aligned}}$$

The derivation of this is somewhat tedious. It is carried out systematically in Exercises 3.5.5–3.5.7.

Exercises

Exercise 3.5.1: Note that of the four indices on $R_j{}^i{}_{kl}$, i and j are bundle (that is, π^*TM) indices while k, l are manifold (namely, $TM \setminus 0$) indices. The combination

$$dR_j{}^i{}_{kl} - R_t{}^i{}_{kl}\, \omega_j{}^t + R_j{}^t{}_{kl}\, \omega_t{}^i - R_j{}^i{}_{tl}\, \omega_k{}^t - R_j{}^i{}_{kt}\, \omega_l{}^t$$

we encountered above suggests that all four indices have been treated as bundle indices, not by us willfully but by the exterior calculus. How does the bundle connection $\omega_j{}^i$ even know what to do with the manifold indices on R ?

Exercise 3.5.2: Use identity (3.5.3) to derive the following:

$$R^i{}_{kl|t} + R^i{}_{lt|k} + R^i{}_{tk|l} = - \dot{A}^i{}_{ku}\, R^u{}_{lt} - \dot{A}^i{}_{lu}\, R^u{}_{tk} - \dot{A}^i{}_{tu}\, R^u{}_{kl} \, .$$

Exercise 3.5.3:

(a) By contracting (3.5.5) with ℓ^j, prove that

$$\boxed{P_j{}^i{}_{kl} - P_l{}^i{}_{kj} = \dot{A}^i{}_{kj;l} - \dot{A}^i{}_{kl;j}} \; .$$

(b) Using this and the constitutive relation (3.4.11), derive the intriguing formula

$$\boxed{\begin{aligned} \dot{A}^i{}_{kj;l} - \dot{A}^i{}_{kl;j} = \quad & A^i{}_{kj|l} - A^i{}_{kl|j} \\ & + (A^i{}_{ju}\, \dot{A}^u{}_{kl} + \dot{A}^i{}_{ju}\, A^u{}_{kl}) \\ & - (A^i{}_{lu}\, \dot{A}^u{}_{kj} + \dot{A}^i{}_{lu}\, A^u{}_{kj}) \,. \end{aligned}}$$

Exercise 3.5.4: Prove that the following three statements are equivalent:

- $\dot{A}_{ijk} = 0$ (this is the definition of a **Landsberg space**).
- $A_{ijk|l}$ is totally symmetric in all *four* of its indices.
- $P_{jikl} = -A_{jik|l}$.

Hint: you will need to use (3.4.11) and Exercise 3.5.3.

Exercise 3.5.5:

(a) Show that $\ell^j\, R_j{}^i{}_{kl;t} = R^i{}_{kl;t} - R_t{}^i{}_{kl} + \ell_t\, R^i{}_{kl}$.

(b) By contracting (3.5.4) with ℓ^j and relabeling, prove that

$$\begin{aligned} R_j{}^i{}_{kl} = \quad & R^i{}_{kl;j} + \ell_j\, R^i{}_{kl} \\ & - (\dot{A}^i{}_{jl|k} - \dot{A}^i{}_{jk|l} + \dot{A}^i{}_{uk}\, \dot{A}^u{}_{jl} - \dot{A}^i{}_{ul}\, \dot{A}^u{}_{jk}) \,. \end{aligned}$$

Note that the quantity inside the parentheses is like a curvature!

Exercise 3.5.6:

(a) Show that $\ell^u\, R^i{}_{ku;l} = R^i{}_{k;l} + R^i{}_k\, \ell_l - R^i{}_{kl}$.

(b) Contract part (b) of Exercise 3.5.5 with ℓ^l to obtain

$$R_j{}^i{}_{kl}\, \ell^l = R^i{}_{k;j} + 2\, R^i{}_k\, \ell_j + R^i{}_{jk} + \dot{A}_j{}^i{}_{k|l}\, \ell^l \,.$$

(c) Without doing any more computation, explain how one could get

$$R_k{}^i{}_{lj}\, \ell^l = - R^i{}_{j;k} - 2\, R^i{}_j\, \ell_k + R^i{}_{jk} - \dot{A}_k{}^i{}_{j|l}\, \ell^l \,.$$

Exercise 3.5.7:

(a) How would one obtain $R_j{}^i{}_{kl}\, \ell^l + R_k{}^i{}_{lj}\, \ell^l + R^i{}_{jk} = 0$?

(b) Into this, substitute parts (b) and (c) of Exercise 3.5.6. Show that one gets

$$\boxed{R^i{}_{kl} = \frac{1}{3} (R^i{}_{k;l} - R^i{}_{l;k}) + \frac{2}{3} (R^i{}_k\, \ell_l - R^i{}_l\, \ell_k)} \,.$$

(c) Using this and part (b) of Exercise 3.5.5, derive the constitutive relation (3.5.6) for $R_j{}^i{}_{kl}$.

3.6 Interchange Formulas or Ricci Identities

Let

$$T^p{}_q \, \frac{\partial}{\partial x^p} \otimes dx^q$$

be a smooth section of $\pi^*TM \otimes \pi^*T^*M$. Its covariant differential ∇T can be written out two ways:

$$dT^p{}_q \; + \; T^j{}_q \, \omega_j{}^p \; - \; T^p{}_j \, \omega_q{}^j \; = \; T^p{}_{q|j} \, dx^j \; + \; T^p{}_{q;j} \, \frac{\delta y^j}{F} \, .$$

By taking the exterior derivative of this equation, one can deduce the following **Ricci identities** or **interchange formulas**:

$$(3.6.1) \qquad \boxed{T^p{}_{q|j|i} \; - \; T^p{}_{q|i|j} \; = \; T^s{}_q \, R_s{}^p{}_{ij} \; - \; T^p{}_s \, R_q{}^s{}_{ij} \; - \; T^p{}_{q;s} \, R^s{}_{ij}} \, ,$$

$$(3.6.2) \qquad \boxed{T^p{}_{q;j|i} \; - \; T^p{}_{q|i;j} \; = \; T^s{}_q \, P_s{}^p{}_{ij} \; - \; T^p{}_s \, P_q{}^s{}_{ij} \; + \; T^p{}_{q;s} \, \dot{A}^s{}_{ij}} \, ,$$

$$(3.6.3) \qquad \boxed{T^p{}_{q;j;i} \; - \; T^p{}_{q;i;j} \; = \; T^p{}_{q;j} \, \ell_i \; - \; T^p{}_{q;i} \, \ell_j} \, .$$

The exercises provide some minimal guidance, should the reader decide to fill in the details.

These are versatile formulas. In the Riemannian case, only (3.6.1) is nonvacuous, and its right-hand side reduces to two terms.

Exercises

Exercise 3.6.1:

(a) Explain why, when we encounter $dT^p{}_{q|j}$, we can replace it by the expression

$$- \; T^i{}_{q|j} \, \omega_i{}^p \; + \; T^p{}_{i|j} \, \omega_q{}^i \; + \; T^p{}_{q|i} \, \omega_j{}^i \; + \; T^p{}_{q|j|i} \, dx^i \; + \; T^p{}_{q|j;i} \, \frac{\delta y^i}{F} \, .$$

(b) Likewise, explain why $dT^p{}_{q;j}$ can be replaced by

$$- \; T^i{}_{q;j} \, \omega_i{}^p \; + \; T^p{}_{i;j} \, \omega_q{}^i \; + \; T^p{}_{q;i} \, \omega_j{}^i \; + \; T^p{}_{q;j|i} \, dx^i \; + \; T^p{}_{q;j;i} \, \frac{\delta y^i}{F} \, .$$

Exercise 3.6.2: Deduce the interchange formulas (3.6.1)–(3.6.3).

3.7 Lie Brackets among the $\frac{\delta}{\delta x}$ and the $F\frac{\partial}{\partial y}$

Given any local vector fields $\hat{X}$, $\hat{Y}$ on $TM \smallsetminus 0$, and any 1-form ω, the **Cartan formula** says that

$$(3.7.1) \quad \boxed{(d\omega)(\hat{X},\hat{Y}) \;=\; d[\omega(\hat{Y})](\hat{X}) \;-\; d[\omega(\hat{X})](\hat{Y}) \;-\; \omega([\hat{X},\hat{Y}])} \;.$$

Let $W := W^j \frac{\partial}{\partial x^j}$ be any local section of π^*TM. Using the Cartan formula, one can prove that

$$(3.7.2) \quad \boxed{W^j\, \Omega_j^{\ i}(\hat{X},\hat{Y})\, \frac{\partial}{\partial x^i} \;=\; \left(\nabla_{\hat{X}} \nabla_{\hat{Y}} \;-\; \nabla_{\hat{Y}} \nabla_{\hat{X}} \;-\; \nabla_{[\hat{X},\hat{Y}]}\right) W} \;.$$

Here, the covariant derivative operator ∇ is the one we encountered in §2.4 and §2.5.

Recall from §2.3 that the $\frac{\delta}{\delta x}$ span the **horizontal distribution**, and that the $F\frac{\partial}{\partial y}$ span the **vertical distribution**. These notions make sense with respect to the Sasaki (type) metric on $TM \smallsetminus 0$. Their Lie brackets can be computed with the help of the Cartan formula.

- Formula (3.7.2), together with the fact that ℓ and F are both covariantly constant along all horizontal directions, will show that

$$(3.7.3) \quad \boxed{\left[\frac{\delta}{\delta x^k}, \frac{\delta}{\delta x^l}\right] \;=\; -\, \ell^j\, R_j^{\ i}{}_{kl}\, F\frac{\partial}{\partial y^i}} \;.$$

  See Exercise 3.7.3. Since the Lie bracket of the horizontal vector fields $\frac{\delta}{\delta x^k}$, $\frac{\delta}{\delta x^l}$ is strictly vertical, the horizontal distribution is not involutive, hence not integrable.

- Next, (3.7.2) and primarily (3.4.10) can be used to deduce that

(3.7.4)

$$\boxed{\left[\frac{\delta}{\delta x^k}, F\frac{\partial}{\partial y^l}\right] \;=\; \left\{ \dot{A}^i{}_{kl} \;+\; \frac{\ell^i}{F}\,(F\,\ell_k)_{x^l} \;-\; \ell^i\, \frac{N_{kl}}{F} \right\} F\frac{\partial}{\partial y^i}} \;.$$

  This one involves considerable detail. See Exercise 3.7.4.

- Finally, (3.7.2) and an appropriate interchange formula from §3.6 will lead to

$$(3.7.5) \quad \boxed{\left[F\frac{\partial}{\partial y^k}, F\frac{\partial}{\partial y^l}\right] \;=\; (\ell_k\, \delta^i{}_l \;-\; \ell_l\, \delta^i{}_k)\, F\frac{\partial}{\partial y^i}} \;.$$

  See Exercise 3.7.5. This, together with the identity

$$[f\hat{X}, g\hat{Y}] \;=\; fg\,[\hat{X},\hat{Y}] \;+\; f\,(dg)(\hat{X})\,\hat{Y} \;-\; g\,(df)(\hat{Y})\,\hat{X}\;,$$

  will show that the vertical distribution is involutive, hence integrable.

Exercises

Exercise 3.7.1: Verify the Cartan formula (3.7.1).

Exercise 3.7.2: Derive (3.7.2) in detail.

Exercise 3.7.3: Let $\hat{X}$, $\hat{Y}$ be $\frac{\delta}{\delta x^k}$, $\frac{\delta}{\delta x^l}$, respectively.

(a) In (3.7.2), set W equal to ℓ. Show that one obtains

$$\nabla_{[\frac{\delta}{\delta x^k}, \frac{\delta}{\delta x^l}]} \ell \ = \ - \ \ell^j \ R_j{}^i{}_{kl} \ \frac{\partial}{\partial x^i} \ .$$

Formula (2.5.14) says that ℓ is covariantly constant in all horizontal directions. So the Lie bracket here cannot possibly be horizontal, unless it vanishes.

(b) With the help of Exercise 2.3.5, check that F is constant along $[\frac{\delta}{\delta x^k}, \frac{\delta}{\delta x^l}]$.

(c) Using the above, together with $\frac{\delta y^i}{F} = (\nabla \ell)^i + \ell^i \, d(\log F)$, show that

$$\left(\frac{\delta y^i}{F}\right)\left(\left[\frac{\delta}{\delta x^k} \, , \, \frac{\delta}{\delta x^l} \right] \right) \ = \ - \ \ell^j \ R_j{}^i{}_{kl} \ .$$

(d) In the Cartan formula, set $\omega := dx^i$. Check that it gives

$$(dx^i)\left(\left[\frac{\delta}{\delta x^k} \, , \, \frac{\delta}{\delta x^l} \right] \right) \ = \ 0 \ .$$

We can now conclude that

$$\left[\frac{\delta}{\delta x^k} \, , \, \frac{\delta}{\delta x^l} \right] \ = \ - \ \ell^j \ R_j{}^i{}_{kl} \ F \frac{\partial}{\partial y^i} \ .$$

In particular, the Lie bracket of $\frac{\delta}{\delta x^k}$ with $\frac{\delta}{\delta x^l}$ is *strictly vertical* !

Exercise 3.7.4: Return to (3.7.2). This time, set

$$\hat{X} \ := \ \frac{\delta}{\delta x^k} \qquad \text{and} \qquad \hat{Y} \ := \ F \frac{\partial}{\partial y^l} \ .$$

(a) Show that the curvature formula (3.7.2), together with (3.4.10), gives

$$\nabla_{[\frac{\delta}{\delta x^k}, F\frac{\partial}{\partial y^l}]} \ell \ = \ \dot{A}^i{}_{kl} \ \frac{\partial}{\partial x^i} \ .$$

(b) With the help of
  * $\frac{\delta F}{\delta x} = 0$ (Exercise 2.3.5) and $\ell_{l|k} = 0$ (2.5.16),
  * the relationship (2.5.3) between $\frac{\delta}{\delta x^k} \ell_l$ and $\ell_{l|k}$,
  * formulas (2.4.9), (2.3.2) twice, (2.3.1), (1.4.1) for Γ, N, γ, g,
  * Euler's theorem (Theorem 1.2.1) applied to $(\frac{1}{2}F^2)_{y^k}$,

show that:

$$d(\log F)\left(\left[\frac{\delta}{\delta x^k}\,,\,F\frac{\partial}{\partial y^l}\right]\right) = \frac{1}{F}\,(F\,\ell_k\,)_{x^l} - \frac{N_{kl}}{F}\,.$$

(c) Using the above, together with $\frac{\delta y^i}{F} = (\nabla \ell)^i + \ell^i\, d(\log F)$, show that

$$\left(\frac{\delta y^i}{F}\right)\left(\left[\frac{\delta}{\delta x^k}\,,\,F\frac{\partial}{\partial y^l}\right]\right) = \dot{A}^i{}_{kl} + \frac{\ell^i}{F}\,(F\,\ell_k\,)_{x^l} - \ell^i\,\frac{N_{kl}}{F}\,.$$

(d) Deduce from the Cartan formula that

$$(dx^i)\left(\left[\frac{\delta}{\delta x^k}\,,\,F\frac{\partial}{\partial y^l}\right]\right) = 0\,.$$

Now conclude that

$$\left[\frac{\delta}{\delta x^k}\,,\,F\frac{\partial}{\partial y^l}\right] = \left\{\dot{A}^i{}_{kl} + \frac{\ell^i}{F}\,(F\,\ell_k\,)_{x^l} - \ell^i\,\frac{N_{kl}}{F}\right\}F\frac{\partial}{\partial y^i}\,.$$

So, the Lie bracket of $\frac{\delta}{\delta x^k}$ with $F\frac{\partial}{\partial y^l}$ is *vertical.*

Exercise 3.7.5: In (3.7.2), let us set

$$\hat{X} := F\frac{\partial}{\partial y^k}\,, \qquad \hat{Y} := F\frac{\partial}{\partial y^l}\,.$$

We now complement Exercises 3.7.3 and 3.7.4 by computing the remaining Lie brackets.

(a) With the help of an interchange formula on ℓ^i, show that

$$\nabla_{[F\frac{\partial}{\partial y^k}\,,\,F\frac{\partial}{\partial y^l}]}\,\ell = (\ell_k\,\delta^i{}_l - \ell_l\,\delta^i{}_k)\,\frac{\partial}{\partial x^i}\,.$$

(b) By applying an interchange formula to $\log F$, check that it is constant along $[F\frac{\partial}{\partial y^k}\,,\,F\frac{\partial}{\partial y^l}]$.

(c) Use the Cartan formula to verify that

$$(dx^i)\left(\left[F\frac{\partial}{\partial y^k}\,,\,F\frac{\partial}{\partial y^l}\right]\right) = 0\,.$$

Imitate the trend of thought in Exercise 3.7.4 to conclude that

$$\left[F\frac{\partial}{\partial y^k}\,,\,F\frac{\partial}{\partial y^l}\right] = (\ell_k\,\delta^i{}_l - \ell_l\,\delta^i{}_k)\,F\frac{\partial}{\partial y^i}\,.$$

Thus the Lie bracket of $F\frac{\partial}{\partial y^k}$ with $F\frac{\partial}{\partial y^l}$ is *vertical.*

3.8 Derivatives of the Geodesic Spray Coefficients G^i

In Exercise 2.3.3, we introduced certain quantities

$$\boxed{G^i \ := \ \gamma^i{}_{jk}\, y^j\, y^k} \ , \tag{3.8.1}$$

where the $\gamma^i{}_{jk}$ are the formal Christoffel symbols of the fundamental tensor. (Our G^i is twice that in [AIM].) In that exercise, it was shown that

$$\boxed{\frac{1}{2}\,\frac{\partial G^i}{\partial y^j} \ = \ N^i{}_j} \ , \tag{3.8.2}$$

where $N^i{}_j$ is the nonlinear connection [see (2.3.2)].

Constant speed Finslerian geodesics are (see §5.3) curves in M that obey the coupled system of quasilinear second order ODEs $\ddot{x}^i + G^i = 0$, where in G^i we set $y^i := \dot{x}^i$ (that is, $\frac{dx^i}{dt}$). The vector field

$$y^k\,\frac{\delta}{\delta x^k} \ = \ y^k \left(\frac{\partial}{\partial x^k} \ - \ N^i{}_k\,\frac{\partial}{\partial y^i} \right) \ = \ y^i\,\frac{\partial}{\partial x^i} \ - \ G^i\,\frac{\partial}{\partial y^i}$$

on $TM \setminus 0$ is called the **geodesic spray** (see [AIM], [AB]). In the second equality, we have used the fact that G^i is homogeneous of degree two in y, and Euler's theorem (1.2.1): $y^j\,\frac{\partial G^i}{\partial y^j} = 2G^i$.

As we have seen, the first y-derivative of $\frac{1}{2}G^i$ gives the nonlinear connection $N^i{}_j$. We show in the following exercises that successive y-derivatives continue to yield quantities which are both interesting and geometrically significant. The following is a quick summary. We have

$$\boxed{\frac{1}{2}\,(G^i)_{y^j y^k} \ = \ \Gamma^i{}_{jk} \ + \ \dot{A}^i{}_{jk}} \ . \tag{3.8.3}$$

The quantity on the right is known as the **Berwald connection** ${}^b\Gamma^i{}_{jk}$. It has two *a priori* nonzero curvature tensors just like the Chern connection does. The "*hv*" one is given by

$$\boxed{{}^bP_j{}^i{}_{kl} \ = \ P_j{}^i{}_{kl} \ - \ \dot{A}^i{}_{jk;l} \ = \ - \ \frac{F}{2}\,(G^i)_{y^j y^k y^l}} \ . \tag{3.8.4}$$

This has the consequence [part (c) of Exercise 3.8.5]:

$$\boxed{\dot{A}_{jkl} \ = \ -\frac{1}{4}\,y_i\,(G^i)_{y^j y^k y^l}} \ . \tag{3.8.5}$$

Finally, the "*hh*" quantities $R^i{}_k := \ell^j\,R_j{}^i{}_{kl}\,\ell^l$ are intimately related to the flag curvature. We make precise the relation in §3.9. Exercise 3.3.4

demonstrates that $R^i{}_k$ can be expressed rather simply in terms of the nonlinear connection $N^i{}_j$. Namely,

$$\boxed{R^i{}_k \;=\; \ell^j \left(\frac{\delta}{\delta x^k} \frac{N^i{}_j}{F} \;-\; \frac{\delta}{\delta x^j} \frac{N^i{}_k}{F} \right)} \,, \tag{3.8.6}$$

where $\frac{\delta}{\delta x^k} := \frac{\partial}{\partial x^k} - N^i{}_k \frac{\partial}{\partial y^i}$. Through (3.8.2), the curvatures $R^i{}_k$ can then be expressed entirely in terms of the x and y derivatives of the geodesic spray coefficients G^i. This was carried out in Exercise 2.5.7. The result reads:

$$\boxed{\begin{aligned} F^2\, R^i{}_k \;=&\;\; 2\,(\bar{G}^i)_{x^k} \;-\; (\bar{G}^i)_{y^j}\,(\bar{G}^j)_{y^k} \\ &- y^j\,(\bar{G}^i)_{y^k x^j} \;+\; 2\,\bar{G}^j\,(\bar{G}^i)_{y^k y^j} \end{aligned}} \,, \tag{3.8.7}$$

where

$$\boxed{\bar{G}^s \;:=\; \frac{G^s}{2}} \,.$$

Exercises

Exercise 3.8.1:

(a) Check that

$$G_i \;=\; (g_{ij})_{x^k}\, y^j\, y^k \;-\; \frac{1}{2}\,(g_{jk})_{x^i}\, y^j\, y^k \,.$$

(b) Then show that

$$G_i \;=\; \frac{1}{2}\, y^j\,(F^2)_{x^j y^i} \;-\; \frac{1}{2}\,(F^2)_{x^i} \,.$$

Exercise 3.8.2: Verify that $G^i = \Gamma^i{}_{jk}\, y^j\, y^k$, where the $\Gamma^i{}_{jk}$ are the Chern connection coefficients.

Exercise 3.8.3:

(a) With the help of Exercise 2.3.3, prove that

$$\frac{1}{2}\,(G^i)_{y^j y^k} \;=\; \Gamma^i{}_{jk} \;+\; \dot{A}^i{}_{jk} \,.$$

Hint: review the derivation of (3.3.3) and (3.4.9).

(b) Why is this statement potentially very useful?

Exercise 3.8.4: Revisit part (a) of Exercise 3.8.3. There, the left-hand side of the equation defines another torsion-free linear connection on π^*TM.

This connection is known as the **Berwald connection**. Its connection coefficients are

$$ {}^b\Gamma^i{}_{jk} \;:=\; \frac{1}{2}\,(\,G^i\,)_{y^j y^k} \;=\; \Gamma^i{}_{jk} \;+\; \dot{A}^i{}_{jk} \;. $$

Note that it differs from the Chern connection by a simple $\dot{A}$ term.

(a) Express the curvature tensors of the Berwald connection in terms of the R, P, Q of the Chern connection, together with suitable derivatives of the Cartan tensor. Show that one has:

$$ \begin{aligned} {}^bR_j{}^i{}_{kl} &= R_j{}^i{}_{kl} \;+\; \left[\dot{A}^i{}_{jl|k} \;+\; \dot{A}^i{}_{sk}\,\dot{A}^s{}_{jl} \;-\; \begin{pmatrix} k \to l \\ l \to k \end{pmatrix} \right] \;, \\ {}^bP_j{}^i{}_{kl} &= P_j{}^i{}_{kl} \;-\; \dot{A}^i{}_{jk;l} \;, \\ {}^bQ_j{}^i{}_{kl} &= Q_j{}^i{}_{kl} \;=\; 0 \;. \end{aligned} $$

(b) Prove that

$$ {}^bP_j{}^i{}_{kl} \;=\; -\,F\,\frac{\partial\, {}^b\Gamma^i{}_{jk}}{\partial y^l} \;=\; P_j{}^i{}_{kl} \;-\; \dot{A}^i{}_{jk;l} \;=\; -\,\frac{F}{2}\,(\,G^i\,)_{y^j y^k y^l} \;. $$

Hint: you may want to borrow the strategy in §3.3.

(c) Why is the Berwald connection torsion-free?

Exercise 3.8.5: So far, we have seen that taking

* one y derivative of $\frac{1}{2}G^i$ yields the nonlinear connection,
* two y derivatives gives the Berwald connection, and
* three y derivatives produces $\frac{-1}{F}$ times the Berwald hv-curvature ${}^bP$.

(a) Verify that $y_i := g_{ij}\,y^j = F\ell_i = F\,F_{y^i}$.

(b) Use (3.4.8) to show that $\ell^i\,P_{jikl} = -\,\ell^i\,P_{ijkl}$.

(c) Using part (b) of Exercise 3.8.4, together with the Bianchi identity (3.4.9), and (2.5.17), prove that

$$ \dot{A}_{jkl} \;=\; -\,\frac{1}{4}\,y_i\,(\,G^i\,)_{y^j y^k y^l} \;. $$

3.9 The Flag Curvature

We now illustrate the usefulness of some of our Bianchi identities in the study of the flag curvature. This is a geometrical invariant that generalizes the sectional curvature of Riemannian geometry. Furthermore, the flag curvature is insensitive to whether one is using the Chern-, Cartan-, Berwald-, or Hashiguchi-connection.

3.9 A. Its Definition and Its Predecessor

We begin with the notion of a flag on M. The act of installing a flag at $x \in M$ necessitates a nonzero $y \in T_x M$ which serves as the **flagpole**. The actual **flag** itself is described by one edge (for instance, ℓ) along the flagpole and another **transverse edge**, say $V := V^i \frac{\partial}{\partial x^i}$. The flag curvature that we define does not depend on the actual "length" of the edge along the flagpole. This is made precise in (3.9.2). See also Exercise 3.9.1.

Now that we have a flag, we can associate with it a number $K(y, V)$. It is obtained by carrying out the following computation at the point $(x, y) \in TM \setminus 0$, and viewing y, V as sections of the pulled-back bundle $\pi^* TM$:

$$K(y,V) \ := \ \frac{V^i \ (y^j \ R_{jikl} \ y^l \ ) \ V^k}{g(y,y) \ g(V,V) \ - \ [g(y,V)]^2} \, . \tag{3.9.1}$$

As a reminder,

$$g \ := \ g_{ij\,(x,y)} \ dx^i \otimes dx^j \ := \ \left(\frac{1}{2} F^2 \right)_{y^i y^j} \ dx^i \otimes dx^j$$

is a Riemannian metric on the pulled-back bundle $\pi^* TM$.

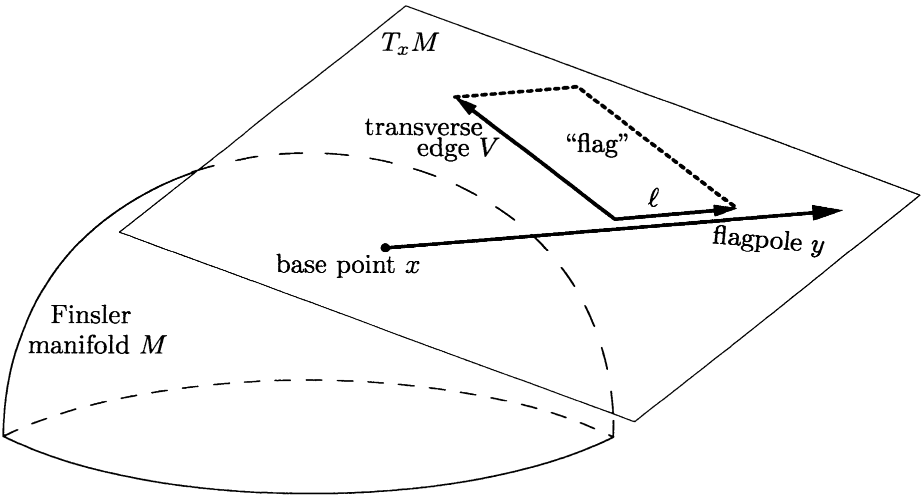

Figure 3.1

A typical flag, based at the point x on a Finsler manifold M. The flagpole is y, and the "cloth" part of the flag is $\ell \wedge V$. The entire flag lies in the tangent space $T_x M$.

The right-hand side of (3.9.1) is unchanged if we divide both numerator and denominator by $F^2(x,y)$. Thus

$$K(y,V) \; = \; K(\ell,V) \,, \tag{3.9.2}$$

where

$$\begin{aligned} K(\ell,V) \; &:= \; \frac{V^i \, (\, \ell^j \, R_{jikl} \, \ell^l \,) \, V^k}{g(\ell,\ell) \, g(V,V) \; - \; [g(\ell,V)]^2} \\ &= \; \frac{V^i \, R_{ik} \, V^k}{g(V,V) \; - \; [g(\ell,V)]^2} \; . \end{aligned} \tag{3.9.3}$$

Here, we have used the fact (2.1.7) that $g(\ell,\ell) = 1$. The number we have just defined is called the **flag curvature** for the flag $y \wedge V$ or $\ell \wedge V$.

There is no loss of generality in choosing only those transverse edges V that are g-orthogonal to the flagpole y. Indeed, any arbitrary V can be decomposed as $V = W + \xi\,\ell$, where $g(\ell,W) = 0$ and ξ is a scalar multiple. Part (c) of Exercise 3.4.4 tells us that

$$\ell^i \, R_{ik} \; = \; 0 \; = \; R_{ik} \, \ell^k \; .$$

Hence

* $V^i \, R_{ik} \, V^k \; = \; W^i \, R_{ik} \, W^k .$
* $g(V,V) - [g(\ell,V)]^2 \; = \; g(W,W).$

There is no need for the $[g(\ell,W)]^2$ term because it is zero. Consequently

$$K(\ell,V) \; = \; K(\ell,W) \; .$$

Although the above calculation was carried out on π^*TM, it admits an interpretation on T_xM as follows. The flagpole $y \in T_xM$ singles out an inner product g as defined immediately after (3.9.1). That inner product may be viewed as living on T_xM, and lets us make sense of the $(n-1)$-dimensional g-orthogonal complement of y. It is from this subspace in T_xM that we should pick our transverse edges V. For further discussions, see Exercises 3.9.4 and 3.9.5.

The following object is algebraically a **predecessor of the flag curvature**:

$$\begin{aligned} K(\ell,V,W) \; &:= \; \frac{V^i \, (\, \ell^j \, R_{jikl} \, \ell^l \,) \, W^k}{g(V,W) \; - \; g(\ell,V) \, g(\ell,W)} \\ &= \; \frac{V^i \, R_{ik} \, W^k}{g(V,W) \; - \; g(\ell,V) \, g(\ell,W)} \; . \end{aligned} \tag{3.9.4}$$

- It is clear that $K(\ell,V,V) = K(\ell,V)$.
- The fact (3.4.7) that R_{ik} is symmetric tells us that

$$K(\ell,W,V) \; = \; K(\ell,V,W) \; . \tag{3.9.5}$$

- One can derive the **polarization identity**

$$K(\ell, V, W) = \frac{1}{4} K(\ell, V+W) - \frac{1}{4} K(\ell, V-W) . \tag{3.9.6}$$

3.9 B. An Interesting Family of Examples of Numata Type

We now calculate the flag curvature for a family of examples that are valid in all dimensions. These examples are interesting because their flag curvatures do not depend on the transverse edges V at all. Moreover, the flag curvatures in question have a simple nonconstant dependence on the basepoint x and the flagpole y.

To describe our family, begin with a smooth function $f = f(x)$ defined on an open subset of $\mathbb{R}^n$. We either require that f has at least one critical point, or scale it down so that the open subset

$$M := \{ x \in \mathbb{R}^n : \sqrt{\delta^{ij} f_{x^i} f_{x^j}} < 1 \}$$

is nonempty. In other words, M is the open set on which the gradient of f has Euclidean length strictly less than 1. On TM, define

$$\boxed{F(x,y) := \sqrt{\delta_{ij} y^i y^j} + f_{x^i} y^i =: \alpha + (df)(y)} . \tag{3.9.7}$$

This is a Randers metric. Its underlying Riemannian metric is the Euclidean one on $\mathbb{R}^n$, and its drift 1-form is df. As stated in §1.3C, and proven in §11.1, the fact that $\| df \| < 1$ on M enables F to satisfy all the axioms of a Finsler structure.

With the help of

$$\ell_i := \frac{y_i}{F} = F_{y^i} = \frac{\delta_{is} y^s}{\alpha} + f_{x^i} , \tag{3.9.8}$$

one can check that

$$G_i := \gamma_{ijk} y^j y^k = \ell_i f_{x^j x^k} y^j y^k . \tag{3.9.9}$$

Hence

$$G^i := \gamma^i{}_{jk} y^j y^k = \frac{y^i}{F} f_{x^j x^k} y^j y^k . \tag{3.9.10}$$

As always, our G^i is twice that of [AIM]'s. Formula (3.9.10) says that

$$G^i = y^i p , \tag{3.9.11}$$

where

$$p := \frac{1}{F} f_{x^j x^k} y^j y^k . \tag{3.9.12}$$

Let us digress to point out something. Note that the analogous $\tilde{G}^i$ of the Euclidean metric is zero. We may ([AIM], §3.3) thus interpret (3.9.11) as

$$G^i - \tilde{G}^i = p y^i .$$

This says that the Finsler structure F is **projectively related** to the Euclidean structure. The function p is *our* **projective factor**.

We turn to the calculation of $F^2 R^i{}_k$. Let us begin with formula (3.8.7). It reads:

$$F^2 R^i{}_k = 2 (\bar{G}^i)_{x^k} - (\bar{G}^i)_{y^j} (\bar{G}^j)_{y^k} - y^j (\bar{G}^i)_{y^k x^j} + 2 \bar{G}^j (\bar{G}^i)_{y^k y^j} ,$$

where $\bar{G}^s := \frac{G^s}{2}$. Into this formula we substitute (3.9.11). Since p is positively homogeneous of degree 1 in y, Euler's theorem (Theorem 1.2.1) says that

$$y^i p_{y^i} = p , \qquad y^i p_{y^j y^i} = 0 .$$

These simplify the result to

$$F^2 R^i{}_k = (\bar{p}^2 - y^j \bar{p}_{x^j}) \delta^i{}_k + y^i (2 \bar{p}_{x^k} - \bar{p} \bar{p}_{y^k} - y^j \bar{p}_{x^j y^k}) ,$$

where $\bar{p}$ is the projective factor that [AIM] would have used:

$$\text{(3.9.13)} \qquad \boxed{\bar{p} := \frac{1}{2} p := \frac{1}{2F} f_{x^j x^k} y^j y^k} .$$

We are now ready to compute the flag curvature. Formula (3.9.3) implies that

$$F^2 K(\ell, V) = \frac{V_i (F^2 R^i{}_k) V^k}{g(V,V) - [g(\ell, V)]^2} .$$

In §3.9A, we explained why one only needs to use transverse edges V that are g-orthogonal to the flagpole y. Consequently, the $g(\ell, V)$ term drops out, and we have

$$F^2 K(\ell, V) = \frac{V_i (F^2 R^i{}_k) V^k}{g(V,V)} .$$

Into this we substitute our formula for $F^2 R^i{}_k$, and use

$$\delta^i{}_k V_i V^k = V_i V^i = g(V,V) , \qquad y^i V_i = g(y, V) = 0 .$$

These maneuvers give

$$K = \frac{1}{F^2} (\bar{p}^2 - y^j \bar{p}_{x^j}) .$$

Inputting formula (3.9.13) for $\bar{p}$ and computing briefly, we get

(3.9.14)

$$\boxed{K = K(x,y) = \frac{3}{4F^4} (f_{x^i x^j} y^i y^j)^2 - \frac{1}{2F^3} (f_{x^i x^j x^k} y^i y^j y^k)} .$$

Here, the second- and third-order partials on f are functions of x only. As promised, the flag curvatures have no dependence on the transverse edges

V. Furthermore, they have an explicit simple dependence on the basepoint x and the flagpole y. That dependence on y is quite obviously variable.

As a concrete example, take f to be a quadratic polynomial $c_{ij}\, x^i x^j$, where the c_{ij} are constants. Formula (3.9.14) then simplifies to

$$K \;=\; \frac{3}{4\,F^4}\, \left(c_{ij}\; y^i\, y^j \right)^2 \;.$$

In this case, the flag curvatures K are strictly positive if and only if the matrix (c_{ij}) is positive-definite.

Let us put our family of examples into the proper perspective. It is a theorem of Numata's [Nu] that the flag curvatures of

$$(*) \qquad (x\,,\, y) \;\mapsto\; \sqrt{q_{ij}(y)\; y^i\, y^j} \;+\; \tilde{b}_i(x)\; y^i \;,$$

where $\tilde{b}$ is a *closed* 1-form, depend only on the basepoint x and the flagpole $y \in T_xM$, but not on the transverse edges V. See Matsumoto's account in [AIM]. We hasten to point out that the dominant term here is a *Minkowski* norm of a special type, rather than the norm of a Riemannian metric. The latter would look like $\sqrt{\tilde{a}_{ij}(x)\, y^i y^j}$ instead. Thus, the above [namely, those given by (*)] Finsler **metrics of Numata type** are in general *not* Randers metrics. However, the ones we considered in our calculations are of both Numata and Randers type because their q_{ij} are constants. For a companion investigation that focuses on Randers metrics rather than those of Numata type, see Exercise 3.9.8.

Exercises

Exercise 3.9.1: Show that as far as the flag curvature is concerned, the actual dimensions of the flag are irrelevant. In other words, neither the size of the edge along the flagpole, nor that of the transverse edge, matters. Mathematically, this means that

$$K(\,\alpha\, y\,,\, \beta\, V\,) \;=\; K(y, V)$$

for all constants α and β.

Exercise 3.9.2: Derive the polarization identity (3.9.6).

Exercise 3.9.3: Check that

(a) $R_{ii} = (g_{ii} - \ell_i\,\ell_i)\;\; K(\ell,\, \frac{\partial}{\partial x^i}\,)$.

(b) $R_{ik} = (g_{ik} - \ell_i\,\ell_k)\;\; K(\ell,\, \frac{\partial}{\partial x^i}\,,\, \frac{\partial}{\partial x^k}\,)$.

Exercise 3.9.4: Fix a flagpole $0 \neq y \in T_xM$. There are $n-1$ linearly independent transverse edges. Using the Riemannian metric $g_{ij\,(x,y)}\, dx^i \otimes dx^j$, we can normalize these transverse edges and require them to be perpendicular to each other, as well as to the flagpole. Denote the resulting collection by $\{e_\alpha\}$. Adding $e_n := \ell$ to this collection produces a special $g_{(x,y)}$ orthonormal basis $\{e_a\}$ for T_xM, exactly the type discussed in §2.2. Show that with respect to this basis:

(a) $R_{\alpha\alpha} = K(\ell, e_\alpha)$.

(b) $R_{nn} = 0$.

Here, R_{ac} equals $u_a^{\ i} R_{ik} u_c^{\ k}$, a consequence of Exercise 3.3.1.

Exercise 3.9.5: The **Ricci scalar** is defined as

$$Ric := g^{ik} R_{ik} .$$

(a) Show that it is a sum of $n-1$ appropriate flag cuvatures.

(b) Identify the flags behind these flag curvatures.

(c) Explain why Ric can actually be obtained from $R_j^{\ i}{}_{kl}$ without ever using the fundamental tensor or its inverse.

We would like to remark that:

- In two dimensions, the Ricci scalar is the same as (the Finslerian analogue of) the Gaussian curvature. See §4.4 and especially Exercises 4.4.7, 4.4.8.
- The Ricci scalar plays a central role in the concept of Finsler–Einstein structures. An elementary discussion of this scalar is given in §7.6.

Exercise 3.9.6: Using (3.5.6), one can express R_{jikl} in terms of R_{ik}, A_{ijk} and their derivatives. In this exercise, let us derive an alternative formula in which the dependence on derivatives is eliminated. The price we pay is that the algebraic dependence on R_{ik}, A_{ijk} is made much more complicated than before.

In Riemannian geometry, the full curvature tensor is expressible as a sum of sectional curvatures (each multiplied by the area of the corresponding parallelogram or "flag"). An explicit formula can be found in Cheeger–Ebin [CE]. It is the goal of this exercise to generalize such a statement to Finsler geometry, and to correct two typographical errors in the above reference.

Let W, V, X, Y be local sections of π^*TM over $TM \setminus 0$. Define the quantities

$$(W,V,X,Y) := W^j V^i X^k Y^l R_{jikl} ,$$
$$(Y,X) := (Y,X,X,Y) = Y^j X^i X^k Y^l R_{jikl} .$$

(a) Let $R(X,Y)\,W := W^j X^k Y^l R_j^{\ i}{}_{kl} \frac{\partial}{\partial x^i}$. Check that

$$(W,V,X,Y) = g(\,R(X,Y)\,W\,,\,V\,) ,$$
$$(Y,X) = g(\,R(X,Y)\,Y\,,\,X\,) .$$

Hence $R_{ik} = g(\,R(b_k,\ell)\,\ell\,,\,b_i\,) = (\ell, b_i, b_k, \ell)$.

(b) In order to avoid clutter, let us introduce the abbreviation

$$B(WVXY) \ := \ -A(W,\, V,\, R(X,Y)\,\ell) \ .$$

This B is symmetric in its first two arguments, and skew-symmetric on the last two. It vanishes when the Finsler structure is Riemannian, or when one of its first two arguments is proportional to the canonical global section ℓ. Show that the symmetry properties (3.2.4), (3.1.3), (3.4.4), (3.4.5) of R_{jikl} can be re-expressed as

$$(W,V,X,Y) \ + \ (X,V,Y,W) \ + \ (Y,V,W,X) \ = \ 0 \ ,$$

$$(W,V,Y,X) \ = \ -(W,V,X,Y) \ ,$$

$$(V,W,X,Y) \ + \ (W,V,X,Y) \ = \ 2\, B(WVXY) \ ,$$

$$\begin{aligned} (X,Y,W,V) - (W,V,X,Y) \ = \ & [\,B(XYWV) \ - \ B(WVXY)\,] \\ & +[\,B(XVYW) \ + \ B(YWXV)\,] \\ & +[\,B(VYWX) \ + \ B(WXVY)\,] \ . \end{aligned}$$

(c) Verify that

$$(X,Y) \ - \ (Y,X) \ = \ -2\, B(YXXY) \ .$$

Thus when the Finsler structure is of Riemannian type, we have $(X,Y) = (Y,X)$. Nevertheless, for Finsler geometry in general, we still have $(X,\ell) = (\ell,X)$.

(d) Lastly, prove that $6\,(W,V,X,Y)$ is equal to

$$\begin{aligned} & (X+V,Y+W) \ - \ (X+W,Y+V) \\ & + (X+W,Y) \ + \ (X+W,V) \ + \ (X,Y+V) \ + \ (W,Y+V) \\ & - (X+V,Y) \ - \ (X+V,W) \ - \ (X,Y+W) \ - \ (V,Y+W) \\ & + (X,W) \ + \ (V,Y) \ + \ (V,W) \\ & - (W,Y) \ - \ (X,V) \ - \ (W,V) \\ & + 6\, B(WVXY) \\ & + 4\,[\,B(YWVX) \ - \ B(YVWX)\,] \\ & - 2\,[\,B(XYWV) \ + \ B(XWVY) \ + \ B(XVYW)\,] \ . \end{aligned}$$

When the space is Riemannian, all B terms vanish and (X,Y) is symmetric. Then the above reduces to the formula intended by Cheeger–Ebin. Their typographical mistakes consist of the following: instead of $(X+W,Y)$ and $-(X,V)$, they had $(V+W,Y)$ and $-(X,Y)$.

(e) Rewrite that last formula as $R_{jikl} \ = \ \dots$.

Exercise 3.9.7: Consider

$$F(x,y) \ := \ \sqrt{\delta_{ij}\ y^i\, y^j} \ + \ \tilde{b}_i\ y^i \ =: \ \alpha \ + \ \tilde{b}(y)\ ,$$

where $\tilde{b}$ is a **closed** 1-form. Restrict to an open subset M of $\mathbb{R}^n$, so that on M we have

$$\sqrt{\delta^{ij}\ \tilde{b}_i\, \tilde{b}_j} \ < \ 1\ .$$

(a) Show that the flag curvatures of this F are given by

$$K(x,y) \ = \ \frac{3}{4\,F^4}\ \left([\,\tilde{b}_i\,]_{x^j}\ y^i\, y^j \right)^2 \ - \ \frac{1}{2\,F^3}\ \left([\,\tilde{b}_i\,]_{x^j x^k}\ y^i\, y^j\, y^k \right)\ .$$

(b) We insist that the F considered here is *not* more general than that studied in the section proper. Why?

Exercise 3.9.8: Finsler metrics of Numata type were discussed at the end of §3.9. Let us now consider Randers metrics

$$F(x,y) \ := \ \sqrt{\tilde{a}_{ij}(x)\ y^i\, y^j} \ + \ \epsilon\ \tilde{b}_i(x)\ y^i\ ,$$

where $0 < \epsilon < 1$ is a constant. We assume that:

* The Riemannian metric $\tilde{a}$ has **constant** sectional curvature.
* The 1-form $\tilde{b}$ has Riemannian norm < 1 everywhere, and is **closed**.

(a) Explain why, for sufficiently small ϵ, the flag curvatures of F have the same sign as the sectional curvatures of $\tilde{a}$.

(b) Show that the flag curvatures of F do not depend on the transverse edges V of the flags.

(c) Now suppose $\tilde{a}$ has constant *positive* sectional curvature, and $\tilde{b}$ is not identically zero. For small enough ϵ, and in dimension $\geqslant 3$, is it true that the flag curvatures of F are positive numbers that *must* depend on the flagpole?

3.10 Schur's Lemma

If rotation of the transverse edge V about the flagpole y leaves the flag curvature unchanged, we say that our Finsler space has **scalar flag curvature**. In [M2] and [AIM], such spaces are said to be "*of scalar curvature.*" Denote this scalar by $\lambda = \lambda(x,y)$. For instance, every Finsler surface has scalar flag curvature. We have also seen (§3.9B) examples of this vintage in higher dimensions. In those examples, the dependence of λ on the flagpole y is explicit and *nonconstant.*

To effect a less cluttered presentation below, let us introduce two abbreviations:

(3.10.1) $$\boxed{h_{ij} \ := \ g_{ij} \ - \ \ell_i\ \ell_j}\ ,$$

the so-called **angular metric**, and

$$\boxed{h_{ijk} := g_{ij}\,\ell_k - g_{ik}\,\ell_j} \,. \tag{3.10.2}$$

Proposition 3.10.1. *Let (M,F) be a Finsler manifold and let R be the hh-curvature tensor of the Chern connection. The following four statements are equivalent:*

(a) $R_{ii} = \lambda_{(x,y)}\, h_{ii}$; *that is, (M,F) has scalar flag curvature $\lambda(x,y)$.*

(b) $R_{ik} = \lambda_{(x,y)}\, h_{ik}$.

(c) $R_{ikl} = \lambda_{(x,y)}\, h_{ikl} + \frac{1}{3}\,(h_{ik}\,\lambda_{;l} - h_{il}\,\lambda_{;k})$.

(d) *The full curvature tensor R_{jikl} has the formula*

$$\begin{aligned} &\lambda_{(x,y)}\,(g_{ik}\,g_{jl} - g_{il}\,g_{jk}) \\ &- (\dot{A}_{ijl|k} - \dot{A}_{ijk|l} + \dot{A}_{isk}\,\dot{A}^s{}_{jl} - \dot{A}_{isl}\,\dot{A}^s{}_{jk}) \\ &+ \lambda_{;j}\,h_{ikl} + \frac{1}{3}\,[\lambda_{;k}\,(\ell_i\,h_{jl} + h_{ijl}) - \lambda_{;l}\,(\ell_i\,h_{jk} + h_{ijk})] \\ &+ \frac{1}{3}\,(h_{ik}\,\lambda_{;l;j} - h_{il}\,\lambda_{;k;j}) \,. \end{aligned}$$

And, given any of them, we have

$$\ddot{A}_{ijk} + \lambda_{(x,y)}\,A_{ijk} + \frac{1}{3}\,(\lambda_{;i}\,h_{jk} + \lambda_{;j}\,h_{ki} + \lambda_{;k}\,h_{ij}) = 0 \,, \tag{3.10.3}$$

where

$$\ddot{A}_{ijk} := \dot{A}_{ijk|s}\,\ell^s = (A_{ijk|r}\,\ell^r)_{|s}\,\ell^s = A_{ijk|r|s}\,\ell^r\,\ell^s \,. \tag{3.10.4}$$

Remark: In case the above λ has no dependence on either x or y, the Finsler manifold in question is said to have **constant flag curvature**. We study such manifolds in Chapter 12.

Proof.

- The equivalence between (a) and (b) is a consequence of the polarization identity (3.9.6).
- Given the R_{ik} in (b), we can reconstruct $R^i{}_{kl}$ using the formula in Exercise 3.5.7, and then lower the index i to get (c). Conversely, (c) and the fact that $\lambda_{;s}\,\ell^s = 0$ (see Exercise 3.10.2) gives (b).
- Given (c), we use (3.5.6) to reconstruct R_{jikl}, thereby obtaining (d). The converse follows readily with the help of $\lambda_{;s}\,\ell^s = 0$ and $\lambda_{;r;s}\,\ell^s = 0$ (Exercise 3.10.2 again).

As for (3.10.3), first compute $R_{jikl} + R_{ijkl}$ two ways, once using (3.4.4) and the other using formula (d) here. Then we compare the two answers to deduce that

$$\begin{aligned}
& \dot{A}_{ijl|k} \ - \ \dot{A}_{ijk|l} \\
= \ & A_{ijk}\,(\lambda_{;l} \ + \ \lambda\,\ell_l\,) \ - \ A_{ijl}\,(\lambda_{;k} \ + \ \lambda\,\ell_k\,) \\
+ \ & \frac{1}{2}\,(\lambda_{;i}\,h_{jkl} \ + \ \lambda_{;j}\,h_{ikl}\,) \ + \ \frac{1}{3}\,h_{ij}\,(\lambda_{;k}\,\ell_l \ - \ \lambda_{;l}\,\ell_k\,) \\
+ \ & \frac{1}{6}\,[\,(h_{ik}\,\lambda_{;l;j} \ + \ h_{jk}\,\lambda_{;l;i}\,) \ - \ (h_{il}\,\lambda_{;k;j} \ + \ h_{jl}\,\lambda_{;k;i}\,)\,] \ .
\end{aligned}$$

Finally, we contract this with ℓ^l to get (3.10.3). □

For n-dimensional Finsler manifolds of scalar flag curvature:

* Let us contract the second Bianchi identity (3.5.3) with ℓ^j and ℓ^t, and lower the index i. We get

$$R_{ikl|t}\,\ell^t \ + \ R_{il|k} \ - \ R_{ik|l} \ = \ - \ \dot{A}_{ik}{}^{s}\,R_{sl} \ + \ \dot{A}_{il}{}^{s}\,R_{sk} \ .$$

* Into this we substitute formulas (b), (c) of the above proposition, and then use the fact that both h_{ij} and h_{ijk} have zero horizontal covariant derivatives (see Exercise 3.10.3).
* Contract with g^{ik} and relabel. The result is:

$$(n-2)\,\left(\dot{\lambda}\,\ell_i \ - \ \lambda_{|i} \ + \ \frac{1}{3}\,\lambda_{;i|r}\,\ell^r\right) \ = \ 0 \ , \tag{3.10.5}$$

where

$$\dot{\lambda} \ := \ \lambda_{|r}\,\ell^r \ . \tag{3.10.6}$$

Within the class of Finsler manifolds whose flag curvature has no dependence on the transverse edge, we now consider those for which there is no dependence on the flagpole y either. In that case, the flag curvature is simply a function $\lambda(x)$ of x only. All Riemannian surfaces have this property, where $\lambda(x)$ is simply the Gaussian curvature function. It turns out that the situation in higher dimensions is constrained by a good amount of rigidity, as described in the following **Schur's lemma**. Its proof is adapted from the one given in Matsumoto [M2]. We have also been informed by Lilia del Riego that this was proved in her thesis [delR].

Lemma 3.10.2. *Suppose*:

- *(M, F) is a connected Finsler manifold.*
- *Its dimension n is at least* 3.
- *Its flag curvature depends neither on the transverse edge nor on the flagpole; in other words, it is a function λ of x only.*

Then λ must in fact be constant.

Remarks:

* In dimension two, the Gaussian curvature of every Riemannian metric is a function $K(x)$ of the position x only. And there are many Riemannian metrics for which this function is nonconstant. Thus, the second hypothesis in Schur's lemma is sharp.
* In §3.9B, we have encountered a particular family of examples of Numata type, for *all* dimensions. In those examples, the flag curvatures have no dependence on the transverse edges, but they do manifest a variable dependence on the flagpole. For this reason, the third hypothesis in Schur's lemma cannot be weakened.

Proof. The key hypotheses here are $\lambda_{;i} = 0$ (namely, λ has no dependence on y) and $n \geqslant 3$. They reduce (3.10.5) to

$$(**) \qquad \lambda_{|i} \; = \; \dot{\lambda} \, \ell_i \; ,$$

which implies that

$$\lambda_{|i;j} \; = \; \dot{\lambda}_{;j} \, \ell_i \; + \; \dot{\lambda} \, h_{ij} \; .$$

Since λ is a scalar, and is independent of y, we see from (2.5.3), (2.5.5) that

$$\lambda_{|i} \; = \; \frac{\delta \lambda}{\delta x^i} \; = \; \frac{\partial \lambda}{\partial x^i} \; ,$$

which is yet another function of x alone. Thus

$$\lambda_{|i;j} \; := \; F \frac{\partial}{\partial y^j} \, \lambda_{|i} \; = \; 0 \; .$$

So the above becomes

$$(***) \qquad \dot{\lambda}_{;j} \, \ell_i \; + \; \dot{\lambda} \, h_{ij} \; = \; 0 \; .$$

Since M is higher than 1-dimensional, in each fiber of π^*TM there is at least one nonzero U which is g-orthogonal to ℓ. Contracting (***) with U^i gives $\dot{\lambda} \, U_j = 0$. That is, $\dot{\lambda} \; U = 0$. Hence

$$\dot{\lambda} \; = \; 0 \; ,$$

which by (**) gives $\lambda_{|i} = 0$. Since $\lambda = \lambda(x)$, the vanishing of its horizontal derivatives is synonymous with

$$\frac{\partial \lambda}{\partial x^i} \; = \; 0 \; .$$

So λ is constant on the *connected* M. □

Exercises

Exercise 3.10.1: Explain why every Finsler surface has scalar flag curvature.

Exercise 3.10.2: Suppose (M, F) has scalar flag curvature $\lambda(x, y)$.

(a) Check that $\lambda_{;s}\, \ell^s = 0$.

(b) Apply vertical covariant differentiation to the above statement; then use a Ricci identity (§3.6). Show that $\lambda_{;r;s}\, \ell^s = 0$. Explain why this conclusion can also be obtained through a much simpler way.

Exercise 3.10.3: Verify the following statements:

(a) $h_{is}\, \ell^s = 0$.

(b) $h_{ijs}\, \ell^s = h_{ij}$.

(c) $g^{ij}\, h_{ik} = \delta^j_{\ k} - \ell^j\, \ell_k$.

(d) $g^{ij}\, h_{ij} = n - 1$.

(e) $g^{ij}\, h_{ijk} = (n-1)\, \ell_k$.

(f) $h_{ij|k} = 0$.

(g) $h_{ijk|l} = 0$.

Exercise 3.10.4: We merely outlined the proof of Proposition 3.10.1. Give the details of the omitted computations.

Exercise 3.10.5: For Finsler manifolds of scalar flag curvature, derive the important equation (3.10.5) by following the given guidelines.

Exercise 3.10.6: In the proof of Lemma 3.10.2, we saw that if λ depends only on x, then $\lambda_{|i;j} = 0$. Now suppose λ is a function of y only, does the quantity $\lambda_{;j|i}$ have to vanish ?

Exercise 3.10.7:

(a) Show that the concept of flag curvature does not depend on whether one is using the Berwald or the Chern connection. (Nor does it depend on the Cartan and Hashiguchi connections mentioned in §2.4. Furthermore, it can be obtained through a **Jacobi endomorphism** in Foulon's [Fou] dynamical systems approach to Finsler geometry.)

(b) In part (a) of Exercise 3.8.4, we encountered the *hh*-curvature ${}^bR$ of the Berwald connection. With the help of part (d) in Proposition 3.10.1, prove that whenever a Finsler space has constant flag curvature λ, the Berwald curvature ${}^bR$ must have the form

$$ {}^bR_{jikl} \;=\; \lambda\, (\, g_{ik}\, g_{jl} \;-\; g_{il}\, g_{jk}\,) \;. $$

For this reason, the Berwald connection is particularly suited for the study of Finsler spaces of constant flag curvature.

Exercise 3.10.8:

(a) Show that for a Riemannian manifold, the Berwald and Chern connections both reduce to the Christoffel symbols of the Riemannian metric.

(b) A Riemannian manifold is said to have constant sectional curvature λ if, as a Finsler space, it has constant flag curvature λ. Explain why, in that case, the curvature tensor R_{jikl} (see the end of §3.3) of the Christoffel symbols $\gamma^i{}_{jk}$ has the form

$$R_{jikl} = \lambda\,(g_{ik}\,g_{jl} - g_{il}\,g_{jk})\ .$$

Exercise 3.10.9: Here is an interesting result given in [AIM].

Suppose M is connected, F is of scalar curvature $\lambda(x,y)$, and $\lambda_{|i} = 0$. Then λ must in fact be constant.

We may assume that λ is not identically zero; otherwise there is nothing to prove. Note also that there is no restriction on the dimension of M.

(a) Use the Ricci identity (3.6.1) to check that $\lambda_{;s}\,R^s{}_{ij} = 0$.

(b) Prove, from part (a) here and part (c) of Proposition 3.10.1, that

$$\lambda_{;i}\,\left(\frac{1}{3}\lambda_{;j} + \lambda\,\ell_j\right) - \lambda_{;j}\,\left(\frac{1}{3}\lambda_{;i} + \lambda\,\ell_i\right) = 0\ .$$

(c) Part (b) implies that two particular sections of π^*T^*M are linearly dependent at every point (x,y). Explain why there is a scalar field ξ such that $\lambda_{;i} = \xi\,(\frac{1}{3}\lambda_{;i} + \lambda\ell_i)$, and why contracting with ℓ^i gives $\xi\,\lambda = 0$.

(d) On the interior of the support of λ, use part (c) to show that λ has no dependence on y. Then use $\lambda_{|i} = 0$ to show that it has no dependence on x either. Finally, invoke the continuity of λ.

References

[AB] P. L. Antonelli and R. H. Bradbury, *Volterra–Hamilton Models in the Ecology and Evolution of Colonial Organisms*, World Scientific, 1996.

[AIM] P. L. Antonelli, R. S. Ingarden, and M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology*, FTPH 58, Kluwer Academic Publishers, 1993.

[CE] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, North Holland/American Elsevier, 1975.

[delR] L. del Riego, *Tenseurs de Weyl d'un spray de directions*, Theses, Université Scientifique et Medicale de Grenoble, 1973.

[Fou] P. Foulon, *Géométrie des équations différentielles du second ordre*, Ann. Inst. Henri Poincaré **45(1)** (1986), 1–28.

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

[Nu] S. Numata, *On the torsion tensors R_{jhk} and P_{hjk} of Finsler spaces with a metric $ds = (g_{ij}(dx)\,dx^i dx^j)^{1/2} + b_i(x)\,dx^i$*, Tensor, N.S. **32** (1978), 27–31.

Chapter 4

Finsler Surfaces and a Generalized Gauss–Bonnet Theorem

4.0 Prologue

So far, our treatment has emphasized the use of natural coordinates. At the beginning of Chapter 2, we stated our policy that in important computations, we only use objects which are invariant under positive rescaling in y. Consequently, our treatment using natural coordinates on $TM \setminus 0$ can be regarded as occurring on the (projective) sphere bundle SM, in the context of homogeneous coordinates.

Nevertheless, it is also useful to learn to compute with orthonormal frames on SM. We would like to illustrate how that can be carried out for Finsler surfaces. In this relatively simple setting, all the important geometrical invariants manifest themselves as two pseudo-scalars and one scalar rather than tensors.

Dealing exclusively with surfaces and not higher-dimensional manifolds has freed us from the sometimes cumbersome tensor calculus. Recall the orthonormal frame $\{\omega^a , \omega^{n+a}\}$ for $T^*(TM \setminus 0)$. We encountered that in §2.3. For $T^*(SM)$, we delete the last member $\omega^{n+n} := d(\log F)$. In the case where M is a surface, its sphere bundle SM is 3-dimensional, and the frame in question consists simply of three 1-forms ω^1, ω^2, ω^3. We explicitly write down their formulas in a later section.

The geometry of Finsler surfaces is completely controlled by two pseudo-scalars (whose sign changes if we reverse the orientation of a certain basis) and one scalar, all living on SM. Specifically:

- The Cartan (or Main) "scalar" I. This is the only *a priori* nonvanishing component of the Cartan tensor, when the latter is expressed in terms of a special orthonormal basis. It is a pseudo-scalar.
- The Landsberg "scalar" J. This is also a pseudo-scalar. We derive a Bianchi identity which says that J is simply the directional derivative of I along a distinguished horizontal direction on SM.
- The Gaussian curvature K. Like I and J, this function *a priori* lives on SM. But unlike those two, K is a true scalar. It has a familiar analogue that goes by the same name in elementary differential geometry ([doC1], [On]). The main difference between the two is that the K here does not live on M in general.

In order to simplify our prose, let us refer to all three as scalars. These appear prominently in the structural equations

$$\begin{aligned} d\omega^1 &= -I\,\omega^1 \wedge \omega^3 + \omega^2 \wedge \omega^3 \\ d\omega^2 &= -\,\omega^1 \wedge \omega^3 \\ d\omega^3 &= K\,\omega^1 \wedge \omega^2 - J\,\omega^1 \wedge \omega^3 . \end{aligned}$$

We carry out a systematic review of these structural equations, together with their associated Bianchi identities and interchange formulas. The vanishing of I, J, K, respectively, characterizes Riemannian surfaces, Landsberg surfaces, and flat surfaces (in the Finslerian sense).

4.1 Minkowski Planes and a Useful Basis

A **Minkowski plane** is the number space $\mathbb{R}^2$ equipped with a Minkowski norm F (§1.2). The strong convexity assumption, which says that

$$g_{ij} := \left(\frac{1}{2} F^2 \right)_{y^i y^j}$$

is positive-definite at all $0 \neq y \in \mathbb{R}^2$, is included in our definition of Minkowski norms. As a result, the **indicatrix**

$$\text{(4.1.1)} \qquad \mathrm{S} := \{ y \in \mathbb{R}^2 : F(y) = 1 \}$$

is a closed, strictly convex, smooth curve that surrounds but never passes through the origin. Here, y^1, y^2 are the canonical coordinates on $\mathbb{R}^2$. See Figure 4.1.

The entire plot
occurs in the tangent plane
$T_x M$

$\|$

$\left\{ y^1 \frac{\partial}{\partial x^1} + y^2 \frac{\partial}{\partial x^2} \; : y^1, y^2 \in \mathbb{R} \right\}$

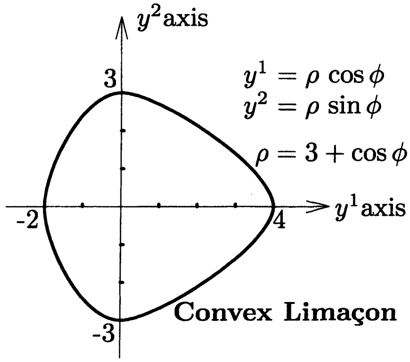

Figure 4.1

A typical indicatrix: the convex limaçon. It is strongly convex, hence strictly convex. See §4.1B and Exercises 4.1.4, 1.2.6, 1.2.7.

The Minkowskian norm F induces a Riemannian metric $\hat{g}$ on the punctured plane $\mathbb{R}^2 \setminus 0$ by

$$\hat{g} \; := \; g_{ij} \; dy^i \otimes dy^j \; . \tag{4.1.2}$$

- One can check that the Gaussian curvature of $\hat{g}$ vanishes identically.
- With respect to $\hat{g}$, the vector $\frac{y^i}{F(y)} \frac{\partial}{\partial y^i}$ is the outward pointing unit normal to the level curves of F, as was pointed out in §1.4.

4.1 A. Rund's Differential Equation and Its Consequence

Let h denote the **induced Riemannian metric** of $\hat{g}$ on the indicatrix S. Let $y(t)$ be a unit speed (with respect to h) parametrization of S. We have

$$y(t) \; = \; (\, y^1(t) \, , \, y^2(t) \,) \; , \tag{4.1.3}$$

where

$$g_{ij\,(y)} \; y^i \, y^j \; = \; F^2(y) \; = \; 1 \; , \tag{4.1.4}$$

$$g_{ij\,(y)} \; \frac{dy^i}{dt} \frac{dy^j}{dt} \; = \; 1 \; . \tag{4.1.5}$$

The **Cartan scalar** $I : \mathrm{S} \to \mathbb{R}$ is provisionally defined here by

$$I(y) \; := \; A_{ijk\,(y)} \; \frac{dy^i}{dt} \frac{dy^j}{dt} \frac{dy^k}{dt} \; . \tag{4.1.6}$$

The definition is extended to all of $\mathbb{R}^2 \setminus 0$ by requiring that I be constant along each ray that emanates from the origin of $\mathbb{R}^2$. Notice that F is Euclidean if and only if $I = 0$.

Proposition 4.1.1 (Rund) [R]. *Every unit speed parametrization $y(t)$ of the indicatrix (S, h) must satisfy the following ODE:*

$$\boxed{\ddot{y} + I\,\dot{y} + y = 0} \,. \tag{4.1.7a}$$

That is,

$$\ddot{y}^i + I\,\dot{y}^i + y^i = 0 \quad \text{for } i = 1, 2 \,. \tag{4.1.7b}$$

Here, the abbreviations $\dot{y} = \frac{dy}{dt}$ and $\ddot{y} = \frac{d^2y}{dt^2}$ are used in order to avoid clutter.

Remark: The quantity $\frac{\partial g_{ij}}{\partial y^k}$ is totally symmetric on all three indices. Recall from (1.4.5) that as a consequence of Euler's theorem, one has

$$y^j \,\frac{\partial g_{ij}}{\partial y^k} = 0 \,.$$

We use this basic fact without explicit mention.

Proof. Differentiating (4.1.4) once with respect to t yields

$$g_{ij\,(y)}\,\dot{y}^i\,y^j = 0 \,, \tag{4.1.8}$$

which says that the position vector y and the velocity vector $\dot{y}$ are $\hat{g}$-orthogonal. Note that in (4.1.4), differentiating g_{ij} produces the term

$$\left[\frac{1}{2}\,F^2\right]_{y^i y^j y^k} \dot{y}^k\,y^i\,y^j = \left[\frac{1}{2}\,F^2\right]_{y^k y^i y^j} y^j\,\dot{y}^k\,y^i \,.$$

This vanishes by Euler's theorem and the fact that $[\frac{1}{2}F^2]_{y^k y^i}$ is homogeneous of degree zero.

Next, differentiation of (4.1.8), followed by the unit speed condition [namely (4.1.5)], gives

$$g_{ij\,(y)}\,\ddot{y}^i\,y^j = -1 \,. \tag{4.1.9}$$

Likewise, differentiating (4.1.5) and using the definition (4.1.6) of I yields

$$g_{ij\,(y)}\,\ddot{y}^i\,\dot{y}^j = -I \,. \tag{4.1.10}$$

To make sure there is no misunderstanding: the right-hand side of (4.1.9) is minus one, whereas that of (4.1.10) is the negative of the Cartan scalar.

Let us use the above to resolve the acceleration $\ddot{y}$ into two $\hat{g}$-orthogonal components:

$$\ddot{y} = \alpha\,y + \beta\,\dot{y} \,.$$

* To determine α, we take the $\hat{g}$ inner product of this equation with y. With the help of (4.1.9), (4.1.4), and (4.1.8), we get $\alpha = -1$.
* Similarly, taking the $\hat{g}$ inner product with $\dot{y}$ and using (4.1.10), (4.1.5) yields $\beta = -I$.

Thus $\ddot{y} = (-1)\,y + (-I)\,\dot{y}$, which is the desired (4.1.7). □

The following corollary invokes the concept of the **Riemannian arc length** L of the indicatrix S. Recall that S is a submanifold of the punctured Riemannian manifold $(\mathbb{R}^2 \setminus 0\,,\,\hat{g}\,)$, and has inherited a Riemannian metric h from the simple restriction of $\hat{g}$. The said length of the simple closed curve S is measured with respect to this h. As mentioned in (1.4.8), an element of arc along S has length

$$ds \;=\; \frac{\sqrt{g}}{F}\,(\,y^1\,\dot{y}^2 \;-\; y^2\,\dot{y}^1\,)\;dt\,. \tag{4.1.11}$$

Here, $\sqrt{g} := \sqrt{\det(g_{ij})}$ and we have left the factor of F in the denominator for conceptual purpose only. Also, along S there is no distinction between the above ds and the form

$$d\theta \;:=\; \frac{\sqrt{g}}{F^2}\,(\,y^1\,dy^2 \;-\; y^2\,dy^1\,)\,.$$

Historically, the parameter θ is called the **Landsberg angle.**

Formula (4.1.11) is valid as long as the parametrization (4.1.3) traces S out in a direction deemed positive by ds. For a unit speed parametrization, (4.1.11) reduces simply to $ds = dt$ because, in that case, we are supposed to have $\frac{ds}{dt} = 1$. Anyway, the Riemannian arc length of the indicatrix is

$$L \;:=\; \int_{\mathrm{S}} ds\;, \tag{4.1.12}$$

and it is typically *not* equal to 2π. See §4.5.

* In [BS], a 1-parameter family of *absolutely* homogeneous Minkowski norms was considered. The Riemannian arc length of the corresponding indicatrices was found to decrease from 2π down to about 5.4414.
* In [BL1], a 1-parameter family of *positively* homogeneous Minkowski norms was considered. Those were not absolutely homogeneous. Interestingly, the Riemannian arc length of the corresponding indicatrices was found to increase from 2π to ∞.

See also the paper [M3] by Matsumoto.

Corollary 4.1.2. *Let F be a Minkowski norm on $\mathbb{R}^2$. (It is smooth and strongly convex on $\mathbb{R}^2 \setminus 0$.) Then the average value of the Cartan scalar, over the indicatrix S, is zero. In particular, for any unit speed parametrization of S, we have*

$$\boxed{\int_0^L I(t)\;dt \;=\; 0}\;, \tag{4.1.13}$$

where $I(t) := I[\,y(t)\,]$ and L is the length of (S, h).

Proof. As before, let $y(t) = (y^1(t), y^2(t))$, $0 \leqslant t \leqslant L$ be a unit speed parametrization of S. Since the indicatrix is a closed curve in $\mathbb{R}^2 \setminus 0$, the functions $y^i(t)$ are periodic with period L.

In components, (4.1.7) reads $\ddot{y}^i + I\,\dot{y}^i + y^i = 0$. Using this, one checks that the L-periodic quantity

$$\chi := y^1\, \dot{y}^2 - y^2\, \dot{y}^1$$

satisfies the ODE

$$\dot{\chi} + I\chi = 0 .$$

Here, $I(t)$ is continuous because F is (by hypothesis) smooth (hence at least C^3) away from the origin. Thus

$$\chi(t) = \chi(0)\, e^{\,[\int_0^t I(\tau)\, d\tau]} .$$

Since our parametrization has unit speed, (4.1.11) tells us that

$$\frac{\sqrt{g}}{F}\,(y^1\, \dot{y}^2 - y^2\, \dot{y}^1) = 1 .$$

In particular, $\chi(t)$ is nowhere zero. Therefore the condition $\chi(L) = \chi(0)$ implies (4.1.13). $\square$

4.1 B. A Criterion for Checking Strong Convexity

In the previous subsection, we have seen the efficacy of the basis $\{y, \dot{y}\}$. Here, we further explore its versatility. We use it to help us derive an elegant criterion for checking the strong convexity, or the lack of such, of any given candidate F.

So, let us be given a nonconstant F, defined on $\mathbb{R}^2$. The associated indicatrix S is the level curve on which F has constant value 1. We begin by parametrizing the indicatrix S as

$$y(t) = (y^1(t), y^2(t)) .$$

Here:

- The parametrization does *not* have to have unit speed.
- Also, it does *not* have to be counterclockwise.

Since $y(t)$ lies on the indicatrix, we have $F(y) = 1$. Equivalently,

$$(*) \qquad g_{ij\,(y)}\, y^i\, y^j = 1 .$$

Differentiating this with respect to t gives (4.1.8), which we reproduce here for convenience:

$$(**) \qquad g_{ij\,(y)}\, \dot{y}^i\, y^j = 0 .$$

As explained before, the term containing a derivative on g_{ij} is absent because of Euler's theorem. Yet another differentiation gives the analogue of (4.1.9) for parametrizations that are not necessarily of unit speed. The result reads:

(4.1.14) $$g_{ij\,(y)}\ \ddot{y}^i\ y^j \ = \ -\ g_{ij\,(y)}\ \dot{y}^i\ \dot{y}^j\ .$$

Again, the derivative on g_{ij} does not contribute on account of Euler's theorem.

Let us turn to the main issue, which is the strong convexity, or the lack of such, of F. Being strongly convex means that at any $y \neq 0$ (not necessarily on the indicatrix), we must have $g_{ij\,(y)}\, w^i w^j > 0$ whenever w is nonzero. Keep in mind that w is a tangent vector on the linear space $\mathbb{R}^2$, and that $g_{ij\,(y)}$ is invariant under positive rescaling in y. A moment's thought then convinces us that nothing is missed by checking strong convexity only at points y which lie on the indicatrix.

With this realization, we can make good use of the basis $\{y, \dot{y}\}$, just as we did in the proof of Proposition 4.1.1. In view of (**), this basis is $\hat{g}$-orthogonal. Let us express w as a linear combination of y and $\dot{y}$,

$$w \ = \ \alpha\, y \ + \ \beta\, \dot{y}\ .$$

Using this and (*), (**), (4.1.14), we get

(4.1.15) $$g_{ij\,(y)}\ w^i\, w^j \ = \ \alpha^2 \ - \ \beta^2\ g_{ij\,(y)}\ \ddot{y}^i\ y^j\ .$$

By applying Euler's theorem to the quantity $[\frac{1}{2}F^2]_{y^i}$, which is homogeneous of degree 1, we have

$$g_{ij\,(y)}\ y^j \ = \ \left[\frac{1}{2}\ F^2 \right]_{y^i} \ = \ F(y)\ F_{y^i} \ = \ F_{y^i}\ .$$

Thus, at y on the indicatrix,

$$g_{ij\,(y)}\ \ddot{y}^i\ y^j \ = \ F_{y^i}\ \ddot{y}^i \ = \ F_{y^1}\ \ddot{y}^1 \ + \ F_{y^2}\ \ddot{y}^2\ .$$

Now,

$$y^1\ F_{y^1} \ + \ y^2\ F_{y^2} \ = \ 1$$

because F is homogeneous of degree one, and $F(y) = 1$. Also, one gets

$$\dot{y}^1\ F_{y^1} \ + \ \dot{y}^2\ F_{y^2} \ = \ 0$$

by applying the chain rule to the statement $F(y^1, y^2) = 1$. Solving these two equations for F_{y^1}, F_{y^2} gives

$$F_{y^1} \ = \ \frac{-\ \dot{y}^2}{\dot{y}^1\ y^2 \ - \ y^1\ \dot{y}^2}\ , \qquad F_{y^2} \ = \ \frac{\dot{y}^1}{\dot{y}^1\ y^2 \ - \ y^1\ \dot{y}^2}\ .$$

These then lead to

(4.1.16) $$-\ g_{ij\,(y)}\ \ddot{y}^i\ y^j \ = \ \frac{\ddot{y}^1\ \dot{y}^2 \ - \ \dot{y}^1\ \ddot{y}^2}{\dot{y}^1\ y^2 \ - \ y^1\ \dot{y}^2}\ .$$

Note that according to (4.1.14), the right-hand side of (4.1.16) gives an explicit formula for the $g_{ij\,(y)}$ length-*squared* of the velocity $\dot{y}$ of our parametrization of S. We show below that the positive-definiteness of $g_{ij\,(y)}$ is equivalent to $\dot{y}$ having a positive length-squared. In other words, it is neither null nor time-like, in the language of Lorentzian geometry.

Substituting (4.1.16) into (4.1.15), we see that at any $y \in \mathrm{S}$,

$$\boxed{g_{ij\,(y)}\, w^i\, w^j \;=\; \alpha^2 \,+\, \beta^2\, \frac{\ddot{y}^1\, \dot{y}^2 \,-\, \dot{y}^1\, \ddot{y}^2}{\dot{y}^1\, y^2 \,-\, y^1\, \dot{y}^2}}\;, \tag{4.1.17}$$

where $w = \alpha y + \beta \dot{y}$. It is now immediate that the following two statements are equivalent **at any point y on the indicatrix of F**:

* $g_{ij\,(y)}\, w^i w^j > 0$ **for every nonzero w.**
* $\frac{\ddot{y}^1\, \dot{y}^2 - \dot{y}^1\, \ddot{y}^2}{\dot{y}^1\, y^2 - y^1\, \dot{y}^2} = g_{ij\,(y)}\, \dot{y}^i\, \dot{y}^j$ **is positive at y.**

In particular, given any parametrization $\big(\, y^1(t), y^2(t)\,\big)$ of the indicatrix:
(4.1.18)

$$\boxed{F \;\text{ is strongly convex } \;\Leftrightarrow\; \frac{\ddot{y}^1\, \dot{y}^2 \,-\, \dot{y}^1\, \ddot{y}^2}{\dot{y}^1\, y^2 \,-\, y^1\, \dot{y}^2} \;>\; 0 \;\text{ everywhere on S}}\;.$$

This is the criterion we seek. We emphasize that it is applicable to an *arbitrary* parametrization of the indicatrix S. See Exercise 4.1.4 for a concrete illustration.

Exercises

Exercise 4.1.1: In (4.1.2), we defined a Riemannian metric $\hat{g}$ on the punctured plane $\mathbb{R}^2 \setminus 0$. It induces the metric h on the indicatrix S.

(a) Show that the Gaussian curvature of $\hat{g}$ is identically zero.

(b) Suppose $\mathbb{R}^2 \setminus 0$ is identified with $(0, \infty) \times \mathrm{S}$ by

$$y \;\mapsto\; \left(F(y)\,,\, \frac{y}{F(y)} \right)\;.$$

Verify that $\hat{g}$ admits the block decomposition

$$\hat{g} \;=\; dr \otimes dr \;+\; r^2\, h\;.$$

Exercise 4.1.2: Suppose the indicatrix is not parametrized to have speed (relative to h) 1 but some nowhere zero function $v = v(t)$. Show that the analogue of (4.1.7) is

$$\ddot{y} \;+\; \left(Iv - \frac{\dot{v}}{v} \right)\, \dot{y} \;+\; v^2\, y \;=\; 0\;.$$

Exercise 4.1.3: Suppose the Minkowski norm is given by

$$F(y^1, y^2) \;:=\; \sqrt{(y^1)^2 \,+\, (y^2)^2} \;+\; B y^1\;,$$

where $0 \leqslant B < 1$ is a constant parameter. Introduce polar coordinates on $\mathbb{R}^2$, so that $y^1 = r\cos\phi$ and $y^2 = r\sin\phi$.

(a) Check that the polar equation of the indicatrix S is $r = \frac{1}{1+B\cos\phi}$.

(b) This indicatrix is an ellipse of eccentricity B. Identify its center, the two foci, as well as the semimajor and semiminor axes.

(c) Check that the differential $d\theta$ of the Landsberg angle is given by $d\theta = \frac{1}{\sqrt{1+B\cos\phi}}\,d\phi$.

(d) With the substitution $\phi = 2\mu$, prove that the Riemannian arc length of S is given by

$$L = \frac{4}{\sqrt{1+B}} \int_0^{\frac{\pi}{2}} \frac{d\mu}{\sqrt{1-k^2\sin^2\mu}} ,$$

where $k := \sqrt{\frac{2B}{1+B}}$. Note that the above involves a complete elliptic integral of the first kind.

(e) Get numerical answers of L for the following values of the parameter B: 0.0, 0.3, 0.9, 0.999, 0.999999. As $B \to 1^-$, what becomes of the value of L ?

Exercise 4.1.4: This exercise uses criterion (4.1.18) to verify the strong convexity of the Finsler function, whose indicatrix is the convex limaçon. In polar coordinates for $\mathbb{R}^2$, the convex limaçon is given by

$$\rho = 3 + \cos\phi , \qquad 0 \leqslant \phi \leqslant 2\pi .$$

The explicit formula for F has been recovered using Okubo's technique; see Exercise 1.2.8. The power of criterion (4.1.18), however, is that what F explicitly looks like is totally irrelevant. All we need is *some* parametrization of the indicatrix.

(a) Check that a parametrization at hand is

$$\begin{aligned}\big(y^1(t)\,,\,y^2(t)\big) &= \big(\,[3+\cos t]\,\cos t\;,\;[3+\cos t]\,\sin t\,\big) \\ &= \frac{1}{2}\,\big(\,6\,\cos t + \cos 2t + 1\;,\;6\,\sin t + \sin 2t\,\big)\;.\end{aligned}$$

(b) Show that

$$\frac{\ddot{y}^1\,\dot{y}^2 - \dot{y}^1\,\ddot{y}^2}{\dot{y}^1\,y^2 - y^1\,\dot{y}^2} = 2\left(\frac{11 + 9\,\cos t}{19 + 12\cos t + \cos 2t}\right) .$$

(c) Explain why the above ratio is positive everywhere on the convex limaçon.

4.2 The Equivalence Problem for Minkowski Planes

Consider two Minkowskian planes $(\mathbb{R}^2, F)$ and $(\mathbb{R}^2, \tilde{F})$, with their corresponding indicatrices S, $\tilde{\text{S}}$. Let $y(t)$ and $\tilde{y}(t)$ be unit speed parametrizations of the indicatrices S and $\tilde{\text{S}}$, respectively. As in (4.1.6), we have the Cartan scalars I and $\tilde{I}$. When restricted to the respective indicatrices, they give rise to functions $I(t)$ and $\tilde{I}(t)$ of the parameter t. The following result gives a criterion for the **global equivalence** between two Minkowski planes.

Proposition 4.2.1. *Suppose the unit speed parametrizations $y(t)$, $\tilde{y}(t)$ trace out the indicatrices S, $\tilde{\text{S}}$ in a positive manner with respect to the canonical orientation of $\mathbb{R}^2$. Then the following two statements are equivalent:*

- *The two Minkowski planes are equivalent; that is, there exists a linear orientation-preserving map $L : \mathbb{R}^2 \to \mathbb{R}^2$ such that*

(a) $$\tilde{F} \circ L = F .$$

- *The Cartan scalars restricted to the corresponding indicatrices satisfy*

(b) $$\tilde{I}(t + t_o) = I(t) \quad \text{for some} \quad t_o .$$

Proof.

That (a) ⇒ (b) :

Suppose the two Minkowski planes are equivalent in the sense (a). Then L maps S into $\tilde{\text{S}}$ and therefore $L[\,y(t)\,]$ parametrizes the indicatrix $\tilde{\text{S}}$. Next, at all $v := (y^1, y^2) \in \mathbb{R}^2$, differentiating $\tilde{F}^2[\,L(v)\,] = F^2(v)$ twice with respect to the natural coordinates y^i gives

$$\frac{\partial^2 \tilde{F}^2}{\partial \tilde{y}^a \partial \tilde{y}^b} \, \frac{\partial L^a}{\partial y^i} \, \frac{\partial L^b}{\partial y^j} = \frac{\partial^2 F^2}{\partial y^i \partial y^j} ,$$

where we have dropped the term proportional to $\frac{\partial^2 L^a}{\partial y^i \partial y^j}$ because the components of L are (by hypothesis) linear functions of the y^i. The above equation is equivalent to

(4.2.1) $$\tilde{g}_{ab\,[L(v)]} \, \frac{\partial L^a}{\partial y^i} \, \frac{\partial L^b}{\partial y^j} = g_{ij\,(v)} .$$

We deduce from what has been said that $L[\,y(t)\,]$ is, just like $\tilde{y}(t)$, a unit speed parametrization of the indicatrix $\tilde{\text{S}}$. Moreover, both of them have the same orientation. Hence

(4.2.2) $$L[\,y(t)\,] = \tilde{y}(t + t_o) \quad \text{for some} \quad t_o , \quad \text{and}$$

(4.2.3) $$L_*[\,\dot{y}(t)\,] = \dot{\tilde{y}}(t + t_o) .$$

Differentiating (4.2.1) and using again the linearity of L, we obtain

$$\tilde{A}_{abc\,[L(v)]}\ \frac{\partial L^a}{\partial y^i}\ \frac{\partial L^b}{\partial y^j}\ \frac{\partial L^c}{\partial y^k} \;=\; A_{ijk\,(v)}\ .$$

Contracting both sides with $\dot{y}\otimes\dot{y}\otimes\dot{y}$ and using (4.2.2), (4.2.3) on the left, we get

$$\tilde{I}(t+t_o) \;=\; I(t)\ ,$$

which is (b).

That (b) $\Rightarrow$ (a) :

Suppose (b) holds. Recall from (4.1.8) that the nonzero vectors y and $\dot{y}$ are $\hat{g}$ orthogonal at all times. In particular, they form a basis for $\mathbb{R}^2$ at time $t=0$. We construct a linear map $L:\mathbb{R}^2\to\mathbb{R}^2$ by stipulating that

$$L[\,y(0)\,] \;:=\; \tilde{y}(t_o)\ , \qquad L[\,\dot{y}(0)\,] \;:=\; \dot{\tilde{y}}(t_o)\ , \tag{4.2.4}$$

where t_o is the "phase" promised by statement (b). Since the curves $y(t)$, $\tilde{y}(t)$ are both traced out in a positive manner, the canonical orientations of the domain $\mathbb{R}^2$ and the range $\mathbb{R}^2$ coincide, respectively, with those given by the ordered bases $\{\,y\,,\,\dot{y}\,\}$ and $\{\,\tilde{y}\,,\,\dot{\tilde{y}}\,\}$. Thus L is orientation preserving by construction.

By Proposition 4.1.1, $y(t)$ satisfies the ODE $\ddot{y}+I\,\dot{y}+y=0$. Apply L to both sides of this equation and use the following:

* $L(0)=0$
* $L[\,\dot{y}\,]=\frac{d}{dt}\,(L\circ y)$
* $L[\,\ddot{y}\,]=\frac{d^2}{dt^2}\,(L\circ y)$
* $\tilde{I}(t+t_o)=I(t)$ for some t_o .

Here, the first three are due to the linearity of L, while the last one is our hypothesis (b). Consequently, $L\circ y$ satisfies

$$\frac{d^2}{dt^2}\,(L\circ y)(t) \;+\; \tilde{I}(t+t_o)\,\frac{d}{dt}\,(L\circ y)(t) \;+\; (L\circ y)(t) \;=\; 0\ ,$$

with initial data (4.2.4). But Proposition 4.1.1 tells us that this initial value problem already has a solution, namely, the map $t\mapsto\tilde{y}(t+t_o)$. Thus by uniqueness we must have

$$L[\,y(t)\,] \;=\; \tilde{y}(t+t_o)\ .$$

We have just shown that our linear L maps the indicatrix S into the indicatrix $\tilde{\mathrm{S}}$. In particular, $\tilde{F}[\,L(y)\,]=F(y)$ for all $y(t)\in S$. Multiplying this by positive λ, and using the fact that $\tilde{F}$, F are both positively homogeneous of degree one, we conclude that

$$\tilde{F}[\,L(v)\,] \;=\; F(v)$$

for all nonzero $v \in \mathbb{R}^2$. Of course such statement also holds at $v = 0$. We have therefore obtained (a). □

4.3 The Berwald Frame and Our Geometrical Setup on SM

Having studied Minkowski planes, we turn our attention to orientable Finsler surfaces. A Finsler structure on a surface M is a function $F : TM \to [0, \infty)$ that is C^∞ on $TM \setminus 0$ and whose restriction to each tangent plane T_xM is a Minkowskian norm.

Let SM be the quotient of $TM \setminus 0$ under the following equivalence relation: $(x, y) \sim (x, \tilde{y})$ if and only if y, $\tilde{y}$ are positive multiples of each other. In other words, SM is the bundle of all directions or rays, and is called the (projective) **sphere bundle**. A moment's thought shows that it is diffeomorphic to the **indicatrix bundle** $\{(x, y) \in TM : F(x, y) = 1\}$, which is a subbundle of $TM \setminus 0$. They are in fact isometric to each other; see [BS]. So, no real harm is done even if one inadvertently confuses the two.

Local coordinates x^1, x^2 on M induce global coordinates y^1, y^2 on each fiber T_xM, through the expansion $y = y^i \frac{\partial}{\partial x^i}$. Thus $(x^i; y^i)$ is a coordinate system on SM, with the y^i regarded as homogeneous coordinates (in the projective space sense).

- The **base manifold** here is the sphere bundle SM.
- Using the canonical projection $p : SM \to M$, we pull the tangent bundle TM back so that it sits over SM. This **pulled-back bundle**, denoted p^*TM, is a vector bundle (of fiber dimension 2) over the 3-manifold SM.

The vector bundle p^*TM has a global section $\ell := \frac{y^i}{F(y)} \frac{\partial}{\partial x^i}$. It also has a natural Riemannian metric

$$g \ := \ g_{ij} \, dx^i \otimes dx^j \ .$$

By Euler's theorem, ℓ has norm 1. One can complete ℓ into a positively oriented g-orthonormal frame $\{e_1, e_2\}$ for p^*TM, with $e_2 := \ell$. Explicitly,

$$\text{(4.3.1)} \qquad \boxed{e_1 \ = \ \frac{F_{y^2}}{\sqrt{g}} \frac{\partial}{\partial x^1} - \frac{F_{y^1}}{\sqrt{g}} \frac{\partial}{\partial x^2} \ =: \ m^1 \frac{\partial}{\partial x^1} + m^2 \frac{\partial}{\partial x^2}} \ ,$$

$$\text{(4.3.2)} \qquad \boxed{e_2 \ = \ \frac{y^1}{F} \frac{\partial}{\partial x^1} + \frac{y^2}{F} \frac{\partial}{\partial x^2} \ = \ \ell^1 \frac{\partial}{\partial x^1} + \ell^2 \frac{\partial}{\partial x^2}} \ .$$

Here,

$$\sqrt{g} \ := \ \sqrt{\det(g_{ij})}$$

and, as usual, F_{y^i} abbreviates the partial derivative $\frac{\partial F}{\partial y^i}$. It is a peculiarity of the 2-dimensional case that the **Berwald frame** $\{e_1, e_2\}$ is a *globally* defined g-orthonormal frame field for p^*TM. Our m^i in (4.3.1) are opposite in sign to those in [AIM] and [M2]. However, they list our e_2 as their first vector in the basis. Consequently, our Berwald frame and theirs have the same *orientation*.

The natural dual of ℓ is the Hilbert form $\omega := F_{y^i}\, dx^i$, which is a global section of p^*T^*M. The coframe corresponding to $\{e_1, e_2\}$ is $\{\omega^1, \omega^2\}$, where

$$\boxed{\omega^1 \ = \ \frac{\sqrt{g}}{F}\,(y^2\,dx^1 \ - \ y^1\,dx^2) \ =: \ m_1\,dx^1 \ + \ m_2\,dx^2} \ , \tag{4.3.3}$$

$$\boxed{\omega^2 \ = \ F_{y^1}\ dx^1 \ + \ F_{y^2}\,dx^2 \ = \ \ell_1\,dx^1 \ + \ \ell_2\,dx^2} \ . \tag{4.3.4}$$

Recall that the formal Christoffel symbols of the fundamental tensor are

$$\gamma^l{}_{jk} \ := \ \frac{1}{2}\ g^{li}\left(\frac{\partial g_{ij}}{\partial x^k} \ - \ \frac{\partial g_{jk}}{\partial x^i} \ + \ \frac{\partial g_{ki}}{\partial x^j}\right) .$$

Let us also recall the nonlinear connection

$$\frac{N^i{}_j}{F} \ := \ \gamma^i{}_{jk}\,\ell^k \ - \ A^i{}_{jk}\,\gamma^k{}_{rs}\,\ell^r\,\ell^s \ ,$$

and

$$\frac{\delta}{\delta x^i} \ := \ \frac{\partial}{\partial x^i} \ - \ N^s{}_i\,\frac{\partial}{\partial y^s} \ .$$

$\{\frac{\delta}{\delta x^i}, F\frac{\partial}{\partial y^i}\}$ is a local field of bases for $T(TM \setminus 0)$; it is naturally dual to $\{dx^i, \frac{\delta y^i}{F}\}$, where

$$\delta y^i \ := \ dy^i \ + \ N^i{}_s\,dx^s \ .$$

The sphere bundle SM is a 3-dimensional Riemannian manifold equipped with the **Sasaki** (type) **metric**

$$\omega^1 \otimes \omega^1 \ + \ \omega^2 \otimes \omega^2 \ + \ \omega^3 \otimes \omega^3 \ , \tag{4.3.5}$$

where

$$\boxed{\omega^3 \ := \ \frac{\sqrt{g}}{F}\left(y^2\,\frac{\delta y^1}{F} \ - \ y^1\,\frac{\delta y^2}{F}\right) \ = \ m_1\,\frac{\delta y^1}{F} \ + \ m_2\,\frac{\delta y^2}{F}} \ . \tag{4.3.6}$$

Here, we are exercising a slight abuse of notation by regarding ω^1, ω^2 as sections of p^*T^*M as well as 1-forms on SM.

The collection $\{\omega^1, \omega^2, \omega^3\}$ is a globally defined orthonormal basis for $T^*(SM)$. Its natural dual is $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$, where

$$\boxed{\hat{e}_1 \ = \ \frac{F_{y^2}}{\sqrt{g}}\,\frac{\delta}{\delta x^1} \ - \ \frac{F_{y^1}}{\sqrt{g}}\,\frac{\delta}{\delta x^2} \ = \ m^1\,\frac{\delta}{\delta x^1} \ + \ m^2\,\frac{\delta}{\delta x^2}} \ , \tag{4.3.7}$$

(4.3.8)
$$\boxed{\hat{e}_2 = \frac{y^1}{F}\frac{\delta}{\delta x^1} + \frac{y^2}{F}\frac{\delta}{\delta x^2} = \ell^1 \frac{\delta}{\delta x^1} + \ell^2 \frac{\delta}{\delta x^2}} ,$$

(4.3.9)
$$\boxed{\hat{e}_3 = \frac{F_{y^2}}{\sqrt{g}} F\frac{\partial}{\partial y^1} - \frac{F_{y^1}}{\sqrt{g}} F\frac{\partial}{\partial y^2} = m^1 F\frac{\partial}{\partial y^1} + m^2 F\frac{\partial}{\partial y^2}} .$$

These three vector fields on SM form a global orthonormal frame for $T(SM)$. The first two are **horizontal** while the third one is **vertical**.

- The objects ω^1, ω^2, ω^3 and $\hat{e}_1$, $\hat{e}_2$, $\hat{e}_3$ are defined in terms of objects that live on the slit tangent bundle $TM \setminus 0$. But the definitions are invariant under positive rescaling in y. Therefore they give bonafide objects on the sphere bundle SM.
- Conversely, any object on SM can be viewed as objects on $TM \setminus 0$. They are then necessarily invariant under the rescaling $y \mapsto \lambda y$, with $\lambda > 0$.

Exercise

Exercise 4.3.1: Show that $\{\omega^1, \omega^2, \omega^3\}$ and $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ consist of well defined objects on SM, and that they are indeed naturally dual as claimed.

Exercise 4.3.2: Use Exercise 2.3.5 to review why $(dF)(\hat{e}_1) = 0 = (dF)(\hat{e}_2)$.

(a) Check that $(dF)(\hat{e}_3) = 0$.

(b) Explain why $\hat{e}_3$ must be tangent to each indicatrix.

(c) In (4.1.11), we used the arc length form $\mu := (\sqrt{g}/F)(y^1\, dy^2 - y^2\, dy^1)$ to obtain ds and its companion $d\theta$, where θ is the Landsberg angle. Show that $\mu(\hat{e}_3) = -1$. This means that $\hat{e}_3$ points *opposite* to the direction which is deemed 'positive' by the arc length form.

Exercise 4.3.3: *Directly* verify the following formulas.

(a) $m_i = g_{ij}\, m^j$ and $\ell_i = g_{ij}\, \ell^j$.

(b) $m_i\, m^i = 1 = \ell_i\, \ell^i$.

(c) $m_i\, \ell^i = 0$.

(d) $\ell^i\, \ell_k + m^i\, m_k = \delta^i{}_k$.

Exercise 4.3.4:

(a) Let $h_{ij} := g_{ij} - \ell_i\, \ell_j$ be the components of the **angular metric** h. Explain why $g = \omega^1 \otimes \omega^1 + \omega^2 \otimes \omega^2$, $h = \omega^1 \otimes \omega^1$.

(b) How can one deduce that $F\, F_{y^i y^j} = m_i\, m_j$?

(c) Verify that

$$\boxed{\begin{aligned} g_{ij} &= \ell_i\,\ell_j \;+\; m_i\,m_j\;, \\ g^{ij} &= \ell^i\,\ell^j \;+\; m^i\,m^j\;, \\ \det(g_{ij}) &= (\ell_1\,m_2 \;-\; \ell_2\,m_1)^2\;. \end{aligned}}$$

4.4 The Chern Connection and the Invariants I, J, K

With respect to the special g-orthonormal basis $\{e_a\}$ for p^*TM, the Cartan tensor takes the form

$$A \;=\; A_{abc}\,\omega^a \otimes \omega^b \otimes \omega^c\;.$$

We know that when A is contracted in any slot with the distinguished global section ℓ, the answer is zero. Here, this means that the component A_{abc} vanishes whenever any of its three indices equals n, which is 2 in the case of surfaces. Thus the only *a priori* nonvanishing component of A is A_{111}. Let us call

$$\boxed{I \;:=\; A_{111} \;=\; A(e_1\,,\,e_1\,,\,e_1)} \tag{4.4.1}$$

the **Cartan scalar**. Here, e_1 has the formula (4.3.1).

- A provisional definition of I was given in (4.1.6). Although the two definitions are readily shown to be equivalent, the definition here is the preferred one.
- The vanishing of I is equivalent to our Finsler structure being Riemannian; that is, $F(x,y) = \sqrt{g_{ij}(x)\,y^i y^j}$ in some local coordinate system.
- It was pointed out to us, in particular by Antonelli and Lackey, that I is not a true scalar. The reason is that if the orientation of the Berwald frame $\{e_1, e_2\}$ were reversed, our e_1 would have to be replaced by $-e_1$. And (4.4.1) tells us that I would then undergo a sign change. Nevertheless, in order to simplify our prose, let us continue to call I a scalar.

The Chern connection forms for the special g-orthonormal basis were considered in Exercise 2.4.5. In particular, the "almost" g-compatibility criterion (one of the structural equations) was re-expressed as

$$\omega_{ab} \;+\; \omega_{ba} \;=\; -2\,A_{abc}\,\omega^{n+c}\;.$$

Now, our base manifold is the 3-dimensional sphere bundle SM rather than the 4-dimensional $TM \smallsetminus 0$. So conceptually we need to delete the form $\omega^4 := d(\log F)$. Pragmatically, that term cannot contribute anyway because its coefficient $-2A_{ab2}$ vanishes. No matter which viewpoint we take, the above statement reduces to

$$\omega_{ab} \;+\; \omega_{ba} \;=\; -\,2\,A_{ab1}\,\omega^3\ .$$

We see immediately that:

* $\omega_{11} \;=\; -\,I\,\omega^3\ .$
* $\omega_{12} \;=\; -\,\omega_{21}\ .$
* $\omega_{22} \;=\; 0\ .$

Also, Exercise 2.4.7, which secretly invokes the Chern connection's torsion freeness, says that

* $\omega_{21} \;=\; {\omega_2}^1 \;=\; \omega^3\ ,$

where ω^3 is given by (4.3.6). Thus the **Chern connection matrix** has been completely determined, and it has the rather simple structure:

$$\text{(4.4.2a)}\qquad \boxed{\begin{pmatrix}\omega_{11} & \omega_{12}\\ \omega_{21} & \omega_{22}\end{pmatrix} \;=\; \begin{pmatrix}{\omega_1}^1 & {\omega_1}^2\\ {\omega_2}^1 & {\omega_2}^2\end{pmatrix} \;=\; \begin{pmatrix}-I\,\omega^3 & -\omega^3\\ \omega^3 & 0\end{pmatrix}}\ .$$

Explicitly,

$$\text{(4.4.2b)}\qquad \boxed{{\omega_1}^2 \;=\; \frac{\sqrt{g}}{F}\left(y^1\,\frac{\delta y^2}{F} \;-\; y^2\,\frac{\delta y^1}{F}\right)}\ ,$$

where $\delta y^i = dy^i + N^i{}_j\,dx^j$.

More information can be extracted from the torsion-freeness criterion among the structural equations. According to Exercise 2.4.5, it says that

$$d\omega^a \;=\; \omega^b \wedge {\omega_b}^a\ .$$

This and (4.4.2) then implies that

$$\text{(4.4.3)}\qquad \boxed{d\omega^1 \;=\; -I\,\omega^1\wedge\omega^3 \;+\; \omega^2\wedge\omega^3}\ ,$$

$$\text{(4.4.4)}\qquad \boxed{d\omega^2 \;=\; -\,\omega^1\wedge\omega^3}\ .$$

Since $\{\omega^1,\omega^2,\omega^3\}$ is an orthonormal basis for $T^*(SM)$, let us complete the above by taking the exterior differential of ω^3. We get, *a priori*,

$$d\omega^3 \;=\; K\,\omega^1\wedge\omega^2 \;-\; J\,\omega^1\wedge\omega^3 \;+\; L\,\omega^2\wedge\omega^3\ .$$

Now, applying d to (4.4.4) gives $0 = L\,\omega^1\wedge\omega^2\wedge\omega^3$. Thus L vanishes and

$$\text{(4.4.5)}\qquad \boxed{d\omega^3 \;=\; K\,\omega^1\wedge\omega^2 \;-\; J\,\omega^1\wedge\omega^3}\ .$$

The choice of a minus sign in front of J is strictly cultural. It is done here in order to agree with a convention in [Ch2].

The scalar K is called the **Gaussian curvature** of the Finsler surface. If our F happens to be Riemannian, K will reduce to the usual Gaussian curvature (for example, that defined in [doC1]). We say a bit more about the **Landsberg scalar** J after we set up the following notation.

Given any function f on SM, we can expand its differential in terms of its directional derivatives along $\hat{e}_1$, $\hat{e}_2$, $\hat{e}_3$, getting:

$$df \;=\; f_1\,\omega^1 \;+\; f_2\,\omega^2 \;+\; f_3\,\omega^3\;. \tag{4.4.6}$$

Let us now give the **Bianchi identities** that exhibit the relationships among the three scalars I, J, K. We remind ourselves that these are all functions on SM, not M.

- Differentiating (4.4.3) yields

$$\begin{aligned} 0 \;&=\; -(dI)\wedge\omega^1\wedge\omega^3 \;-\; J\,\omega^1\wedge\omega^2\wedge\omega^3 \\ &=\; (\,I_2 \;-\; J\,)\,\omega^1\wedge\omega^2\wedge\omega^3\;. \end{aligned}$$

  Thus

$$\boxed{J \;=\; I_2 \;=\; \frac{1}{F}\left(y^1\,\frac{\delta I}{\delta x^1} \;+\; y^2\,\frac{\delta I}{\delta x^2}\right)}\;. \tag{4.4.7}$$

  As observed at the beginning of this section, I changes sign under a change of orientation of p^*TM. In view of (4.4.7), the same applies to J. Thus J is not a true scalar either, although we continue to refer to it as such.

- Differentiating (4.4.5) gives

$$0 \;=\; (\,K_3 \;+\; K\,I \;+\; J_2\,)\,\omega^1\wedge\omega^2\wedge\omega^3\;.$$

  So

$$\boxed{K_3 \;+\; K\,I \;+\; J_2 \;=\; 0}\;. \tag{4.4.8}$$

Exercises

Exercise 4.4.1: Show that the two definitions of the Cartan scalar, given respectively in (4.1.6) and (4.4.1), are indeed equivalent.

Exercise 4.4.2: Supply the details in the derivation of (4.4.5), (4.4.7), and (4.4.8).

Exercise 4.4.3: Suppose $I_3 = 0$ at *all* points of a particular indicatrix

$$\mathrm{S}_x \;:=\; \{\,y\in T_xM:\; F(x,y)=1\,\}\;.$$

(a) Explain why I must then be constant on that S_x.

(b) Use Corollary 4.1.2 to show that I must in fact vanish identically on that S_x.

Exercise 4.4.4: Show that exterior differentiation of (4.4.6) yields

$$\begin{aligned} 0 \;=\; & (-f_{12} + f_{21} + K\, f_3)\, \omega^1 \wedge \omega^2 \\ & + (-f_{13} - I\, f_1 - f_2 + f_{31} - J\, f_3)\, \omega^1 \wedge \omega^3 \\ & + (f_1 - f_{23} + f_{32})\, \omega^2 \wedge \omega^3 \,. \end{aligned}$$

This then gives rise to the following **Ricci identities**:

$$\boxed{\begin{aligned} f_{21} - f_{12} &= -K\, f_3 \,, \\ f_{32} - f_{23} &= -f_1 \,, \\ f_{31} - f_{13} &= I\, f_1 + f_2 + J\, f_3 \,. \end{aligned}}$$

In these identities, all indicated derivatives are directional derivatives. For example, f_{12} means first differentiate f along $\hat{e}_1$ to get a new function f_1, whose directional derivative along $\hat{e}_2$ is f_{12}. See [BCS3] for applications.

Exercise 4.4.5: The Cartan formula was given in (3.7.1). Using it and (4.4.5), show that

$$\boxed{K \;=\; -\,\omega^3\,([\,\hat{e}_1\,,\,\hat{e}_2\,])}\,.$$

Exercise 4.4.6: In our orthonormal coframe $\{\omega^1, \omega^2, \omega^3\}$, the curvature 2-forms are

$$\Omega_b^{\ a} \;=\; \frac{1}{2}\, R_b^{\ a}{}_{cd}\, \omega^c \wedge \omega^d \;+\; P_b^{\ a}{}_{cd}\, \omega^c \wedge \omega^{n+d} \,.$$

According to Exercise 3.3.1, $R_b^{\ a}{}_{cd}$ is related to the curvature $R_j^{\ i}{}_{kl}$ in the natural basis via

$$R_b^{\ a}{}_{cd} \;=\; u_b^{\ j}\, v^a_{\ i}\, u_c^{\ k}\, u_d^{\ l}\, R_j^{\ i}{}_{kl} \,.$$

The same relation holds between $P_b^{\ a}{}_{cd}$ and $P_j^{\ i}{}_{kl}$. Now restrict to the case of Finsler surfaces.

(a) With the help of Exercise 3.3.2, show that

$$\Omega_b^{\ a} \;=\; R_b^{\ a}{}_{12}\, \omega^1 \wedge \omega^2 \;+\; P_b^{\ a}{}_{c1}\, \omega^c \wedge \omega^3 \,.$$

What happened to the factor of $\frac{1}{2}$?

(b) Using the symmetry (3.2.3) of P and the Bianchi identity (3.4.9), reduce the above further to

$$\Omega_b^{\ a} \;=\; R_b^{\ a}{}_{12}\, \omega^1 \wedge \omega^2 \;+\; P_b^{\ a}{}_{11}\, \omega^1 \wedge \omega^3 \;-\; \dot{A}^a{}_{b1}\, \omega^2 \wedge \omega^3 \,.$$

(c) Check that in the surface case,

$$\Omega_b^{\ a} \;=\; d\omega_b^{\ a} \,.$$

In other words, the quadratic term $\omega_b^{\ c} \wedge \omega_c^{\ a}$ drops out.

Therefore:

$$d\omega_b{}^a \;=\; R_b{}^a{}_{12}\,\omega^1\wedge\omega^2 \;+\; P_b{}^a{}_{11}\,\omega^1\wedge\omega^3 \;-\; \dot{A}^a{}_{b1}\,\omega^2\wedge\omega^3 \;.$$

Exercise 4.4.7: All statements in this exercise are made with respect to the orthonormal frame $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ and its dual $\{\omega^1, \omega^2, \omega^3\}$.

(a) Explain why, up to a sign, there are only two *a priori* nonvanishing components in R_{bacd}. They are R_{1112} and R_{1212}.

(b) Explain why, up to a sign, there are only two *a priori* nonvanishing components in P_{bac1}. They are P_{2111} (which is $-\dot{A}_{111}$) and P_{1111}. Hint: you will need to use (3.4.8) and Exercise 3.3.2.

(c) Is $\dot{A}_{111}$ the only *a priori* nonzero component of $\dot{A}$?

(d) Specialize the conclusion of Exercise 4.4.6 to $d\omega_2{}^1$. By comparing with (4.4.5), show that

$$\boxed{\begin{aligned} K &= R_2{}^1{}_{12} = -\,R_{1212}\,, \\ J &= -\,P_2{}^1{}_{11} = \dot{A}^1{}_{11} = \dot{A}_{111}\,. \end{aligned}}$$

(e) Specialize Exercise 4.4.6 to $d\omega_1{}^1$ and apply exterior differentiation. Prove that

$$\boxed{\begin{aligned} R_{1112} &= -\,I\,K\,, \\ P_{1111} &= I\,J \,-\, I_1\,. \end{aligned}}$$

Hence we have expressed all the components of R and P in terms of I, J, K and a directional derivative (namely, that along $\hat{e}_1$) of I. Also, let us not forget that, according to the Bianchi identity (4.4.7), J itself is equal to the directional derivative I_2.

Exercise 4.4.8: Part (d) of Exercise 4.4.7 demonstrates that, with respect to the orthonormal frame $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$, one has $K = R_2{}^1{}_{12}$. It can be checked that in this orthonormal frame, $R_2{}^a{}_{b2}$ is actually $R^a{}_b$, the predecessor of the flag curvature. In particular, $K = R^1{}_1$. Also, the same trend of thought gives $R^2{}_2 = 0$ (and $R^1{}_2 = 0 = R^2{}_1$). **For the rest of this exercise, we turn our attention to the components $R^i{}_k$ in natural coordinates.**

(a) The Ricci scalar *Ric* is defined as the trace $R^i{}_i$. With our automatic summation convention, there is an implied sum on the natural coordinate index i. We first encountered *Ric* in Exercise 3.9.5, in the equivalent form $g^{ik}\,R_{ik}$. Check that

$$\boxed{K \;=\; Ric \;=\; R^1{}_1 \;+\; R^2{}_2}\;.$$

(b) Show that $R^i{}_k\,y^k = 0$ and $R^i{}_k\,y_i = 0$. Here, $y_i := g_{ij}\,y^j$.

(c) Where do the following five equations come from?

$$\begin{aligned} {R^1}_1 \;+\; {R^2}_2 \;&=\; K && (1)\\ {R^1}_1\, y^1 \;+\; {R^1}_2\, y^2 \;&=\; 0 && (2)\\ {R^2}_1\, y^1 \;+\; {R^2}_2\, y^2 \;&=\; 0 && (3)\\ {R^1}_1\, y_1 \;+\; {R^2}_1\, y_2 \;&=\; 0 && (4)\\ {R^1}_2\, y_1 \;+\; {R^2}_2\, y_2 \;&=\; 0\,. && (5) \end{aligned}$$

(d) Use (1) to express ${R^1}_1$ in terms of ${R^2}_2$, substitute that into (2), and pair the resulting equation with (5). This gives two equations for the two "unknowns" ${R^1}_2$ and ${R^2}_2$. Likewise, we can use (1) to express ${R^2}_2$ in terms of ${R^1}_1$, substitute that into (3), and pair the resulting equation with (4). That gives two equations for the two "unknowns" ${R^1}_1$ and ${R^2}_1$. Derive the following formulas for the components ${R^i}_k$ **in natural coordinates**:

$$\boxed{\begin{pmatrix} {R^1}_1 & {R^1}_2 \\ {R^2}_1 & {R^2}_2 \end{pmatrix} \;=\; \frac{K}{F^2} \begin{pmatrix} y^2\, y_2 & -\,y^1\, y_2 \\ -\,y^2\, y_1 & y^1\, y_1 \end{pmatrix}}\,.$$

(e) Using $\mathcal{L} := \frac{1}{2}F^2$, re-express your conclusions in part (d) as

$$\boxed{\begin{pmatrix} {R^1}_1 & {R^1}_2 \\ {R^2}_1 & {R^2}_2 \end{pmatrix} \;=\; \frac{K}{F^2} \begin{pmatrix} y^2\, \mathcal{L}_{y^2} & -\,y^1\, \mathcal{L}_{y^2} \\ -\,y^2\, \mathcal{L}_{y^1} & y^1\, \mathcal{L}_{y^1} \end{pmatrix}}\,.$$

Use Euler's theorem (Theorem 1.2.1) to verify that $y^1 \mathcal{L}_{y^1} + y^2 \mathcal{L}_{y^2} = F^2$, so that the diagonal entries are indeed consistent with part (a).

Part (a), along with (d) or (e), offers five different formulas for machine computations of the Gaussian curvature K. For instance,

$$K \;=\; \frac{F^2}{y^1\, \mathcal{L}_{y^1}}\; {R^2}_2 \qquad \text{wherever } \; y^1 \neq 0\,.$$

The one in (a), namely, $K = {R^1}_1 + {R^2}_2$, may *appear* to be more efficient because it involves less symbolic division. Strangely, our actual experience indicates otherwise. In any case, these formulas are useful because each specific component of ${R^i}_k$ is fairly accessible through the following sequence of objects:

- $g_{ij} := \mathcal{L}_{y^i y^j}$.
- $\gamma_{ijk} := \frac{1}{2}\left(g_{ij,x^k} - g_{jk,x^i} + g_{ki,x^j} \right)$.
- ${\gamma^i}_{jk} := g^{is}\, \gamma_{sjk}$.
- $G^i := {\gamma^i}_{jk}\, y^j y^k$.
- ${N^i}_j = \frac{1}{2}\,(G^i)_{y^j}$, derived in Exercise 2.3.3.
- ${R^i}_k = \ell^j \left(\frac{\delta}{\delta x^k} \frac{{N^i}_j}{F} - \frac{\delta}{\delta x^j} \frac{{N^i}_k}{F} \right)$, derived in Exercise 3.3.4.

Here, $\frac{\delta}{\delta x^k} := \frac{\partial}{\partial x^k} - N^i{}_k \frac{\partial}{\partial y^i}$. See §12.5 for more on this matter.

Exercise 4.4.9: (Suggested by Brad Lackey)

(a) Show that $(d \log \sqrt{g})(\hat{e}_3) = \frac{1}{2} g^{ij} (dg_{ij})(\hat{e}_3) = I$.

(b) Use this to give a new proof of the fact that I has zero average on each indicatrix (Corollary 4.1.2).

4.5 The Riemannian Arc Length of the Indicatrix

Let $(\mathbb{R}^2, F)$ be a Minkowski plane. We saw in §4.1 that F endows the punctured plane $\mathbb{R}^2 \setminus 0$ with a natural Riemannian metric $g_{ij}\, dy^i \otimes dy^j$. We encountered the **indicatrix**

$$\mathrm{S} := \{ y \in \mathbb{R}^2 : F(y) = 1 \}$$

in §4.1. It is a submanifold of the punctured plane. It was pointed out in §4.1 that the **Riemannian arc length** of the indicatrix S is

$$L = \int_{\mathrm{S}} \frac{\sqrt{g}}{F} (y^1\, dy^2 - y^2\, dy^1), \tag{4.5.1}$$

and the factor of F is left there for conceptual purpose.

As we can see, the domain of integration depends on F. Out of curiosity, one might wonder whether the integration can be carried out on the **standard unit circle**

$$\boxed{\mathbb{S}^1 := \{ y \in \mathbb{R}^2 : (y^1)^2 + (y^2)^2 = 1 \}}$$

instead, say at the expense of having to use a more complicated integrand. This can indeed be done and, happily, the new integrand is almost as simple as the original. Here are the steps involved:

- Since $F = 1$ on S, we can divide the integrand in (4.5.1) by an additional factor of F without affecting the outcome. Thus

$$\boxed{L = \int_{\mathrm{S}} \frac{\sqrt{g}}{F^2} (y^1\, dy^2 - y^2\, dy^1)}\,. \tag{4.5.2}$$

- Note that the revised integrand

$$\frac{\sqrt{g}}{F^2} (y^1\, dy^2 - y^2\, dy^1)$$

  is constant along any ray that emanates from the origin. In other words, it is invariant under positive rescaling in y. Given that, we might *expect* its integrals over S and $\mathbb{S}^1$ to give the same answer.

- Our intuition is borne out by the fact that the new integrand is a closed 1-form on $\mathbb{R}^2 \setminus 0$, together with an application of Stokes' theorem. See Exercise 4.5.2 for more elaborations.

- This remarkable integrand is sometimes denoted by $d\theta$. As we mentioned in §4.1, the quantity θ is known as the **Landsberg angle.**

Therefore

$$\boxed{L \;=\; \int_{\mathbb{S}^1} \frac{\sqrt{g}}{F^2}\,(y^1\,dy^2 \;-\; y^2\,dy^1)} \; . \tag{4.5.3}$$

Each tangent plane T_xM of a Finsler surface (M, F) is a Minkowski plane. The coordinates y^1, y^2 come from local coordinates x^1, x^2, through the expansion $y = y^i\,\frac{\partial}{\partial x^i}$. At each $x \in M$, we can calculate the Riemannian arc length $L(x)$ of the indicatrix

$$\mathrm{S}_x \;:=\; \{\, y \in T_xM : \; F(x,y) = 1 \,\}\ ,$$

using formula (4.5.3). A quick glance at (4.4.2b) (remember: $\delta y^i = dy^i + N^i{}_j\,dx^j$) tells us that the integrand in (4.5.3) is none other than the dy terms of ${\omega_1}^2$. Thus,

$$\boxed{L(x) \;=\; \int_{\mathbb{S}^1} \text{the pure } dy \text{ part of } {\omega_1}^2} \; . \tag{4.5.4}$$

If the Finsler surface (M, F) is Riemannian, $L(x)$ has the constant value 2π on M. In general, $L(x)$ is nonconstant. With the help of an unexpected but useful technical formula (see Exercise 4.5.6), and the fact that $d{\omega_1}^2 = {\Omega_1}^2$ (Exercise 4.4.6), we find that

$$\frac{\partial L}{\partial x^i} \;=\; \int_{\mathbb{S}^1} \text{the pure } dy \text{ part of } \left\{\, {\Omega_1}^2\Big(\frac{\partial}{\partial x^i}\,,\,\cdot\,\Big) \right\} .$$

We simplify the above integrand to the pure dy part of ${P_1}^2{}_{c1}\,\omega^c(\frac{\partial}{\partial x^i})\,\omega^3$, which can then be further reduced to

$$-\,J\,\omega^1\Big(\frac{\partial}{\partial x^i}\Big)\,\frac{\sqrt{g}}{F^2}\,(y^1\,dy^2 \;-\; y^2\,dy^1)\ .$$

This is done with the help of parts (b) and (d) of Exercise 4.4.7. Hence

$$\boxed{\begin{aligned} \frac{\partial L}{\partial x^1} \;&=\; -\int_{\mathbb{S}^1} J\,(\ell^2\sqrt{g})\,\frac{\sqrt{g}}{F^2}\,(y^1\,dy^2 \;-\; y^2\,dy^1)\ , \\ \frac{\partial L}{\partial x^2} \;&=\; +\int_{\mathbb{S}^1} J\,(\ell^1\sqrt{g})\,\frac{\sqrt{g}}{F^2}\,(y^1\,dy^2 \;-\; y^2\,dy^1)\ . \end{aligned}} \tag{4.5.5}$$

On a locally Minkowskian surface, J vanishes because both R and P do. (Again, see Exercise 4.4.7.) Thus on such a surface, the Riemannian arc length function $L(x)$ is constant. However, unlike Riemannian surfaces, the value of that constant is typically not 2π. We remarked about that right after (4.1.12), and referenced examples which are testimonial to this phenomenon. See Matsumoto [M2], [M3], and [BS], [BL1] for further discussions.

Exercises

Exercise 4.5.1: Suppose the Minkowski norm F on $\mathbb{R}^2$ is derived from an inner product $\langle\,,\,\rangle$, in the sense that $F(y) = \sqrt{\langle y\,,\,y\rangle}$. Prove that the Riemannian arc length L [defined in (4.5.1)] of the indicatrix must have the value 2π.

Exercise 4.5.2:

(a) Prove that

$$\boxed{d\,\sqrt{g} \;=\; \frac{1}{2}\,\sqrt{g}\;g^{ij}\;dg_{ij}}\;.$$

In particular, $\frac{\partial\sqrt{g}}{\partial y^k} = \frac{1}{2}\,\sqrt{g}\;g^{ij}\,\frac{\partial g_{ij}}{\partial y^k}$.

(b) Introduce the abbreviation $A_k := g^{ij}\,A_{ijk}$. Check that

$$\frac{\partial\sqrt{g}}{\partial y^k} \;=\; \frac{\sqrt{g}}{F}\;A_k\;.$$

(c) Show that $\frac{\sqrt{g}}{F^2}\,(y^1\,dy^2 - y^2\,dy^1)$ is a *closed* 1-form on $\mathbb{R}^2 \setminus 0$.

(d) Use Stokes' theorem to demonstrate that integrating this 1-form over S and $\mathbb{S}^1$ give the same answer. Explain why the same conclusion holds if the standard unit circle $\mathbb{S}^1$ is replaced by some simple closed curve with winding number 1 around the origin.

Exercise 4.5.3: Suppose, instead of the indicatrix S, we want to compute the Riemannian arc length of an arbitrary level curve of F, say

$$\mathrm{S}(r) \;:=\; \{\,y \in \mathbb{R}^2 :\; F(y) = r\,\}\;.$$

(a) By beginning with (1.4.8), show that the answer is

$$r\int_{\mathbb{S}^1}\frac{\sqrt{g}}{F^2}\;(\,y^1\;dy^2 \;-\; y^2\;dy^1\,)\;.$$

(b) Is the Riemannian arc length of $\mathrm{S}(r)$ simply r times that of the indicatrix? If so, is there a more elementary reason for that?

Exercise 4.5.4: In this exercise, let us use the temporary abbreviations

$$p \;:=\; y^1\;, \qquad q \;:=\; y^2\;.$$

Parametrize the indicatrix S as $t \mapsto (\,p(t)\,,\,q(t)\,)$. Abbreviate the components of the velocity vector as $\dot p = \dot y^1$ and $\dot q = \dot y^2$.

(a) Explain why $\dot p\,F_p + \dot q\,F_q = 0$. Then conclude that

$$\dot p \;=\; -\,\alpha\;F_q\;, \qquad \dot q \;=\; +\,\alpha\;F_p$$

for some function $\alpha(t)$.

(b) With the help of $\det(F_{y^i y^j}) = 0$, show that

$$\det(g_{ij}) = F\,[\,F_{pp}\,(F_q)^2 - 2\,F_{pq}\,F_p\,F_q + F_{qq}\,(F_p)^2\,]\,.$$

(c) With the help of Euler's theorem (§1.2) and (a), check that

$$\frac{\det(g_{ij})}{F^4}\,(p\,\dot q - q\,\dot p)^2 = \frac{\det(g_{ij})}{F^2}\,\alpha^2 = \det(g_{ij})\,\alpha^2\,.$$

(d) By a direct computation and part (a), prove that

$$g_{ij\,(y)}\,\dot y^i\,\dot y^j = \det(g_{ij})\,\alpha^2\,.$$

Thus

$$\boxed{\frac{\sqrt{g}}{F^2}\,(y^1\,\dot y^2 - y^2\,\dot y^1) = \sqrt{g_{ij\,(y)}\,\dot y^i\,\dot y^j}}\,.$$

This again explains why (4.5.3) is measuring the **Riemannian arc length** of the indicatrix.

Exercise 4.5.5: The **Finslerian arc length** of the indicatrix S is given by the integral

$$\int_{\mathrm{S}} \sqrt{g_{ij\,(\dot y)}\,\dot y^i\,\dot y^j}\;dt\,.$$

Here, the indicatrix is treated as a parametric curve, and $\dot y$ is the instantaneous velocity vector. Note that unlike Exercise 4.5.4, the g_{ij} here is evaluated at $\dot y$, not at y.

(a) As a contrast, explain why its Riemannian arc length is given by $\int_{\mathrm{S}} \sqrt{g_{ij\,(y)}\,\dot y^i\,\dot y^j}\;dt$ instead.

(b) Show that the indicatrix's Finslerian arc length can be re-expressed as $\int_{\mathrm{S}} F(\dot y)\,dt$.

(c) What condition must be imposed on the Minkowski norm F in order for the Riemannian and Finslerian arc lengths to coincide?

Exercise 4.5.6: Suppose

- ω is any 1-form on SM,
- b^r are functions of x only; that is, they live on M.

Then:

$$\boxed{\begin{aligned} & b^r \frac{\partial}{\partial x^r} \int_{y\in\mathbb{S}^1} \text{the pure } dy \text{ part of } \omega \\ &= \int_{y\in\mathbb{S}^1} \text{the pure } dy \text{ part of } \Big\{ (d\omega)\Big(b^r \frac{\partial}{\partial x^r}\,,\ \cdot\ \Big) \Big\} \end{aligned}}\,.$$

The proof amounts to a careful scrutiny of two cases:

(a) Let ω be a 1-form on SM that is of the type $\omega_i \, dy^i$. Check that both sides of our assertion reduce to $\int_{y \in \mathbb{S}^1} b^r \, \frac{\partial \omega_i}{\partial x^r} \, dy^i$.

(b) Next, let ω be of the special type $\omega_i \, dx^i$. The left-hand side of the asserted statement is of course zero. Show that its right-hand side equals

$$\int_{y \in \mathbb{S}^1} - b^r \, \frac{\partial \omega_r}{\partial y^s} \, dy^s \ .$$

Verify that the integrand is an exact differential in the y variable. Then explain why the displayed integral vanishes.

Exercise 4.5.7: Fill in the details in the derivation of (4.5.5).

4.6. A Gauss–Bonnet Theorem for Landsberg Surfaces

A Finsler surface (M, F) is said to be of **Landsberg** type if $J = 0$. Both Riemannian and locally Minkowskian surfaces belong to this category. The reason is as follows. We observed at the end of §3.3 that $P_j{}^i{}_{kl} = 0$ (in natural coordinates) for these spaces. Since P is a tensor, we must have $P_b{}^a{}_{cd} = 0$ (in any g-orthonormal basis) as well. By part (d) of Exercise 4.4.7, J must therefore vanish.

On a Landsberg surface, we see from (4.5.5) that the Riemannian arc lengths of its indicatrices are all equal to a constant value L.

The theorem that we present below is the generalization of the classical **Gauss–Bonnet theorem** to Landsberg surfaces. Matsumoto informs us that a result of this vintage was first anticipated by Berwald [Ber1]. However, our method of proof is through Chern's transgression, which is an intrinsic argument. See [Ch2] and [BC2]. It has the advantage that the Gauss–Bonnet theorem and the Poincaré–Hopf index theorem are established *simultaneously*.

The proof of this Gauss–Bonnet theorem involves the concept of metric distance. We give a cursory definition of that here. A detailed study is undertaken in §6.2. Let $\sigma : [a, b] \to M$ be a piecewise C^∞ curve with velocity $\frac{d\sigma}{dt} = \frac{d\sigma^i}{dt} \frac{\partial}{\partial x^i} \in T_{\sigma(t)}M$. Its **integral length** $L(\sigma)$ is defined as

$$L(\sigma) \ := \ \int_a^b F\Big(\sigma \, , \, \frac{d\sigma}{dt}\Big) \, dt \ .$$

For p, $x \in M$, denote by $\Gamma(p, x)$ the collection of all piecewise C^∞ curves $\sigma : [a, b] \to M$ with $c(a) = p$ and $c(b) = x$. Define the metric distance from p to x by

$$d(p \, , \, x \,) \ := \ \inf_{\Gamma(p,x)} \ L(\sigma) \ .$$

Note that since F is typically positively homogeneous (of degree 1) but not absolutely homogeneous, $d(p, x)$ is possibly *not* equal to $d(x, p)$.

Theorem 4.6.1 (Gauss–Bonnet). *Let (M, F) be a compact connected Landsberg surface without boundary. Denote the common value of the Riemannian arc lengths of all its indicatrices by L. Then:*

$$\boxed{\frac{1}{L} \int_M K \sqrt{g}\, dx^1 \wedge dx^2 \;=\; \chi(M)} \,.$$

Here:

- *The product $K\sqrt{g}$ depends only on x, even though K and $\sqrt{g}$ individually may have y dependences.*
- *$\chi(M)$ is the Euler characteristic of M.*

Proof.

Fix any arbitrary continuous vector field V on M, and suppose its zeroes are all isolated. Since M is compact, the number of zeroes of such a V must be finite, possibly none. Denote each of its zeroes by a generic symbol $\mathfrak{z}$.

For each small $\epsilon > 0$, we remove the metric discs

$$\{x \in M : \; d(\mathfrak{z}\,, x) \;<\; \epsilon\}$$

from M. The resulting compact connected manifold with boundary is denoted by M_ϵ. Its boundary consists of circles

$$\sigma^\epsilon_{\mathfrak{z}}(\,-t\,)\;, \qquad 0 \leqslant t \leqslant 2\pi$$

of metric radius ϵ. Each of them is centered at $\mathfrak{z}$ and is parametrized in a *clockwise* manner (that's why we put that minus sign in front of t). Normalizing our vector field V produces a lift (Figure 4.2) of M_ϵ into SM:

$$x \;\overset{U}{\longmapsto}\; \frac{V(x)}{F(\,V(x)\,)} \,.$$

For a Landsberg surface, (4.4.5) tells us that

$$- \,K\, \omega^1 \wedge \omega^2 \;=\; d\omega_1{}^2 \,, \tag{4.6.1}$$

where all quantities involved are *globally* defined on SM. Equivalently (see Exercise 4.6.1),

$$- \,K \sqrt{g}\; dx^1 \wedge dx^2 \;=\; d\omega_1{}^2 \,. \tag{4.6.2}$$

Exterior differentiation on SM (but treating y^1, y^2 as homogeneous coordinates) then gives

$$\frac{\partial}{\partial y^i}\,(\,K \sqrt{g}\,) \;=\; 0 \,. \tag{4.6.3}$$

So the product $K\sqrt{g}$ depends only on x. In other words, the left-hand sides of (4.6.1) and (4.6.2) live on M whenever (M, F) is of Landsberg type.

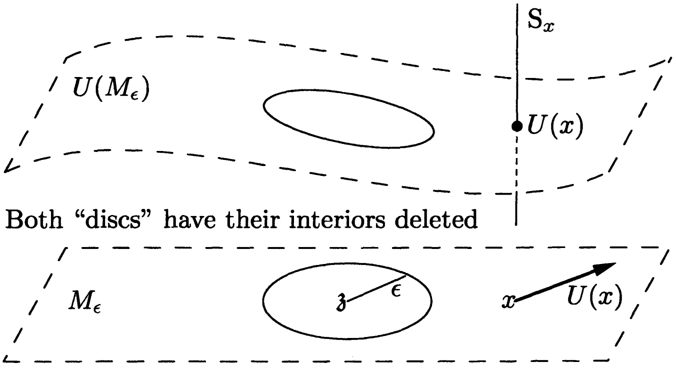

Figure 4.2

$U(M_\epsilon)$ is a codimension 1 submanifold of SM. The form ${\omega_1}^2$ is globally defined on SM, but only makes sense locally on M. This is so because its definition requires an orthonormal frame, which may fail to be continuously extendible past certain points of M.

Let us integrate (4.6.2) over the 2-dimensional submanifold $U(M_\epsilon)$ of SM, apply Stokes' theorem, and take the limit as $\epsilon \to 0$. The result is:

$$\int_M - K\sqrt{g}\, dx^1 \wedge dx^2 \;=\; \sum_{\mathfrak{z}} \lim_{\epsilon\to 0} \int_{U[\sigma^\epsilon_{\mathfrak{z}}(-t)]} {\omega_1}^2 \,.$$

In other words,

$$\int_M K\sqrt{g}\, dx^1 \wedge dx^2 \;=\; \sum_{\mathfrak{z}} \lim_{\epsilon\to 0} \int_{U[\sigma^\epsilon_{\mathfrak{z}}(t)]} {\omega_1}^2 \,. \tag{4.6.4}$$

Here, the circles

$$\sigma^\epsilon_{\mathfrak{z}}(t)\,, \qquad 0 \leqslant t \leqslant 2\pi$$

are manifestly traced out in a *counterclockwise* manner. Also, recall from (4.4.2b) that

$${\omega_1}^2 \;=\; \frac{\sqrt{g}}{F}\left(y^1\,\frac{\delta y^2}{F} \;-\; y^2\,\frac{\delta y^1}{F} \right),$$

where $\delta y^i = dy^i + N^i{}_j\, dx^j$.

Focus on a fixed zero $\mathfrak{z}$:

* Consider its corresponding integral

$$\lim_{\epsilon\to 0} \int_{U[\sigma^\epsilon_{\mathfrak{z}}(t)]} \frac{\sqrt{g}}{F}\left(y^1\,\frac{\delta y^2}{F} \;-\; y^2\,\frac{\delta y^1}{F} \right)$$

on the right-hand side of (4.6.4). As we traverse the small circle $\sigma_{\mathfrak{z}}^{\epsilon}$ (in a counterclockwise way), the dx terms in this integral keep track of our spatial displacements, while the terms $(y^1\, dy^2 - y^2\, dy^1)$ measure the net change in directions of the unit vector field U.

* We are continually shrinking the metric radius of the circle centered at $\mathfrak{z}$. So the dx terms in the above integrand do not contribute in the long run. Hence the integral is the same as

$$\lim_{\epsilon \to 0} \int_{U[\sigma_{\mathfrak{z}}^{\epsilon}(t)]} \frac{\sqrt{g}}{F^2} \left(y^1\, dy^2 \;-\; y^2\, dy^1 \right) .$$

Here, we have again used the fact that $\delta y^i = dy^i + N^i{}_j\, dx^j$.

* As we walk once around an arbitrarily *small* counterclockwise circle $\sigma_{\mathfrak{z}}^{\epsilon}$, we note that the direction field U traces out the indicatrix $\mathrm{S}_{\mathfrak{z}}$ (in the tangent plane $T_{\mathfrak{z}}M$) a number of times, counted algebraically. Namely, $+1$ for each counterclockwise tracing of $\mathrm{S}_{\mathfrak{z}}$, and -1 for each clockwise tracing. This number $I_{\mathfrak{z}}$ is called the **index** of V at the zero $\mathfrak{z}$. See Figure 4.3. With this understanding, we realize that the above limit is equal to

$$I_{\mathfrak{z}} \int_{\mathrm{S}_{\mathfrak{z}}} \frac{\sqrt{g}}{F^2} \left(y^1\, dy^2 \;-\; y^2\, dy^1 \right) .$$

By (4.5.2), this is simply $I_{\mathfrak{z}}\, L$.

Repeating the described procedure at every zero of V converts (4.6.4) to

$$\frac{1}{L} \int_M K \sqrt{g}\; dx^1 \wedge dx^2 \;=\; \sum_{\mathfrak{z}} I_{\mathfrak{z}} \,. \tag{4.6.5}$$

Formula (4.6.5) is interesting:

- It says that the sum of the indices of V gives a number which is independent of V. A creative choice of V reveals that this sum is simply the Euler characteristic of M. See Exercise 4.6.2. Thus, the **Poincaré–Hopf index theorem** is a *corollary* of (4.6.5)! It says that on a compact manifold without boundary, the total index of any vector field with isolated zeroes is equal to a topological invariant, namely, the Euler characteristic.
- Updating (4.6.5) with the above realization, it reads

$$\frac{1}{L} \int_M K \sqrt{g}\; dx^1 \wedge dx^2 \;=\; \chi(M) \,,$$

which is the Gauss–Bonnet theorem we seek. □

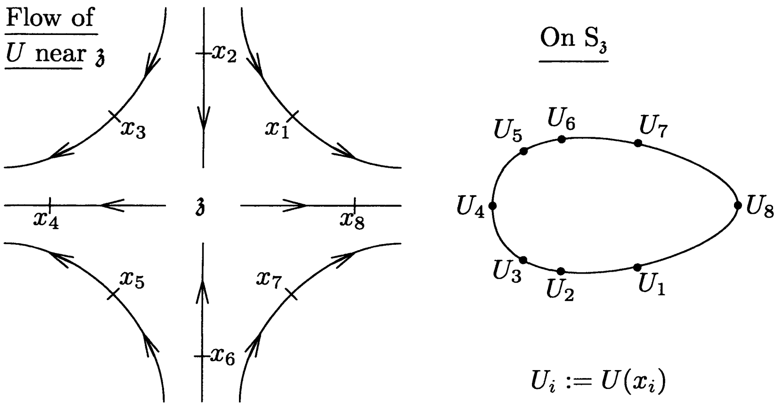

Figure 4.3

The sample points $x_1, \ldots, x_8$ are taken from a *small* counterclockwise metric circle centered at $\mathfrak{z}$. The sample values $U_i := U(x_i) \in S_{x_i} \approx S_{\mathfrak{z}}$. Since the $U(x_i)$ in this example trace out $S_{\mathfrak{z}}$ once *clockwise*, the index $I_{\mathfrak{z}}$ is -1. If the sample values $U(x_i)$ were to trace out $S_{\mathfrak{z}}$ once *counterclockwise*, the index would be $+1$.

Exercises

Exercise 4.6.1:

(a) Show that $\omega^1 \wedge \omega^2 = \sqrt{g}\, dx^1 \wedge dx^2$.

(b) Explain carefully how (4.6.2) gives (4.6.3).

(c) Attempt to find an explicit Finsler surface (M, F) for which K and $\sqrt{g}$ do *not* live on M, but their product does.

Exercise 4.6.2: Let V be a continuous vector field.

(a) Prove that its indices at a source, sink, and saddle are, respectively, $+1$, $+1$, and -1.

(b) Take any triangulation of a compact surface M. Show that there is a V on M which has a source at each vertex, a saddle at the midpoint of each edge, and a sink at the barycenter of each face. Thus the sum of the indices of this V is #vertices − #edges + #faces, which is the **Euler characteristic** of M.

References

[AIM] P. L. Antonelli, R. S. Ingarden, and M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology*, FTPH 58, Kluwer Academic Publishers, 1993.

[BC2] D. Bao and S. S. Chern, *A note on the Gauss–Bonnet theorem for Finsler spaces*, Ann. Math. **143** (1996), 233–252.

[BCS3] D. Bao, S. S. Chern, and Z. Shen, *Rigidity issues on Finsler surfaces*, Rev. Roumaine Math. Pures Appl. **42** (1997), 707–735.

[BL1] D. Bao and B. Lackey, *Randers surfaces whose Laplacians have completely positive symbol*, Nonlinear Analysis **38** (1999), 27–40.

[BS] D. Bao and Z. Shen, *On the volume of unit tangent spheres in a Finsler manifold*, Results in Math. **26** (1994), 1–17.

[Ber1] L. Berwald, *Atti Congresso Internal dei Mate., Bologna 3–10, Sept.* (1928).

[Ch2] S. S. Chern, *Historical remarks on Gauss–Bonnet*, Analysis et Cetera, volume dedicated to Jürgen Moser, Academic Press, 1990, pp. 209–217.

[doC1] M. P. do Carmo, *Differential Forms and Applications*, Springer-Verlag, 1994.

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

[M3] M. Matsumoto, *Theory of curves in tangent planes of two-dimensional Finsler spaces*, Tensor, N.S. **37** (1982), 35–42.

[On] B. O'Neill, *Elementary Differential Geometry*, 2nd ed., Academic Press, 1997.

[R] H. Rund, *The Differential Geometry of Finsler Spaces*, Springer-Verlag, 1959.

Chapter 5

Variations of Arc Length, Jacobi Fields, the Effect of Curvature

5.1 The First Variation of Arc Length

In this section, we use the method of differential forms to describe the first variation. There is another approach which uses vector fields and covariant differentiation. That is explored in a series of guided exercises at the end of §5.2. (Those exercises involve the second variation as well.) A systematic self-contained account can also be found in [BC1].

Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a regular piecewise C^∞ curve in M. Let $0 =: t_o < t_1 < \cdots < t_k := r$ be a partition of $[0, r]$ such that σ is C^∞ on each closed subinterval $[t_{s-1}, t_s]$. The **integral length** of σ is defined as

$$L(\sigma) \;:=\; \sum_{s=1}^{k} \int_{t_{s-1}}^{t_s} F\Big(\frac{d\sigma}{dt}\Big)\, dt \;.$$

Consider the rectangle

$$\square \;:=\; \{\,(t,u) : \; 0 \leqslant t \leqslant r \;,\; -\epsilon < u < +\epsilon\,\} \;.$$

A **piecewise C^∞ variation** of $\sigma(t)$ is a continuous map $\sigma(t,u)$ from $\square$ into M which is smooth on each $[t_{s-1}, t_s] \times (-\epsilon, \epsilon)$, and such that $\sigma(t, 0)$ reduces to the given **"base" curve** $\sigma(t)$. The object $\sigma(t, u = \text{constant})$ is known as a **t-curve**, while $\sigma(t = \text{constant}, u)$ is called a **u-curve**. Their velocity fields give rise, respectively, to two vector fields:

$$T := \sigma_* \frac{\partial}{\partial t} = \frac{\partial \sigma}{\partial t} , \quad U := \sigma_* \frac{\partial}{\partial u} = \frac{\partial \sigma}{\partial u} . \tag{5.1.1}$$

The quantity U is called the **variation vector field**, and is defined over all of $\Box$. On the other hand, note that at each (t_s, u), we have two velocities

$$T(t_s^-, u) := \lim_{t \to t_s^-} T(t,u) , \quad T(t_s^+, u) := \lim_{t \to t_s^+} T(t,u) ,$$

and these may not coincide. Thus T is possibly not defined at the points $t_s \times (-\epsilon, \epsilon)$ of $\Box$.

The map $\sigma(t,u)$ admits a **canonical lift** $\hat{\sigma} : \Box \to TM \setminus 0$, defined by

$$\boxed{\hat{\sigma}(t,u) := (\sigma(t,u) , T(t,u))} . \tag{5.1.2}$$

The map $\hat{\sigma}$ is possibly undefined at the points $t_s \times (-\epsilon, \epsilon)$ of $\Box$. In order that $\hat{\sigma}$ be well defined everywhere else, we assume from now on that all the t-curves in the variation $\sigma(t,u)$ are regular; that is, they have nowhere zero velocity fields.

Corresponding to $\hat{\sigma}$, one gets the following vector fields that are defined over $\Box$, except perhaps at the points (t_s, u):

$$\hat{T} := \hat{\sigma}_* \frac{\partial}{\partial t} = \frac{\partial \hat{\sigma}}{\partial t} , \quad \hat{U} := \hat{\sigma}_* \frac{\partial}{\partial u} = \frac{\partial \hat{\sigma}}{\partial u} . \tag{5.1.3}$$

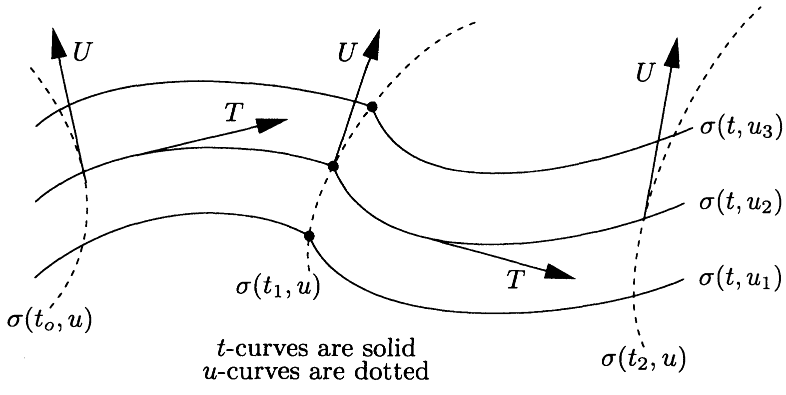

Figure 5.1

A piecewise C^∞ variation: its t-curves, u-curves, together with the vector fields T and U. Note that the kinks (3 large solid dots) all occur at the same parameter value $t = t_1$. Also, at the kinks, U is defined but T only has left and right limits.

Each $\hat{T}$ or $\hat{U}$ is a tangent vector of $TM \setminus 0$, based at the point $\hat{\sigma}$. On the other hand, each T or U is a tangent vector of M; as such, they are elements of $T_\sigma M$. Without being unreasonable, we can also regard T and U as elements in the fiber of π^*TM over the point $\hat{\sigma}$. In other words, the vectors $T^k \frac{\partial}{\partial x^k}_{|\sigma}$, $U^k \frac{\partial}{\partial x^k}_{|\sigma}$ on M can be identified with the elements $T^k \frac{\partial}{\partial x^k}_{|\hat{\sigma}}$, $U^k \frac{\partial}{\partial x^k}_{|\hat{\sigma}}$ of π^*TM. In particular, note that

$$T_{(\hat{\sigma})} = F(T)\,\ell\,. \tag{5.1.4}$$

Take a special g-orthonormal frame field $\{e_i\}$ for π^*TM, namely, one with $e_n := \ell$. Its coframe $\{\omega^i\}$ then has ω^n equal to the Hilbert form. Let us digress to explain what it means to express T and U in terms of $\{e_i\}$.

* T and U are elements of the tangent space $T_\sigma M$. At the point $\hat{\sigma}$ on $TM \setminus 0$, we have the g-orthonormal basis $\{e_i(\hat{\sigma})\}$ for the fiber of π^*TM over that point. But the fiber in question is simply a transplanted copy of $T_\sigma M$, to which the e_i (evaluated at $\hat{\sigma}$) belong. So we may regard $\{e_i(\hat{\sigma})\}$ as living on M and expand T and U in terms of it.
* Alternatively, think of T and U as belonging to the fiber of π^*TM over the point $\hat{\sigma}$. That fiber has the g-orthonormal basis $\{e_i(\hat{\sigma})\}$. We use this basis to expand T and U.

These two equivalent interpretations allow us to write, for later use,

$$T = T^i\, e_i\ , \quad U = U^i\, e_i\ .$$

With respect to the coframe $\{\omega^i\}$, we obtain the Chern connection 1-forms $\omega_j^{\ i}$, which are differential forms on $TM \setminus 0$. By our usual abuse of notation, the forms ω^i can also be regarded as differential forms on $TM \setminus 0$. Strictly speaking, they are sections of π^*T^*M.

These 1-forms can be pulled back to $\Box$ by $\hat{\sigma}^*$. We write

$$\hat{\sigma}^*\omega^i = a^i\, dt + b^i\, du\ , \tag{5.1.5}$$

$$\hat{\sigma}^*\omega_j^{\ i} = a_j^{\ i}\, dt + b_j^{\ i}\, du\ . \tag{5.1.6}$$

Since ω^i has no dy terms,

$$a^i = (\hat{\sigma}^*\omega^i)\Big(\frac{\partial}{\partial t}\Big) = \omega^i(\hat{T}) = \omega^i(T) = \omega^i[F(T)\,\ell] = F(T)\,\delta^i_{\ n}\ .$$

Hence

$$\boxed{a^\alpha = 0\ , \quad a^n = F(T)}\ . \tag{5.1.7}$$

Similarly,

$$\boxed{b^i = U^i}\ , \tag{5.1.8}$$

$$\boxed{a_j^{\ i} = \omega_j^{\ i}(\hat{T})\ , \quad b_j^{\ i} = \omega_j^{\ i}(\hat{U})}\ . \tag{5.1.9}$$

In order to avoid confusion with $a^{\cdot}$, $a_{\cdot}{}^{\cdot}$ and $b^{\cdot}$, $b_{\cdot}{}^{\cdot}$, we are temporarily using i, j, k, etc. to denote indices with respect to our special g-orthonormal frame. Previously, these midrange lower case Latin indices were exclusively reserved for natural coordinate bases. Here, they are shared by both types of bases. For instance, we write $\{e_i\}$ instead of $\{e_a\}$, and continue to write $\{\frac{\partial}{\partial x^k}\}$.

Next:

- The almost metric-compatibility of the Chern connection [part (b) of Exercise 2.4.5] implies that $\omega_{ni} = -\omega_{in}$, hence

$$\omega_n{}^n = 0 \text{ and } \omega_\alpha{}^n = -\,\delta_{\alpha\beta}\,\omega_n{}^\beta\,,$$

where we remind ourselves that lower case Greek indices run from 1 to $n-1$. These are equivalent to the statements

$$\boxed{a_n{}^n = 0\,, \quad b_n{}^n = 0}\,, \tag{5.1.10}$$

$$\boxed{a_\alpha{}^n = -\,\delta_{\alpha\beta}\,a_n{}^\beta\,, \quad b_\alpha{}^n = -\,\delta_{\alpha\beta}\,b_n{}^\beta}\,. \tag{5.1.11}$$

- Using $\hat{\sigma}^*$ to pull back the torsion-free criterion [part (a) of Exercise 2.4.5] and plugging in (5.1.5), (5.1.6) give

$$\boxed{-\frac{\partial a^i}{\partial u} + \frac{\partial b^i}{\partial t} = a^j\,b_j{}^i - b^j\,a_j{}^i}\,. \tag{5.1.12}$$

In particular, we have

$$\boxed{\frac{\partial b^\alpha}{\partial t} = a^n\,b_n{}^\alpha - b^j\,a_j{}^\alpha} \tag{5.1.13}$$

and

$$\boxed{\frac{\partial a^n}{\partial u} = \frac{\partial b^n}{\partial t} + b^\alpha\,a_\alpha{}^n}\,. \tag{5.1.14}$$

The length $L(u)$ of any t-curve in the variation $\sigma(t,u)$ is given by

$$L(u) = \sum_{s=1}^{k} \int_{t_{s-1}}^{t_s} F(T)\,dt = \sum_{s=1}^{k} \int_{t_{s-1}}^{t_s} a^n\,dt\,. \tag{5.1.15}$$

Differentiate this with respect to the parameter u and input (5.1.14). We get:

$$\boxed{L'(u) = b^n\,\Big|_0^r - \sum_{s=1}^{k-1} b^n\,\Big|_{t_s^-}^{t_s^+} + \sum_{s=1}^{k} \int_{t_{s-1}}^{t_s} b^\alpha\,a_\alpha{}^n\,dt}\,. \tag{5.1.16}$$

This is the formula for the **first variation of arc length** in Finsler geometry. See also Exercise 5.2.4.

* The use of t_s^- and t_s^+ is necessary in dealing with piecewise smooth paths.
* Since $b^i = U^i$ (the components of the variation vector field U), we see that b^n vanishes at both 0 and r if the variation of $\sigma(t)$ keeps its endpoints fixed.
* It is also helpful to realize that

(5.1.17)
$$b^n = g_T\left(U , \frac{T}{F(T)} \right) ,$$

where

(5.1.18)
$$\boxed{g_T := g_{ij\,(\sigma,T)}\, dx^i \otimes dx^j}$$

is an inner product induced on the tangent space $T_{\sigma(t,u)}M$ by the fundamental tensor. Consequently, the second group of terms on the right-hand side of (5.1.16) is now better understood:

(5.1.19)
$$b^n \Big|_{t_s^-}^{t_s^+} = g_T\left(U(t_s) , \frac{T}{F(T)} \Big|_{t_s^-}^{t_s^+} \right) .$$

The difference $\frac{T}{F}(t_s^+) - \frac{T}{F}(t_s^-)$ represents a sudden change in direction at t_s . In other words, it is a "kink" along the t-curve in question. Curves that are C^1 do not have kinks.

Proposition 5.1.1. *Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a regular piecewise C^∞ curve in a Finsler manifold (M, F). The following two statements are equivalent:*

(a) *$L'(0) = 0$ for all piecewise C^∞ variations of σ that keep its endpoints fixed. Any curve σ with this property is called a Finslerian* **geodesic**.

(b) *σ is C^1 (that is, has no kinks) on the entire $[0, r]$ and satisfies*

$$a_\alpha{}^n(t, 0) = 0 ; \quad \text{equivalently,} \quad a_n{}^\alpha(t, 0) = 0 .$$

Remarks:

• This proposition reaffirms the intuition that a curve (with fixed endpoints) cannot be critical in arc length unless all its kinks are eliminated.

• There are other characterizations of geodesics besides part (b) of the above proposition. In the Exercise section of §5.2, we develop the concept of covariant differentiation, and deduce in Exercise 5.2.4

the geodesic equation

$$D_T \left[\frac{T}{F(T)} \right] = 0 \text{ with reference vector } T := \frac{d\sigma}{dt} .$$

For those interested in a a description using only ODEs, see Exercise 5.3.1. There, we find that Finslerian geodesics are solutions of

$$\frac{d^2\sigma^i}{dt^2} + \frac{d\sigma^j}{dt}\frac{d\sigma^k}{dt} \, (\gamma^i{}_{jk})_{(\sigma,T)} = \frac{d}{dt}[\log F(T)] \, \frac{d\sigma^i}{dt} .$$

None of the descriptions mentioned ever assumes that the parametrization gives constant Finslerian speed. In other words, the quantity $F(T) := F(\frac{d\sigma}{dt})$ need *not* be constant.

Proof of Proposition 5.1.1. Consider only variations in which all the t-curves share the same endpoints. For these, the variation vector field vanishes at $t_o = 0$ and $t_k = r$. After restricting (5.1.16) to the base curve, we get:

$$(*) \quad L'(0) = - \sum_{s=1}^{k-1} g_T\left(U(t_s) \, , \, \frac{T}{F(T)} \Big|_{t_s^-}^{t_s^+} \right) + \sum_{s=1}^{k} \int_{t_{s-1}}^{t_s} b^\alpha \, a_\alpha{}^n \, dt .$$

That (b) ⇒ (a) :

This is immediate from the remarks we just made about kinks, together with formula (*) above.

That (a) ⇒ (b) :

Suppose (a) holds. Construct a variation that has variation vector field $U^\alpha(t) = f(t)\, a_n{}^\alpha$ and $U^n(t) = 0$ along the base curve $\sigma(t)$. Here, $f(t)$ is chosen to be zero at $t_o, \ldots, t_k$ and positive on each (t_{s-1}, t_s). Since by hypothesis $L'(0) = 0$, formula (*) becomes

$$0 = - \sum_{s=1}^{k} \int_{t_{s-1}}^{t_s} f(t) \, \delta_{\alpha\beta} \, a_\alpha{}^n \, a_\beta{}^n \, dt .$$

Thus $a_\alpha{}^n = 0$ on each open subinterval (t_{s-1}, t_s).

Given that, formula (*) reduces to

$$L'(0) = - \sum_{s=1}^{k-1} g_T\left(U(t_s) \, , \, \frac{T}{F(T)} \Big|_{t_s^-}^{t_s^+} \right) .$$

This time, construct a variation of σ with endpoints fixed, and such that the variation vector field satisfies

$$U(t_s) = \frac{T}{F(T)} \Big|_{t_s^-}^{t_s^+}$$

for $s = 1, \ldots, k-1$. Again, $L'(0) = 0$ by hypothesis and we have

$$0 = - \sum_{s=1}^{k-1} g_T \left(\frac{T}{F(T)} \Big|_{t_s^-}^{t_s^+} , \frac{T}{F(T)} \Big|_{t_s^-}^{t_s^+} \right) .$$

Thus the kinks must all vanish, and σ is C^1 on the entire interval $[0, r]$. This implies that the coefficients $a_\alpha^{\ n}$ are well defined at each t_s, and vanish there too. □

Exercises

Exercise 5.1.1: The tangent space $T_{\sigma(t,u)}M$ has an induced inner product

$$g_T := g_{ij\,(\sigma,T)}\, dx^i \otimes dx^j .$$

(a) Check that $g_T(T,T) = F^2(T)$.

(b) With the help of the fundamental inequality (1.2.3), show that

$$g_T(U,U) < F^2(U) + (F\, F_{y^i y^j})_{(\sigma,T)}\, U^i U^j .$$

(c) Explain why that second term on the right is positive. Hint: try to obtain that through a careful Cauchy–Schwarz type argument on the inner product g_T. Consult (1.2.9) only as a last resort.

Exercise 5.1.2: Let ω be any section of $\pi^* T^* M$. Regarding T and U as $\pi^* TM$-valued, the pairings $\omega(T)$ and $\omega(U)$ make sense. On the other hand, ω can be viewed as a 1-form of $TM \setminus 0$, in which case $\omega(\hat{T})$ and $\omega(\hat{U})$ make sense. Explain why

$$\omega(T) = \omega(\hat{T}) \text{ and } \omega(U) = \omega(\hat{U}) .$$

Hint: find out how $\hat{T}$ differs from T in natural coordinates, and keep in mind that ω has no dy components.

Exercise 5.1.3:

(a) Give the details in the derivation of (5.1.12).

(b) Establish the first variation formula (5.1.16), following the guidelines stated in the section proper.

Exercise 5.1.4: Let $\sigma(t,u)$ be any piecewise C^∞ variation of a *geodesic* $\sigma(t)$, $0 \leqslant t \leqslant r$, with velocity T. However, do *not* assume that the variation leaves the endpoints of $\sigma(t)$ fixed. Show that

$$L'(0) = g_T \left(U , \frac{T}{F(T)} \right) \Big|_0^r .$$

Exercise 5.1.5: Let $\sigma(t)$ be a C^1 regular curve in M, with velocity T. Let $\hat{\sigma} := (\sigma, T)$ denote its canonical lift into $TM \setminus 0$. Prove that the following three statements are mutually equivalent.

(a) The curve $\sigma(t)$ is a geodesic.

(b) Its canonical lift satisfies the exterior differential system:

$$\boxed{\begin{aligned} \omega^\alpha(\hat{T}) &= \omega^\alpha(T) = 0 \,, \\ \omega^n(\hat{T}) &= F(T) \,, \\ \omega_n{}^\alpha(\hat{T}) &= 0 \,. \end{aligned}}$$

(c) The velocity $\hat{T}$ of its canonical lift is given by the formula

$$\boxed{\hat{T} = F(T) \left[\ell^i \frac{\delta}{\delta x^i} \right] + \left[\frac{d}{dt} \log F(T) \right] \left(\ell^i \, F \frac{\partial}{\partial y^i} \right)} \,.$$

Hint: review part (a) of Exercise 2.3.5, and part (c) of Exercise 2.4.7.

Exercise 5.1.6: As a continuation of Exercise 5.1.5, show that the following two statements are equivalent.

(a) $\sigma(t)$ is a geodesic parametrized to have constant speed.

(b) Its canonical lift $\hat{\sigma} := (\sigma, T)$ is an integral curve of $F\,\hat{\ell}$.

Here, $\hat{\ell}$ abbreviates $\ell^i \frac{\delta}{\delta x^i}$. Can you now explain why constant speed geodesics are necessarily C^∞?

Exercise 5.1.7: The previous exercise shows that every constant speed geodesic of M generates, through the canonical lift, an integral curve of $F\,\hat{\ell}$. It does *not* claim that every integral curve of $F\,\hat{\ell}$ must arise this way. Nevertheless, such is indeed the case.

(a) Show that if (x_t, y_t) is an integral curve of $F\,\hat{\ell}$, then it is equal to the canonical lift of its projection x_t. Hint: compute the velocity field of x_t.

(b) Use Exercise 5.1.6 to explain why the curve x_t must be a constant speed geodesic.

Exercise 5.1.8: Instead of using the route proposed in Exercise 5.1.7, there is a more direct way. Again, let (x_t, y_t), now abbreviated simply as (x, y), be an integral curve of the vector field $F\,\hat{\ell}$.

(a) Where do the equations $\dot{x}^j = y^j$ and $\dot{y}^j = -N^j{}_i$ come from? Here, N is the nonlinear connection introduced in §2.3. Hint: list the components of $\hat{\ell}$ in terms of $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$.

(b) Use part (a) of Exercise 2.4.6 to help show that $\ddot{x}^j = -\Gamma^j{}_{ik}\, y^i\, y^k$, where the $\Gamma^i{}_{jk}$ are the components of the Chern connection in natural coordinates. Check that the right-hand side further reduces to $-\gamma^j{}_{ik}\, y^i\, y^k$, where the $\gamma^i{}_{jk}$ are the fundamental tensor's Christoffel symbols of the second kind.

Therefore, the projected curve x_t satisfies $\ddot{x}^j + \gamma^j{}_{ik}\, \dot{x}^i\, \dot{x}^k = 0$. As we show in §5.3, this second order system of ODEs describes the component functions of constant speed Finslerian geodesics.

5.2 The Second Variation of Arc Length

The setup is just like what we had in §5.1, except now the base curve $\sigma(t) = \sigma(t, 0)$ in our piecewise C^∞ variation

$$\sigma(t,u)\ , \quad \text{with}\quad (t,u)\ \in\ \square\ :=\ [0,r] \times (-\epsilon, \epsilon)\ ,$$

is a geodesic. As before, the continuous map $\sigma(t,u)$ is smooth on each $[t_{s-1}, t_s] \times (-\epsilon, \epsilon)$, and $0 =: t_o < t_1 < \cdots < t_k := r$. Also, all the t-curves in the variation are assumed to have nowhere zero velocity fields $T(t,u)$. Thus the canonical lift $\hat{\sigma} := (\sigma, T)$, a map from $\square$ into $TM \setminus 0$, is well defined.

Exercise 3.2.1 says that

$$d\omega_j{}^i\ -\ \omega_j{}^k \wedge \omega_k{}^i\ =\ \frac{1}{2}\, R_j{}^i{}_{kl}\; \omega^k \wedge \omega^l\ +\ P_j{}^i{}_{kl}\; \omega^k \wedge \omega^{n+l}\ .$$

With the help of $P_j{}^i{}_{kn} = 0$ (Exercise 3.3.2) and the relations $\omega^{n+\alpha} = \omega_n{}^\alpha$, $\omega_n{}^n = 0$ (see Exercise 2.4.7), the above becomes

$$d\omega_j{}^i\ -\ \omega_j{}^k \wedge \omega_k{}^i\ =\ \frac{1}{2}\, R_j{}^i{}_{kl}\; \omega^k \wedge \omega^l\ +\ P_j{}^i{}_{kl}\; \omega^k \wedge \omega_n{}^l\ .$$

Use σ to pull this back to $\square$. After some simplifications, we obtain

$$\boxed{\begin{aligned} \frac{\partial b_j{}^i}{\partial t} - \frac{\partial a_j{}^i}{\partial u} &= a_j{}^k\, b_k{}^i\ -\ b_j{}^k\, a_k{}^i \\ &\quad + R_j{}^i{}_{nl}\; a^n\, b^l \\ &\quad + P_j{}^i{}_{n\beta}\; a^n\, b_n{}^\beta\ -\ P_j{}^i{}_{k\beta}\; b^k\, a_n{}^\beta\ . \end{aligned}} \tag{5.2.1}$$

As always, n is *not* a summation index. Now specialize j to α and i to n. We also use the fact that $P_\alpha{}^n{}_{n\beta} = -P^n{}_{\alpha n\beta} = \dot{A}_{\alpha n\beta} = 0$, which comes from some of the Bianchi identities in §3.4. The result is:

$$\boxed{\begin{aligned} \frac{\partial a_\alpha{}^n}{\partial u} &= \frac{\partial b_\alpha{}^n}{\partial t} - a_\alpha{}^k\, b_k{}^n\ +\ b_\alpha{}^k\, a_k{}^n \\ &\quad - R_\alpha{}^n{}_{nl}\; a^n\, b^l \\ &\quad + P_\alpha{}^n{}_{k\beta}\; b^k\, a_n{}^\beta\ . \end{aligned}} \tag{5.2.2}$$

We are ready to compute the second variation of arc length. Begin with

$$L''(u)\ =\ \sum_{s=1}^{k} \int_{t_{s-1}}^{t_s} \frac{\partial^2 a^n}{\partial u^2}\; dt\ .$$

Recall that

$$\frac{\partial a^n}{\partial u} = \frac{\partial b^n}{\partial t} + b^\alpha \, a_\alpha^{\ n} .$$

To calculate $\frac{\partial^2 a^n}{\partial u^2}$, we use (5.2.2), followed by

$$b^\alpha \, \frac{\partial b_\alpha^{\ n}}{\partial t} = \frac{\partial}{\partial t}(b^\alpha \, b_\alpha^{\ n}) - \frac{\partial b^\alpha}{\partial t} \, b_\alpha^{\ n} ,$$

and get the intermediate formula

$$\begin{aligned}\frac{\partial^2 a^n}{\partial u^2} = \ & \frac{\partial}{\partial t}\left(\frac{\partial b^n}{\partial u} + b^\alpha \, b_\alpha^{\ n} \right) - \left(\frac{\partial b^\alpha}{\partial t} \, b_\alpha^{\ n} + b^\alpha \, a_\alpha^{\ k} \, b_k^{\ n} \right) \\ & + a_\alpha^{\ n} \left(\frac{\partial b^\alpha}{\partial u} + b^\beta \, b_\beta^{\ \alpha} - b^\beta \, b^k \, P_{\beta\ k}^{\ n\ \alpha} \right) \\ & - a^n \, b^\alpha \, b^l \, R_{\alpha\ nl}^{\ n} .\end{aligned}$$

Our work in §5.1 tell us that

$$\begin{aligned} - \left(\frac{\partial b^\alpha}{\partial t} \, b_\alpha^{\ n} + b^\alpha \, a_\alpha^{\ k} \, b_k^{\ n} \right) &= - a_\alpha^{\ n} \, b^n \, \delta^{\beta\alpha} \, b_\beta^{\ n} - a^n \, b_n^{\ \alpha} \, b_\alpha^{\ n} \\ &= - a_\alpha^{\ n} \, b^n \, \delta^{\alpha\beta} \, b_\beta^{\ n} + a^n \, \delta_{ij} \, b_n^{\ i} \, b_n^{\ j} ,\end{aligned}$$

and

$$b^\alpha \, b_\alpha^{\ n} = - \delta_{ij} \, b^i \, b_n^{\ j} .$$

Also, the R term can be rewritten: $R_{\alpha\ nl}^{\ n} = R_{\alpha nnl} = R_{n\alpha ln} = R_{\alpha l}$. Consequently,

$$b^\alpha \, b^l \, R_{\alpha\ nl}^{\ n} = b^\alpha \, R_{n\alpha ln} \, b^l = b^i \, R_{nijn} \, b^j = b^i \, R_{ij} \, b^j .$$

These maneuvers improve the intermediate formula to:
(5.2.3)

$$\boxed{\begin{aligned}\frac{\partial^2 a^n}{\partial u^2} = \ & \frac{\partial}{\partial t}\left(\frac{\partial b^n}{\partial u} - \delta_{ij} \, b^i \, b_n^{\ j} \right) \\ & + a^n \, (\delta_{ij} \, b_n^{\ i} \, b_n^{\ j} - b^i \, R_{ij} \, b^j) \\ & + a_\alpha^{\ n} \left(\frac{\partial b^\alpha}{\partial u} + b^\beta \, b_\beta^{\ \alpha} - b^\beta \, b^k \, P_{\beta\ k}^{\ n\ \alpha} - b^n \, \delta^{\alpha\beta} \, b_\beta^{\ n} \right) .\end{aligned}}$$

Now:

* Evaluate (5.2.3) at $u = 0$,
* Use the fact that the base curve $\sigma(t)$ is a geodesic, so $a_\alpha^{\ n}(t, 0) = 0$.
* Being a geodesic, $\sigma(t)$ has no kinks (Proposition 5.1.1), thus the quantity $\frac{\partial b^n}{\partial u} - \delta_{ij} \, b^i \, b_n^{\ j}$ has no jump discontinuity at any t_s.

These observations give

$$
\text{(5.2.4)} \qquad \boxed{\begin{aligned} L''(0) \;=\; & \left(\frac{\partial b^n}{\partial u} \;-\; \delta_{ij}\; b^i\; b_n^{\;j} \right) \Bigg|_{t=0}^{t=r} \\ & + \int_0^r a^n \; (\, \delta_{ij}\; b_n^{\;i}\; b_n^{\;j} \;-\; b^i\; R_{ij}\; b^j \,)\; dt \;. \end{aligned}}
$$

This is the formula for the **second variation of arc length** in Finsler geometry.

- One should keep in mind that the base curve is a geodesic here.
- According to (3.9.3), the term $b^i\, R_{ij}\, b^j$ is proportional to a flag curvature. The flagpole in question is the velocity T, and the transverse edge of the flag is the variation vector field U.
- Note that the second Chern curvature P is conspicuously absent.

For the second variation of the *energy functional* in Finsler geometry, see Dazord [Daz] and Matsumoto [M2].

Exercises

In these exercises, we explore the variational calculus using vector fields and covariant derivatives, instead of differential forms. We show that the covariant derivative approach has its merits too.

Let $\sigma(t)$ be a smooth regular curve in M, with velocity field T. Let $W(t) := W^i(t)\, \frac{\partial}{\partial x^i}$ be a vector field along σ. The expression

$$
\left[\frac{dW^i}{dt} \;+\; W^j\; T^k\; \Gamma^i{}_{jk} \right] \; \frac{\partial}{\partial x^i}_{|\sigma(t)}
$$

would have defined the covariant derivative $D_T W$, had Γ not had a directional y-dependence. This dilemma is resolved by defining two versions of the said covariant derivative:

- If T is plugged into the direction slot y, we get

  $$
  \left[\frac{dW^i}{dt} \;+\; W^j\; T^k\; (\Gamma^i{}_{jk})_{(\sigma,T)} \right] \; \frac{\partial}{\partial x^i}_{|\sigma(t)} \;.
  $$

  We call it $D_T W$ **with reference vector** T.

- If W is plugged into the direction slot y, we get

  $$
  \left[\frac{dW^i}{dt} \;+\; W^j\; T^k\; (\Gamma^i{}_{jk})_{(\sigma,W)} \right] \; \frac{\partial}{\partial x^i}_{|\sigma(t)} \;.
  $$

  We call it $D_T W$ **with reference vector** W.

Exercise 5.2.1: Let T and U be the vector fields defined in (5.1.1). They are associated with any piecewise C^∞ variation. Prove that

$$\boxed{D_T U \;=\; D_U T\;, \quad \text{both with reference vector } T}\;.$$

Exercise 5.2.2: Let $\sigma(t)$ be a smooth regular curve in M, with velocity field T. Let $\hat{T}$ denote the velocity field of its canonical lift $\hat{\sigma} := (\sigma, T)$. Recall the almost g-compatibility criterion (2.4.6) of the Chern connection. By evaluating this criterion on $\hat{T}$, show that

$$\boxed{\begin{aligned} \frac{d}{dt}\, g_T(V\,, W) \;=\;& \; g_T(D_T V\,, W) \;+\; g_T(V\,, D_T W) \\ & + 2\, A(V, W, a_n{}^\alpha\, e_\alpha)\;. \end{aligned}}$$

This is an important formula. Embodied in it are the special circumstances under which the derivative of $g_T(V, W)$ obeys a "product rule." Those are enumerated in Exercise 5.2.3. In the displayed formula:

* $g_T := g_{ij\,(\sigma,T)}\, dx^i \otimes dx^j$; see (5.1.18).
* V and W are two arbitrary vector fields along σ.
* Both $D_T V$ and $D_T W$ are defined with reference vector T.
* A is the Cartan tensor, evaluated at (σ, T).
* $a_n{}^\alpha := \omega_n{}^\alpha(\hat{T})$, as in (5.1.9).

Exercise 5.2.3: As a continuation of Exercise 5.2.2, show that

$$\boxed{\frac{d}{dt}\, g_T(V\,, W) \;=\; g_T(D_T V\,, W) \;+\; g_T(V\,, D_T W)}$$

whenever *one* of the following three conditions is met:

- V or W is proportional to T.
- σ is a geodesic.
- A vanishes along our curve σ.

In other words, the coveted product rule holds under these circumstances.

Exercise 5.2.4:

(a) Exploit the relationship between $D_T[\frac{T}{F(T)}]$ and $\nabla_{\hat{T}}\,\ell$. Do that in natural coordinates.

(b) Write out $\nabla_{\hat{T}}\,\ell$ in our special g-orthonormal frame for π^*TM. Check that it equals $a_n{}^\alpha e_\alpha$ (automatically summed on α from 1 to $n-1$). In particular, there is no $e_n := \ell$ component. With this information at hand, rewrite the **first variation** (5.1.16) as

$$\boxed{\begin{aligned} L'(u) \;=\;& g_T\left(U\,,\,\frac{T}{F(T)}\Big|_0^r\right) \;-\; \sum_{s=1}^{k-1} g_T\left(U\,,\,\frac{T}{F(T)}\Big|_{t_s^-}^{t_s^+}\right) \\ & -\sum_{s=1}^{k}\int_{t_{s-1}}^{t_s} g_T\left(U\,,\,D_T\left[\frac{T}{F(T)}\right]\right)\,dt\,. \end{aligned}}$$

(c) Recall that **Finslerian geodesics** are characterized by the conditions $a_n{}^\alpha = 0$. Show that they are equally well described by the following equation:

$$\boxed{D_T\left[\frac{T}{F(T)}\right] \;=\; 0\;, \quad \text{with reference vector } T}\;.$$

(d) Show that the **constant speed geodesics** are precisely the solutions of

$$\boxed{D_T T \;=\; 0\;, \quad \text{with reference vector } T}\;.$$

Exercise 5.2.5: Establish the following correspondence for the terms that appear in the second variation of arc length:

$$\begin{aligned} \frac{\partial b^n}{\partial u} \;-\; \delta_{ij}\, b^i\, b_n^{\,j} \;&=\; g_T\left(D_U U\,,\,\frac{T}{F(T)}\right)\,, \\ a^n\,\delta_{ij}\,b_n^{\,i}\,b_n^{\,j} \;&=\; \frac{1}{F(T)}\left[\,g_T(D_T U\,,\,D_T U) \;-\; \left(\frac{\partial F(T)}{\partial u}\right)^2\,\right]\,, \\ -\,a^n\,b^i\,R_{ij}\,b^j \;&=\; -\,\frac{1}{F(T)}\,g_T(\,R(U,T)T\,,\,U\,)\,. \end{aligned}$$

Here:

- *All* covariant derivatives have T as reference vector.
- Exercise 5.2.1 is needed for the second statement.
- $\boxed{R(U,T)T \;:=\; (T^j\,R_j{}^i{}_{kl}\,T^l\,)\,U^k\,\frac{\partial}{\partial x^i}}\;.$

Exercise 5.2.6: Consider the component of U that is g_T-orthogonal to the velocity T. Namely,

$$\boxed{U_\perp \;:=\; U \;-\; g_T\left(U\,,\,\frac{T}{F(T)}\right)\frac{T}{F(T)}} \;=:\; U \;-\; \lambda\,\ell\,.$$

The abbreviations λ and ℓ are temporarily used to reduce clutter.

(a) With the help of Exercises 5.2.3, 5.2.4, and 5.2.1, show that $D_T\lambda$ is actually equal to $-\frac{\partial}{\partial u}F(T)$.

(b) Next, show that

$$g_T(D_T U_\perp, D_T U_\perp) = g_T(D_T U, D_T U) - (D_T\lambda)^2 .$$

Thus

$$g_T(D_T U, D_T U) - \left(\frac{\partial F(T)}{\partial u}\right)^2 = g_T(D_T U_\perp, D_T U_\perp) .$$

(c) Use the properties of $R_j{}^i{}_{kl}$ to check that

$$g_T(R(U_\perp, T)T, U_\perp) = g_T(R(U,T)T, U) .$$

Exercise 5.2.7:

(a) Show that the second variation can be expressed as

$$\boxed{L''(0) = I(U_\perp, U_\perp) + g_T\left(D_U U, \frac{T}{F(T)}\right)\Bigg|_0^r}$$

and

$$\boxed{\begin{aligned} L''(0) = \; & I(U,U) + g_T\left(D_U U, \frac{T}{F(T)}\right)\Bigg|_0^r \\ & - \int_0^r \frac{1}{F(T)} \left(\frac{\partial F(T)}{\partial u}\right)^2 dt , \end{aligned}}$$

where

$$\boxed{\begin{aligned} I(V,W) := \int_0^r \frac{1}{F(T)} [\; & g_T(D_T V, D_T W) \\ & - g_T(R(V,T)T, W) \;] \, dt \end{aligned}}$$

is the **index form**. All covariant derivatives here use T as the reference vector.

(b) In case the variation does keep the endpoints of σ fixed, explain why the formula for $L''(0)$ reduces to

$$L''(0) = I(U_\perp, U_\perp)$$

and

$$L''(0) = I(U,U) - \int_0^r \frac{1}{F(T)} \left(\frac{\partial F(T)}{\partial u}\right)^2 dt .$$

(c) Is the index form symmetric? Namely, is it true that

$$I(W,V) = I(V,W) \; ?$$

5.3 Geodesics and the Exponential Map

We assume that all our geodesics $\sigma(t)$ have been parametrized to have constant Finslerian speed. That is, the length $F(T)$ is constant, where T is the velocity field. According to Exercise 5.2.4, these geodesics are characterized by the equation

$$D_T T = 0 \ , \quad \text{with reference vector } T \ .$$

Namely, they are **autoparallels**. Since

$$T = \frac{d\sigma^i}{dt} \frac{\partial}{\partial x^i} \ ,$$

this equation says that

$$\frac{d^2\sigma^i}{dt^2} + \frac{d\sigma^j}{dt} \frac{d\sigma^k}{dt} \, ({\Gamma^i}_{jk})_{(\sigma,T)} = 0 \ . \tag{5.3.1}$$

The second term on the left can be reduced, using the explicit formula (2.4.9) of the Chern connection coefficients ${\Gamma^i}_{jk}$. It turns out that only the formal Christoffel symbols ${\gamma^i}_{jk}$ [see (2.3.1)] contribute to that double contraction. Thus the differential equations that describe **constant speed geodesics** are:

$$\boxed{\frac{d^2\sigma^i}{dt^2} + \frac{d\sigma^j}{dt} \frac{d\sigma^k}{dt} \, ({\gamma^i}_{jk})_{(\sigma,T)} = 0} \ . \tag{5.3.2}$$

It turns out that the system of ODEs is equally elegant when the geodesic has variable speed. In that case, one begins with

$$D_T \left[\frac{T}{F(T)} \right] = 0 \ , \quad \text{with reference vector } T \ .$$

See Exercise 5.3.1.

Let us now mention two analytical properties of constant speed geodesics:

- At each point of M, we have a precompact coordinate neighborhood U and an $\epsilon > 0$ such that, given any $x \in U$ and $y \in T_xM$ with $0 < F(x,y) < \epsilon$, there exists a *unique* geodesic $\sigma_{x,y}(t)$, $-2 < t < 2$ which passes through x at $t = 0$ with velocity y. Furthermore, $\sigma_{x,y}(t)$ is C^∞ in t, and in its initial data x, $y \neq 0$.

  The existence aspect can be seen as follows. Take any coordinate neighborhood U that has compact closure. For each x in the closure $\overline{U}$, consider the finite-dimensional vector space T_xM. The coordinate basis $\{\frac{\partial}{\partial x^i}\}$ induces an x-dependent Euclidean norm $|y| := \sqrt{\delta_{ij}\, y^i y^j}$. The quotient $F(y)/|y|$ is constant along rays that emanate from the origin of T_xM. It defines a continuous positive function on the compact set given by the portion of the indicatrix bundle over $\overline{U}$. Its absolute minimum $\mathfrak{m}$ and absolute maximum $\mathcal{M}$ are both positive. Choose $c > 1$ such that one has $\frac{1}{c} < \mathfrak{m} \leqslant \mathcal{M} < c$.

Then $\frac{1}{c}|\,y\,| \leqslant F(y) \leqslant c\,|\,y\,|$ for all $y = y^i\,\frac{\partial}{\partial x^i} \in T_xM$ and $x \in \overline{U}$. The asserted existence can now be obtained from standard ODE theory.

- Suppose we are given the above $\sigma_{x,y}(t)$ and any positive constant λ. The chain rule tells us that the curve $\sigma_{x,y}(\lambda t)$, $\frac{-2}{\lambda} < t < \frac{2}{\lambda}$ is also a geodesic which passes through x at time 0, but the velocity at that moment is λy. By uniqueness, the said curve must be $\sigma_{x,\lambda y}(t)$. We express this as

$$\sigma(t,x,\lambda y) \;=\; \sigma(\lambda t,x,y)\ ,\quad \lambda > 0\ . \tag{5.3.3}$$

Define the **exponential map** as follows:

$$\exp(x,y) \;:=\; \begin{cases} \sigma(1,x,y)\ , & y \neq 0 \\ x\ , & y = 0\ . \end{cases} \tag{5.3.4}$$

The above properties have the following to say about the exponential map:

- **exp is defined on an open neighborhood of the zero section of TM, and is C^∞ away from the zero section.**
- Statement (5.3.3) implies that

$$\boxed{\exp(x,\lambda y) \;=\; \sigma(\lambda,x,y)\ ,\quad \lambda > 0}\ . \tag{5.3.5}$$

The derivative of exp exists at the zero section of TM, and is the identity map. In other words,

$$\boxed{\frac{\partial \exp^i}{\partial y^j}(x,0) \;=\; \delta^i{}_j}\ . \tag{5.3.6}$$

This follows from the following two facts:

- * $\lim_{\lambda\to 0^-} \frac{\exp^i(x,\lambda v) - x^i}{\lambda} = v^i$.
- * $\lim_{\lambda\to 0^+} \frac{\exp^i(x,\lambda v) - x^i}{\lambda} = v^i$.

For the first one, we have

$$\begin{aligned} \lim_{\lambda\to 0^-} \frac{\exp^i(x,\lambda v) \;-\; x^i}{\lambda} &= +\lim_{\lambda\to 0^+} \frac{\exp^i(x,\lambda[-v]) \;-\; x^i}{-\lambda} \\ &= -\lim_{\lambda\to 0^+} \frac{\sigma^i(\lambda,x,-v) \;-\; \sigma^i(0,x,-v)}{\lambda} \\ &= -\frac{\partial \sigma^i}{\partial t}(0,x,-v) \\ &= -(-v)^i \;=\; v^i\ . \end{aligned}$$

The second limit can be established similarly.

We would also like to show that **the exponential map is C^1 at the zero section of TM.** To this end:

* Relabel y as v in (5.3.5) and differentiate with respect to v. The chain rule gives

$$\frac{\partial \exp^i}{\partial y^k}(x, \lambda v)\ \lambda\ \delta^k{}_j \ = \ \frac{\partial \sigma^i}{\partial y^j}(\lambda, x, v)\ .$$

That is,

$$\frac{\partial \exp^i}{\partial y^j}(x, \lambda v) \ = \ \frac{1}{\lambda}\ \frac{\partial \sigma^i}{\partial y^j}(\lambda, x, v)\ , \quad \lambda > 0\ .$$

* Note that $\sigma(0, x, v) = x$ for all v, hence $\frac{\partial \sigma^i}{\partial y^j}(0, x, v) = 0$. Thus the above can be rewritten as

$$\frac{\partial \exp^i}{\partial y^j}(x, \lambda v) \ = \ \frac{1}{\lambda}\ \left[\frac{\partial \sigma^i}{\partial y^j}(\lambda, x, v) \ - \ \frac{\partial \sigma^i}{\partial y^j}(0, x, v) \right]\ .$$

* Letting $\lambda \to 0^+$, the right-hand side becomes $\frac{\partial^2 \sigma^i}{\partial t \partial y^j}(0, x, v)$, which equals $\frac{\partial^2 \sigma^i}{\partial y^j \partial t}(0, x, v)$. The latter, upon the use of $\frac{\partial \sigma}{\partial t} = y$, gives $\delta^i{}_j$. Therefore

$$\lim_{\lambda \to 0^+} \frac{\partial \exp^i}{\partial y^j}(x, \lambda v) \ = \ \delta^i{}_j\ .$$

* Since the indicatrix $F(x, v) = 1$ is compact, it can be shown that the above limit is uniform in all v which have norm 1. In other words, for each $\epsilon > 0$, there exists a $\delta > 0$ independent of v [as long as $F(v) = 1$] such that

$$0 < \lambda < \delta \ \Rightarrow \ \left| \frac{\partial \exp^i}{\partial y^j}(x, \lambda v) \ - \ \delta^i{}_j \right| < \epsilon\ .$$

We have just demonstrated that

$$\text{(5.3.7)} \qquad \boxed{\lim_{y \to 0} \frac{\partial \exp^i}{\partial y^j}(x, y) \ = \ \delta^i{}_j \ = \ \frac{\partial \exp^i}{\partial y^j}(x, 0)}\ .$$

This, together with the fact that exp is already C^∞ at all x, gives the asserted C^1 smoothness.

Let us summarize what we have accomplished in this section:

- We have shown that for Finsler manifolds in general, **the exponential map exp is (only) C^1 at the zero section of TM (and C^∞ away from it). Its derivative at the zero section is the identity map.** This is a result of Whitehead's [W].
- In Exercise 5.3.5, the analysis here is extended to derive a result of Akbar-Zadeh's [AZ]. It says that **exp is C^2 at the zero section if and only if the Finsler structure is of Berwald type** (as defined in Exercise 5.3.3). In that case, the exponential map exp is actually C^∞ throughout TM.

Exercises

Exercise 5.3.1:

(a) Show that the system of differential equations which describes geodesics (not necessarily of constant speed) is:

$$\boxed{\frac{d^2\sigma^i}{dt^2} + \frac{d\sigma^j}{dt}\frac{d\sigma^k}{dt}\,(\gamma^i{}_{jk})_{(\sigma,T)} = \frac{d}{dt}[\log F(T)]\,\frac{d\sigma^i}{dt}}\ .$$

Hint: begin with part (c) of Exercise 5.2.4.

(b) Using the theory of ordinary differential equations, show that geodesics (hence C^1, according to Proposition 5.1.1) must in fact be C^∞ in t, in x, and in $y \neq 0$. Here, $\sigma(0) = x$ and $\frac{d\sigma}{dt}(0) = y$ are the initial data of the geodesic.

The formula in part (a) is useful because it is not always to our advantage to parametrize geodesics to have constant Finslerian speed. For example, we show in §11.3 that when a Randers space is of Douglas type, a parametrization with constant Riemannian speed is much more telling.

Exercise 5.3.2: Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a geodesic in a Finsler manifold. Define $\zeta(s) := \sigma(\alpha s + \beta)$, where α, β are constants.

(a) If $\alpha > 0$, show that ζ is again a geodesic.

(b) If $(\Gamma^i{}_{jk})_{(x,-y)} = (\Gamma^i{}_{jk})_{(x,y)}$, show that ζ is a geodesic for all values of α and β.

Exercise 5.3.3: Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a geodesic. Its **reverse** is the curve $\zeta(s) := \sigma(r - s)$. Show that the reverse of σ is again a geodesic if *one* of the following conditions is satisfied:

- The Finsler structure F is absolutely homogeneous. For example, this is the case on Riemannian manifolds.
- In natural coordinates, the Chern connection coefficients $\Gamma^i{}_{jk}$ depend only on x. Manifolds of this type are called **Berwald spaces**.

Exercise 5.3.4: In a locally Minkowskian space, let $\sigma(t)$ be a geodesic that passes through the point x at $t = 0$ with velocity v. Prove that in some local coordinate system, our geodesic has the description $\sigma^i(t) = x^i + t\,v^i$.

Exercise 5.3.5: This exercise characterizes the situation in which exp is C^2 at the zero section of TM. It addresses a result of Akbar-Zadeh's [AZ].

(a) Suppose the exponential map is C^2 at the zero section. Differentiate (5.3.5) with respect to t twice. Show that one gets

$$v^j\,v^k\,\frac{\partial^2 \exp^i}{\partial y^j \partial y^k}(x, \lambda v) = \frac{\partial^2 \sigma^i}{\partial t^2}(\lambda, x, v)\ , \quad \lambda > 0\ .$$

(b) Substitute this into the geodesic equation (5.3.1) and take the limit as $\lambda \to 0^+$. With the help of the C^2 hypothesis, show that

$$y^j\, y^k \left[\frac{\partial^2 \exp^i}{\partial y^j \partial y^k}(x,0) \;+\; ({\Gamma^i}_{jk})_{(x,y)} \right] \;=\; 0\ .$$

(c) Note that since exp is C^2 at $(x, 0)$, the first term inside the brackets is, like the Chern connection coefficients, symmetric in j, k. Differentiate the above with respect to y^l to get

$$2\, y^j \left[\frac{\partial^2 \exp^i}{\partial y^j \partial y^l}(x,0) \;+\; ({\Gamma^i}_{jl})_{(x,y)} \right] \;+\; y^j\, y^k\, \frac{\partial {\Gamma^i}_{jk}}{\partial y^l} \;=\; 0\ .$$

(d) Use (3.3.3) and (3.4.9) to help you show that $y^j\, y^k\, \frac{\partial {\Gamma^i}_{jk}}{\partial y^l}$ is in fact zero. Thus

$$y^j\, ({\Gamma^i}_{jk})_{(x,y)} \;=\; -\, \frac{\partial^2 \exp^i}{\partial y^j \partial y^k}(x,0)\, y^j\ ,$$

which says that the left-hand side depends only linearly on y.

(e) Now use Exercise 3.4.6 to show that ${\Gamma^i}_{jk}$ does not depend on y. That is, the Finsler structure is of Berwald type.

(f) Conversely, check that if ${\Gamma^i}_{jk}$ has no y-dependence, then the exponential map is C^∞ (and in particular, C^2) at the zero section of TM.

5.4 Jacobi Fields

Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a geodesic with velocity field T. To keep things simple, let us assume that σ has been parametrized to have constant speed $F(T) = c$. The defining equation for σ is then

$$D_T T \;=\; 0\ ,$$

the criterion for being an autoparallel.

For the rest of this chapter, all covariant differentiations D_T and D_U are to be carried out with reference vector T.

In Exercise 5.2.7, we have encountered the index form $I(V, W)$, whose arguments V and W are piecewise C^∞ vector fields along σ. Let $0 =: t_o < t_1 < \cdots < t_k := r$ be a partition of $[0, r]$ such that V, W are both C^∞ on each closed subinterval $[t_{s-1}, t_s]$. Using integration by parts, one can re-express the index form as

$$\begin{aligned} I(V,W) \;:=\;& \frac{1}{c}\, g_T(\, D_T V \,,\, W\,) \Big|_0^r \\ & - \frac{1}{c} \sum_{s=1}^{k-1} g_T(\, D_T V \,,\, W\,) \Big|_{t_s^-}^{t_s^+} \\ & - \frac{1}{c} \int_0^r g_T\Big(\, D_T D_T V \;+\; R(V,T)T \;,\; W \,\Big)\, dt \;. \end{aligned}$$

A vector field J along σ is said to be a **Jacobi field** if it satisfies the equation

$$(5.4.1) \qquad \boxed{\, D_T D_T J \;+\; R(J,T)T \;=\; 0 \,}\;.$$

The following abbreviation is customary:

$$J'' \;:=\; D_T D_T J \;.$$

Standard ODE theory tells us that given the initial data

$$J(t_o) \quad \text{and} \quad J'(t_o) \;:=\; (D_T J)(t_o) \;,$$

there exists a unique C^∞ solution $J(t)$ along the geodesic $\sigma(t)$.

Jacobi fields naturally arise through geodesic variations. Namely:

> **Given any piecewise C^∞ variation (*not* necessarily with fixed endpoints) in which *all* the t-curves are geodesics, the variation vector field U is necessarily a Jacobi field.**

A sketch of the proof is as follows:

- In (5.2.1), set $j = n$. With the help of $a_n{}^k(t,u) = 0$, we obtain

$$(*) \qquad \frac{\partial b_n{}^i}{\partial t} \;+\; b_n{}^k\; a_k{}^i \;=\; -\; a^n\; R^i{}_k\; b^k \;.$$

- The left-hand side, when summed with e_i , gives $D_T(\, b_n{}^i\, e_i\,)$. Now, $b_n{}^i\, e_i = b_n{}^\alpha\, e_\alpha$, and it is straightforward to show that

$$b_n{}^\alpha\; e_\alpha \;=\; \frac{1}{F(T)} \left(D_U T \;-\; \frac{\partial \log F(T)}{\partial u}\; T \right) \;.$$

- On the other hand,

$$(\,-\; a^n\; R^i{}_k\; b^k\,)\; e_i \;=\; -\; \frac{1}{F(T)}\; R(U,T)T \;.$$

- Substitute these into (*), and use the fact that
  * $D_T F(T) = 0$ (constant speed),
  * $D_T T = 0$ (all t-curves are geodesics),
  * $D_U T = D_T U$ (Exercise 5.2.1).

We obtain

$$D_T D_T U \ + \ R(U,T)T \ = \ 0 \ ,$$

which says that the variation vector field satisfies the Jacobi equation.

Let us now consider the case of "wedge-shaped" geodesic variations.

- Restrict the exponential map exp to a fixed tangent space $T_x M$, and denote it by $\exp_x$. Take $\exp_x(tT)$, $0 \leqslant t \leqslant r$ as our base geodesic, where $T \in T_x M$. This geodesic emanates from the point x with initial velocity T.
- Let $W^i \frac{\partial}{\partial y^i}$ be a *tangent vector* emanating from the point tT along the straight ray. Since $T_x M$ is a linear manifold, we can slide that tangent vector to the origin, thereby getting an element $W^i \frac{\partial}{\partial x^i}$ of $T_x M$. Let us, by a slight abuse of notation, refer to both as W.
- For conceptual clarity, we only think about those W that are transverse to T. In the Minkowski space $T_x M$, we have the variation $t\,(\,T + u\,W\,)$ of the straight ray $t\,T$. If the parameter u is kept small, say $-\epsilon < u < +\epsilon$, and if r is made smaller if necessary, the definition

(5.4.2) $$\boxed{\sigma(t,u) \ := \ \exp_x[\ t\ (T \ + \ u\ W\ )\ ]}$$

makes sense. It gives a variation by geodesics of our base curve $\exp_x(tT)$, $0 \leqslant t \leqslant r$.
- For each value of u, the corresponding geodesic emanates from x at time $t = 0$, and has constant speed $F(\,x\,,\,T + uW\,)$.

The variation vector field $U(t)$ of this $\sigma(t,u)$ is, through the chain rule, given by

(5.4.3) $$\boxed{U(t) \ = \ \exp_{x*(tT)}(tW)} \ .$$

For Finsler geometry in general, we learned in §5.3 that $\exp_x$ is only C^1 at the origin of $T_x M$. Thus the $U(t)$ here is *a priori* only continuous *from the right* at $t = 0$ and C^∞ on $0 < t \leqslant r$. Because it arises from a variation by geodesics, $U(t)$ is a Jacobi field. Strictly speaking though, it is only one on $(0, r]$, since Jacobi fields are supposed to be C^∞ after all.

In the following lemma, we contend that $U(t)$ is actually C^∞ at $t = 0$. And its initial data are $U(0) = 0$, $U'(0) = W$. Keep in mind that by ODE theory, our base geodesic $\exp_x(tT)$, $0 \leqslant t \leqslant r$ always admits an extension to a (slightly) larger time interval.

Lemma 5.4.1. *Suppose*

* *(M, F) is a Finsler manifold, where F is C^∞ on $TM \setminus 0$ but not necessarily symmetric. That is, F is positively (but perhaps not absolutely) homogeneous of degree one.*

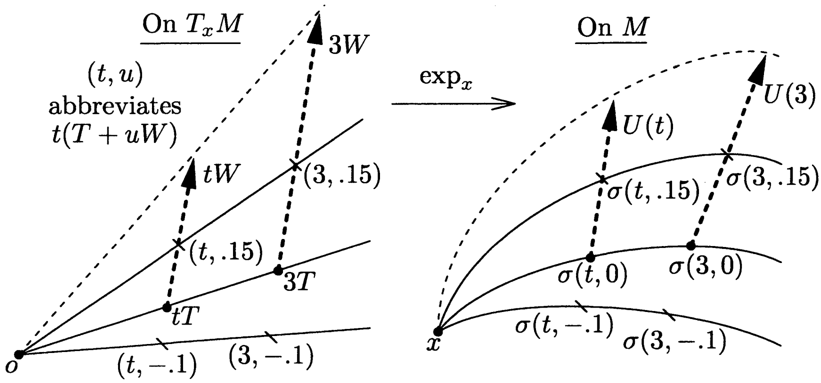

Figure 5.2

The straight ray tT and its variation in T_xM. Exponentiating this gives the variation $\sigma(t,u)$, whose variation vector field is $U(t) = \exp_{x*(tT)}(tW)$. The straight ray, under $\exp_x$, becomes the base geodesic $\sigma(t,0)$.

* *$\gamma(t)$, $-\delta < t < r+\delta$ (with $\delta > 0$) is any extension of the base geodesic $\exp_x(tT)$, $0 \leqslant t \leqslant r$.*
* *W is a fixed (but arbitrary) element of T_xM.*

Then:

* *The Jacobi field $U(t) = \exp_{x*(tT)}(tW)$, $0 < t \leqslant r$ extends uniquely to a C^∞ Jacobi field $J(t)$, $-\delta < t < r+\delta$ (along γ).*
* *J has initial data $J(0) = 0$ and $J'(0) := (D_TJ)(0) = W$.*

Proof.

Fix an arbitrary $t_o \in (0,r)$ and compute $U(t_o)$, $U'(t_o)$. Let $J(t)$, $-\delta < t < r+\delta$ be the unique C^∞ Jacobi field along the extended geodesic $\gamma(t)$, satisfying the initial data $J(t_o) = U(t_o)$, $J'(t_o) = U'(t_o)$.

Since $U(t)$ is C^∞ for $0 < t \leqslant r$, the uniqueness theorem in ODE theory tells us that J must agree with U on $(0,r]$. This agreement also holds at $t = 0$ because

$$J(0) \;=\; \lim_{t\to 0^+} J(t) \;=\; \lim_{t\to 0^+} U(t) \;=\; U(0) \;=\; 0\,.$$

Thus $J(t)$ is the unique C^∞ extension of $U(t)$ to γ.

Abbreviate the velocity field $T(t)$ of γ as T. From the fact that $J(0) = 0$ and $J' := [\, \frac{dJ^i}{dt} + J^j\, T^k\, ({\Gamma^i}_{jk})_{(\gamma,T)}\,]\, \frac{\partial}{\partial x^i}$, we see that

$$J'(0) \;=\; \frac{dJ^i}{dt}(0)\, \frac{\partial}{\partial x^i}_{|x}\,.$$

However,

$$\frac{dJ^i}{dt}(0) \;=\; \lim_{t\to 0+} \frac{J^i(t) - J^i(0)}{t} \;=\; \lim_{t\to 0+} \frac{J^i(t)}{t} \;=\; \lim_{t\to 0+} \frac{U^i(t)}{t}\,.$$

Using (5.4.3), followed by the continuity of $\exp_{x*}$ at the origin o of T_xM, and the fact that it is the identity map there, we have

$$\lim_{t\to 0+} \frac{U(t)}{t} \;=\; \lim_{t\to 0+} \exp_{x*(tT)} W \;=\; \exp_{x*(o)} W \;=\; W\,.$$

So

$$J'(0) \;=\; W\,. \qquad \square$$

In view of this lemma, we can say that:

- The $U(t)$ given by (5.4.3), namely, $\exp_{x*(tT)}(tW)$, is truly a C^∞ Jacobi field on $0 \leqslant t \leqslant r$.
- It has initial data

$$(5.4.4) \qquad \boxed{U(0) \;=\; 0\,, \quad U'(0) \;=\; W}\,.$$

Exercises

Exercise 5.4.1:

(a) We gave an outline for the proof that geodesic variations give rise to Jacobi fields. Fill in all the details.

(b) Does the converse hold?

Exercise 5.4.2: Consider the geodesic ray $\exp_x(tT)$, $t \in [0,r]$ that emanates from x with initial velocity T. ODE theory tells us that there exists a C^∞ extension $\gamma(t)$, $-\delta < t < r+\delta$ (for some $\delta > 0$) of our geodesic ray.

(a) Use the fact that $\exp_x$ is at least C^1 at the origin of T_xM to conclude that $\gamma(t) = \exp_x(tT)$ on $[0, r+\delta)$.

(b) For negative t, show that $\exp_x(tT)$, $-\delta < t < 0$ is the *reverse* of a certain geodesic. Identify the initial velocity of that geodesic. Hint: consider the expression $\exp_x[(-t)(-T)]$.

(c) Explain why, *if* the Finsler structure F is absolutely homogeneous (which includes the Riemannian case) or of Berwald type, then $\gamma(t) = \exp_x(tT)$ on the interval $(-\delta, 0)$.

(d) Explain why $\exp_x(-tT)$, $0 \leqslant t \leqslant r$ is a geodesic for some r. What is its initial velocity?

Exercise 5.4.3: Suppose J is a Jacobi field along the geodesic $\sigma(t)$ with velocity T.

(a) Use Exercise 5.2.3 to show that $[\,g_T(J,T)\,]'' = g_T(J'',T) = 0$.

(b) Explain why $g_T(J',T)$ is constant.

(c) Fix any t_o in the domain of $\sigma(t)$. Show that

$$\boxed{g_T(J,T) \;=\; g_T(\,J(t_o)\,,\,T(t_o)\,) \;+\; (t-t_o)\;g_T(\,J'(t_o)\,,\,T(t_o)\,)}\;.$$

(d) Deduce that J is g_T-orthogonal to σ at two points if and only if it is g_T-orthogonal to σ everywhere.

Exercise 5.4.4: Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a geodesic with velocity T in a Finsler manifold (M,F). Let J be a Jacobi field along σ. Prove that if any one of the following conditions holds,

$$\begin{aligned} J(0) &= 0 = J(r) \\ J'(0) &= 0 = J(r) \\ J(0) &= 0 = J'(r) \end{aligned}$$

then J and $J' = D_T J$ must both be g_T-orthogonal to T at all times.

Exercise 5.4.5: Let J be a Jacobi field along the geodesic $\sigma(t)$ with velocity T. Let c denote the constant speed of σ. Define

$$J_{\|} \;:=\; g_T(J,T)\,\frac{T}{c^2}\,, \qquad J_{\perp} \;:=\; J \;-\; J_{\|}\;.$$

(a) Prove that $J_{\|}$ and $J_{\perp}$ are both Jacobi fields along σ.

(b) If (M,F) has scalar flag curvature $\lambda(x,y)$ (see §3.10), prove that

$$\boxed{J_{\perp}'' \;+\; \lambda\, c^2\, J_{\perp} \;=\; 0}\;.$$

Hint: calculate $g_T(\,V,\,R(J_{\perp},T)T\,)$ for arbitrary V. In this equation and henceforth, we use λ to abbreviate $\lambda(\sigma,T)$.

(c) Suppose $J_{\perp}$ happens to have a special form, say, a function f multiplying a *parallel* (see definition below) vector field E which is g_T-orthogonal to σ. Demonstrate how the equation in (b) can be reduced to a **scalar Jacobi equation**

$$\ddot{f} \;+\; \lambda\, c^2\, f \;=\; 0\,.$$

Here, E is obtained by parallel translation along σ, with reference vector T. It satisfies the equation $D_T E = 0$ with reference T.

(d) Let $\{E_i : i = 1,\ldots,n\}$ be a moving basis of parallel vector fields (again, with reference vector T) along σ. Then $J_{\perp}$ can always be expressed as $J_\perp = f^i E_i$ (with automatic summation on i). Suppose (M, F) has scalar flag curvature $\lambda(x, y)$. Explain why each of the component functions f^i must satisfy the scalar Jacobi equation. Namely,

$$\boxed{\ddot{f}^i + \lambda\, c^2\, f^i = 0} \quad \text{for } i = 1, \dots, n\,.$$

Exercise 5.4.6: Let U, W be Jacobi fields along a geodesic $\sigma(t)$ with velocity T. Prove that

$$\boxed{g_T(U', W) - g_T(U, W') = \text{some constant}}\,.$$

This is known as the **Lagrange identity**.

5.5 How the Flag Curvature's Sign Influences Geodesic Rays

The purpose of this section is simple. Imagine a family of geodesic rays emanating from the point x. We would like to show that:

- **If the flag curvature is positive at x, then these geodesic rays will appear to "bunch together."**
- **If the flag curvature at x is negative, then these geodesic rays will appear to "disperse."**

But in order to prove these statements, we must first make precise what we mean by "bunching together" and "dispersing."

Let us begin with the setup discussed in the last section. Take $T, W \in T_xM$, where W is transverse to T. In order to simplify things, let us choose $T \in T_xM$ to have Finsler norm $F(T) = 1$, and use $g_{ij\,(x,T)}\, dx^i \otimes dx^j$ to ferret out those $W \in T_xM$ that have length one and are orthogonal to T. Any such W would be a legitimate choice. Using T and W, we generate a family of geodesics that emanate from x at time $t = 0$. This family is indexed by a parameter $u \in (-\epsilon, +\epsilon)$, and the "$u$th" member is $\exp_x[\,t\,(T + u\,W\,)\,]$, $t \in [0, r]$. These curves are viewed as variations of the base geodesic $\sigma(t) := \exp_x(tT)$, whose velocity field we abbreviate as T.

The length squared

$$\text{(5.5.1)} \qquad \boxed{\|\, U(t)\,\|^2 := g_T(\,U(t)\,,\, U(t)\,)}$$

of the variation vector field $U(t) = \exp_{x*(tT)}(tW)$ measures the rate (in units of length squared per unit change in u) with which the "uth" geodesic is deviating from $\sigma(t)$. If this function of t climbs to a maximum and then decreases, the geodesics in question must be showing a focusing behavior. On the other hand, if this function is monotonically increasing, then the geodesics must be diverging. This type of detailed information, however, is beyond the technical reach of the present chapter. What we do in this section is the next best thing.

Consider the situation in the tangent space T_xM. There, the ray tT is the analogue of our base geodesic σ. The map $(t,u) \mapsto t(T + u\,W)$, with $u \in (-\epsilon, +\epsilon)$ and $t \in [0,r]$, is the variation of the ray tT. One can see by inspection that the corresponding variation vector field has the value tW at time t along the said ray. With respect to the inner product $g_{ij\,(x,T)}\, dx^i \otimes dx^j$ on T_xM, the length squared of this variation vector field is simply t^2 because the length of W was stipulated to be 1. This t^2 describes a rate with which the "uth" ray is deviating from the base ray tT.

We now have two variation vector fields at hand. The one on T_xM is tW, and the one on M is $U(t)$. Let us compare the Riemannian length squared $\| U(t) \|^2$ with that of tW, which is t^2.

* If $\| U(t) \|^2 < t^2$, we say that the geodesic rays emanating from x on M are **bunching together.**
* If $\| U(t) \|^2 > t^2$, we say that the geodesic rays emanating from x on M are **dispersing.**

This comparison need only be made for small t, which means that we can get by with the first few terms in the power series expansion of $\| U(t) \|^2$. Information which is valid on long time intervals can also be obtained, but that requires a more delicate comparison technique. See Chapter 9.

Let us calculate (as in [CE]) the first four t derivatives of the function $\| U(t) \|^2$. To reduce clutter, we abbreviate that as $f(t)$ and suppress (whenever possible) the t dependence on $U(t)$ and its successive covariant derivatives. Since the base curve $\sigma(t)$ along which $U(t)$ is defined is a geodesic with velocity T, and since the inner product in question is g_T, we can apply Exercise 5.2.3 repeatedly to get

$$\begin{aligned}
f^{(1)}(t) &= 2\, g_T(U, U^{(1)})\,, \\
f^{(2)}(t) &= 2\, g_T(U^{(1)}, U^{(1)}) + 2\, g_T(U, U^{(2)})\,, \\
f^{(3)}(t) &= 6\, g_T(U^{(2)}, U^{(1)}) + 2\, g_T(U, U^{(3)})\,, \\
f^{(4)}(t) &= 8\, g_T(U^{(3)}, U^{(1)}) + 6\, g_T(U^{(2)}, U^{(2)}) \\
&\quad + 2\, g_T(U, U^{(4)})\,,
\end{aligned}$$

where

$$\begin{aligned}
U^{(1)} &:= D_T U\,, \\
U^{(2)} &:= D_T D_T U = -\,R(U,T)T\,, \\
U^{(3)} &:= D_T D_T D_T U = -\,[D_T R](U,T)T - R(U',T)T\,, \\
U^{(4)} &:= D_T D_T D_T D_T U\,.
\end{aligned}$$

All covariant derivatives here are taken with reference vector T.

Now evaluate the derivatives of f at $t = 0$ with the help of (5.4.4). Note that we do not need $U^{(4)}(0)$ because, in $f^{(4)}(0)$, it is paired with $U(0)$ which happens to be zero. Anyway:

$$\begin{aligned} U(0) &= 0 , \\ U^{(1)}(0) &= U'(0) = W \in T_xM , \\ U^{(2)}(0) &= 0 , \\ U^{(3)}(0) &= - R(W,T)T . \end{aligned}$$

Therefore

$$\begin{aligned} f(0) &= 0 , \\ f^{(1)}(0) &= 0 , \\ f^{(2)}(0) &= 2 \, g_T(W,W) , \\ f^{(3)}(0) &= 0 , \\ f^{(4)}(0) &= - 8 \, g_T(R(W,T)T , W) . \end{aligned}$$

So we have the following power series expansion in t:

$$(5.5.2) \quad \| U(t) \|^2 = g_T(W,W) \, t^2 - \frac{1}{3} \, g_T(R(W,T)T , W) \, t^4 + O(t^5) .$$

The curvature operator was introduced in Exercise 5.2.5. Definition chasing tells us that $g_T(R(W,T)T , W)$ is equal to

$$K(T,W) \, \{ \, g_T(T,T) \, g_T(W,W) - [g_T(T,W)]^2 \, \} ,$$

where $K(T,W)$ is a flag curvature. Recall that at the outset, we have normalized $T \in T_xM$ to have Finsler norm $F(T) = 1$, and have used $g_{ij\,(x,T)} \, dx^i \otimes dx^j$ to choose only those $W \in T_xM$ that have length one and are orthogonal to T. Therefore (5.5.2) simplifies to

$$(5.5.3) \quad \boxed{\| U(t) \|^2 = t^2 - \frac{1}{3} \, K(T,W) \, t^4 + O(t^5)} \, .$$

Note that:

* If the flag curvature at x is positive, then $\| U(t) \|^2 < t^2$ for small t. Thus geodesic rays emanating from x are spreading apart more slowly than the corresponding rays in T_xM. In other words, the said geodesic rays are bunching together.
* If the flag curvature at x is negative, then $\| U(t) \|^2 > t^2$ for small t. Thus geodesic rays emanating from x are spreading apart more quickly than the corresponding rays in T_xM. In other words, the said geodesic rays are dispersing.

These conclusions are illustrated in Figure 5.3.

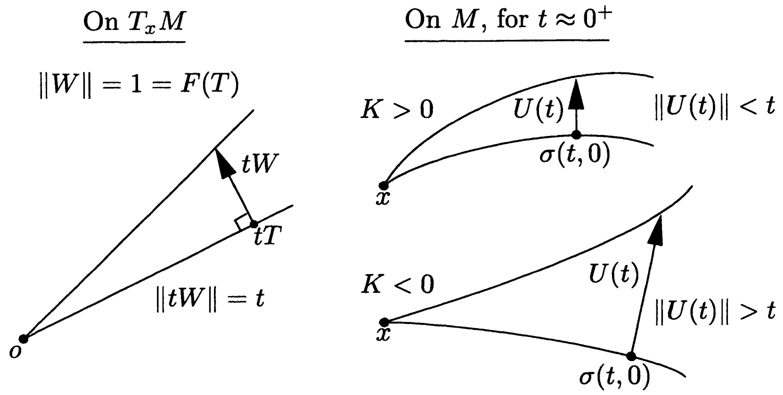

Figure 5.3

The straight ray tT and its variation $t(T + uW)$ in T_xM. Exponentiating this gives the geodesic ray $\sigma(t) := \exp_x(tT)$ and its variation $\exp_x[t(T + uW)]$ on M. The two are compared by examining the length squared of the respective variation vector fields. If the flag curvature at x is positive, then geodesic rays emanating from x appear to bunch together. If the flag curvature at x is negative, then geodesic rays emanating from x appear to disperse. This "appearance" is gauged by comparing the variation vector field $U(t)$ with its generator tW. Their Riemannian lengths squared are, respectively, $\| U(t) \|^2$ and t^2.

References

[AZ] H. Akbar-Zadeh, *Sur les espaces de Finsler à courbures sectionnelles constantes*, Acad. Roy. Belg. Bull. Cl. Sci. (5) **74** (1988), 281–322.

[BC1] D. Bao and S. S. Chern, *On a notable connection in Finsler geometry*, Houston J. Math. **19** (1993), 135–180.

[CE] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, North Holland/American Elsevier, 1975.

[Daz] P. Dazord, *Propriétés globales des géodésiques des Espaces de Finsler*, Theses, Université de Lyon, 1969.

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

[W] J. H. C. Whitehead, *Convex regions in the geometry of paths*, Quarterly J. Math. Oxford, Ser. 3 (1932), 33–42.

Chapter 6

The Gauss Lemma and the Hopf–Rinow Theorem

6.1 The Gauss Lemma

Fix $x \in M$. In T_xM, we define the **tangent spheres**

(6.1.1) $$\mathrm{S}_x(r) := \{ y \in T_xM : F(x,y) = r \}$$

and open **tangent balls**

(6.1.2) $$\mathrm{B}_x(r) := \{ y \in T_xM : F(x,y) < r \}$$

of radii r. The exponential map $\exp_x$ is a local diffeomorphism at the origin of T_xM because its derivative there is the identity; see §5.3. Thus, for r

small enough, not only does $\exp_x[\,\mathrm{S}_x(r)\,]$ makes sense, it is also diffeomorphic to $\mathrm{S}_x(r)$. The image set

$$\exp_x[\ \mathrm{S}_x(r)\ ]$$

is called a **geodesic sphere** in M centered at x. We later show why it can be said to have radius equal to r.

The punctured tangent space $T_xM \smallsetminus 0$ is a Riemannian manifold with metric

$$\hat{g}_x \ := \ g_{ij\,(x,y)}\ dy^i \otimes dy^j \ .$$

Fix $y \in T_xM \smallsetminus 0$; say $F(x,y) = r$.

- Since F is constant on $\mathrm{S}_x(r)$, we have

  $$0 \ = \ w^i\ F_{y^i} \ = \ w^i\ g_{ij}\ \frac{y^j}{r}$$

  for any $w^i \frac{\partial}{\partial y^i}$ tangent to $\mathrm{S}_x(r)$. In particular, any constant multiple of $y^i \frac{\partial}{\partial y^i}$ is $\hat{g}_x$-orthogonal to $\mathrm{S}_x(r)$.
- Consider the ray ty, $t \geqslant 0$. What we have just said, together with a moment's thought, implies that at any instant $t > 0$, the (ray's) velocity $y^i \frac{\partial}{\partial y^i}$ is always $\hat{g}_x$-orthogonal to the tangent sphere $\mathrm{S}_x(tr)$ that the ray has just pierced. We record that symbolically as

  $$\boxed{\ ty \ \perp \ \mathrm{S}_x(tr) \quad \text{with respect to } \ \hat{g}_x\ } \ .$$

6.1 A. The Gauss Lemma Proper

Apply $\exp_x$ to the above setup. Namely, for r small, we take $y \in \mathrm{S}_x(r)$ and exponentiate the ray segment ty, $0 \leqslant t \leqslant 1$. The resulting radial geodesic $\exp_x(ty)$, $0 \leqslant t \leqslant 1$ has velocity field $T(t)$, with $T(0) = y$. It intersects all geodesic spheres of radii not exceeding r and centered at x. The fact that T is g_T-orthogonal to all the geodesic spheres that $\exp_x(ty)$ pierces is the essence of the following lemma.

Lemma 6.1.1 (The Gauss Lemma).

- *Fix $y \in T_xM \smallsetminus 0$ and set $r := F(x,y)$. Suppose $\exp_x y$ is defined.*
- *Let T denote the velocity field of the geodesic $\exp_x(ty)$, $0 \leqslant t \leqslant 1$ emanating from x. Call this a* **radial geodesic**.

  Technical statement: *Fix any instant $t \in [0,1]$. Take any vector V in the tangent space of $\mathrm{S}_x(tr)$. Then, at the time t location of our radial geodesic, we have the orthogonality relation:*

  $$\boxed{\ g_T(\exp_{x*}V\,,T) \ = \ 0\ } \ .$$

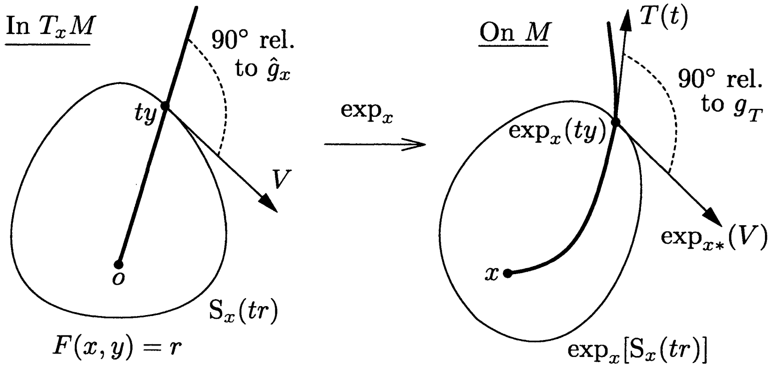

Figure 6.1

For r small, take $y \in S_x(r)$. With respect to the "punctured" Riemannian metric $\hat{g}_x$, the ray ty intersects the tangent sphere $S_x(tr)$ at right angles. Next, exponentiate the ray segment ty, $0 \leqslant t \leqslant 1$. The resulting radial geodesic $\exp_x(ty)$, $0 \leqslant t \leqslant 1$ has velocity field $T(t)$, with $T(0) = y$. It is g_T-orthogonal to all the geodesic spheres of radii not exceeding r and centered at x.

Geometric statement: *Suppose y is short enough that it belongs to some open tangent ball on which the exponential map is a diffeomorphism. Then, at each instant t, our radial geodesic (or more precisely its velocity) is g_T-orthogonal to the geodesic sphere $\exp_x[\,S_x(tr)\,]$ that it is piercing. Symbolically:*

$$\boxed{\exp_x(ty) \;\perp\; \exp_x[\,S_x(tr)\,] \quad \text{with respect to } g_T} \,.$$

Remarks:

* Recall that along any curve σ with velocity field T,

$$g_T \;:=\; g_{ij\,(\sigma,T)}\; dx^i \otimes dx^j \;.$$

* In the tangent space, we have deduced that the ray ty is always $\hat{g}_x$-orthogonal to the tangent spheres $S_x(tr)$. According to this lemma, the exponential image of this ray is g_T-orthogonal to the geodesic spheres $\exp_x[\,S_x(tr)\,]$. This phenomenon is a bit of a surprise because the exponential map is *not* known to be a local isometry, even in the Riemannian setting.

Proof.

For the technical statement:

Fix an instant $\tau \in [0,1]$. Then $\tau y \in \mathrm{S}_x(\tau r)$. Let $c(u)$, $-\epsilon < u < +\epsilon$ be any curve on $\mathrm{S}_x(\tau r)$ that passes through the point τy at $u = 0$. Denote its velocity at that moment by V. Consider the rectangle $\{(t,u) : 0 \leqslant t \leqslant 1, -\epsilon < u < +\epsilon\}$. Define a variation of the geodesic $\sigma(t) := \exp_x(t\tau y)$, $0 \leqslant t \leqslant 1$ by

$$\sigma(t,u) := \exp_x[\, t\, c(u)\,] .$$

* Every t-curve in this variation is a geodesic which emanates from x and has constant speed τr. Thus their lengths are equal and we have $L'(0) = 0$.
* The variation vector field U has $U(0) = 0$, $U(1) = \exp_{x*} V$, and the latter is tangent to the geodesic sphere $\exp_x[\, \mathrm{S}_x(\tau r)\,]$.
* Since the velocity of $\exp_x(ty)$ is T, that of $\sigma(t) := \exp_x(t\tau y)$ must be τT.

Exercise 5.1.4 deals with a special case of the first variation of arc length. The variation we have just described is of that type. Inputting the above observations leads to $0 = g_{_T}(\exp_{x*} V \,,\, \tau T\,)$, where we have used the fact that $g_{(\sigma,\tau T)} = g_{(\sigma,T)}$. Hence T is $g_{_T}$-orthogonal to $\exp_{x*} V$, as asserted.

For the geometric statement:

When r is sufficiently small as described, $\exp_x$ is a diffeomorphism from each $\mathrm{S}_x(tr)$ onto its image. Thus, every tangent vector on the geodesic sphere $\exp_x[\, \mathrm{S}_x(tr)\,]$ can be expressed as $\exp_{x*} V$ for some V that is tangent to $\mathrm{S}_x(tr)$. The conclusion now follows from the technical statement we have just established. □

6.1 B. An Alternative Form of the Lemma

The Gauss lemma can be recast into a form that *appears* more general. As before, fix $y \in T_x M \smallsetminus 0$ and suppose $\exp_x y$ is defined. Fix any instant $t \in [0,1]$. Take any vector W in the tangent space $T_{ty}(T_x M \smallsetminus 0)$. Note that the velocity of the ray ty is y, which also belongs to $T_{ty}(T_x M \smallsetminus 0)$. Then:

$$\text{(6.1.3)} \qquad \boxed{\, g_{_T}\Big((\exp_{x*})_{(ty)} W \,,\, T(t) \Big) \;=\; \hat{g}_{x\,(ty)}(W, y) \,} \,.$$

Both $\exp_{x*}$ and $\hat{g}_x$ are evaluated at the point ty. But we suppress this dependence whenever possible, in order to reduce clutter.

To derive (6.1.3), resolve W into $V + \alpha y$, where V belongs to the tangent space of $\mathrm{S}_x(tr)$, and αy is along the straight ray. Since T is the velocity field of $\exp_x(ty)$, we have $(\exp_{x*})_{(ty)} y = T(t)$ and $T(0) = y$. Thus

$$\begin{aligned} g_{_T}(\,\exp_{x*} \alpha y \,,\, T\,) &= \alpha\; g_{_T}(\,T(t) \,,\, T(t)\,) \\ \text{(6.1.4)} \qquad &= \alpha\; g_{_T}(\,T(0) \,,\, T(0)\,) \;=\; \hat{g}_x(\alpha y, y) \,. \end{aligned}$$

Also, the Gauss lemma says that

$$g_T(\exp_{x*}V\,,\,T) \;=\; 0 \;=\; \hat{g}_x(V,y)\;.$$

Adding the last two statements produces (6.1.3).

Let us digress to point out a subtlety in the last step of (6.1.4).

* In $g_T(T(0)\,,\,T(0)\,)$, that $T(0)$ refers to the initial velocity y of our geodesic $\exp_x(ty)$. This y is a point in the punctured manifold $T_xM \setminus 0$. Thus $g_T = g_{ij\,(x,y)}\,dx^i \otimes dx^j$ and $y = y^i \frac{\partial}{\partial x^i}$.
* In $\hat{g}_x(y,y)$, that y refers to the time t velocity of the straight ray ty. As such, it is given by the tangent vector $y^i \frac{\partial}{\partial y^i}$ which emanates from the point ty. On the other hand, $\hat{g}_x$ means $g_{ij\,(x,ty)}\,dy^i \otimes dy^j$, which is numerically the same as $g_{ij\,(x,y)}\,dy^i \otimes dy^j$, and the y here denotes the point $y^i \frac{\partial}{\partial x^i}$ of $T_xM \setminus 0$.

6.1 C. Is the Exponential Map Ever a Local Isometry?

Now take any two vectors, say W_1 and W_2, from the same tangent space $T_{ty}(T_xM \setminus 0)$. One might wonder whether, for $0 < t \leqslant 1$,

$$(6.1.5) \qquad g_T\Big(\,(\exp_{x*})_{(ty)}W_1\,,\,(\exp_{x*})_{(ty)}W_2\,\Big) \;=\; \hat{g}_{x\,(ty)}(W_1,W_2) \quad ??$$

To analyze this, decompose W_1 as $V_1{+}\alpha_1 y$, and W_2 as $V_2{+}\alpha_2 y$, where V_1, V_2 are tangent to $\mathrm{S}_x(tr)$. With the help of the Gauss lemma and a calculation very much like (6.1.4), it can be checked that (6.1.5) is equivalent to the following question: for $0 < t \leqslant 1$,

$$(6.1.6) \qquad g_T\Big(\,(\exp_{x*})_{(ty)}V_1\,,\,(\exp_{x*})_{(ty)}V_2\,\Big) \;=\; \hat{g}_{x\,(ty)}(V_1,V_2) \quad ??$$

The more general (in appearance only) statement of the Gauss lemma, namely (6.1.3), evidently prompted these questions. One can check that:

* In the Riemannian setting: (6.1.5) makes sense and holds at $t = 0$; if it also holds for $0 < t \leqslant 1$ and all y, then the exponential map is a local isometry between T_xM and M.
* In the non-Riemannian case, even if (6.1.5) were to hold at all y and all $0 < t \leqslant 1$, the exponential map would still fail to be a local isometry between $T_xM \setminus 0$ and $M \setminus x$. See Exercise 6.1.4.

It turns out that for Finsler manifolds of constant flag curvature λ, and tangent vectors V_1, V_2 which are $\hat{g}_x$-orthogonal to the straight ray ty in $T_xM \setminus 0$, one has, for $0 < t \leqslant 1$:

(6.1.7)

$$\boxed{\,t^2\, g_T\Big(\,(\exp_{x*})_{(ty)}V_1\,,\,(\exp_{x*})_{(ty)}V_2\,\Big) \;=\; \mathfrak{s}^2_{\lambda r^2}(t)\;\hat{g}_{x\,(ty)}(V_1,V_2)\,}\;.$$

Here, $r := F(x,y)$ is the constant speed of the geodesic ray $\exp_x(ty)$, $0 \leqslant t \leqslant 1$, and

$$\mathfrak{s}_{\lambda r^2}(t) := \left\{ \begin{array}{c} \dfrac{1}{\sqrt{\lambda}\, r} \sin(\sqrt{\lambda}\, r\, t) \\ t \\ \dfrac{1}{\sqrt{-\lambda}\, r} \sinh(\sqrt{-\lambda}\, r\, t) \end{array} \right\} \text{ resp., for } \left\{ \begin{array}{c} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \end{array} \right\} .$$

In particular, for the $\lambda = 0$ case, (6.1.6) holds. The derivation of (6.1.7) is carried out in the guided Exercises 6.1.5 and 6.1.6.

Exercises

Exercise 6.1.1:

(a) Explain why for small r, $\exp_p$ is a C^1-diffeomorphism from $\mathrm{B}_p(r)$ onto $\exp_p[\,\mathrm{B}_p(r)\,]$.

(b) What can you say about the Finsler structure if the said diffeomorphism is of class $C^{k>1}$?

Exercise 6.1.2: Suppose $F(x,y) = r$ and the exponential map $\exp_x$ is a diffeomorphism on some open tangent ball containing y. Let $V \in T_y(T_xM)$. Explain why the following two statements are equivalent:

(a) V is tangent to $\mathrm{S}_x(r)$.

(b) $\exp_{x*}V$ is tangent to the geodesic sphere $\exp_x[\,\mathrm{S}_x(r)\,]$.

Exercise 6.1.3: Use the Gauss lemma to help show that (6.1.5) and (6.1.6) are equivalent questions.

Exercise 6.1.4:

(a) Suppose our Finsler structure is Riemannian. Explain why, if (6.1.5) holds, then the exponential map $\exp_x$ is a local isometry.

(b) Suppose our Finsler structure is non-Riemannian. Explain why, even if (6.1.5) holds, the exponential map $\exp_x$ is still typically not a local Finslerian isometry between $T_xM \setminus 0$ and $M \setminus x$. Hint: is the norm induced by g_T the same as the Finslerian norm?

Exercise 6.1.5: Fix a nonzero $y \in T_xM$ at which $\exp_x y$ is defined. Set $r := F(x,y)$. This fixed y singles out an inner product $g_{ij\,(x,y)}\, dx^i \otimes dx^j$ on T_xM. Use it to choose and fix any $V := V^i \frac{\partial}{\partial x^i}$ that is **orthogonal to** y.

(a) At any instant t along the straight ray ty, construct the tangent vector $V^i \frac{\partial}{\partial y^i}$ and denote it also as V. For any $t \neq 0$, check that V is $\hat{g}_x$-orthogonal to the straight ray. What are the initial data of the Jacobi field $\exp_{x*(ty)}(tV)$?

(b) Let T denote the velocity field of the geodesic ray $\exp_x(ty)$, $0 \leqslant t \leqslant 1$. It has constant speed $r := F(x,y)$. Use Exercise 5.4.3 to help show that our Jacobi field is everywhere g_T-orthogonal to the said geodesic ray.

(c) Now suppose that our Finsler manifold has constant flag curvature λ. Parallel transport $V := V^i \frac{\partial}{\partial x^i}$, with direction vector T, to generate a vector field $V(t)$ along $\exp_x(ty)$. Use Exercise 5.2.3 to check that $V(t)$ is everywhere g_T-orthogonal to the geodesic ray. Then use part (b) of Exercise 5.4.5 to show that $\mathfrak{s}_{\lambda r^2}(t)\, V(t)$ is a Jacobi field. Identify its initial data.

(d) Explain why the two Jacobi fields discussed in parts (a) and (c) are equal. That is, on a Finsler manifold of constant flag curvature λ, we have

$$\boxed{\exp_{x*(ty)}(tV) \ = \ \mathfrak{s}_{\lambda r^2}(t)\ V(t)}\ .$$

Here, V is orthogonal to y in the sense that $g_{ij\,(x,y)}\, V^i y^j = 0$.

Exercise 6.1.6: We now derive (6.1.7) for Finsler spaces of constant flag curvature λ. In Exercise 6.1.5, choose and fix any two V, say V_1 and V_2. Using part (d) of Exercise 6.1.5, and then Exercise 5.2.3, show that

$$t^2\, g_T\Big(\, (\exp_{x*})_{(ty)} V_1 \,,\, (\exp_{x*})_{(ty)} V_2 \Big) \ = \ \mathfrak{s}^2_{\lambda r^2}(t)\ g_T\Big(\, V_1(0) \,,\, V_2(0) \Big)\ .$$

Now check that

$$g_T\Big(\, V_1(0) \,,\, V_2(0) \Big) \ = \ g_{ij\,(x,y)}\ V_1^i\ V_2^j \ = \ \hat{g}_{x\,(ty)}(V_1, V_2)\ .$$

In the term involving $\hat{g}_x$, V_1 means $V_1^i \frac{\partial}{\partial y^i}$ and likewise for V_2.

6.2 Finsler Manifolds and Metric Spaces

Let (M,F) be a Finsler manifold, where F is positively homogeneous of degree 1. Let $\sigma : [a,b] \to M$ be a piecewise C^∞ curve with velocity $\frac{d\sigma}{dt} = \frac{d\sigma^i}{dt} \frac{\partial}{\partial x^i} \in T_{\sigma(t)}M$. Its **integral length** $L(\sigma)$ is defined as

$$L(\sigma) \ := \ \int_a^b F\Big(\sigma \,,\, \frac{d\sigma}{dt}\Big)\ dt\ .$$

For $x_o, x_1 \in M$, denote by $\Gamma(x_o, x_1)$ the collection of all piecewise C^∞ curves $\sigma : [a,b] \to M$ with $c(a) = x_o$ and $c(b) = x_1$. Define a map $d : M \times M \to [0,\infty)$ by

$$d(x_o, x_1) \ := \ \inf_{\Gamma(x_o,x_1)} \ L(\sigma)\ .$$

It can be shown that (M,d) satisfies the first two axioms of a metric space. Namely:

(1) $d(x_o, x_1) \geqslant 0$, where equality holds if and only if $x_o = x_1$.

(2) $d(x_o, x_2) \ \leqslant \ d(x_o, x_1) + d(x_1, x_2)$.

If the Finsler structure F is absolutely homogeneous, then one also has

(3) $d(x_o, x_1) = d(x_1, x_o)$.

In that case, (M, d) is a genuine metric space. We emphasize that generically, the distance function d on a Finsler manifold does *not* have the symmetry property (3).

6.2 A. A Useful Technical Lemma

In order to establish (1), we need the following technical lemma.

Lemma 6.2.1. *Let (M, F) be a Finsler manifold. At every point $p \in M$, there exists a local coordinate system $\varphi : \overline{U} \longrightarrow \mathbb{R}^n$ that has the following properties:*

- *The closure of U is compact, $\varphi(p) = 0$, and φ maps U diffeomorphically onto an open ball of $\mathbb{R}^n$.*
- *There is a constant $c > 1$ such that*

$$\boxed{\frac{1}{c}\,|\,y\,| \ \leqslant \ F(y) \ \leqslant \ c\,|\,y\,|} \quad \text{and} \quad \boxed{F(-y) \ \leqslant \ c^2\, F(y)}$$

for all $y = y^i \frac{\partial}{\partial x^i} \in T_x M$ and $x \in \overline{U}$. Here, $|\,y\,| := \sqrt{\delta_{ij}\ y^i\, y^j}$.

- *Given any $x_o,\ x_1 \in U$, we have*

$$\boxed{\frac{1}{c}\,|\,\varphi(x_1) - \varphi(x_o)\,| \ \leqslant \ d(x_o, x_1) \ \leqslant \ c\,|\,\varphi(x_1) - \varphi(x_o)\,|} \ .$$

- *For every pair of points $x_o,\ x_1 \in U$, we have*

$$\boxed{\frac{1}{c^2}\ d(x_1, x_o) \ \leqslant \ d(x_o, x_1) \ \leqslant \ c^2\ d(x_1, x_o)} \ .$$

Proof. Take any local coordinate system $\varphi : W \longrightarrow \mathbb{R}^n$ at p with $\varphi(p) = 0$. Let $\mathbb{B}^n(r)$ denote the standard open ball $\{(v^i) \in \mathbb{R}^n : \sqrt{\delta_{ij}\ v^i\, v^j} < r\}$. Take any $r > 0$ such that $\overline{\mathbb{B}^n(r)} \subset \varphi(W)$. Its corresponding inverse image $\varphi^{-1}[\mathbb{B}^n(r)]$ is a precompact U that satisfies the first claim of our lemma.

For each x in the closure (in the manifold topology) $\overline{U}$, consider the tangent space $T_x M$. On this finite-dimensional vector space, we have the Minkowski norm $y \mapsto F(x, y)$. It is positive-definite, continuous, and positively homogeneous of degree one. Using the coordinate basis $\{\frac{\partial}{\partial x^i}\}$, we also obtain an x-dependent Euclidean norm $y \mapsto |y|$. The quotient $F(y)/|y|$ is well defined for $y \neq 0$, and is invariant under positive rescaling.

Consider the portion of the indicatrix bundle over $\overline{U}$. It is a compact set because $\overline{U}$ and the indicatrix $\{y \in T_xM : F(x,y) = 1\}$ are both compact. The said quotient is a positive continuous function on this compact set. So its absolute minimum $\mathfrak{m}$ and absolute maximum $\mathcal{M}$ exist and are both positive. Now choose $c > 1$ such that $\frac{1}{c} < \mathfrak{m} \leqslant \mathcal{M} < c$. In particular, for *nonzero* elements y in T_xM, with $x \in \overline{U}$, one has $\frac{1}{c} < \frac{F(y)}{|y|} < c$. Equivalently, for *all* $y \in T_xM$, $x \in \overline{U}$, we have

$$(*) \qquad \boxed{\frac{1}{c}\,|\,y\,| \ \leqslant\ F(y) \ \leqslant\ c\,|\,y\,|}\,,$$

which immediately gives $F(-y) \leqslant c^2\, F(y)$. The second conclusion of our lemma has thus been established.

Before tackling the third statement, we digress to make two technical observations.

* Observation (I):

For any $x_o, x_1 \in U$, one has $d(x_o, x_1) \leqslant c\,|\varphi(x_1) - \varphi(x_o)|$.

Indeed, let γ be the path in U whose image under φ is the line segment joining $\varphi(x_o)$ and $\varphi(x_1)$. Abbreviate the velocity field of γ as γ'. Then

$$\begin{aligned} d(x_o, x_1) \ \leqslant\ L_F(\gamma) \ &=\ \int_0^1 F(\gamma')\ dt \\ &\leqslant\ c \int_0^1 |\gamma'|\ dt \ =\ c\,|\,\varphi(x_1) - \varphi(x_o)\,|\ , \end{aligned}$$

where the second inequality follows from (*).

* Observation (II):

Let $r_o := \frac{r}{5c^2}$, $\epsilon_o := \frac{r}{5c}$, and $U_o := \varphi^{-1}[\mathbb{B}^n(r_o)] \subset U$. Take any two points $x_o, x_1 \in U_o$. Let $\gamma : [0,1] \to M$ be a piecewise C^∞ curve with $\gamma(0) = x_o$, $\gamma(1) = x_1$. Suppose $L_F(\gamma) \leqslant d(x_o, x_1) + \epsilon_o$; then the curve γ is necessarily contained in U.

This can be deduced as follows. Note that by Observation (I), $d(x_o, x_1) \leqslant \frac{2r}{5c}$ for all x_o, $x_1 \in U_o$. Thus the described curve γ must satisfy

$$(**) \qquad L_F(\gamma) \ \leqslant\ \frac{3r}{5c}\ .$$

Suppose γ is not contained in U. Let $0 < t_o < 1$ be the first instant γ reaches the boundary ∂U, say, at the point $q := \gamma(t_o)$. Observe that $|\varphi(q)| = r$. Then:

$$
\begin{aligned}
L_F(\gamma) \geqslant L_F(\gamma_{|[0,t_o]}) &= \int_0^{t_o} F(\gamma')\, dt \\
&\geqslant \frac{1}{c} \int_0^{t_o} |\gamma'|\, dt \\
&\geqslant \frac{1}{c}\, |\, \varphi(q) - \varphi(x_o) \,| \\
&\geqslant \frac{1}{c}\, (r - r_o) \,,
\end{aligned}
$$

where the second inequality follows from (*). But this lower bound equals $\frac{(5c^2-1)r}{5c^3}$ which, since $c > 1$, is $\geqslant \frac{4r}{5c}$. A comparison with (**) produces a contradiction. So γ must in fact be contained in U.

We are now ready to examine the third assertion of the lemma. In view of the first technical observation, we only need to derive the lower bound on $d(x_o, x_1)$, where x_o, $x_1 \in U_o$. This is because the first two conclusions of our lemma remain valid when restricted to the smaller open subset U_o. So, at the end, we simply replace the old U by this U_o.

By the definition of metric distance, given any $0 < \epsilon < \epsilon_o$, the two points x_o, x_1 can be joined by a piecewise C^∞ curve $\gamma : [0,1] \to M$ with integral length

$$L_F(\gamma) \leqslant d(x_o, x_1) + \epsilon \,.$$

Our second technical observation assures us that this γ must lie in U. A calculation similar to the one above then yields

$$L_F(\gamma) \geqslant \frac{1}{c}\, |\, \varphi(x_1) - \varphi(x_o) \,| \,.$$

Hence

$$\frac{1}{c}\, |\, \varphi(x_1) - \varphi(x_o) \,| \leqslant d(x_o, x_1) + \epsilon \,.$$

Letting $\epsilon \to 0$ gives the desired estimate.

The final assertion now follows from what we have just shown, through the symmetry of the Euclidean norm. Specifically,

$$|\, \varphi(x_1) - \varphi(x_o) \,| = |\, \varphi(x_o) - \varphi(x_1) \,| \,. \qquad \square$$

One of the consequences of this lemma is that the topology defined by the distance d is equivalent to the original manifold topology of M. See §6.2C.

6.2 B. Forward Metric Balls and Metric Spheres

We defined the notion of a tangent ball $\mathrm{B}_p(r)$ and a tangent sphere $\mathrm{S}_p(r)$ in the beginning of §6.1. These objects have radii r and center p, and live in the tangent space $T_x M$. Their counterparts on M are the **forward metric balls** $\mathcal{B}_p^+(r)$ and the **forward metric spheres** $\mathcal{S}_p^+(r)$:

$$\mathcal{B}_p^+(r) \ := \ \{ x \in M : \ d(p,x) \ < \ r \} \ , \tag{6.2.1}$$

$$\mathcal{S}_p^+(r) \ := \ \{ x \in M : \ d(p,x) \ = \ r \} \ . \tag{6.2.2}$$

They are said to have center p. As we show in §6.3, if r is small, $\mathcal{B}_p^+(r)$ and $\mathcal{S}_p^+(r)$ have metric radii r.

* The metric distance function is in general *not* symmetric in its two arguments. This mirrors the fact that generic Finsler structures F do *not* satisfy $F(x,-y) = F(x,y)$. For example, non-Riemannian Randers spaces are positively homogeneous but not absolutely homogeneous.
* Since the distances in the above definitions are always measured from p, we tagged on "forward" as an adjective.

6.2 C. The Manifold Topology Versus the Metric Topology

Our goal here is to show that the manifold topology coincides with that generated by the forward metric balls. For this purpose, the second last inequality of Lemma 6.2.1 is of direct relevance. It says that:

* Given any $p \in M$, there is a coordinate map φ defined on the closure of some precompact open subset U containing p.
* φ maps U diffeomorphically onto the open Euclidean ball $\mathbb{B}^n(r)$, and sends p to the origin.
* Furthermore, there is a constant $c > 1$, depending only on p and U, such that

$$\boxed{\frac{1}{c} \, | \varphi(x_o) \ - \ \varphi(x_1) \, | \ \leqslant \ d(x_o, x_1) \ \leqslant \ c \, | \varphi(x_o) \ - \ \varphi(x_1) \, |} \tag{6.2.3}$$

  for any x_o, x_1 in U. Here, $| \ \ |$ denotes the Euclidean norm on $\mathbb{R}^n$.

- The right half of (6.2.3) says that points which are close to each other in the manifold topology are also close in the metric topology. Equivalently, the metric topology is contained in the manifold topology. The technical argument is given below, and is entirely straightforward.
- The left half of (6.2.3) seems to say that points deemed close by the metric topology are also close in the manifold topology. We are sounding a bit tentative here because, in order to use (6.2.3), one must ascertain that the points in question actually lie inside the (possibly) small coordinate neighborhood U. This is indeed the case. But as we demonstrate next, the requisite technical argument is less straightforward than the other inclusion described above.

Every forward metric ball is an open set:

Take any point p in the forward metric ball $\mathcal{B}^+_x(s)$. The number ϵ given by $s - d(x,p)$ is therefore positive. Associated with p is an open set U, a coordinate map φ, an open Euclidean ball $\mathbb{B}^n(r)$, and a constant $c > 1$ with the properties described in (6.2.3). By passing to a larger c if necessary, we may assume without any loss of generality that $\frac{\epsilon}{c} \leqslant r$. This, together with the fact that φ is a diffeomorphism, implies that

$$\mathcal{O} := \varphi^{-1}\left[\mathbb{B}^n\left(\frac{\epsilon}{c}\right)\right]$$

is a well-defined open set in the manifold topology. It contains the point p.

For each $q \in \mathcal{O}$, the right half of (6.2.3) implies that

$$d(p,q) \leqslant c\,|\varphi(q)| < c\,\frac{\epsilon}{c} = \epsilon\,.$$

Thus

$$d(x,q) \leqslant d(x,p) + d(p,q) < (s-\epsilon) + \epsilon = s\,.$$

In other words, $\mathcal{O}$ is contained in $\mathcal{B}^+_x(s)$.

Carrying out the above procedure for every point p in the forward metric ball $\mathcal{B}^+_x(s)$ shows that the latter is expressible as a union of manifold open sets. This is the straightforward half.

Every open set is a union of forward metric balls:

Now the less straightforward half. Let $\mathcal{O}$ be any open set in the manifold topology of M. Take any $p \in \mathcal{O}$. Associated with p is a coordinate neighborhood U, a coordinate map φ, an open Euclidean ball $\mathbb{B}^n(r)$, and a constant $c > 1$ with the properties described in (6.2.3). Shrink r, if necessary, so that U is contained in $\mathcal{O}$. Let us demonstrate that

$$(***) \qquad \mathcal{B}^+_p\left(\frac{r}{2c}\right) \subset U\,.$$

Since we have already arranged to have $U \subset \mathcal{O}$, the said metric ball will then be contained in $\mathcal{O}$. Carrying out such a process for each $p \in \mathcal{O}$ would imply that the latter is a union of forward metric balls.

So, it suffices to establish (***). Suppose there is a point q in the said metric ball that lies outside U. We derive a contradiction by reasoning as follows:

* Since $d(p,q) < \frac{r}{2c}$, there is a continuous piecewise smooth curve $\gamma(t)$, $0 \leqslant t \leqslant 1$ that goes from p to q, and which is shorter than $\frac{r}{2c}$. In particular, the entire curve γ must lie in the forward metric ball $\mathcal{B}^+_p(\frac{r}{2c})$.
* Now, γ starts at p but its destination q is outside U. Let $\gamma(t_o)$ be the point at which it first leaves U. Note that for $0 \leqslant t < t_o$, $\gamma(t)$ lies in U and $\mathcal{B}^+_p(\frac{r}{2c})$. Hence the left half of (6.2.3) is applicable. It

tells us that

$$\frac{1}{c}\,|\,\varphi(\gamma(t))\,| \;\leqslant\; d(p\,,\gamma(t)) \;<\; \frac{r}{2c}\,.$$

That is,

$$|\,\varphi(\gamma(t))\,| \;<\; \frac{r}{2}$$

for $0 \leqslant t < t_o$.

* On the other hand, since $\gamma(t_o)$ is at the boundary of U, we have

$$|\,\varphi(\gamma(t_o))\,| \;=\; r\,.$$

These last two conclusions violate the continuity of the composite map $\varphi \circ \gamma$. This is the contradiction we seek.

Let us summarize:

The topology generated by the forward metric balls coincides with the underlying manifold topology.

6.2 D. Forward Cauchy Sequences, Forward Completeness

- A sequence $\{x_i\}$ in M is said to **converge** to $x \in M$ if, given any open set $\mathcal{O}$ containing x, there exists a positive integer N (depending on $\mathcal{O}$) such that

$$i \geqslant N \;\Longrightarrow\; x_i \in \mathcal{O}\,.$$

- A sequence $\{x_i\}$ in M is called a **forward** (resp., **backward**) **Cauchy sequence** if, for all $\epsilon > 0$, there exists a positive integer N (depending on the ϵ) such that

$$N \leqslant i < j \;\Longrightarrow\; d(x_i, x_j) \;<\; \epsilon \quad \big[\text{resp., } d(x_j, x_i) \;<\; \epsilon\big]\,.$$

Consider the last inequality in Lemma 6.2.1. It says that for each fixed p, there is a coordinate neighborhood U containing p, together with a constant $c > 1$ depending only on p and U, such that

$$\boxed{\frac{1}{c^2}\,d(x_1, x_o) \;\leqslant\; d(x_o, x_1) \;\leqslant\; c^2\,d(x_1, x_o)} \tag{6.2.4}$$

for all x_o, x_1 in U.

Using (6.2.4), together with what we have just said about the manifold topology and the metric topology, one can show that the following three statements

$$x_i \to x\;, \qquad d(x, x_i) \to 0\;, \qquad d(x_i, x) \to 0$$

are equivalent. This is the subject of Exercise 6.2.5.

One can also use (6.2.4) to check that among the *convergent* Cauchy sequences, there is no distinction between forward Cauchy and backward Cauchy. See Exercises 6.2.6.

A Finsler manifold (M, F) is said to be **forward complete** with respect to its metric distance function d if every forward Cauchy sequence converges in M. It is **backward complete** with respect to d if every backward Cauchy sequence converges. See our example in §12.6, especially §12.6D.

- Exercise 6.2.11 shows that a *compact* Finsler manifold is at the same time both forward and backward complete, whether d is symmetric or not.
- If the Finsler structure F is *absolutely* homogeneous of degree one, then d is symmetric. In which case, forward completeness is equivalent to backward completeness. Thus, for example, one only needs to speak of completeness when dealing with Riemannian manifolds.
- Absolute homogeneity of F (equivalently, symmetry of d) is *not* a prerequisite for having both forward and backward completeness. For instance, the Finsler functions of Minkowski spaces are in general only positively homogeneous of degree one. Yet, these spaces are always forward and backward complete. This issue is addressed more leisurely in Exercise 14.1.1.

Exercises

Exercise 6.2.1:

(a) Suppose the Finsler structure F is absolutely homogeneous; explain why we must then have $d(x_o, x_1) = d(x_1, x_o)$.

(b) Which fact about F correlates with the triangle inequality satisfied by the distance function d ? Is absolute homogeneity needed?

Exercise 6.2.2:

(a) In the proof of Lemma 6.2.1, we sketched the arguments for deriving the second and the third [namely (6.2.3)] conclusions. Fill in the details.

(b) Use (6.2.3) to deduce inequality (6.2.4).

(c) Show that $d(x_o, x_1)$ is positive unless $x_o = x_1$.

Exercise 6.2.3: Given a closed subset of Euclidean $\mathbb{R}^n$, we know from topology that it is compact if and only if it is bounded. Explain why Lemma 6.2.1 does *not* imply this property for the "metric" space (M, d).

Exercise 6.2.4: Suppose $\exp_p$ is defined on all of $\mathrm{B}_p(r)$.

(a) Explain why $\exp_p[\,\mathrm{B}_p(r)\,] \subseteq \mathcal{B}_p^+(r)$.

(b) At this moment, can you tell whether the same relationship holds between $\exp_p[\,\mathrm{S}_p(r)\,]$ and $\mathcal{S}_p^+(r)$?

Exercise 6.2.5: Let $\{x_i\}$ be any sequence in a Finsler manifold (M, F). Prove that the following three statements are equivalent.

(a) $\{x_i\}$ converges to x under the manifold topology of M.

(b) $d(x, x_i) \to 0$.

(c) $d(x_i, x) \to 0$.

Exercise 6.2.6: Use (6.2.4) to prove the following statements.

(a) A convergent sequence is at the same time both forward Cauchy and backward Cauchy.

(b) $\{x_i\}$ is a *convergent* backward Cauchy sequence if and only if it is a *convergent* forward Cauchy sequence.

Exercise 6.2.7: Let M be a C^∞ manifold equipped with a *metric* distance function d. Suppose there is a map $F : TM \to [0, \infty)$ such that, given any two C^1 curves σ_1, σ_2 which emanate from x with initial velocities y_1, y_2, we have

$$d(\, \sigma_1(t)\, ,\, \sigma_2(t)\,) \;=\; |t|\; F(\, x\, ,\, y_1 - y_2\,) \;+\; o(t)\; .$$

Show that this F must have the following properties:

(a) $F(x\, ,\, \lambda\, y) = |\lambda|\, F(x, y)$ for all real numbers λ.

(b) $F(x\, ,\, y_1 + y_2) \leqslant F(x\, ,\, y_1) + F(x\, ,\, y_2)$.

(c) For any C^1 curve σ that emanates from x with initial velocity y, one has

$$F(x, y) \;=\; \lim_{t \to 0} \frac{d(\, x\, ,\, \sigma(t)\,)}{|t|}\; .$$

This is in essence the **Busemann–Mayer theorem** [BuMa] for absolutely homogeneous F. For its counterpart in the positively homogeneous (but not absolutely homogeneous) case, see Exercise 6.3.4.

Exercise 6.2.8: In this exercise, we illustrate another application of the inequality (6.2.4). To this end, let $c > 1$ and U (a coordinate neighborhood centered at p) be as described in Lemma 6.2.1. Let $\epsilon > 0$ be such that the forward metric ball $\mathcal{B}_p^+(\, [1 + c^2]\, \epsilon\,)$ is contained in U. Prove that:

> If γ is any minimal geodesic that emanates from a point (say, z) in the small ball $\mathcal{B}_p^+(\epsilon)$ and terminates at p, then it must be contained entirely in the larger ball $\mathcal{B}_p^+(\, [1 + c^2]\, \epsilon\,)$. In case the distance function is symmetric, the said geodesic γ must actually lie entirely in $\mathcal{B}_p^+(\epsilon)$.

Here, a geodesic from x_o to x_1 is said to be **minimal** if its length equals the metric distance $d(x_o, x_1)$. Also, whenever the distance function is symmetric (equivalently, when the Finsler function is absolutely homogeneous), the decoration $^+$ on $\mathcal{B}_p^+$ will be superfluous.

Suppose there was a minimal geodesic γ from some $z \in \mathcal{B}_p^+(\epsilon)$ to p that violated our claim. Derive a contradiction as follows:

(a) **When the distance function is possibly nonsymmetric.** Since γ started inside $\mathcal{B}_p^+(\epsilon)$ and was supposed to have wandered out of the larger metric ball $\mathcal{B}_p^+(\,[1+c^2]\,\epsilon\,)$, it must first cross the boundary of $\mathcal{B}_p^+(\epsilon)$ at some point A, and then the boundary of $\mathcal{B}_p^+(\,[1+c^2]\,\epsilon\,)$ at some point B. Give the portion of γ from A to B a name, say σ. By using a triangle inequality on $d(p,B)$, explain why the length of the segment σ must be at least $c^2\epsilon$. Hence the length of γ must exceed $c^2\epsilon$. On the other hand, the length of γ equals $d(z,p)$. Use (6.2.4) to show that this is in turn $< c^2\epsilon$.

(b) **When the distance function is symmetric.** Show that if γ wandered outside $\mathcal{B}_p^+(\epsilon)$, its length would have to exceed ϵ. On the other hand, this length equals $d(z,p)$; explain why it must therefore be $< \epsilon$.

Exercise 6.2.9: Let us give yet another application of (6.2.4). As usual, the Finsler structure F is only assumed to be positively homogeneous of degree one. We want to show that:

Each of the following two families of functions

$$\{\, d(\,\cdot\,,q) : \; q \in M \,\} \quad \text{and} \quad \{\, d(q\,,\cdot\,) : \; q \in M \,\}$$

is equicontinuous at every $p \in M$.

To this end, let $\epsilon > 0$ be given.

(a) Consider any $p \in M$. Take the coordinate neighborhood U and the constant $c > 1$ guaranteed by Lemma 6.2.1. Explain why, by passing to a larger c if necessary, one may assume without any loss of generality that the forward metric ball $\mathcal{B}^+ := \mathcal{B}_p^+(\frac{\epsilon}{1+c^2})$ is contained in U. Hint: review §6.2C.

(b) For $x \in \mathcal{B}^+$, use a straightforward triangle inequality to deduce that $d(p,q) < \epsilon + d(x,q)$. Also, using an appropriate triangle inequality, together with part (a) and (6.2.4), show that $d(x,q) < \epsilon + d(p,q)$. Thus, given any $\epsilon > 0$, there exists a $\mathcal{B}^+$ (depending on ϵ and p but *not* on q) such that

$$x \in \mathcal{B}^+ \quad \Rightarrow \quad |\, d(x,q) \; - \; d(p,q) \,| \; < \; \epsilon \, .$$

(c) For $x \in \mathcal{B}^+$, use an appropriate triangle inequality, together with part (a) and (6.2.4), to deduce that $d(q,p) < d(q,x) + \epsilon$. Next, using a straightforward triangle inequality, show that $d(q,x) < d(q,p) + \epsilon$. So, given any $\epsilon > 0$, there exists a $\mathcal{B}^+$ (depending on ϵ and p but *not* on q) such that

$$x \in \mathcal{B}^+ \quad \Rightarrow \quad |\, d(q,x) \; - \; d(q,p) \,| \; < \; \epsilon \, .$$

Since we have seen in §6.2C that forward metric balls are open sets, we are done. We see that the "equi-" part of the continuity is directly attributable to the triangle inequality and the remarkable property (6.2.4).

Exercise 6.2.10: Suppose (M, F) is a Finsler manifold, where F is positively (but perhaps not absolutely) homogeneous of degree one. Let us define the **backward metric balls** as

$$\mathcal{B}_p^- \;:=\; \{\, x \in M : \; d(x,p) \;<\; r \,\} \,.$$

(a) Show that every backward metric ball is an open set of M.

(b) Prove that every open set of M is a union of backward metric balls.

In other words, **the topology generated by the backward metric balls is precisely the underlying manifold topology.**

Exercise 6.2.11: Let d be the possibly nonsymmetric metric distance function of a Finsler manifold. Prove that:

> If M is compact, then all forward and backward Cauchy sequences (with respect to d) must converge in M. In other words, compact Finsler spaces are automatically both forward complete and backward complete. This holds whether the Finsler structure is absolutely homogeneous or only positively homogeneous.

6.3 Short Geodesics Are Minimizing

We now give a substantive application of the Gauss lemma. Recall that geodesics are defined as critical points of the arc length functional. It is then a natural question to wonder whether any such critical point is an absolute minimum. As we show, the affirmative answer for short geodesics is a consequence of the Gauss lemma and a fundamental inequality [namely (1.2.3)] in Finsler geometry. So, every short geodesic indeed minimizes the arc length functional, among all piecewise C^∞ curves that share its endpoints. The precise statement carries much more information than that, and is given in the following theorem.

Theorem 6.3.1.

* *Fix a point p in a Finsler manifold (M, F), where F is positively (but perhaps not absolutely) homogeneous of degree one in y.*
* *Suppose for some r and ϵ, both positive, $\exp_p$ is a C^1-diffeomorphism from the tangent ball $B_p(r+\epsilon)$ onto its image.*

Then:

(a) *Each radial geodesic $\exp_p(tv)$, $0 \leqslant t \leqslant r$, $F(p,v) = 1$, will minimize distance (which is r) among all piecewise C^∞ curves in M that share its endpoints.*

(b) *Any piecewise C^∞ curve in M that has the same arc length and endpoints as the geodesic $\exp_p(tv)$, $0 \leqslant t \leqslant r$, $F(p,v) = 1$, must lie inside $\exp_p[\, B_p(r) \cup S_p(r)\,]$ and is in fact a reparametrization of that geodesic.*

(c) *The following hold:*

$$\text{(6.3.1)} \qquad \boxed{\begin{aligned} \exp_p[\, B_p(r)\,] &= \mathcal{B}_p^+(r) \ , \\ \exp_p[\, S_p(r)\,] &= \mathcal{S}_p^+(r) \ . \end{aligned}}$$

Remarks:

- The hypothesis is satisfied by small positive r and ϵ, in view of the Inverse Function theorem and (5.3.6).
- The statement $\exp_p[\,\mathrm{S}_p(r)\,] = \mathcal{S}_p^+(r)$ justifies our calling the object $\exp_p[\,\mathrm{S}_p(r)\,]$ a geodesic sphere of radius r.

Proof.

A special case of (a):

Let us first prove that,

> If $c(u)$, $0 \leqslant u \leqslant 1$ is a piecewise C^∞ curve with endpoints $c(0) = p$ and $c(1) = \exp_p(rv)$, and if c lies inside $\exp_p[\,\mathrm{B}_p(r) \cup \mathrm{S}_p(r)\,]$, then its arc length is at least r.

Note that r is the arc length of the unit speed geodesic $\exp_p(tv)$, $0 \leqslant t \leqslant r$, $F(p,v) = 1$. The argument we present is directly fashioned after the one given by Bao and Chern in [BC1].

Since c lies inside $\exp_p[\,\mathrm{B}_p(r) \cup \mathrm{S}_p(r)\,]$, we can express it as

$$\text{(6.3.2)} \qquad c(u) = \exp_p[\, t(u)\, v(u)\,] \ , \quad \text{with} \quad F(p\,,\, v(u)\,) = 1 \ .$$

The endpoints of c are $c(0) = p$ and $c(1) = \exp_p(rv)$, thus

$$\text{(6.3.3)} \qquad t(0) = 0 \ , \quad t(1) = r \ ,$$

$$\text{(6.3.4)} \qquad v(1) = v \ .$$

For the moment, digress to consider the following variation of our geodesic $\exp_p(tv)$:

$$\sigma(t,u) := \exp_p[\, t\, v(u)\,] \ ,$$

where $0 \leqslant t \leqslant r$, $0 \leqslant u \leqslant 1$. As usual, employ the abbreviations

$$T := \sigma_* \frac{\partial}{\partial t} \ ,$$

$$U := \sigma_* \frac{\partial}{\partial u} \ .$$

For each fixed u, T represents the velocity vectors of the unit speed geodesic $\exp_p[\,t\,v(u)\,]$. The chain rule also tells us that

$$U \;=\; \exp_{p*}\left[\,t\,\frac{d\,v(u)}{du}\,\right] ,$$

where $\frac{d\,v(u)}{du}$ is tangent to the indicatrix $F(p,v)=1$. By the Gauss lemma, we see that

$$g_T(U\,,T\,) \;=\; 0\,. \tag{6.3.5}$$

Returning to the curve c, we relate its representation (6.3.2) to the above variation. One has

$$c(u) \;=\; \sigma(\,t(u)\,,\,u\,)$$

which, through the chain rule, gives

$$\frac{dc}{du} \;=\; \frac{\partial\sigma}{\partial t}\,\frac{dt}{du} \;+\; \frac{\partial\sigma}{\partial u}\,.$$

In other words (see Figure 6.2),

$$\text{the velocity } \frac{dc}{du} \text{ of } c \;=\; \frac{dt}{du}T \;+\; U\,. \tag{6.3.6}$$

With this decomposition, we can obtain a lower bound for the length of c.

To this end, we use the fundamental inequality (1.2.3), namely,

$$w^i\;F_{y^i}(x,y) \;\leqslant\; F(x,w) \;\text{ at all }\; y\neq 0\,.$$

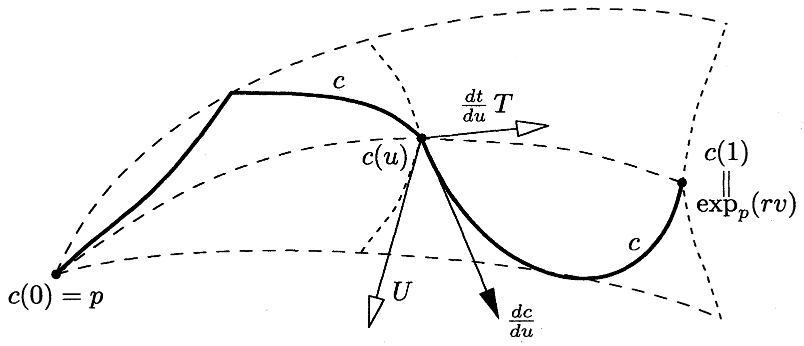

Figure 6.2

The curve $c(u) := \exp_p[\,t(u)\,v(u)\,]$ gives a 1-parameter family of (possibly repetitious) directions $v(u)$. Those in turn define a geodesic variation $\sigma(t,u) := \exp_p[\,t\,v(u)\,]$, with velocities T for the t-curves, and velocities U for the u-curves. The velocity of the curve $c(u)$ is $\frac{dc}{du}$. By the chain rule, we find that $\frac{dc}{du} = \frac{dt}{du}\,T + U$.

It is an equality if and only if $w = \alpha y$ for some $\alpha \geqslant 0$. Recall from the discussion in §1.2.C that this can be interpreted as a kind of Cauchy–Schwarz inequality for normed vector spaces. Anyway, into this inequality we substitute $c(u)$ for x, $\frac{dc}{du}$ for w, and T for y. In conjunction with (6.3.6), we obtain

$$(6.3.7) \qquad F\Big(x\,,\,\frac{dc}{du}\Big) \;\geqslant\; F_{y^i}(c,T)\,\frac{dt}{du}\,T^i \;+\; F_{y^i}(c,T)\,U^i\,.$$

By (1.4.3) and (6.3.5), we have

$$F_{y^i}(c,T)\,U^i \;=\; \frac{1}{F}\,g_{ij\,(c,T)}\,T^j\,U^i \;=\; 0\,.$$

Using Euler's theorem together with the fact that for fixed u, each geodesic $\exp_p[\,t\,v(u)\,]$ has (constant) unit speed and initial velocity $v(u)$, we get

$$F_{y^i}(c,T)\,T^i\,\frac{dt}{du} \;=\; F(c,T)\,\frac{dt}{du} \;=\; F(p\,,\,v(u))\,\frac{dt}{du} \;=\; \frac{dt}{du}\,.$$

Inputting these last two statements into (6.3.7), we see that

$$F\Big(x\,,\,\frac{dc}{du}\Big) \;\geqslant\; \frac{dt}{du}\,.$$

Hence,

$$(*) \quad L(c) \;=\; \int_0^1 F\Big(x\,,\,\frac{dc}{du}\Big)\,du \;\geqslant\; \int_0^1 \frac{dt}{du}\,du \;=\; t(1)-t(0) \;=\; r\,.$$

Thus we have shown that if c lies in $\exp_p[\,\mathrm{B}_p(r)\cup \mathrm{S}_p(r)\,]$ and shares the same endpoints with the geodesic $\exp_p(tv)$ $[0\leqslant t\leqslant r,\ F(p,v)=1]$, then it cannot be shorter than the geodesic in question.

The rest of (a) and the first half of (b):

What if c wanders outside $\exp_p[\,\mathrm{B}_p(r)\cup \mathrm{S}_p(r)\,]$? Well, in that case, let $u_o < 1$ be the first instant it intersects the geodesic sphere $\exp_p[\,\mathrm{S}_p(r)\,]$. The argument presented above shows that the portion $c(u)$, $0\leqslant u\leqslant u_o$ already has arc length at least r, hence that of c must exceed r. Thus part (a) is completely proved, and so is the first half of part (b).

The second half of (b):

Since we now know that any c which has the same arc length and endpoints as the unit speed geodesic $\exp_p(tv)$, $0\leqslant t\leqslant r$ must lie inside $\exp_p[\,\mathrm{B}_p(r)\cup \mathrm{S}_p(r)\,]$, the representation (6.3.2)–(6.3.4) holds. Also, by hypothesis we have $L(c)=r$. Thus (*) becomes an equality. A bit of argument-chasing shows that this can happen if and only if the fundamental inequality is actually an equality. Such is in turn valid if and only if

$$\frac{dc}{du} \;=\; \alpha\,T \quad \text{for some } \alpha\geqslant 0\,.$$

Comparing this with (6.3.6), we conclude that the "transverse" component U of $\frac{dc}{du}$ must vanish. In view of our formula for U, this means that

$$\exp_{p*}\left[t(u)\, \frac{dv(u)}{du} \right] \;=\; 0 \, .$$

But $\exp_p$ is by assumption nonsingular on $\mathrm{B}_p(r) \cup \mathrm{S}_p(r)$, so this last statement is equivalent to

$$\frac{dv(u)}{du} \;=\; 0 \, .$$

In other words,

$$v(u) \;=\; v(1) \;\equiv\; v$$

and

$$c(u) \;=\; \exp_p[\, t(u)\, v \,] \, , \quad 0 \leqslant u \leqslant 1 \, ,$$

with

$$t(0) \;=\; 0 \quad \text{and} \quad t(1) \;=\; r \, .$$

This is manifestly a reparametrization of the geodesic $\exp_p(tv)$, $0 \leqslant t \leqslant r$. Part (b) is now completely proved.

The first half of (c):

By Exercise 6.2.4, we have $\exp_p[\,\mathrm{B}_p(r)\,] \;\subseteq\; \mathcal{B}_p^+(r)$.

> Suppose for the sake of argument that $\exp_p[\,\mathrm{B}_p(r)\,]$ is a *proper* subset of the forward metric ball $\mathcal{B}_p^+(r)$.

Then there exists a piecewise C^∞ curve $c(u)$, $0 \leqslant u \leqslant 1$ such that $c(0) = p$, $c(1) = q$, $L(c) < r$, and $q \notin \exp_p[\,\mathrm{B}_p(r)\,]$. Since the destination of c lies outside $\exp_p[\,\mathrm{B}_p(r)\,]$, it must intersect $\exp_p[\,\mathrm{S}_p(r)\,]$ at least once during its journey. Let $u_o \leqslant 1$ be the first instant that happens. By part (a), we conclude that the portion $c(u)$, $0 \leqslant u \leqslant u_o$ must have arc length at least r, which contradicts the statement $L(c) < r$. Thus the above supposition is false and we have

$$\exp_p[\,\mathrm{B}_p(r)\,] \;=\; \mathcal{B}_p^+(r)$$

instead.

The second half of (c):

What we have just deduced says that $\exp_p[\,\mathrm{B}_p(r+\delta)\,] \;=\; \mathcal{B}_p^+(r+\delta)$ for small δ. In particular, every point on $\mathcal{S}_p^+(r)$ can be reached from p through a unit speed radial geodesic $\exp_p(tv)$, $0 \leqslant t \leqslant r$. Thus $\mathcal{S}_p^+(r) \;\subseteq\; \exp_p[\,\mathrm{S}_p(r)\,]$. A moment's thought shows that the inclusion in the other direction is a consequence of part (a). So

$$\exp_p[\, S_p(r) \,] \;=\; \mathcal{S}_p^+(r) \, . \quad \square$$

Exercises

Consider a piecewise C^∞ curve $\gamma : [a,b] \to M$. We say that **short pieces of γ are minimizing** if, for every $t_o \in [a,b]$, there exists a subinterval

$$[a_1\,,\,b_1] \;:=\; [\,t_o - \epsilon\,,\,t_o + \epsilon\,] \;\cap\; [a,b]$$

such that

$$L(\,\gamma_{|[a_1,b_1]}\,) \;\leqslant\; L(c)$$

for *all* piecewise C^∞ curves c which share the same endpoints with $\gamma_{|[a_1,b_1]}$.

Exercise 6.3.1: Show that short pieces of a geodesic are indeed minimizing, as the title of this section asserts.

Exercise 6.3.2: Show that if short pieces of a piecewise C^∞ curve γ are minimizing, then γ can in fact be reparametrized as a constant speed geodesic. Hint: your argument may depend on the smooth extension of curves that satisfy second order ODEs.

Exercise 6.3.3: Analogous to the "forward" metric balls, we have also defined the "backward" metric balls

$$\mathcal{B}_p^- \;:=\; \{\,x \in M:\; d(x,p) \;<\; r\,\}\;.$$

See Exercise 6.2.10. Here, our goal is to show that:

> For every point p in a Finsler manifold (M,F), there exists a small $r > 0$ such that every pair of points q_o, q_1 in $\mathcal{B}_p^+(r) \cap \mathcal{B}_p^-(r)$ can be joined by a unique minimizing geodesic from q_o to q_1.

Note that it is the intersection that we want here, not the union.

(a) First apply the Inverse Function theorem to the map

$$(q,v) \;\mapsto\; (\,q\,,\,\exp_q v\,)\;.$$

Conclude that at every $p \in M$, there exists an $r > 0$ (depending only on p) such that for all q close to p, the map $\exp_q$ is a C^1-diffeomorphism from $\mathrm{B}_q(2r)$ onto $\mathcal{B}_q^+(2r)$.

(b) By shrinking r if necessary, we assume without loss of generality that all points q in $\mathcal{B}_p^+(r) \cap \mathcal{B}_p^-(r)$ may be considered close enough to p so that the aforementioned property of $\exp_q$ holds. Show that this neighborhood of p has the feature we are supposed to establish.

(c) Also, check that the resulting minimizing geodesics all lie in $\mathcal{B}_p^+(3r)$.

Exercise 6.3.4: Let (M,F) be a Finsler manifold, where F is positively (but perhaps not absolutely) homogeneous of degree one. Let $\sigma(t)$, $0 \leqslant t < \epsilon$ be any *short* (hence ϵ must be small) C^1 curve that emanates from p with initial velocity $v := v^i\,\frac{\partial}{\partial x^i}$.

(a) Show that σ admits the representation $\sigma(t) = \exp_p(y_t)$, where y_t is a curve in T_pM that emanates from the origin with initial velocity

v. Strictly speaking, the local coordinates x^i on M induce global coordinates y^i on T_pM. Therefore, we should have asserted that the initial velocity of y_t is $v^i \frac{\partial}{\partial y^i}$ rather than $v^i \frac{\partial}{\partial x^i}$. However, this is a forgivable confusion on linear spaces, of which T_pM is one.

(b) Use Theorem 6.3.1 to explain why $d(p, \sigma(t)) = F(p, y_t)$.

(c) Check that $v = \lim_{t \to 0^+} \frac{1}{t} y_t$.

(d) Finally, use the continuity of F to help you deduce that

$$F(p,v) = \lim_{t\to 0^+} \frac{d(p, \sigma(t))}{t} .$$

This is the **Busemann–Mayer theorem** [BuMa] for positively homogeneous functions F.

Exercise 6.3.5: The following argument seems to have bypassed the use of the fundamental inequality (1.2.3) in proving part (a) of Theorem 6.3.1.

Begin with (6.3.6): $\frac{dc}{du} = \frac{dt}{du} T + U$. By the Gauss lemma, $g_T(U,T) = 0$. So

$$g_T\left(\frac{dc}{du}, \frac{dc}{du}\right) = \left|\frac{dt}{du}\right|^2 g_T(T,T) + g_T(U,U) \geqslant \left|\frac{dt}{du}\right|^2 .$$

Here, T is the velocity field of the unit speed geodesics $\exp_p[t\, v(u)]$, and we have used the fact that $g_T(T,T) = 1$. Thus

$$\left\|\frac{dc}{du}\right\| := \sqrt{g_T\left(\frac{dc}{du}, \frac{dc}{du}\right)} \geqslant \left|\frac{dt}{du}\right| .$$

Upon integration, we get

$$\int_0^1 \left\|\frac{dc}{du}\right\| du \geqslant \int_0^1 \left|\frac{dt}{du}\right| du \geqslant \left|\int_0^1 \frac{dt}{du}\, du\right| = |t(1) - t(0)| = r .$$

Therefore $L(c) \geqslant r$.

(a) Can you identify the flaw in this argument?

(b) Explain why the above argument is valid in the Riemannian case.

6.4 The Smoothness of Distance Functions

6.4 A. On Minkowski Spaces

We begin with a Minkowski space $(\mathbb{R}^n, F)$. The distance from the origin of any $y \in \mathbb{R}^n$ is simply $F(y)$. Recall that the Minkowski norm F is by hypothesis C^∞ on $\mathbb{R}^n \smallsetminus 0$. As a matter of fact, its *square* F^2 is differentiable and C^1 at $y = 0$, and the derivative there is zero. Indeed:

* Note that

$$F^2(tv) = t^2 F^2(v) , \quad t > 0 ,$$
$$F^2(tv) = t^2 F^2(-v) , \quad t < 0 .$$

So all directional (hence partial) derivatives of F^2 are zero at $y = 0$. By contrast, those of F do not exist at the origin, even if F is absolutely homogeneous of degree one.

* One can derive the identity

$$\frac{\partial F^2}{\partial y^i}(y) \;=\; 2\, y^j\, g_{ij}(y)\ , \quad \text{for all} \quad y \neq 0\ .$$

Since the g_{ij} are constant along rays, we conclude that (g_{ij}) is a bounded positive matrix. Therefore $\frac{\partial F^2}{\partial y^i}(y) \to 0$ as $y \to 0$.

In the Riemannian setting, $F^2(y) = g_{ij}\, y^i\, y^j$, where the g_{ij} are constants. Thus F^2 is C^2 (in fact, C^∞) at $y = 0$. It turns out that the Riemannian case is the only scenario in which F^2 can be C^2 or smoother at the origin. To see this, suppose F^2 is C^2 at $y = 0$, which then enables us to do a second-order Taylor expansion. Since F^2 and its derivative both vanish at $y = 0$, we have

$$F^2(v) \;=\; \frac{1}{2}\, \frac{\partial^2 F^2}{\partial y^i \partial y^j}(0)\ v^i\, v^j \;+\; O\left(\, |v|^3 \,\right)\ .$$

But $F^2(v)$ is homogeneous of degree 2 in v, so all the higher-order terms must vanish, and our claim follows.

Let us summarize:

Lemma 6.4.1. *Let $(\mathbb{R}^n, F)$ be a Minkowski space, where F is positively (but perhaps not absolutely) homogeneous of degree one. The square F^2 of the distance function is:*

(a) *C^∞ away from the origin.*

(b) *C^1 at the origin and has a zero derivative there.*

(c) *C^2 at the origin if and only if F is Riemannian.*

6.4 B. On Finsler Manifolds

Fix a point $p \in M$. Let $d_p := d(p\,, \cdot\,)$ be the distance function as measured from p, using the Finsler structure and integrating along curves. See the beginning of §6.2.

* Note that for x close to p, we can write $x = \exp_p(v)$. The correspondence $x \leftrightarrow v$ is only C^1 at p, although it is C^∞ away from p. We learned that in §5.3.

* In view of Theorem 6.3.1, the geodesic $\exp_p(tv)$, $0 \leqslant t \leqslant 1$ is minimizing. It also has constant speed $F(v)$. Hence

$$d_p(x) \;=\; d(p, x) \;=\; F(v)\ .$$

These, together with Lemma 6.4.1, now imply that d_p^2 is C^∞ in a punctured neighborhood of p, is C^1 at p, and its derivative there is zero.

If F happens to be Riemannian, then the correspondence $x \leftrightarrow v$ is C^∞ even at p. In that case, the function d_p^2 is C^∞ (in particular, C^2) at p.

Conversely, suppose d_p^2 is C^2 at p. We perform a second-order Taylor expansion of d_p^2 in an arbitrary (C^∞) coordinate chart $\varphi : U \subset M \to \mathbb{R}^n$ with $\varphi(p) = 0$. Since d_p^2 vanishes at p and has a zero derivative there,

$$d_p^2(x) \;=\; \frac{1}{2}\,\frac{\partial^2 d_p^2}{\partial x^i \partial x^j}(p)\; x^i\, x^j \;+\; O\left(\,|x|^3\,\right)\ .$$

Given any $y = y^i\,\frac{\partial}{\partial x^i} \in T_pM$, we can use the components y^i to define a useful curve $\sigma(t) := \varphi^{-1}[t\,(y^i)]$ in the coordinate patch U. This curve passes through the point p with initial velocity y. By Exercise 6.3.4,

$$F(p,y) \;=\; \lim_{t\to 0^+} \frac{1}{t}\; d_p(\,\sigma(t)\,)\ .$$

Using our Taylor expansion, we have

$$\frac{d_p(\,\sigma(t)\,)}{t} \;=\; \frac{1}{t}\sqrt{\frac{1}{2}\,\frac{\partial^2 d_p^2}{\partial x^i \partial x^j}(p)\; ty^i\, ty^j \;+\; O\left(\,|ty|^3\,\right)}\ .$$

Therefore

$$F(p,y) \;=\; \sqrt{\frac{1}{2}\,\frac{\partial^2 d_p^2}{\partial x^i \partial x^j}(p)\; y^i\, y^j}\ ,$$

which says that F is Riemannian in T_pM.

We have just derived the exact analogue of Lemma 6.4.1:

Proposition 6.4.2. *Let (M,F) be a Finsler manifold, where F is C^∞ on $TM \setminus 0$, and is positively (but perhaps not absolutely) homogeneous of degree one. Fix any $p \in M$. The function d_p^2 that measures metric distance squared from p is:*

(a) *C^∞ in a punctured neighborhood of p.*

(b) *C^1 at p and has a zero derivative there.*

(c) *C^2 at p if and only if F is Riemannian in T_pM.*

Exercises

Exercise 6.4.1: Let $(\mathbb{R}^n, F)$ be a Minkowski space. This exercise concerns the strict convexity of the open balls

$$\mathrm{B}(r) \;:=\; \{\, y \in \mathbb{R}^n :\; F(y) < r \,\}\ .$$

Namely, every line segment with endpoints in $\mathrm{B}(r)$ must be completely contained in $\mathrm{B}(r)$.

(a) Parametrize the line segment in question as $\gamma(t) = y + t\xi$, $a \leqslant t \leqslant b$. If γ passes through the origin, deduce the conclusion from the homogeneity of F.

(b) Suppose γ does not pass through the origin; that is, $y \neq 0$ and ξ is not collinear with y. Consider the function $h(t) := F(y + t\xi)$, $a \leqslant t \leqslant b$. By hypothesis, $h(a)$ and $h(b)$ are both less than r. Use the chain rule (which is applicable because $y+t\xi$ never goes through the origin) to show that

$$h''(t) = \frac{\partial^2 F}{\partial y^i \partial y^j}_{|(y+t\xi)} \xi^i \xi^j .$$

Then explain, with the help of (1.2.9), why the graph of $h(t)$ is concave up throughout $[a, b]$. Why is that relevant to the problem at hand?

Exercise 6.4.2: Let p be any point in a Minkowski space $(\mathbb{R}^n, F)$. Define

$$\mathrm{B}(p, r) := \{ y \in \mathbb{R}^n : F(y - p) < r \} .$$

Show that $\mathrm{B}(p, r)$ is strictly convex.

Exercise 6.4.3: The above statements extend from Minkowski spaces to Finsler manifolds, and are results of Whitehead's [W]:

> Let K be a compact subset of a Finsler manifold (M, F). Then there exists a positive number ϵ such that every $\mathcal{B}_p^+(r)$ with $p \in K$ and $r \leqslant \epsilon$ is strictly convex. Namely, given any geodesic segment with endpoints in $\mathcal{B}_p^+(r)$, it must stay entirely in $\mathcal{B}_p^+(r)$.

Would you *attempt* to give a proof of this theorem?

Exercise 6.4.4: Let (M, F) be a Finsler manifold. Suppose F is C^∞ on $TM \setminus 0$, and is possibly only positively homogeneous of degree 1. Fix $q \in M$. Do the conclusions of Proposition 6.4.2 hold for the function $d^2(\cdot\,, q)$ that measures metric distance squared *towards* q?

6.5 Long Minimizing Geodesics

In this section and the next, we study various characterizations of forward complete Finsler manifolds (M, F) through the Hopf–Rinow theorem. Two rather striking features emerge.

- We find that the behavior of geodesics controls, and is controlled by, the metric space structure.
- It is shown that on a metrically forward complete Finsler manifold, any two points can be joined by a minimizing geodesic. The fact that a connecting geodesic exists at all, let alone a (possibly "long") minimizing one, is itself a surprising phenomenon.

The following proposition is the driving force behind what we just discussed. Since our Finsler structure is only assumed to be positively homogeneous of degree 1, the metric distance function d may not be symmetric. This introduces a subtle twist (which is happily resolved by Exercise 6.2.9) in the standard proof. Nevertheless, we owe our style of presentation to O'Neill [ON].

Proposition 6.5.1. *Let p be any point in a Finsler manifold (M, F), where F is positively (but perhaps not absolutely) homogeneous of degree one. Suppose*

* *M is connected, and*
* *$\exp_p$ is defined on all of T_pM.*

Then:

- *For any $q \in M$, there exists a minimizing unit speed geodesic $\exp_p(tv)$ from p to q.*
- *The direction of the unit initial velocity v is chosen in the following way. Among all points on a sufficiently small metric sphere $\mathcal{S}_p^+(r)$ centered at p, the point $\exp_p(rv)$ is closest, although perhaps not uniquely so, to the point q.*

Remark: We soon show that the hypothesis of "$\exp_p$ being defined on all of T_pM" is equivalent to the forward completeness of M, whether regarded as a "metric" space or as a topological space.

Proof. Fix a sufficiently small r. Then (6.3.1) assures us that

$$\exp_p[\,\mathrm{B}_p(r)\,] \;=\; \mathcal{B}_p^+(r) \quad \text{and} \quad \exp_p[\,\mathrm{S}_p(r)\,] \;=\; \mathcal{S}_p^+(r)\ .$$

Here, S_p and B_p are specified by $F = r$ and $F < r$ in T_pM, while $\mathcal{S}_p^+$ and $\mathcal{B}_p^+$ are defined by $d(p\,,\cdot\,) = r$ and $d(p\,,\cdot\,) < r$. If $q \in \mathcal{B}_p^+(r)$, then Theorem 6.3.1 ensures the desired conclusion.

So suppose $q \notin \mathcal{B}_p^+(r)$. We seek the minimizing geodesic by learning how to aim properly from p towards q. The insight here is that the geodesic in question must cross the metric sphere $\mathcal{S}_p^+(r)$ at a point m that is closest to q. See Figure 6.3. The existence of this m follows from the continuity of $d(\,\cdot\,,\,q\,)$ (see Exercise 6.2.9) and the compactness of $\mathcal{S}_p^+(r)$.

Since q lies outside $\mathcal{B}_p^+(r)$, any curve c from p to q can be broken up into two segments c_1 and c_2, where c_1 ends on $\mathcal{S}_p^+(r)$ and c_2 emanates from $\mathcal{S}_p^+(r)$. By part (a) of Theorem 6.4.1, the length of c_1 will be at least r, which is $d(p,m)$. Also, the definition of m implies that the length of c_2 is at least $d(m,q)$. Thus $L(c) \geqslant d(p,m) + d(m,q)$, from which we deduce that $d(p,q) \geqslant d(p,m) + d(m,q)$. On the other hand, the "$\leqslant$" statement is always

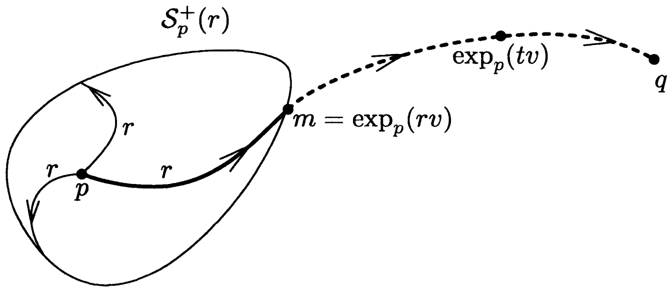

Figure 6.3

The point m minimizes the function $d(\,\cdot\,, q)$ on the forward metric sphere $\mathcal{S}_p^+(r)$. Intuitively, we are minimizing over all possible *directions* issued from p. And $\mathcal{S}_p^+(r)$ can be viewed as parametrizing these directions. By going from p to m through the radial geodesic segment $\exp_p(tv)$, $0 \leqslant t \leqslant r$, we have aimed properly towards q. Once past m, all one needs to do is to keep on going. Namely, keep on extending $\exp_p(tv)$ to $t > r$ until one arrives at q. According to the analytical argument at the end of proving Proposition 6.5.1, the destination q *will* be reached in finite time.

valid because of the triangle inequality. Therefore, proper aiming results in

$$d(p,m) \; + \; d(m,q) \;\; = \;\; d(p,q) \; .$$

Write m as $\exp_p(rv)$, with $F(p,v) = 1$. We claim that the unit speed geodesic $\sigma(t) := \exp_p(tv)$ is precisely the minimizing geodesic being sought. Our hypothesis guarantees that $\sigma(t)$ is defined for all t. It remains to check that when $t = d(p,q)$, we have $\sigma(t) = q$.

To this end, let T be the subset of all $t \in [\,0\,,\, d(p,q)\,]$ that satisfy

$$t \; + \; d(\,\sigma_t\,,\, q\,) \;\; = \;\; d(p,q) \; .$$

Here, we have abbreviated $\sigma(t)$ as σ_t to avoid excessive use of parentheses. Note that if $d(p,q) \in T$, then the defining statement of T implies that $d(\,\sigma_{d(p,q)}\,,\, q\,) = 0$. Hence the desired conclusion

$$\sigma_{d(p,q)} \;\; := \;\; \sigma[\,d(p,q)\,] \;\; = \;\; q$$

would immediately follow.

Suppose for the sake of argument that $d(p,q) \notin T$. This, together with the fact that T is closed (by continuity; again, we need Exercise 6.2.9) and nonempty (because $r \in T$), tells us that $t_o := \operatorname{Max} T$ is strictly less than $d(p,q)$. From $\tilde{p} := \sigma(t_o)$, we aim again at q. That is, q lies outside some forward metric ball $\mathcal{B}_{\tilde{p}}^+(\tilde{r})$ with small $\tilde{r}$, and there exists a certain point $\tilde{m}$

on $\mathcal{S}^+_{\tilde{p}}(\tilde{r})$ such that

$$d(\tilde{p},\tilde{m}) + d(\tilde{m},q) = d(\tilde{p},q) .$$

Since $t_o \in T$, adding it to this equation yields

$$(*) \qquad t_o + d(\tilde{p},\tilde{m}) + d(\tilde{m},q) = d(p,q) .$$

Let $\tilde{\sigma}$ be the geodesic from $\tilde{p}$ to $\tilde{m}$; it has length $d(\tilde{p},\tilde{m})$. Using this and $d(p,q) \leqslant d(p,\tilde{m}) + d(\tilde{m},q)$, equation (*) can be transformed to read

$$L(\sigma_{|[0,t_o]}) + L(\tilde{\sigma}) \leqslant d(p,\tilde{m}) .$$

But the union of $\sigma_{|[0,t_o]}$ with $\tilde{\sigma}$ is itself a curve from p to $\tilde{m}$. This means that

$$d(p,\tilde{m}) \leqslant L(\sigma_{|[0,t_o]}) + L(\tilde{\sigma}) .$$

So we actually have

$$L(\sigma_{|[0,t_o]}) + L(\tilde{\sigma}) = d(p,\tilde{m}) .$$

The said union is therefore minimizing. By Proposition 5.1.1, it must be C^1. Thus the two curves are joined together without kinks to form a single geodesic; namely, $\sigma(t)$, $0 \leqslant t \leqslant t_o + \tilde{r}$. Equation (*) now reads

$$(t_o + \tilde{r}) + d(\sigma_{t_o+\tilde{r}}, q) = d(p,q) ,$$

which says that $(t_o + \tilde{r}) \in T$, a contradiction. This means that $d(p,q) \in T$ after all, and we are done. □

Exercises

Exercise 6.5.1: Consider the conclusion of the proposition we just established. It is evident that we need connectedness. But where exactly did that hypothesis get used in the proof?

Exercise 6.5.2: Proposition 6.5.1 asserts that there is a minimizing geodesic from p to q; but it never claims that there is only one such geodesic.

(a) Give an example on the 2-sphere for which there are at least two minimizing geodesics joining p and q.

(b) Can you think of another example, but not on the 2-sphere?

(c) Is there an example in which that minimizing geodesic is always unique?

Exercise 6.5.3: Consider the following statement:

> If M is connected and if $\exp_p$ is defined on all of T_pM for some $p \in M$, then any two arbitrary points q_1, q_2 in M can be connected by a minimizing geodesic.

Do you think it is true? Do keep in mind that the reverse of any geodesic from p to q_1 may not be a geodesic at all! This rather pathological issue was addressed in Exercise 5.3.3.

6.6 The Hopf–Rinow Theorem

A Finsler manifold (M, F) is said to be **forward geodesically complete** if every geodesic $\gamma(t)$, $a \leqslant t < b$, parametrized to have constant Finslerian speed, can be extended to a geodesic defined on $a \leqslant t < \infty$. The Hopf–Rinow theorem gives several characterizations of this completeness. For its proof, we again borrow the streamlined arguments from [ON]. However, we do have to exercise care because the metric distance in question may not be symmetric. Compare our exposition here with that given by Dazord [Daz].

Theorem 6.6.1 (Hopf–Rinow). *Let (M, F) be any connected Finsler manifold, where F is positively (but perhaps not absolutely) homogeneous of degree one. The following five criteria are equivalent:*

(a) *The "metric" space (M, d) is forward complete.*
(b) *The Finsler manifold (M, F) is forward geodesically complete.*
(c) *At every point $p \in M$, $\exp_p$ is defined on all of T_pM.*
(d) *At some point $p \in M$, $\exp_p$ is defined on all of T_pM.*
(e) *Every closed and forward bounded subset of (M, d) is compact.*

Furthermore, if any of the above holds, then every pair of points in M can be joined by a minimizing geodesic. More details about that geodesic are given in Proposition 6.5.1.

Remarks: The following subtleties arise because the Finsler structure is generically only positively homogeneous of degree 1, hence its associated metric distance d is typically nonsymmetric.

* Forward completeness as a metric space means that every forward Cauchy sequence (§6.2D) is convergent. Exercise 6.2.5 explains why there is no need to attach the adjective "forward" to the notion of convergence. Exercise 6.2.11 shows that **compact Finsler manifolds are always forward complete** (and backward complete).
* The Hopf–Rinow theorem only involves forward geodesic completeness. There are other theorems (see, for instance, §12.4) that also require the notion of **backward geodesic completeness**. Namely, every geodesic $\gamma(t)$, $a < t \leqslant b$, parametrized to have constant Finslerian speed, should be extendible to a geodesic defined on $-\infty < t \leqslant b$. In the Finslerian realm, these two notions are *not* equivalent, unlike Riemannian geometry. See §12.6 for an explicit example.

* **If F is absolutely homogeneous of degree one, then forward and backward geodesic completeness either both hold or both fail.** This is the case for Riemannian metrics. **However, the converse is false!** Namely, forward and backward geodesic completeness can both hold without F having to be absolutely homogeneous. Such a phenomenon is illustrated by locally Minkowskian spaces. There, the geodesics with constant Finslerian speed are *affinely* parametrized straight lines in some suitable coordinate systems (see Exercise 5.3.4). A moment's thought shows that these geodesics are indeed both forward and backward infinitely extendible.

* A subset is said to be **forward bounded** if it is contained in some forward metric ball $\mathcal{B}_p^+(r)$, defined in (6.2.1). That subset is **backward bounded** if it is contained in some backward metric ball, defined in Exercise 6.2.10. In the Finsler setting, forward boundedness and backward boundedness are two distinct concepts. This is also explicitly illustrated in §12.6.

* The concept of a closed subset is best defined in terms of the underlying manifold topology of M. According to §6.2C, the manifold topology coincides with that given by the forward metric balls.

Proof.

- Suppose (a) holds. Let $\gamma(t)$, $a \leqslant t < b$ be a maximally forward extended geodesic. Without loss of generality, let us assume that it has unit speed. If $b \neq \infty$, take an *increasing* sequence $\{t_i\}$ in $[a, b)$ converging to b. The corresponding sequence $\{\gamma(t_i)\}$ is forward Cauchy because, having unit speed implies that for $i \leqslant j$, we have

$$d(\,\gamma(t_i)\,,\,\gamma(t_j)\,) \;\leqslant\; t_j - t_i\;.$$

  Hence by (a) it converges to some $q \in M$. Define $\gamma(b)$ as q. Now ODE theory tells us that $\gamma(t)$ must in fact be defined on a neighborhood of $t = b$, which is a contradiction. Thus $b = \infty$ after all, and we have forward geodesic completeness.

- Suppose (b) holds. Criterion (c) is then automatic. And of course (c) implies (d).

- Suppose (d) holds. Let $\mathcal{A}$ be a closed and forward bounded subset of (M, d). For each $q \in \mathcal{A}$, Proposition 6.5.1 ensures that there exists a minimizing geodesic $\exp_p(tv_q)$, $0 \leqslant t \leqslant 1$ from p to q. The collection of all v_q is a subset A of T_pM. This subset is bounded because $F(p, v_q) = d(p, q)$ which, by the forward boundedness of $\mathcal{A}$, is $\leqslant$ some r independent of q. In other words, A is contained in the compact set $\mathrm{B}_p(r) \cup \mathrm{S}_p(r)$ and $\exp_p\mathrm{A} = \mathcal{A}$. By being closed and sitting inside the compact set

$$\exp_p[\,\mathrm{B}_p(r) \cup \mathrm{S}_p(r)\,] \ ,$$

$\mathcal{A}$ must itself be compact. This establishes (e).

- Finally, suppose (e) holds. Let $\{p_i\}$ be a forward Cauchy sequence in M. By the triangle inequality (see §6.2), it must be forward bounded. The closure of its point set, with respect to the manifold topology, is again forward bounded; see Exercise 6.6.2. This closure, denoted $\mathcal{A}$, must then be compact, in view of the Heine–Borel hypothesis (e).

  We claim that the sequence $\{p_i\}$ therefore contains a convergent subsequence. To see this, suppose it has no convergent subsequence. Then at each $p \in \mathcal{A}$, there exists a small forward metric ball (centered at p) that contains no point of $\{p_i\}$ except perhaps at its center. By §6.2C, these forward metric balls are open in the manifold topology. Since they form an open cover for the compact $\mathcal{A}$, we can extract a finite subcover. By construction, each forward ball in this finite subcover can contain at most one point of our sequence. Consequently, the point set of the said sequence is finite. In which case there surely is at least one convergent subsequence, and that is the contradiction we seek.

  In order to avoid subscripts upon subscripts, let us denote the convergent subsequence as $\{p_\alpha\}$. Say it converges to some $q \in \mathcal{A} \subset M$. We now check that $\{p_i\}$ must converge to q as well. To this end, let $\epsilon > 0$ be given. Since the sequence $\{p_i\}$ is forward Cauchy, there is an N_1 such that if $N_1 \leqslant i < j$, then $d(p_i, p_j) < \frac{\epsilon}{2}$. Also, the subsequence $\{p_\alpha\}$ converges to q. Thus there exists an N_2 such that if $\alpha \geqslant N_2$, then $d(q, p_\alpha) < \frac{\epsilon}{2}$.

  Let N be the larger of N_1 and N_2. By further increasing N if necessary, we may assume without loss of generality that it actually equals some index in the convergent subsequence. And, since N is at least as large as N_2, one sees that $d(q, p_{_N}) < \frac{\epsilon}{2}$. Thus, whenever $i \geqslant N$, we have

$$d(q,p_i) \ \leqslant \ d(q,p_{_N}) \ + \ d(p_{_N},p_i) \ \ < \ \ \frac{\epsilon}{2} + \frac{\epsilon}{2} \ \ = \ \ \epsilon \ .$$

  Given (e), we have just demonstrated that every forward Cauchy sequence is convergent. Hence (M, d) is forward complete.

This finishes the proof of the theorem. □

Compare our treatment of completeness with that in Udriste [U1], [U2]. Keep in mind that, for generic Finsler metrics, there is a distinction between forward and backward completeness.

Exercises

Exercise 6.6.1: This concerns the beginning of our proof of the Hopf–Rinow theorem.

(a) Suppose, instead of an increasing sequence, $\{t_i\}$ is any sequence in $[a, b)$ that converges to b. Explain why $d(\gamma(t_i), \gamma(t_j))$ is *not* necessarily bounded above by $|t_i - t_j|$.

(b) Is the definition of $\gamma(b)$ in the proof independent of our choice of the convergent sequence?

(c) Also, provide the precise ODE statements to which we alluded.

Exercise 6.6.2: Let $\mathcal{A}$ be any subset of a Finsler manifold M, with distance function d induced by the Finsler structure F. A point $x \in M$ is said to be in the closure of $\mathcal{A}$ if there is a sequence $\{x_i\}$ in $\mathcal{A}$ that, under the manifold topology of M, converges to x.

(a) Suppose x is in the closure of $\mathcal{A}$. Prove that for each $\epsilon > 0$, there exists an $x_i \in \mathcal{A}$ such that $d(x_i, x) < \epsilon$. Hint: you may need to consult Exercise 6.2.5.

(b) Use part (a) and the triangle inequality (see §6.2) to show that the closure of a forward bounded subset is again forward bounded.

Exercise 6.6.3:

(a) Show that any two points p, q of a forward complete connected Finsler manifold can be joined by a minimizing geodesic.

(b) Revisit Exercise 6.5.3 and make a better-informed decision.

Exercise 6.6.4: Prove that:

A compact Finsler manifold (M, F) is both forward and backward geodesically complete, whether F is absolutely homogeneous or only positively homogeneous.

Hint: Exercise 6.2.11 may be helpful.

Exercise 6.6.5: Let F be a Finsler structure that is positively (but perhaps not absolutely) homogeneous of degree one. Define

$$\tilde{F}(x, y) := F(x, -y) .$$

(a) Verify that $\tilde{F}$ satisfies all the axioms (§1.1) of a Finsler structure.

(b) Suppose σ is a geodesic of the Finsler structure F. Prove that the reverse of σ is a geodesic of the Finsler structure $\tilde{F}$.

Exercise 6.6.6: Let (M, F) be a connected Finsler manifold, where F is positively (but perhaps not absolutely) homogeneous of degree one. Suppose the exponential map of $\tilde{F}$ (as described in Exercise 6.6.5) is defined on all of some tangent space T_pM. Can you formulate and prove a companion version of Proposition 6.5.1?

Exercise 6.6.7: Can you prove the following companion of the Hopf–Rinow theorem? (Again, see Exercise 6.6.5 for the definition of $\tilde{F}$ in terms of F.)

Let (M, F) be any connected Finsler manifold, where F is positively (but perhaps not absolutely) homogeneous of degree one. The following five criteria are equivalent:

(a) The "metric" space (M, d) is backward complete.

(b) The Finsler manifold (M, F) is backward geodesically complete.

(c) At every $p \in M$, the exponential map of $\tilde{F}$ is defined on all of T_pM.

(d) At some $p \in M$, the exponential map of $\tilde{F}$ is defined on all of T_pM.

(e) Every closed and backward bounded subset of (M, d) is compact.

References

[BC1] D. Bao and S. S. Chern, *On a notable connection in Finsler geometry*, Houston J. Math. **19** (1993), 135–180.

[BuMa] H. Busemann and W. Mayer, *On the foundations of the calculus of variations*, Trans. AMS **49** (1941), 173–198.

[Daz] P. Dazord, *Propriétés globales des géodésiques des Espaces de Finsler*, Theses, Université de Lyon, 1969.

[ON] B. O'Neill, *Semi-Riemannian Geometry, with Applications to Relativity*, Academic Press, 1983.

[U1] C. Udriste, *Appendix 4. Completeness and convexity on Finsler manifolds*, Convex Functions and Optimization Methods on Riemannian Manifolds, MAIA 297, Kluwer Academic Publishers, 1994, pp. 318–330.

[U2] C. Udriste, *Completeness of Finsler manifolds*, Publ. Math. Debrecen **42** (1993), 45–50.

[W] J. H. C. Whitehead, *Convex regions in the geometry of paths*, Quarterly J. Math. Oxford, Ser. 3 (1932), 33–42.

Chapter 7

The Index Form and the Bonnet–Myers Theorem

7.1 Conjugate Points

Let (M, F) be a Finsler manifold, where F is C^∞ on $TM \setminus 0$ and is positively (but perhaps not absolutely) homogeneous of degree 1. Fix $T \in T_pM$. Consider the constant speed geodesic $\sigma(t) = \exp_p(tT)$, $0 \leqslant t \leqslant r$ that emanates from $p = \sigma(0)$ and terminates at $q = \sigma(r)$. If there is no confusion, label its velocity field by T also. Let D_T denote covariant differentiation along σ, with reference vector T. This concept was introduced in the Exercise portion of §5.2.

- Recall from §5.4 that a vector field J along σ is said to be a **Jacobi field** if it satisfies the equation

$$D_T D_T J \;+\; R(J,T)T \;=\; 0\,.$$

- We say that q **is conjugate to** p **along** σ if there exists a nonzero Jacobi field J along σ which vanishes at p and q. More precisely, $J(t)$ vanishes at $t = 0$ and $t = r$. For such a Jacobi field J, Exercise 5.4.6 tells us that J and $J' = D_T J$ must both be g_T-orthogonal to T at all times.

Proposition 7.1.1. *Let $\sigma(t) = \exp_p(tT)$, $0 \leqslant t \leqslant r$ be a constant speed geodesic from $p = \sigma(0)$ to $q = \sigma(r)$. The following five statements are mutually equivalent:*

(a) *The point q is not conjugate to p along σ.*

(b) *Any Jacobi field that vanishes at both points p and q must be identically zero along σ.*

(c) *Take the variation field of any variation of σ by geodesics. If it vanishes at p and q, then it must be identically zero along σ.*

(d) *Given any $v \in T_pM$ and $w \in T_qM$, there exists a unique Jacobi field J along σ that equals v at p and w at q.*

(e) *The derivative $\exp_{p*}$ of the exponential map $\exp_p$ is nonsingular at the location rT in T_pM.*

Proof. By definition, **(b) and (a) are equivalent.**

That (b) and (c) are equivalent:

Suppose (b) holds. Consider any variation whose t-curves are all geodesics. According to §5.4, its variation vector field U is a Jacobi field. Thus, if U is zero at p and q, (b) will force it to vanish identically. This gives (c).

Conversely, suppose (c) holds. Every Jacobi field J along σ is the variation vector field of some variation of σ by geodesics. (This fact was hinted at in Exercise 5.4.1. We now provide guidance for proving it in Exercise 7.1.1.) Thus, if J is zero at p and q, (c) will guarantee that it is identically zero along σ. (Incidentally, the vanishing of the variation field at p and q does *not* imply that the variation keeps the endpoints fixed.)

That (b) and (d) are equivalent:

It is quite transparent that (d) implies (b). So, let us concentrate on the converse. Suppose (b) holds.

Take Jacobi fields $\{V_i\}$ along σ such that $V_i(0) = 0$ and $\{V_i'(0)\}$ are linearly independent. These are nonzero by ODE theory. We claim that $\{V_i(r)\}$ must be linearly independent. Otherwise we have a Jacobi field $c^i V_i$ (for some constants c^i, not all zero) that vanishes at both p, q and is nonzero because $c^i V_i'(0) \neq 0$, thereby violating (b). Since $V_i(0) = 0$ and $\{V_i(r)\}$ are linearly independent, we see that given any $w \in T_qM$, there exists a unique Jacobi field J (built out of a linear combination of the V_i) along σ such that $J(0) = 0$ and $J(r) = w$.

The above conclusion has a "symmetrical" counterpart: given any $v \in T_pM$, there exists a unique Jacobi field J along σ such that $J(0) = v$ and $J(r) = 0$. This, however, does *not* follow from considering the reverse of σ because it may not even be a geodesic. Instead, take nonzero Jacobi fields $\{V_i\}$ along σ such that $V_i(r) = 0$ and $\{V_i'(r)\}$ are linearly independent; then $\{V_i(0)\}$ must be linearly independent by an argument that parallels the one above. Since $\{V_i(0)\}$ are linearly independent and $V_i(r) = 0$, we see that given any $v \in T_pM$, there exists a unique Jacobi field J (again, a linear combination of the V_i) along σ such that $J(0) = v$ and $J(r) = 0$.

Adding the results of these two paragraphs tells us that there is a Jacobi field J along σ such that $J(0) = v$ and $J(r) = w$. The uniqueness of this J follows from yet another application of (b). We have just obtained (d).

That (b) and (e) are equivalent:

Suppose (b) holds. Let W be any element in the null space of $\exp_{p*(rT)}$. Consider the Jacobi field $U(t) := \exp_{p*(tT)}(tW)$ discussed in §5.4. Since this U vanishes at both $t = 0$ and $t = r$, it must be identically zero by (b). Now, (5.4.4) tells us that U has initial data $U(0) = 0$ and $U'(0) = W$. Hence W must be zero, or else $U(t)$ could not vanish identically. This gives criterion (e).

Conversely, suppose (e) holds. Take any Jacobi field J along σ that vanishes at both p and q. Let $W := J'(0)$. In view of (5.4.3), (5.4.4), we know that $U(t) := \exp_{p*(tT)}(tW)$ has the same initial data as $J(t)$. Thus they must be equal by uniqueness. In particular,

$$0 \;=\; J(r) \;=\; U(r) \;=\; \exp_{p*(rT)}(rW)\ .$$

This shows that rW is in the null space of $\exp_{p*(rT)}$. By (e), we have $W = 0$. Therefore U vanishes identically and so must J. □

Exercises

Exercise 7.1.1: Prove that every Jacobi field J along a geodesic $\sigma(t)$ is the variation vector field of a variation by geodesics. Consider the following suggestions.

(a) Find a curve $u \mapsto \gamma(u)$ with velocity $V(u)$ such that $V_{|u=0} = J_{|t=0}$.

(b) Explain why there are vectors $Y(u)$ along γ such that $Y_{|u=0} = T_{|t=0}$ and $(D_V Y)_{|u=0} = J'_{|t=0}$. Here, $J' := D_T J$ is, as usual, covariant differentiation along σ with reference vector T. On the other hand, $D_V Y$ is covariant differentiation along γ with reference vector Y, not V.

(c) Define the variation as $\sigma(t,u) := \exp_{\gamma(u)}[t\,Y(u)]$. Draw a schematic picture of the t-curves in this variation.

(d) Check that the variation vector field U satisfies $U_{|t=0} = J_{|t=0}$ and $(D_T U)_{(0,0)} = (D_U T)_{(0,0)} = (D_V Y)_{|u=0} = J'_{|t=0}$.

(e) Why can you conclude that J is in fact equal to U ?

Exercise 7.1.2: State the contrapositive of Proposition 7.1.1. What does criterion (d) become?

Exercise 7.1.3: We proved Proposition 7.1.1 by checking that statements (a), (c), (d), (e) are individually equivalent to statement (b). Are you able to restructure the thinking to result in $(b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (e) \Rightarrow (b)$ instead?

Exercise 7.1.4: Consider a geodesic σ that goes from p to q. According to Proposition 7.1.1, the conjugacy of q to p is equivalent to:

> Whenever the variation field—of any variation of σ by geodesics—vanishes at p and q, it must be identically zero.

(a) How does the existence of conjugate points relate to the focusing of geodesics issued from p?

(b) Explain why geodesics in spaces with negative flag curvature do not admit conjugate points.

(c) Do geodesics in spaces with positive flag curvature have to have conjugate points?

(d) What about geodesics in spaces with non-positive flag curvature?

Hint: for parts (b) and (c), review §5.5.

7.2 The Index Form

Suppose the geodesic $\sigma(t) = \exp_p(tT)$, $0 \leqslant t \leqslant r$ has constant speed c. As before, let T denote its velocity field as well as its initial velocity. The **index form** was first introduced in Exercise 5.2.7. In the present setting, it reads:

$$\boxed{\begin{aligned} I(V,W) \ := \ \frac{1}{c} \int_0^r [\ & g_T(D_T V \, , \, D_T W) \\ & - g_T(R(V,T)T \, , \, W) \ ] \, dt \end{aligned}} \ . \tag{7.2.1}$$

The vector fields V, W are assumed to be continuous and piecewise C^∞ along σ. As always, all D_T are calculated with reference vector T, and the curvature in

$$\boxed{R(V,T)T \ := \ (T^j \, R_j{}^i{}_{kl} \, T^l) \, V^k \, \frac{\partial}{\partial x^i}}$$

is evaluated at the point (σ, T).

It is a consequence of (3.4.6), (3.4.7) that the object $y^j \, R_{jikl} \, y^l$ is symmetric in the indices i and k. That is,

$$y^j \, R_{jkil} \, y^l \ = \ y^j \, R_{jikl} \, y^l \ .$$

Using this, it is almost immediate that

$$\boxed{I(W,V) \;=\; I(V,W)} \,. \tag{7.2.2}$$

So the index form is a symmetric bilinear form.

Let $0 =: t_o < t_1 < \cdots < t_k := r$ be a partition of $[0,r]$ such that V, W are both C^∞ on each closed subinterval $[t_{i-1}, t_i]$. Integration by parts, which is made possible by Exercise 5.2.3, tells us that

$$\begin{aligned} I(V,W) \;:=\; & \frac{1}{c}\, g_T(\,D_T V\,,\,W)\Big|_0^r \\ & - \frac{1}{c} \sum_{i=1}^{k-1} g_T(\,D_T V\,,\,W\,)\Big|_{t_i^-}^{t_i^+} \\ & - \frac{1}{c} \int_0^r g_T\Big(\,D_T D_T V \;+\; R(V,T)T\,,\,W\Big)\,dt\,. \end{aligned}$$

The sum $\sum_{i=1}^{k-1}$ collapses to zero if V is C^1 on the entire interval $[0,r]$. So

$$\boxed{I(J,W) \;=\; \frac{1}{c}\, g_T(\,J'\,,\,W\,)\Big|_0^r \quad \text{if } J \text{ is a Jacobi field}} \,. \tag{7.2.3}$$

In particular:

$$\boxed{I(J,J) \;=\; \frac{1}{c}\, g_T(\,J'\,,\,J\,)\Big|_0^r \quad \text{for all Jacobi fields } J} \,. \tag{7.2.4}$$

Here, $J' := D_T J$, where T is the velocity field of our constant speed geodesic σ. This covariant derivative is computed with reference vector T. See the exercises of §5.2.

Exercises

Exercise 7.2.1: Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a geodesic. In this exercise, all vector fields $V(t)$, $W(t)$ are *a priori* continuous and piecewise C^∞ along σ. Suppose V vanishes at the endpoints of σ. Show that the following two statements are equivalent:

(a) V is a Jacobi field.
(b) $I(V,W) = 0$ for all fields W that vanish at the endpoints of σ.

Hint: imitate the *spirit* of Proposition 5.1.1's proof.

Exercise 7.2.2: Restrict the index form $I(V,W)$ to continuous piecewise C^∞ vector fields V, W that *vanish at the endpoints* of the geodesic $\sigma(t)$, $0 \leqslant t \leqslant r$. Consider the space of continuous piecewise C^∞ fields V such that $I(V,W)$ vanishes for every W. Explain why:

(a) This space is finite-dimensional.

(b) Its elements are globally C^∞ on $[0,r]$.

(c) Does the dimension of this space have any geometrical significance?

Exercise 7.2.3: Let J be a *Jacobi field* along the constant speed geodesic $\sigma(t)$, $0 \leqslant t \leqslant r$ with speed c. **Suppose** $J(0) = 0$. Use Exercise 5.2.3 to explain why formula (7.2.4) can then be rewritten as

$$\boxed{I(J,J) \;=\; \frac{1}{2c}\,\frac{d}{dt}_{|t=r}\; g_T(J,J)}\ .$$

This says that $I(J,J)$ controls a certain growth rate of the Jacobi field J, provided that the latter starts out with zero length.

Exercise 7.2.4: Let W be any piecewise C^∞ vector field along a geodesic $\sigma(t)$, $0 \leqslant t \leqslant r$. Suppose J is a Jacobi field along σ that happens to have the same boundary values as W. Namely,

$$J(0) \;=\; W(0) \quad \text{and} \quad J(r) \;=\; W(r)\ .$$

Use (7.2.3) to show that

$$I(J,W) \;=\; I(J,J)\ .$$

Exercise 7.2.5: Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a constant speed geodesic with velocity T and speed c. Suppose W is a continuous piecewise C^∞ vector field that is g_T-orthogonal to σ. Show that

$$\boxed{\begin{aligned} I(W,W) \;:=\; \frac{1}{c}\int_0^r [\;\; & g_T(\,D_T W\,,\,D_T W\,) \\ & - K(T,W)\,c^2\,g_T(W,W)\;\,]\,dt \end{aligned}}\ ,$$

where $K(T,W)$ is the flag curvature of the flag with flagpole T and transverse edge W.

Exercise 7.2.6: Assume throughout this exercise that (M,F) has scalar flag curvature $\lambda(x,y)$. This holds, for example, whenever the flag curvature is constant, or if we are working with a Finsler surface. Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a constant speed geodesic with velocity T and speed c. Let E be a *parallel* vector field (constructed using reference vector T) along σ such that $g_T(E,T) = 0$ and $g_T(E,E) = 1$.

(a) Suppose J is a Jacobi field that is g_T-orthogonal to σ and has the form $\mathcal{F}\,E$. Recall from Exercise 5.4.5 that the function $\mathcal{F}$ satisfies the scalar Jacobi equation

$$\mathcal{F}'' \;+\; \lambda\,c^2\,\mathcal{F} \;=\; 0\ .$$

Show that

$$I(J,J) \;=\; \frac{1}{c}\,\Big[\mathcal{F}'(r)\,\mathcal{F}(r) \;-\; \mathcal{F}'(0)\,\mathcal{F}(0)\Big]\ .$$

(b) Let W be a continuous piecewise C^∞ vector field that is g_T-orthogonal to σ and has the form $f\,E$. Show that

$$I(W,W) \;=\; \frac{1}{c}\int_0^r [\,(f')^2 \;-\; \lambda\, c^2\, f^2\,]\; dt\;.$$

7.3 What Happens in the Absence of Conjugate Points?

Consider the following technical setup:

* Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a geodesic that emanates from p. Denote its velocity field by T. Suppose this geodesic contains no point that is conjugate to p.
* Let $\{J_i\}_{i=1}^n$ be Jacobi fields along σ such that $J_i(0) = 0$ and $\{J_i'(0)\}$ are linearly independent. Then $\{J_i(t)\}$ must be linearly independent for each $t \in (0, r]$. Otherwise, there would exist constants c^i (not all zero) and $t_o \in (0, r]$ such that $c^i\, J_i(t_o) = 0$. The Jacobi field $J(t) := c^i\, J_i(t)$ would then vanish at both 0 and t_o, but is nonzero because $J'(0) = c^i\, J_i'(0) \neq 0$. This would violate (see Proposition 7.1.1) the hypothesis that $\sigma(t_o)$ is not conjugate to p.
* Let W be any piecewise C^∞ vector field along σ.

Our plan in this section is as follows:

- Let us first work with the simpler situation in which $W(0) = 0$ so that we can enlist the help of the above $\{J_i\}$. There is an important payoff. It says that among all piecewise C^∞ curves that share the same endpoints, geodesics (without conjugate points) are in a certain sense the "local" minima of the arc length functional.
- After that, we bootstrap our way back up to the case of a general W. There, it is shown that among all piecewise C^∞ vector fields along σ which share the same boundary values, Jacobi fields minimize the quadratic form of the index form.

7.3 A. Geodesics Are Shortest Among "Nearby" Curves

* Suppose $W(0) = 0$. Expand W as

$$W(t) \;=\; f^i(t)\, J_i(t)\;,$$

where the component functions $f^i(t)$ are continuous and piecewise C^∞, and $f^i(0) = 0$. As in Cheeger–Ebin [CE], introduce the temporary abbreviations A, B:

$$W' \;:=\; D_T W \;=\; f^{i\,\prime}\, J_i \;+\; f^i\, J_i' \;=:\; A \;+\; B\;.$$

Note that

$$g_T(W', W') = g_T(A, A) + g_T(B, B) + g_T(A, B) + g_T(B, A) .$$

* The Lagrange identity studied in Exercise 5.4.6 assures us that

$$g_T(J_i', J_j) - g_T(J_i, J_j') = \text{constant} .$$

Evaluating at $t = 0$ shows that the constant in question is zero. So

$$g_T(J_i', J_j) = g_T(J_i, J_j') .$$

This will be needed in the second last step of the calculation we are about to perform.

Using the defining equation of Jacobi fields, Exercise 5.2.3, and the Lagrange identity we just discussed, we have

$$\begin{aligned}
& g_T(B, B) - g_T(R(W,T)T, W) \\
&= f^i f^j \left[g_T(J_i', J_j') + g_T(J_i'', J_j) \right] \\
&= f^i f^j \frac{d}{dt} g_T(J_i', J_j) \\
&= \frac{d}{dt}\left[f^i f^j g_T(J_i', J_j) \right] - f^{i\prime} f^j g_T(J_i', J_j) - f^i f^{j\prime} g_T(J_i', J_j) \\
&= \frac{d}{dt}\left[f^i f^j g_T(J_i', J_j) \right] - f^{i\prime} f^j g_T(J_i, J_j') - f^i f^{j\prime} g_T(J_i', J_j) \\
&= \frac{d}{dt}\left[f^i f^j g_T(J_i', J_j) \right] - g_T(A, B) - g_T(B, A) .
\end{aligned}$$

In the present notation,

$$I(W, W) = \frac{1}{c} \int_0^r \left[g_T(W', W') - g_T(R(W,T)T, W) \right] dt .$$

The above deliberations and computation therefore imply that

$$\begin{aligned}
I(W, W) &= \frac{1}{c} f^i(r) f^j(r) g_T(J_i'(r), J_j(r)) \\
&\quad + \frac{1}{c} \int_0^r g_T(A, A)\, dt ,
\end{aligned} \tag{7.3.1}$$

where $A := f^{i\prime} J_i$.

In particular, if W vanishes at $t = r$ as well, so that $f^i(r) = 0$, then

$$I(W, W) = \frac{1}{c} \int_0^r g_T(A, A)\, dt \geqslant 0 .$$

Note that in this case $I(W, W)$ can equal zero only when $f^{i\prime}(t) = 0$, that is, when each $f^i(t)$ is constant. But $f^i(0) = 0 = f^i(r)$, so every $f^i(t)$ must vanish and hence $W = 0$. We have just established the first part of the following proposition.

Proposition 7.3.1.

- *Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a constant speed geodesic that emanates from some point p in a Finsler manifold (M, F). Denote its constant speed by c.*
- *Suppose no point $\sigma(t)$, $0 < t \leqslant r$ is conjugate to p.*

Then:

(1) *For all piecewise C^∞ vector fields $W(t)$ along σ such that $W(0) = 0 = W(r)$, we have $I(W, W) \geqslant 0$. Equality holds if and only if W is identically zero.*

(2) *The geodesic σ is a minimum among all "nearby" piecewise C^∞ curves from p to $q := \sigma(r)$.*

Remarks: Being a geodesic, σ is a critical curve of the length functional L. An obvious question is whether this critical σ is a minimum or a maximum of L among "nearby" piecewise C^∞ curves from p to q. The comparison with "nearby" curves renders this notion of maximas and minimas a "local" one. Quantifying the meaning of "nearby" requires more work than what we are prepared to do here.

Proof. It remains to prove the second assertion.

Using the index form, the second variation of arc length is succinctly expressed as

$$L''(0) \;=\; I(U_\perp, U_\perp) \;+\; \frac{1}{c}\, g_T(\, D_U U \,,\, T\,)\Big|_0^r \,,$$

with $U_\perp := U - \frac{1}{c^2}\, g_T(U, T)\, T$. This was carried out in Exercise 5.2.7.

Consider only piecewise C^∞ variations in which the t-curves begin at p and end at q. In that case, the variation vector field U satisfies: $U(0, u) = 0 = U(r, u)$ for all u. Hence $D_U U$ vanishes at all $(0, u)$ and (r, u), and the boundary term in $L''(0)$ drops out. In other words,

$$L''(0) \;=\; I(U_\perp, U_\perp) \quad \text{if the variation has fixed endpoints.}$$

Let us work only with variations in which no two t-curves have the same trajectory. By reparametrizing these variations if necessary, we can ensure that $U_\perp$ is not identically zero. In view of the first conclusion of our proposition, the index $I(U_\perp, U_\perp)$ must then be positive. Consequently, the above formula implies that $L''(0) > 0$.

By the second derivative test, the function $L(u)$ has a local minimum at the critical $u = 0$. This means that for small u, the length $L(u)$ of the corresponding t-curve is larger than that of σ . $\square$

7.3 B. A Basic Index Lemma

Return to formula (7.3.1). Since $q := \sigma(r)$ is not conjugate to p, it can be checked that $J(t) := f^i(r)\, J_i(t)$, $0 \leqslant t \leqslant r$ is the unique Jacobi field along σ which has the same boundary values as W. That is, $J(0) = 0$ and $J(r) = W(r)$. Also, (7.2.4) helps us realize that

$$\frac{1}{c}\; f^i(r)\, f^j(r)\; g_T(\, J_i'(r)\,,\; J_j(r)\,) \;=\; I(J,J)\;.$$

Thus (7.3.1) can be summarized as follows:

- Suppose no point along σ is conjugate to p.
- Let $W(t) = f^i(t)\, J_i(t)$ be any piecewise C^∞ vector field along σ that *vanishes* at $t = 0$.
- Let $J(t)$ be the unique Jacobi field along σ that has the same boundary values as $W(t)$.

Then:

$$I(W,W) \;=\; I(J,J) \;+\; \frac{1}{c} \int_0^r g_T(\, f^{i\,\prime}\, J_i\,,\; f^{j\,\prime}\, J_j\,)\; dt\;. \tag{7.3.2}$$

In particular,

$$I(W,W) \;\geqslant\; I(J,J)\;. \tag{*}$$

We now generalize (*) to the case in which $W(0)$ does not necessarily vanish. Of course, J still stands for the unique Jacobi field along σ that has the same boundary values as W. The following is known as a "**basic index lemma.**"

Lemma 7.3.2.

- *Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a geodesic in a Finsler manifold (M,F). Suppose no point $\sigma(t)$, $0 < t \leqslant r$ is conjugate to $p := \sigma(0)$.*
- *Let W be any piecewise C^∞ vector field along σ.*
- *Let J denote the unique Jacobi field along σ that has the same boundary values as W. That is, $J(0) = W(0)$ and $J(r) = W(r)$.*

Then

$$\boxed{\,I(W,W) \;\geqslant\; I(J,J)\,}\;. \tag{7.3.3}$$

Equality holds if and only if W is actually a Jacobi field, in which case the said J coincides with W.

Proof. Since $q := \sigma(r)$ is not conjugate to p along σ, Proposition 7.1.1 guarantees the existence of the said J. Apply the first part of Proposition 7.3.1 to the piecewise C^∞ vector field $W - J$. This requires the hypothesis that no point along σ is conjugate to p. We conclude that

$$I(W - J\,,\, W - J) \;\geqslant\; 0\;,$$

and $I(W - J\,,\, W - J) = 0$ if and only if $W - J = 0$.

To finish the proof, it suffices to note that

$$I(W - J\,,\, W - J) \;=\; I(W,W) - I(J,J)\;.$$

Indeed, $I(W-J\,,\, W-J) \;=\; I(W,W) - 2\,I(J,W) + I(J,J)$; but (7.2.3) and the defining properties of J tell us that $I(J,W)$ is numerically the same as $I(J,J)$. □

Exercises

Exercise 7.3.1: The computation leading up to (7.3.1) could benefit from a few more details. Fill those in.

Exercise 7.3.2: Consider formula (7.3.1). It was derived under the assumption that no point $\sigma(t)$, $0 < t \leqslant r$ is conjugate to p.

(a) Explain why $J(t) := f^i(r)\, J_i(t)$, $0 \leqslant t \leqslant r$ is the *unique* Jacobi field along σ that has the same boundary values as W. That is, $J(0) = 0$ and $J(r) = W(r)$.

(b) With the help of (7.2.4), show that for this Jacobi field,

$$I(J,J) \;=\; \frac{1}{c}\; f^i(r)\, f^j(r)\; g_T\Big(\, J_i'(r)\,,\, J_j(r)\,\Big)\;.$$

Exercise 7.3.3: Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a constant speed geodesic with velocity T and speed c. Suppose it contains no conjugate points.

(a) Let W be any continuous piecewise C^∞ vector field that is g_T-orthogonal to σ. Let J be the unique Jacobi field J that has the same boundary values as W. Explain why J must also be g_T-orthogonal to σ.

(b) Now suppose (M,F) has scalar flag curvature $\lambda(x,y)$. Restrict the basic index lemma to vector fields W that are g_T-orthogonal to σ. In conjunction with Exercise 7.2.6, deduce the following **comparison result for functions**:

Let f be any continuous piecewise C^∞ function on $[0,r]$. Let $\mathcal{F}$ be the unique solution of $\mathcal{F}'' + \lambda\, c^2\, \mathcal{F} = 0$ [here, $\lambda(\sigma,T)$ is a function of t], subject to the boundary data

$$\mathcal{F}(0) \;:=\; f(0)\;,\quad \mathcal{F}(r) \;:=\; f(r)\;.$$

Then

$$\boxed{\Big[\mathcal{F}'(r)\,\mathcal{F}(r) \;-\; \mathcal{F}'(0)\,\mathcal{F}(0)\Big] \;\leqslant\; \int_0^r [\,(f')^2 \;-\; \lambda\, c^2\, f^2\,]\; dt}\;,$$

with equality if and only if f is identical to $\mathcal{F}$.

(c) Suppose, in part (b), we further demand that λ be constant and $f(0) = 0$. Check that the comparison function $\mathcal{F}$ is given as follows:

$$\lambda > 0 : \quad \mathcal{F} = \frac{f(r)}{\sin(\sqrt{\lambda}\, c\, r)} \sin(\sqrt{\lambda}\, c\, t)$$

$$\lambda = 0 : \quad \mathcal{F} = \frac{f(r)}{r}\, t$$

$$\lambda < 0 : \quad \mathcal{F} = \frac{f(r)}{\sinh(\sqrt{-\lambda}\, c\, r)} \sinh(\sqrt{-\lambda}\, c\, t)\ .$$

7.4 What Happens If Conjugate Points Are Present?

The previous section provides us with a fairly good understanding of the quantity $I(W, W)$ and its ramifications when no point $\sigma(t)$, $0 < t \leqslant r$ is conjugate to p. We now investigate what happens when some point $\sigma(t_o)$, $0 < t_o < r$ is conjugate to p. Note that we have specifically excluded the borderline case in which $q := \sigma(r)$ is the *first* point conjugate to p along σ. This special case is addressed in §7.5 and some of its accompanying exercises.

Proposition 7.4.1.

- *Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a constant speed geodesic in a Finsler manifold (M, F). Denote its constant speed by c.*
- *Suppose some point $\sigma(t_o)$, $0 < t_o < r$ is conjugate to $p := \sigma(0)$.*

Then:

(1) *There exists a continuous piecewise C^∞ vector field $U(t)$ along σ, with $U(0) = 0 = U(r)$, such that $I(U, U) < 0$.*

(2) *Among the piecewise C^∞ curves from p to $q := \sigma(r)$, there are some that are "near" σ but have shorter lengths.*

Remark: The proof here mimics that in Cheeger–Ebin [CE].

Proof.

The first assertion:

Since $\sigma(t_o)$ is conjugate to p, there exists a nonzero Jacobi field $J(t)$, $0 \leqslant t \leqslant t_o$ along (a portion of) σ such that $J(0) = 0 = J(t_o)$. Extend J to a piecewise C^∞ vector field $V(t)$, $0 \leqslant t \leqslant r$ by assigning $V(t) = 0$ on $[t_o\, , r]$. We have $I(V, V) = 0$. This is seen by using (7.2.4) on $[0\, , t_o]$ and inspection on $[t_o\, , r]$. We perturb V appropriately to give the desired vector field U.

Choose a small positive δ such that there are no conjugate pairs on $[t_o - \delta\,,\, t_o + \delta]$. By Proposition 7.1.1, there exists a unique Jacobi field $K(t)$, $t_o - \delta \leqslant t \leqslant t_o + \delta$ with the prescribed values $K(t_o - \delta) = V(t_o - \delta) = J(t_o - \delta)$ and $K(t_o + \delta) = V(t_o + \delta) = 0$ at the endpoints. Since V is not smooth at t_o, we know that K has to be different from V.

Define a continuous piecewise C^∞ vector field $U(t)$, $0 \leqslant t \leqslant r$ as follows.

* On $[0\,,\, t_o - \delta]$, set $U(t) := V(t) = J(t)$.
* On $[t_o - \delta\,,\, t_o + \delta]$, set $U(t) := K(t)$; thus on this portion, U is a Jacobi field having the same endpoint values as V.
* On the remainder of the interval, namely, $[t_o + \delta\,,\, r]$, we simply prescribe $U(t) := V(t) = 0$.

This piecewise smooth U is the sought perturbation of V. See Figure 7.1.

Note that:

* On $[0\,,\, t_o - \delta]$, $I(U,U) = I(V,V)$.
* On $[t_o - \delta\,,\, t_o + \delta]$, in view of our basic index Lemma 7.3.2, we have the strict inequality $I(U,U) < I(V,V)$.
* Lastly, both $I(U,U)$ and $I(V,V)$ are trivially equal to zero on the subinterval $[t_o + \delta\,,\, q]$.

Hence $I(U,U) < I(V,V) = 0$ on $[0,r]$, which proves the first part of this proposition.

The second assertion:

Construct a piecewise C^∞ variation $\sigma(t,u)$ of the geodesic $\sigma(t)$, such that the latter's endpoints are kept fixed and that the variation vector field is precisely the U we have just described.

According to Exercise 5.2.7, the second variation simplifies in the present setting to

$$L''(0) \;=\; I(U,U) \;-\; \frac{1}{c} \int_0^r \left[\frac{\partial}{\partial u} F(T) \right]^2 dt \;<\; 0\,.$$

This means that if the parameter u in $\sigma(t,u)$ is kept sufficiently small, then among the family of t-curves considered by that variation, our base geodesic $\sigma(t)$ is actually the *longest*! □

Exercises

Exercise 7.4.1: In the above proof, $L''(0)$ was shown to be negative for *some* variation, not for all variations. Yet, when establishing part (2) of Proposition 7.3.1, we checked that $L''(0)$ was positive for *all* variations with $U_\perp$ not identically zero. Can you explain why the strategies were so different for the two propositions?

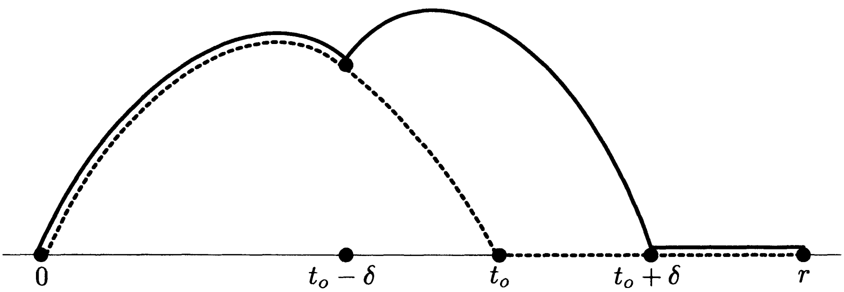

Figure 7.1

A nonzero Jacobi field $J(t)$, $0 \leqslant t \leqslant t_o$ that has zero boundary data. It is graphed here as a thickened dotted arch. The vector field V (its graph is the *entire* thickened dotted pattern) extends J by prolongating the zero data from time t_o to time r. Its index $I(V,V)$ is 0, essentially because of (7.2.4). A clever perturbation of V around the "kink," namely, on $[t_o - \delta\,,\, t_o + \delta]$, gives a piecewise smooth U (the *entire* thickened solid pattern). The basic index lemma is then used to check that $I(U,U) < 0$.

Exercise 7.4.2: Use the standard 2-sphere to illustrate the message of Proposition 7.4.1.

7.5 The Cut Point Versus the First Conjugate Point

Using the index form as a tool, we have arrived at a reasonable answer to the question posed in §7.3.A. Namely:

> Any geodesic σ from p to q is a critical curve of the length functional L. How can one tell whether this critical σ is a minimum or a maximum of L among "nearby" piecewise C^∞ curves from p to q?

The essence of Propositions 7.3.1 and 7.4.1 answers this question as follows:

Theorem 7.5.1. *Let $\sigma(t)$, $0 \leqslant t \leqslant r$ be a geodesic from $p := \sigma(0)$ to $q := \sigma(r)$ in a Finsler manifold (M, F).*

- *If no point $\sigma(t)$, $0 < t \leqslant r$ is conjugate to p, then σ minimizes arc length among "nearby" piecewise C^∞ curves from p to q.*
- *If some point $\sigma(t_o)$, $0 < t_o < r$ is conjugate to p, then among the piecewise C^∞ curves from p to q, there are some that are "near" σ but have shorter lengths. In particular, σ no longer minimizes arc length among* **all**, *nearby or not, piecewise C^∞ curves from p to q.*

Remarks: If $q := \sigma(r)$ is the *first* point conjugate to p along σ, then the conclusion can lean either way. In other words, the above "conjugate point test" is inconclusive.

* There are many examples in which the geodesic σ from p to q still minimizes arc length among nearby piecewise C^∞ curves from p to q. We readily see that on a sphere or on an ellipsoid with two equal semi-axes. These situations are easy to visualize because they possess a lot of symmetry.
* However, it is much harder to come up with examples in which a geodesic σ connecting the conjugate pair p, q fails to minimize arc length among nearby curves from p to q. Ellipsoids (preferably those with distinct semi-axes) and the torus of revolution *seem* to be promising places to look for this phenomenon.

Theorem 7.5.1 is about "long" geodesics whereas Theorem 6.3.1 is about "short" geodesics. Let us compare their messages. Let $\sigma(t) := \exp_p(tT)$, $0 \leqslant t \leqslant r$ be a geodesic that emanates from p with initial velocity T. Generically, $\exp_p$ will map only small tangent Finsler balls centered at the origin of T_pM diffeomorphically onto their images. The latter are aptly called "small" geodesic balls on M centered at p. Our geodesic is "short" if it is contained in one of these small geodesic balls centered at x. Otherwise it is said to be "long."

- Theorem 6.3.1 says, without qualifications, that a "short" geodesic from p to q is always the shortest curve among *all* piecewise C^∞ curves from p to q. Suppose our Finsler manifold is forward geodesically complete, then the geodesic σ can be continued forward indefinitely. Eventually it will no longer be short, although it may still be minimizing. If σ is minimal from p to a certain point $\hat{p}$ but not beyond, then $\hat{p}$ is called the **cut point** of p along the said geodesic. We emphasize that the concept of minimality here is a global one. Namely, our geodesic from p to its cut point $\hat{p}$ must minimize arc length among *all* piecewise C^∞ curves from p to $\hat{p}$.
- Theorem 7.5.1 has two parts. The second part tells us that the cut point of p along our geodesic must occur either before or at the first conjugate point of p, but not beyond. The reason is that beyond the first conjugate point, our geodesic cannot even minimize arc length among "nearby" curves that share its endpoints, so naturally it is no longer a global minimizer.
- Now we come to the first part of Theorem 7.5.1. It concerns a "long" geodesic σ from p to q. If q is *before* the first conjugate point of p, then σ is definitely shortest among "*nearby*" piecewise C^∞ curves from p to q. This notion of minimality involves comparison with nearby curves only, and is local in a tubular neighborhood sense.

For such purpose, it is totally irrelevant whether q has gone past the cut point.

In light of this discussion, one cannot help but get the impression that conjugate points concern the geometry near the given geodesic, whereas cut points embody geometrical/topological data far away. This is in fact the case. See Kobayashi's account [Ko] and references therein.

A somewhat systematic study of the conjugate and cut loci is undertaken in Chapter 8.

Let us summarize this section with a generic picture:

- Any geodesic starts out, say from some point p, by being a global minimizer (Theorem 6.3.1). It enjoys this status up to and including the cut point $\hat{p}$.
- Once it passes the cut point $\hat{p}$, it loses the global minimizer status. However, it still minimizes arc length among "nearby" curves *before* it gets to the first conjugate point. See part (2) of Proposition 7.3.1.

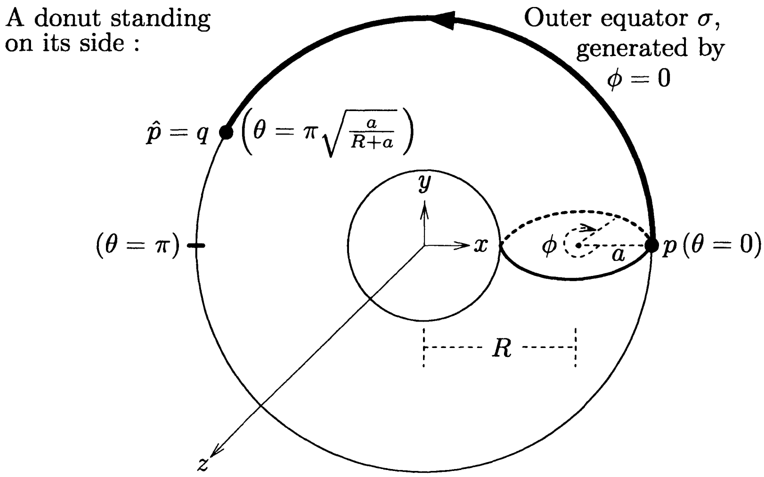

Figure 7.2

Generically, a geodesic σ emanates from p, reaches its cut point $\hat{p}$, and then its first conjugate point q. Although q never precedes $\hat{p}$, it can coincide with $\hat{p}$. For example, on the standard 2-sphere, q and $\hat{p}$ are both equal to the antipodal point of p. As another example, take σ to be the outer equator on the torus of revolution. Given any p on this σ, its first conjugate point q occurs *before* (!) one is even halfway around σ. See Exercise 7.5.5. One learns from [BG] that this q is actually equal to the cut point $\hat{p}$ along σ.

- At the first conjugate point, it may or may not minimize among "nearby" curves that share its endpoints. See Exercises 7.5.1 and 7.5.3.
- Beyond its first conjugate point, our geodesic definitely cannot minimize, even among "nearby" curves. See part (2) of Proposition 7.4.1.

Exercises

Exercise 7.5.1:

(a) Fix any point p on the standard 2-sphere. Let σ be any geodesic that emanates from p and terminates exactly at its first conjugate point q (which is antipodal to p). Show that σ minimizes arc length among all (hence "nearby") piecewise C^∞ curves from p to q.

(b) Formulate an analogous statement on an ellipsoid with two equal semi-axes.

Exercise 7.5.2: In $\mathbb{R}^3$, take a profile curve in the right xz-plane and revolve it around the z-axis. This generates a surface of revolution.

(a) At any given moment during the revolution, the rotated profile curve is called a **meridian**. Show that all meridians are geodesics.

(b) Under the revolution, any given point on the profile curve traces out a circle. Each such circle is called a **parallel**. Single out the ones whose arc lengths are critical when compared to "nearby" parallels. Show that among all parallels, only these are geodesics.

Exercise 7.5.3: Consider ellipsoids with two unequal semi-axes. Decide if there are points p and q on such ellipsoids, and a geodesic σ connecting them, with the following property:

> The point q is the first conjugate point of p along σ. Yet, σ fails to minimize arc length among "nearby" curves from p to q.

Suggestion: as a first step, consult Spivak [Sp4] and doCarmo [doC2] for pictures of conjugate loci on ellipsoids.

Exercise 7.5.4: Circular cylinders and cones are surfaces of revolution.

(a) Identify all the geodesics and check that none contains any conjugate point. Hence every geodesic minimizes arc length among "nearby" curves that share its endpoints.

(b) Produce a geodesic that is *not* a shortest curve joining its endpoints.

(c) Which geodesics do not contain cut points?

Exercise 7.5.5: Consider the torus of revolution. As in Figure 7.2, it is generated by revolving a specific circle around the z-axis. That circle is located in the xz-plane, has radius a, and center $(R, 0, 0)$, with $R > a$. The following parametrization, with $0 \leqslant \theta \leqslant 2\pi$, $0 \leqslant \phi \leqslant 2\pi$, is natural:

$$(\theta, \phi) \mapsto ([R + a \cos\phi] \cos\theta, [R + a \cos\phi] \sin\theta, a \sin\phi) .$$

(a) By working with (4.4.2b) and (4.4.5), or by consulting [On], show that the Gaussian curvature is

$$K = \frac{\cos\phi}{a (R + a \cos\phi)} .$$

(b) The outside equator corresponds to $\phi = 0$. Give it the defining parametrization $\sigma(\theta) := ([R + a] \cos\theta, [R + a] \sin\theta, 0)$. Check that σ is a geodesic with constant speed $R + a$. Show that the vector field ∂_ϕ is parallel along σ.

(c) Let $J(\theta)$ be any nontrivial Jacobi field along σ, with $J(0) = 0 = J(r)$. Use Exercise 5.4.4 to explain why $J(\theta) = f(\theta) \partial_\phi$, where $f(0) = 0 = f(r)$ but f is not identically zero.

(d) With the help of part (c) of Exercise 5.4.5, show that f satisfies the scalar Jacobi equation

$$f_{\theta\theta} + \frac{R + a}{a} f = 0 .$$

(e) Carefully prove that the *first* conjugate point q of $p := (R + a, 0, 0)$ along σ occurs at

$$\theta = \pi \sqrt{\frac{a}{R + a}} .$$

Explain why this is *before* one reaches the halfway point around σ.

(f) Let p, q be as described in (e). Consider the geodesic segment from p to q. Can you explain why it *does* minimize arc length among nearby piecewise C^∞ curves from p to q?

7.6 Ricci Curvatures

There are two notions of Ricci curvature with equivalent mathematical content. We have the Ricci scalar *Ric* which is a real-valued function on $TM \setminus 0$. Then there is the Ricci tensor which is a symmetric covariant 2-tensor on the vector bundle π^*TM that sits over $TM \setminus 0$. We give a cursory treatment of these objects here.

The single most important ingredient that enters into the definition of these two objects is the flag curvature, which we quickly review. In order to erect a flag at any given point $x \in M$, we need a flagpole y that is a nonzero element of the tangent space T_xM. This y singles out a particular inner product, namely,

$$g_{(x,y)} := g_{ij\,(x,y)}\, dx^i \otimes dx^j ,$$

which we use to measure lengths and angles in T_xM. To reduce clutter, suppose the actual "cloth" part of the flag is described by the unit vector $\ell := \frac{y}{F(x,y)}$ along the flagpole, together with another unit vector V (called the transverse edge) which is perpendicular to the flagpole. The flag curvature is then given as

$$(7.6.1a) \qquad K(x,y,\ell \wedge V) \;:=\; V^i\,(\ell^j\,R_{jikl}\,\ell^l\,)\,V^k \;=:\; V^i\,R_{ik}\,V^k\,.$$

If the transverse edge V is orthogonal to the flagpole but not necessarily of unit length, then

$$(7.6.1b) \qquad g_{(x,y)}(V,V)\;\;K(x,y,\ell \wedge V) \;=\; V^i\,R_{ik}\,V^k\,.$$

The case in which V is neither unit length nor orthogonal to ℓ was treated in §3.9.

7.6 A. The Ricci Scalar Ric and the Ricci Tensor Ric_{ij}

We next rotate V about our fixed flagpole. This generates $n-1$ orthonormal transverse edges, say $\{e_\alpha : \alpha = 1, \ldots, n-1\}$. Augmenting these e_α by $e_n := \ell$, we get a g-orthonormal basis for the fiber of π^*TM over the point $(x,y) \in TM \setminus 0$. With respect to this orthonormal basis, one has $K(x,y,\ell \wedge e_\alpha) = R_{\alpha\alpha}$ (no sum). If we add up these $n-1$ flag curvatures, the resulting quantity is a well-defined scalar on $TM \setminus 0$. It is positively homogeneous of degree zero (that is, invariant under positive rescaling) in y. If our Finsler structure F happens to be absolutely homogeneous, $F(x,cy) = |c|\,F(x,y)$, then so is the scalar in question.

The quantity we have just described is called the **Ricci scalar**. Let us denote it by *Ric*. So

$$(7.6.2a) \qquad Ric \;:=\; \sum_{\alpha=1}^{n-1} K(x,y,\ell \wedge e_\alpha) \;=\; \sum_{\alpha=1}^{n-1} R_{\alpha\alpha}\,.$$

To put it simply, *Ric* is a sum of $n-1$ appropriately chosen flag curvatures. It is therefore $n-1$ times the average of these flag curvatures. If the $n-1$ linearly independent transverse edges, say $\{V_\alpha\}$, are orthogonal to the flagpole but are not necessarily of unit length, then

$$(7.6.2b) \qquad Ric \;:=\; \sum_{\alpha=1}^{n-1} K(x,y,\ell \wedge e_\alpha) \;=\; \sum_{\alpha=1}^{n-1} \frac{R_{\alpha\alpha}}{g_{(x,y)}(V_\alpha, V_\alpha)}\,.$$

For computational purposes, it is desirable to have a formula for *Ric* that is valid in any basis. Indeed,

$$(7.6.3) \qquad \boxed{\;Ric \;=\; g^{ik}\,R_{ik} \;=\; R^s{}_s\;}\,.$$

This comes readily from (7.6.2a) if we note that $R_{nn} = 0$. Thus the Ricci scalar is simply the trace of the predecessor of our flag curvature.

In the general Finslerian setting, the notion of a **Ricci tensor** was first introduced by Akbar-Zadeh [AZ]:

$$\text{(7.6.4)} \qquad \boxed{Ric_{ik} \;:=\; \frac{\partial^2(\,\frac{1}{2}\,F^2\,Ric\,)}{\partial y^i \partial y^k} \;=\; \left[\frac{1}{2}\,F^2\,Ric\right]_{y^i y^k}} \,,$$

which is manifestly symmetric and covariant. Compare this with the definition of the fundamental tensor, namely, $g_{ik} := [\,\frac{1}{2}\,F^2\,]_{y^i y^k}$.

Let us summarize:

- We begin with R_{ik}, the predecessor of the flag curvature.
- Taking the trace of R_{ik} gives the Ricci scalar *Ric*.
- The transformation $h \mapsto (\frac{1}{2}\,F^2\,h)_{y^i y^k}$ generates the Ricci tensor Ric_{ik} from the Ricci scalar.

7.6 B. The Interplay between Ric and Ric_{ij}

We would like to express *Ric* in terms of Ric_{ij}. To this end, let us compute the double contraction $\ell^i\,\ell^k\,Ric_{ik}$.

* First, use the fact that $[\,\frac{1}{2}\,F^2\,Ric\,]_{y^i}$ is homogeneous of degree one in y, together with Euler's theorem (1.1.5), to calculate the quantity $y^k\,[\,\frac{1}{2}\,F^2\,Ric\,]_{y^i y^k}$.
* Next, use the fact that $\frac{1}{2}\,F^2\,Ric$ is homogeneous of degree two in y, and again (1.1.5), to calculate $y^i\,[\,\frac{1}{2}\,F^2\,Ric\,]_{y^i}$.

These maneuvers yield the identity

$$\text{(7.6.5a)} \qquad \boxed{Ric \;=\; \ell^i\,\ell^k\,Ric_{ik}} \,.$$

Equivalently,

$$\text{(7.6.5b)} \qquad Ric_{(x,y)} \;=\; \frac{1}{F^2(x,y)}\,[\,y^i\,y^k\,Ric_{ik}\,]\,.$$

In Riemannian geometry, the Ricci tensor is defined as $Ric_{jl} := R_j{}^s{}_{sl}$, and $Ric(U,V) := U^j\,Ric_{jl}\,V^l$. For the *Riemannian* case, a bit of definition-chasing shows that

$$(*) \qquad \text{The Ricci scalar } Ric \;=\; Ric\left(\frac{y}{\|y\|}\,,\,\frac{y}{\|y\|}\right)\,.$$

Here, $\|y\| := \sqrt{g(y,y)}$, where g is the Riemannian metric. Note that the Ricci scalar depends on y even in the Riemannian setting. The above realization puts (7.6.5) in a favorable light, because it now stands out as the natural generalization of (*) to the Finslerian context.

Formula (7.6.5) expresses the Ricci scalar *Ric* in terms of the Ricci tensor Ric_{ik}. Conversely, one already has (7.6.4). But we can make the dependence more explicit.

- Expand the two y derivatives in the defining formula, using the product rule and the notation $_{;i} := F \frac{\partial}{\partial y^i}$. This will produce a useful (albeit not so esthetically pleasant) relation:

$$Ric_{ik} = \frac{1}{2} Ric_{;k;i} + \frac{1}{2} \ell_i Ric_{;k} + \ell_k Ric_{;i} + g_{ik} Ric . \tag{7.6.6}$$

- Next, decompose $\frac{1}{2} Ric_{;k;i}$ into its symmetric and skew-symmetric parts. Then apply the Ricci identity or interchange formula (3.6.2) to the skew-symmetric part.

This computation ultimately gives the explicit formula

$$\boxed{\begin{aligned} Ric_{ik} = &\ \frac{1}{4} (Ric_{;i;k} + Ric_{;k;i}) \\ &+ \frac{3}{4} (\ell_i Ric_{;k} + \ell_k Ric_{;i}) \\ &+ g_{ik} Ric . \end{aligned}} \tag{7.6.7}$$

Exercises

Exercise 7.6.1:

(a) Verify (7.6.3) and the asserted homogeneity of *Ric*.

(b) Check that *Ric* has actually been obtained from the *hh*-curvature $R_j{}^i{}_{kl}$ without the use of the fundamental tensor g_{ik}.

(c) Explain why the same *Ric* is obtained whether one is using the Chern, Cartan, or Berwald connection.

Exercise 7.6.2: Show that if the Finsler structure is Riemannian, the Ricci tensor defined in (7.6.4) coincides with the Ricci tensor in Riemannian geometry. Hint: review the short discussion just before (*).

Exercise 7.6.3: Show that if (M, F) is a Finsler manifold with constant flag curvature c, then:

$$Ric = (n-1)\, c ,$$
$$Ric_{ik} = (n-1)\, c\, g_{ik} .$$

Compare the second statement with the predecessor R_{ik} of the flag curvature, which in this case is $R_{ik} = c\,(g_{ik} - \ell_i \ell_k)$, according to Proposition 3.10.1.

7.7 The Bonnet–Myers theorem

Let (M, F) be a Finsler manifold, where F is positively (but perhaps not absolutely) homogeneous of degree one. Recall from §6.6 that (M, F) is said to be forward geodesically complete if every geodesic, when parametrized to have constant Finslerian speed, is indefinitely forward extendible. Also, recall from §6.2D that (M, F) is said to be forward complete if every forward Cauchy sequence converges. According to the Hopf–Rinow theorem (Theorem 6.6.1), these two notions of completeness are equivalent. As we show in the proof below, geodesic completeness is more appropriate for us to use in the following theorem.

Theorem 7.7.1 (Bonnet–Myers). *Let (M, F) be a forward geodesically complete connected Finsler manifold of dimension n. Suppose its Ricci scalar has the following uniform positive lower bound*

$$Ric \geqslant (n-1)\,\lambda > 0\ .$$

Equivalently, suppose its Ricci tensor satisfies

$$y^i\, y^j\, Ric_{ij\ (x,y)} \geqslant (n-1)\,\lambda\, F^2(x,y)\ , \quad \textit{with}\ \ \lambda > 0\ .$$

Then:

(1) *Along every geodesic, the distance between any two successive conjugate points is at most $\frac{\pi}{\sqrt{\lambda}}$. In other words, every geodesic with length $\frac{\pi}{\sqrt{\lambda}}$ or longer must contain conjugate points.*

(2) *The diameter of M is at most $\frac{\pi}{\sqrt{\lambda}}$.*

(3) *M is in fact compact.*

(4) *The fundamental group $\pi(M, x)$ is finite.*

Remarks:

(a) The equivalence between the hypothesis on Ric and that on Ric_{ij} is a consequence of (7.6.5b).

(b) Conceptually, it is more illuminating to rewrite our inequality on the Ricci scalar as

$$\frac{Ric}{n-1} \geqslant \lambda > 0\ .$$

The reason is that $Ric_{(x,y)}$ is obtained by summing the flag curvatures of $n-1$ mutually $g_{(x,y)}$-orthogonal "unit" flags, all based at $x \in M$ and sharing a common flagpole $y \in T_xM$. This was discussed in detail at the beginning of §7.6. The above inequality says that the *average* flag curvature, corresponding to each selection of base point and flagpole, has a uniform positive lower bound.

(c) The curvature hypothesis is satisfied if the flag curvatures are bounded below uniformly by the positive constant λ. In that case, the result is known as **Bonnet's theorem** whenever $\dim M = 2$.

(d) The essence of our proof comes from that given in Cheeger–Ebin [CE] for the Riemannian case. Compare our treatment here with the ones given by Auslander [Au], Dazord [Daz], and Matsumoto [M2].

Proof.

- **Every geodesic of length $\geqslant \frac{\pi}{\sqrt{\lambda}}$ must contain conjugate points:**

Consider any unit speed geodesic $\sigma(t)$, $0 \leqslant t \leqslant L$ with velocity field $T = T(t)$. As usual, abbreviate $g_{(\sigma,T)}$ by g_T. Use parallel transport (with reference vector T) and Exercise 5.2.3 to generate a moving frame $\{e_i(t)\}$ along σ such that:

* Each e_i is a parallel vector field along σ.
* $\{e_i\}$ is a g_T-orthonormal basis.
* $e_n := T$.

Let

$$S_\lambda(t) := \frac{1}{\sqrt{\lambda}} \sin(\sqrt{\lambda}\, t)$$

and define

$$W_\alpha(t) := S_\lambda(t)\, e_\alpha(t) , \quad \alpha = 1, \ldots, n-1 .$$

Fix a positive $r \leqslant L$ and consider the index form for $\sigma(t)$, $0 \leqslant t \leqslant r$. Since each W_α is g_T-orthogonal to T, we can use the formula depicted in Exercise 7.2.5 to compute $I(W_\alpha , W_\alpha)$:

$$I(W_\alpha , W_\alpha) = \int_0^r \Big\{ \| W_\alpha' \|^2 - \| W_\alpha \|^2 \ K(T , W_\alpha) \Big\}\, dt .$$

Here, we are using the temporary abbreviation

$$\| V \|^2 := g_T(V, V)$$

in order to reduce clutter.

Referring to the properties of e_α, we see that

* $\| W_\alpha'(t) \|^2 = [S_\lambda'(t)]^2 = \cos^2(\sqrt{\lambda}\, t)$.
* $\| W_\alpha(t) \|^2 = S_\lambda^2(t) = \frac{1}{\lambda} \sin^2(\sqrt{\lambda}\, t)$.
* $K(T , W_\alpha) = K(T , e_\alpha)$ (see Exercise 3.9.1) .

Substituting these into the formula for $I(W_\alpha , W_\alpha)$ and then summing over the index α, we get

$$\sum_{\alpha=1}^{n-1} I(W_\alpha , W_\alpha) = \int_0^r \Big\{ (n-1) [S_\lambda'(t)]^2 - S_\lambda^2(t) \sum_{\alpha=1}^{n-1} K(T , e_\alpha) \Big\}\, dt .$$

That sum of flag curvatures is precisely the Ricci scalar *Ric* (see §7.6). And we have hypothesized here that $Ric \geqslant (n-1)\,\lambda$. So the above becomes

$$\sum_{\alpha=1}^{n-1} I(W_\alpha\,,\,W_\alpha) \;\leqslant\; (n-1)\int_0^r \left\{ [\,S'_\lambda(t)\,]^2 \;-\; \lambda\, S^2_\lambda(t) \right\} dt\ .$$

Evaluating that integral explicitly, we get

$$\sum_{\alpha=1}^{n-1} I(W_\alpha\,,\,W_\alpha) \;\leqslant\; (n-1)\,\frac{\sin(\,2\sqrt{\lambda}\,r\,)}{2\sqrt{\lambda}}\ . \tag{7.7.1}$$

Now suppose the length L of our unit speed geodesic is at least $\frac{\pi}{\sqrt{\lambda}}$. In which case we choose our r to be $\frac{\pi}{\sqrt{\lambda}}$, and (7.7.1) becomes

$$\sum_{\alpha=1}^{n-1} I(W_\alpha\,,\,W_\alpha) \;\leqslant\; 0\ .$$

Hence some $I(W_\alpha\,,\,W_\alpha)$ must be nonpositive. Relabel this W_α as W to reduce clutter.

Suppose for the sake of argument that $\sigma(t)$, $0 \leqslant t \leqslant r := \frac{\pi}{\sqrt{\lambda}}$ contains no conjugate points. Then Proposition 7.1.1 tells us that our vector field W, being nowhere zero except at $t=0$ and $t=r$, cannot possibly be a Jacobi field. The same proposition also tells us that the unique Jacobi field $J(t)$ which vanishes at the endpoints of $\sigma(t)$, $0 \leqslant t \leqslant r$ is none other than the identically zero field. So $I(J,J) = 0$. By the basic index lemma (Lemma 7.3.2), we must then have

$$0 \;=\; I(J,J) \;<\; I(W,W) \;\leqslant\; 0\ ,$$

which is a contradiction. Therefore $\sigma(t)$, $0 \leqslant t \leqslant r := \frac{\pi}{\sqrt{\lambda}}$ must contain conjugate points.

• **The fact that diam$(M) \leqslant \frac{\pi}{\sqrt{\lambda}}$:**

Fix a point $x \in M$ and consider all y in the indicatrix S_x. For each such y, issue the unit speed geodesic from x with initial unit velocity y. Since (M,F) is forward geodesically complete, the said geodesic is defined on $0 \leqslant t < \infty$.

Let c_y denote the moment this geodesic reaches the first conjugate point of x, and i_y the moment it reaches the cut point of x. According to the discussion in §7.5, one has

$$i_y \;\leqslant\; c_y$$

because the first conjugate point can never come before the cut point. Furthermore, what we have shown in the previous part may be paraphrased as

$$c_y \;\leqslant\; \frac{\pi}{\sqrt{\lambda}}$$

for all y. Thus

$$\sup_{y \in S_x} i_y \leqslant \sup_{y \in S_x} c_y \leqslant \frac{\pi}{\sqrt{\lambda}} .$$

Since (M, F) is connected and forward geodesically complete, the Hopf–Rinow theorem says that the exponential map is defined on all of $T_x M$. Proposition 6.5.1 then tells us that every point can be reached from x by a unit speed minimizing geodesic. Based on this fact, a moment's thought shows that

$$\operatorname{diam}(M) \leqslant \sup_{y \in S_x} i_y$$

which, in conjunction with the above, gives

$$\operatorname{diam}(M) \leqslant \frac{\pi}{\sqrt{\lambda}} .$$

- **The compactness of M:**

The manifold M is always closed in its own topology. Also, we have just shown that it is forwardly bounded from the above (arbitrarily) fixed x. Since (M, F) is by hypothesis forward geodesically complete, its compactness therefore follows from the Hopf–Rinow theorem (Theorem 6.6.1).

- **About the fundamental group of M:**

Let $\tilde{M}$ be a simply connected covering space (that is, a universal cover) of M, with *smooth* projection $p : \tilde{M} \to M$. Using the smooth map p, we pull the Finsler structure F back to $\tilde{M}$. The resulting $\tilde{F}$ is smooth and strongly convex on $T\tilde{M} \setminus 0$. Since $(\tilde{M}, \tilde{F})$ is locally isometric to (M, F), the Ricci scalar of $\tilde{F}$ inherits the same uniform positive lower bound $(n-1)\,\lambda$.

Our universal cover $(\tilde{M}, \tilde{F})$ is forward geodesically complete because (M, F) is. To see this, let $\tilde{\sigma}(t)$ be any geodesic emanating from some point $\tilde{x} \in \tilde{M}$ at $t = 0$. Its projection on M, denoted $\sigma(t)$, is a geodesic because $(\tilde{M}, \tilde{F})$ and (M, F) are locally isometric. This $\sigma(t)$ emanates from some $x \in M$ at $t = 0$. By the hypothesized forward geodesic completeness on (M, F), $\sigma(t)$ is extendible to all $t \in [0, \infty)$. The said local isometry now implies the same for $\tilde{\sigma}(t)$. Hence the universal cover $(\tilde{M}, \tilde{F})$ is forward geodesically complete, as claimed.

We now know that $(\tilde{M}, \tilde{F})$ is forward geodesically complete, and that its Ricci scalar has the uniform positive lower bound $(n-1)\,\lambda$. Also, its connectedness is part of the *definition* of being simply connected. Therefore we can apply part (3) of the present theorem to $(\tilde{M}, \tilde{F})$ to conclude that it is compact.

By hypothesis, M is connected, so all its fundamental groups $\pi(M, x)$, where the x denotes base points, are isomorphic. Since $\tilde{M}$ is a universal cover, any specific $\pi(M, x)$ is bijective with the collection of isolated points $p^{-1}(x)$. See [ST] for a review. Finally, the compactness of $\tilde{M}$ implies that the closed subset $p^{-1}(x)$ is also compact, hence finite. □

Exercises

Exercise 7.7.1: Verify (7.7.1).

Exercise 7.7.2:

(a) Explain in detail when and why $\text{diam}(M) \leqslant \sup_{y \in S_x} i_y$ is valid.

(b) Does this fail to hold if (M, F) is not assumed to be forward geodesically complete?

Exercise 7.7.3: Suppose (M, F) is forward complete. Can you deduce the forward completeness of its universal cover $(\tilde{M}, \tilde{F})$ directly, using forward Cauchy sequences?

References

[AZ] H. Akbar-Zadeh, *Sur les espaces de Finsler à courbures sectionnelles constantes*, Acad. Roy. Belg. Bull. Cl. Sci. (5) **74** (1988), 281–322.

[Au] L. Auslander, *On curvature in Finsler geometry*, Trans. AMS **79** (1955), 378–388.

[BG] M. Berger and B. Gostiaux, *Differential Geometry: Manifolds, Curves, and Surfaces*, Graduate Texts in Mathematics **115**, Springer-Verlag, 1988.

[CE] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, North Holland/American Elsevier, 1975.

[Daz] P. Dazord, *Propriétés globales des géodésiques des Espaces de Finsler*, Theses, Université de Lyon, 1969.

[doC2] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, 1976.

[Ko] S. Kobayashi, *On conjugate and cut loci*, Global Differential Geometry, S. S. Chern, ed., Math. Assoc. America, 1989, pp. 140–169.

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

[On] B. O'Neill, *Elementary Differential Geometry*, 2nd ed., Academic Press, 1997.

[ST] I. M. Singer and J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry*, Undergraduate Texts in Mathematics, Springer-Verlag, 1976.

[Sp4] M. Spivak, *Differential Geometry*, vol. IV, Publish or Perish, 1975.

Chapter 8

The Cut and Conjugate Loci, and Synge's Theorem

8.1 Definitions

In this chapter, the following assumptions are made throughout:

- The Finsler structure F is positively (but perhaps not absolutely) homogeneous of degree one. Consequently, the associated metric distance function d may not be symmetric.
- **The Finsler manifold (M, F) is forward geodesically complete.** That is, all geodesics, parametrized to have constant Finslerian speed, are indefinitely forward extendible. By the Hopf–Rinow theorem, this is equivalent to saying that all forward Cauchy sequences are convergent. See Theorem 6.6.1 and §6.2D. Also, according to Exercise 6.6.4, the completeness hypothesis is automatically satisfied whenever M is compact.
- Unless we specify otherwise, **all geodesics are parametrized to have unit Finslerian speed.**

Fix a point x in the Finsler manifold (M, F). Let $\sigma_y(t)$ be a unit speed geodesic that passes through x at time $t = 0$, with initial velocity y. Here, $F(x, y) = 1$, so y is an element of the unit tangent Finsler sphere in $T_x M$. In other words, $y \in \mathrm{S}_x$, the indicatrix at x.

- Define the **conjugate value** c_y of y by

$$c_y \;:=\; \sup\{\, r : \text{ no point } \sigma_y(t),\ 0 \leqslant t \leqslant r \text{ is conjugate to } x \,\} .$$

Note that the positive number c_y can equal ∞. If $c_y < \infty$, the point $\sigma_y(c_y)$ is known as the **first conjugate point** of x along σ_y. Otherwise, x is said to have no conjugate point along σ_y.

* The **conjugate radius** at x is defined as

$$\mathfrak{c}_x \;:=\; \inf_{y \in \mathrm{S}_x} c_y \,,$$

while the **conjugate locus** of x is

$$\mathcal{C}on_x \;:=\; \{\, \sigma_y(c_y) :\ y \in \mathrm{S}_x \text{ with } c_y < \infty \,\} .$$

- Likewise, define the **cut value** i_y of y by

$$i_y \;:=\; \sup\{\, r : \text{ the segment } \sigma_{y\,|[0,r]} \text{ is globally minimizing} \,\} .$$

The positive number i_y can equal ∞, in which case the cut point of x does not exist in the direction y. If $i_y < \infty$, the point $\sigma_y(i_y)$ is called the **cut point** of x along σ_y.

* The **injectivity radius** at x is defined as

$$\mathfrak{i}_x \;:=\; \inf_{y \in \mathrm{S}_x} i_y \,,$$

whereas the **cut locus** of x is

$$\mathcal{C}ut_x \;:=\; \{\, \sigma_y(i_y) :\ y \in \mathrm{S}_x \text{ with } i_y < \infty \,\} .$$

Explicit examples of $\mathcal{C}on_x$ and $\mathcal{C}ut_x$ are scarce. See Spivak [Sp4] for those on the ellipsoid, and Berger–Gostiaux [BG] for those on the torus of revolution.

Exercises

Exercise 8.1.1: Given a geodesic $\exp_x(ty)$, $0 \leqslant t \leqslant r$, show that the following two statements are equivalent:

- The point $\exp_x(ry)$ is the *first* conjugate point of x along this geodesic.
- The derivative $\exp_{x*}$ of the exponential map $\exp_x$ is nonsingular at all ty, $t \in [0, r)$, and singular at the point ry.

Hint: consult the proof of Proposition 7.1.1.

Exercise 8.1.2:

(a) Why are c_y and i_y always positive?

(b) Suppose M is compact; explain why c_y and i_y are both finite, for every unit speed geodesic σ_y.

8.2 The Cut Point and the First Conjugate Point

As we discussed in §7.5, every geodesic σ from x starts out by being "short" and therefore minimizing among *all* piecewise C^∞ curves that share its endpoints. However, the moment it gets past the first conjugate point of x, we see from Theorem 7.5.1 that it won't even minimize arc length among "nearby" piecewise C^∞ curves. So the cut point of x along σ must occur either before, or exactly at, the first conjugate point.

Also, we saw that even though σ can no longer globally minimize arc length beyond the cut point of x, it can still continue to do so "locally" (that is, among "nearby" piecewise C^∞ curves) until it hits the first conjugate point of x.

Let us now demonstrate another important property of the cut point.

The first two parts of Theorem 6.3.1 assert that given any "short" unit speed geodesic σ_y from x, it is in fact the *unique* minimizer (up to reparametrization) of arc length among all piecewise C^∞ curves that share its endpoints.

> We show that the distinction of being the unique minimizer is actually sustained until, but perhaps not including, the moment σ_y hits the cut point of x.

To this end, we sketch an argument presented in [Ko]. Let the geodesic in question be $\sigma_y(t)$, $0 \leqslant t \leqslant r$. It goes from x to a point $\tilde{x} := \sigma_y(r)$ that is *strictly before* the cut point $\hat{x} := \sigma_y(i_y)$.

> Suppose there exists another curve γ from x to $\tilde{x}$, having the same arc length as $\sigma_{y\,|[0,r]}$. Suppose for the sake of argument that γ is genuinely different from (that is, not a reparametrization of) σ_y.
>
> * By first traveling along γ from x to $\tilde{x}$, and then along $\sigma_{y\,|[r\,,\,i_y]}$ from $\tilde{x}$ to $\hat{x}$, we trace out a curve c from x to its cut point $\hat{x}$. The defining property of γ implies that it must be a minimal geodesic. Thus c is piecewise C^∞, and has arc length equal to the distance from x to $\hat{x}$, which is i_y. However, c is not a geodesic because it has a "kink" (perhaps an extremely slight one) at $\tilde{x}$.
>
> * Fix a point $\mathcal{P}$ on γ that is slightly before $\tilde{x}$, and a point $\mathcal{Q}$ on σ_y that is slightly beyond $\tilde{x}$. If $\mathcal{P}$ and $\mathcal{Q}$ are close enough, there is always a minimal geodesic connecting them, whether our manifold is forward geodesically complete or not. The triangle inequality for the distance

function can be used to check that this geodesic is shorter than the arc along γ from $\mathcal{P}$ to $\tilde{x}$, followed by that along σ_y from $\tilde{x}$ to $\mathcal{Q}$.

* Therefore c can be replaced by another curve, also going from x to $\hat{x}$, but which has shorter length. This is a contradiction because, as we have seen, the length of c is i_y, which is the distance from x to the cut point $\hat{x}$.

This demonstrates that the supposed scenario is impossible. In other words, every geodesic from x retains its status as the **unique minimizer** of arc length, as long as it does *not* reach the cut point $\hat{x}$.

An immediate question is: what happens to its unique minimizer status if our geodesic σ_y goes from x to the cut point $\hat{x}$? The following proposition claims that:

* If the cut point $\hat{x}$ comes before the first conjugate point, then the unique minimizer status will definitely be lost when σ_y reaches $\hat{x}$.
* If the cut point $\hat{x}$ occurs right at the first conjugate point, then in general there is no definite conclusion when σ_y reaches $\hat{x}$.

Proposition 8.2.1. *Let x be any point in a forward geodesically complete Finsler manifold (M, F), and y an arbitrary element in the indicatrix S_x. Let σ_y be a unit speed geodesic that passes through x with initial velocity y. Then:*

(1) $i_y \leqslant c_y$; *hence* $\mathfrak{i}_x \leqslant \mathfrak{c}_x$.

(2) *The cut point of x along σ_y occurs either before, or exactly at, the first conjugate point.*

(3) *For any $r < i_y$, the geodesic $\sigma_{y\,|[0,r]}$ is, up to reparametrization, the* **unique** *minimizer of arc length among all piecewise C^∞ curves that share its endpoints.*

(4) *This "unique minimizer property" will definitely fail at the cut point* **if** *the latter occurs before the first conjugate point. Namely, whenever $i_y < c_y$, there exists a geodesic that is distinct from (that is, not a reparametrization of) but has the same endpoints and arc length as $\sigma_{y\,|[0\,,\,i_y]}$.*

Proof. Items (1) and (2) have been known since §7.5. And we have just deduced (3). It remains to establish (4).

The portion of σ_y that concerns us has the description $\exp_x(ty)$, $0 \leqslant t \leqslant i_y$. Take a sequence of points $\{\hat{x}_k\}$ along σ_y that are *beyond* the cut point $\hat{x}$ and converging to $\hat{x}$. For concreteness, set

$$\hat{x}_k \ := \ \sigma_y(\,[i_y + \epsilon_k]\,) \ = \ \exp_x([i_y + \epsilon_k]\,y) \ ,$$

where the ϵ_k are positive numbers that decrease monotonically to zero.

For each k, Proposition 6.5.1 gives a minimizing unit speed geodesic $\exp_x(t\,y_k)$, $0 \leqslant t \leqslant L_k$ of length L_k, from x to $\hat{x}_k$. Since $\hat{x}_k$ is beyond the cut point $\hat{x}$, we must have $y_k \neq y$. The continuity of the distance function $d(x\,,\,\cdot)$ assures us that $L_k \to i_y$.

By construction,

$$\exp_x(L_k\,y_k) \;=\; \hat{x}_k \;\;\rightarrow\;\; \hat{x} \;=\; \exp_x(i_y\,y)\ . \tag{8.2.1}$$

The sequence $\{y_k\}$ is contained in the indicatrix S_x, which is a compact set. Thus, by passing to a subsequence if necessary, we may assume that it converges to some $Y \in \mathrm{S}_x$. In particular,

$$L_k\,y_k \;\;\rightarrow\;\; i_y\,Y\ . \tag{8.2.2}$$

Applying $\exp_x$ to this and comparing the result with (8.2.1) gives

$$\exp_x(i_y\,Y) \;=\; \hat{x} \;=\; \exp_x(i_y\,y)\ . \tag{8.2.3}$$

Thus $\exp_x(tY)$, $0 \leqslant t \leqslant i_y$ and our original $\sigma_{y\,|[0\,,\,i_y]}$ are both geodesics from x to $\hat{x}$, and they have the same arc length.

Now we impose the hypothesis that the cut point comes strictly before the first conjugate point. It remains to be checked that under such circumstance, we must have $Y \neq y$.

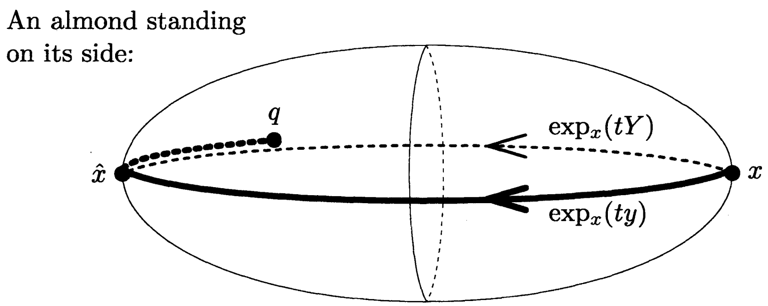

Figure 8.1

An illustration of property (4) in Proposition 8.2.1. Consider for instance the ellipsoid analyzed in Spivak [Sp4]. Its semi-axes are *all unequal*, so the surface looks roughly like that of an almond. We take x to be the East Pole, and $\exp_x(ty)$ (*much* thickened) to be the equator traversed clockwise. The cut point $\hat{x}$ is the West Pole. But the first conjugate point q is, somewhat surprisingly, "beyond" $\hat{x}$ rather than at $\hat{x}$. Note that the equator traversed counterclockwise (*slightly* thickened) gives the "other" geodesic $\exp_x(tY)$ encountered in the proof of Proposition 8.2.1. Namely, for this example, an appropriate Y just happens to be $-y$.

Indeed, $i_y < c_y$ implies (see Exercise 8.1.1) that $\exp_{x*}$ is nonsingular at $i_y\, y$. The Inverse Function theorem then guarantees that $\exp_x$ is injective on a ball B centered at $i_y\, y$. The element $i_y\, Y$ cannot be inside B, or else we would have $L_k\, y_k \in$ B for sufficiently large k, which would contradict the said injectivity because

$$(8.2.4) \qquad \exp_x(L_k\, y_k) \ = \ \hat{x}_k \ = \ \exp_x([i_y + \epsilon_k]\, y)$$

and $y_k \neq y$. Since $i_y\, Y$ lies outside B (with center $i_y\, y$), we see in particular that Y cannot possibly be the same as y. □

Corollary 8.2.2. *Let x be any point in a forward geodesically complete Finsler manifold (M, F), and y an arbitrary element in the indicatrix S_x. Let σ_y be a unit speed geodesic that passes through x with initial velocity y. Assume that i_y is finite and abbreviate the cut point $\sigma_y(i_y\, y)$ as $\hat{x}_y$. Then* **at least one** *of the following two scenarios must hold:*

- *$\hat{x}_y$ is equal to the first conjugate point of x along σ_y.*
- *There exist two distinct geodesics of the same length from x to $\hat{x}_y$.*

Exercises

Exercise 8.2.1: Fill in the details in the proof of part (3) of Proposition 8.2.1.

Exercise 8.2.2:

(a) Suppose the Finsler structure is either absolutely homogeneous (which includes the Riemannian case) or is of Berwald type. Use Corollary 8.2.2 to deduce that: $\hat{x}$ is the cut point of x along σ if and only if x is the cut point of $\hat{x}$ along the reverse of σ.

(b) Why doesn't this statement hold in the general Finsler setting? Can you provide a counterexample?

8.3 Some Consequences of the Inverse Function Theorem

This section consists of four guided exercises. We use them to explore the fact that **nonconjugate points are "stable" under suitable perturbations**.

Exercises

Exercise 8.3.1: Suppose $\exp_{x*}$ is nonsingular at the point $t_o y \in T_xM$, where $F(x, y) = 1$ and $t_o \geqslant 0$. Show that one can find a neighborhood $\mathcal{U}$

of (x,y) in the indicatrix bundle S and a small $\epsilon > 0$ (both depending only on x, y, and t_o) with the following properties:

(a) Whenever $(\tilde{x},\tilde{y}) \in \mathcal{U}$ and $\tilde{t} \in [t_o - \epsilon\,,\, t_o + \epsilon]$, the derivative map $\exp_{\tilde{x}*}$ is nonsingular at the points $\tilde{t}\tilde{y} \in T_{\tilde{x}}M$.

(b) In a local trivialization of S, the neighborhood $\mathcal{U}$ is the Cartesian product of an open ball (for the positions $\tilde{x}$) in $\mathbb{R}^n$ with an open "disc" (for the directions $\tilde{y}$) on the standard unit sphere $\mathbb{S}^{n-1}$.

Hint: by hypothesis, $(\exp_{x*})_{(t_o y)}$ has nonzero determinant or, equivalently, it has rank n.

Exercise 8.3.2: This exercise uses the above but involves a bit more subtlety. Let $(x,y) \in$ S and fix an r such that $0 < r < c_y$. Then $\exp_{x*}$ is nonsingular at all points along the ray ty, $0 \leqslant t \leqslant r$. Show that there is a neighborhood $\mathcal{U}$ of (x,y) in S and a small $\epsilon > 0$ (depending only on x, y, and r) with the following properties:

(a) Whenever $(\tilde{x},\tilde{y}) \in \mathcal{U}$, the derivative map $\exp_{\tilde{x}*}$ is nonsingular at all points along the ray $t\tilde{y}$, $0 \leqslant t \leqslant r + \epsilon$. Hence $c_{\tilde{y}} \geqslant r + \epsilon$.

(b) $\lim_{r \to c_y} \epsilon = 0$.

Try the following strategy:

- For each fixed $t \in [0,r]$, get a neighborhood $\mathcal{U}_t$ and an interval $I_t := [t - \epsilon_t\,,\, t + \epsilon_t]$ with the properties stated in Exercise 8.3.1. Here, both $\mathcal{U}_t$ and ϵ_t also depend on the suppressed (x,y).
- Explain why a finite number of the intervals, say $I_{t_1}, \ldots, I_{t_k}$, cover $[0,r]$. Order the t_i so that $t_1 < \cdots < t_k$. Include 0 and r among the t_i, so that $t_1 = 0$ and $t_k = r$. What is the advantage of this ordering? Is there a real need for having $t_k = r$?
- Construct $\mathcal{U}$ as the intersection of $\mathcal{U}_{t_1}, \ldots, \mathcal{U}_{t_k}$, and perhaps set ϵ equal to ϵ_{t_k} (that is, ϵ_r). Keep in mind that we have to produce property (b). Do you see why $\mathcal{U}$ must shrink as r approaches c_y?

Exercise 8.3.3: Here, we begin with the same hypothesis as that in Exercise 8.3.1, namely, $\exp_{x*}$ is nonsingular at the point $t_o y \in T_x M$, where $F(x,y) = 1$ and $t_o \geqslant 0$, but we extract information from a different perspective.

Consider the map $\text{EXP} : TM \to M \times M$ given by $(x,v) \mapsto (x\,,\, \exp_x[v])$. Show that its derivative matrix EXP_* at the point $(x,v) \in TM$ has the following structure: the two diagonal blocks are, respectively, the $n \times n$ identity matrix and $(\exp_{x*})_{(v)}$, while the upper off-diagonal block is zero.

If $\exp_{x*}$ is nonsingular at the point $t_o y \in T_x M$, demonstrate that:

(a) The matrix $(\text{EXP}_*)_{(x\,,\,t_o y)}$ is nonsingular.

(b) There exists a neighborhood $\mathcal{V}$ of $(x\,,\,t_o y)$ in TM on which EXP is a diffeomorphism.

Exercise 8.3.4: Formulate and prove the statement that **an accumulation point of conjugate points is a conjugate point**.

8.4 The Manner in Which c_y and i_y Depend on y

Let us now turn to the analytical properties of the conjugate and cut value maps. In the following, let S denote the indicatrix bundle; namely,

$$\mathrm{S} := \{ (x,y) \in TM : \ F(x,y) = 1 \} .$$

Proposition 8.4.1. *Let (M,F) be a forward geodesically complete Finsler manifold. Let S be the indicatrix bundle.*

(1) *The function $(x,y) \mapsto c_y$ is lower semicontinuous from S into $(0,\infty]$. That is,*

$$\liminf_{(\tilde{x},\tilde{y})\to(x,y)} c_{\tilde{y}} \geqslant c_y . \tag{8.4.1}$$

(2) *The function $(x,y) \mapsto i_y$ from S into $(0,\infty]$ is continuous.*

Remarks:

- We have to use $(0,\infty]$ instead of $(0,\infty)$ because in some directions y, the quantities i_y and c_y are both equal to ∞. For instance, this is so for all y in Minkowski spaces.
- Whenever c_y is finite, statement (1) means the following:

  Given any $\rho > 0$, we can find a neighborhood of (x,y) in which all the points $(\tilde{x},\tilde{y})$ satisfy

  $$c_{\tilde{y}} > c_y - \rho .$$

Proof.

The lower semicontinuity of $(x,y) \mapsto c_y$:
According to Exercise 8.3.2, for $(\tilde{x},\tilde{y})$ close to (x,y) and any fixed r in $(0,c_y)$, we have $c_{\tilde{y}} \geqslant r+\epsilon$ for some positive ϵ. In particular,

$$\liminf_{(\tilde{x},\tilde{y})\to(x,y)} c_{\tilde{y}} \geqslant r+\epsilon .$$

As $r \to c_y$, we were told that $\epsilon \to 0$. So

$$\liminf_{(\tilde{x},\tilde{y})\to(x,y)} c_{\tilde{y}} \geqslant c_y ,$$

which is part (1).

The continuity of $(x, y) \mapsto i_y$:

Let $\{(x_k, y_k)\}$ be any sequence in S that converges to $(x, y) \in$ S. Hence

$$(8.4.2) \qquad \lim_{k\to\infty} \exp_{x_k}(t\, y_k) \;=\; \exp_x(ty) \quad \text{for all} \;\; t \geqslant 0 \;.$$

The same holds for any subsequence of $\{(x_k, y_k)\}$.

We want to show that $\lim_{k\to\infty} i_{y_k}$ exists and equals i_y. It suffices to check the following statement:

$$(8.4.3) \qquad i_y \;\leqslant\; i_o \;:=\; \liminf_{k\to\infty} i_{y_k} \;\leqslant\; \limsup_{k\to\infty} i_{y_k} \;\leqslant\; i_y \;.$$

First consider the purported inequality on the far right. Without loss of generality, let $i_y < \infty$ (if $i_y = \infty$, there is nothing to prove). Given any $\epsilon > 0$, the number of i_{y_k} that exceed $i_y + \epsilon$ must be finite. Otherwise, for those y_k in question, relabeled as a subsequence $\{(x_j, y_j)\}$, we have

$$d(\, x_j \,,\, \exp_{x_j}\{[i_y + \epsilon]\, y_j\}\,) \;=\; i_y + \epsilon \;.$$

Upon letting $j \to \infty$, this says that the distance from x to $\exp_x([i_y + \epsilon]\, y)$ is $i_y + \epsilon$. (Taking that limit needs Exercise 6.2.9. Try to supply the steps.) In other words, the geodesic $\exp_x(ty)$ continues to minimize arc length beyond the cut point of x, which is a contradiction. Thus, for all but finitely many y_k, we must have $i_{y_k} \leqslant i_y + \epsilon$. Hence $\limsup_{k\to\infty} i_{y_k} \leqslant i_y$.

Next, we establish $i_y \leqslant i_o$. Suppose the contrary; then we have $i_o < i_y \leqslant c_y$ because the cut point never occurs beyond the first conjugate point. Since $\exp_{x*}$ is nonsingular at the point $i_o\, y \in T_xM$, we see from Exercise 8.3.3 that there exists a neighborhood $\mathcal{V}$ of $(x \,,\, i_o\, y)$ in TM on which the map EXP, namely, $(x, v) \mapsto (x \,,\, \exp_x[v])$, is a diffeomorphism.

By the definition of i_o, there is a subsequence $\{(x_j, y_j)\}$ such that i_{y_j} converges to i_o from above. Also, (8.4.1) tells us that $c_y \leqslant \liminf_{k\to\infty} c_{y_k}$. We can therefore choose our subsequence so that each $(x_j \,,\, i_{y_j}\, y_j)$ is in $\mathcal{V}$ and

$$(8.4.4) \qquad i_o \;\leqslant\; i_{y_j} \;<\; i_y \;\leqslant\; c_y \;\leqslant\; c_{y_j} \;.$$

Since $i_{y_j} < c_{y_j}$, the last part of Proposition 8.2.1 guarantees a unit vector $Y_j \neq y_j$ such that

$$(8.4.5) \qquad \exp_{x_j}(i_{y_j}\, Y_j) \;=\; \exp_{x_j}(i_{y_j}\, y_j) \;.$$

Given what was said about EXP, we see that $(x_j \,,\, i_{y_j}\, Y_j)$ cannot be in $\mathcal{V}$.

Since $x_j \to x$, the points x_j are in some compact subset of M. Hence the sequence $\{(x_j, Y_j)\}$ is contained in a compact subset of S, even though S itself is not compact unless M is. By passing to a subsequence if necessary, we may assume that $(x_j, Y_j) \to$ some $(x, Y) \in$ S. Using this and $i_{y_j} \to i_o$ in (8.4.5), we get

$$(8.4.6) \qquad \exp_x(i_o\, Y) \;=\; \exp_x(i_o\, y) \;.$$

Since all the $(x_j \,,\, i_{y_j}\, Y_j)$ are outside $\mathcal{V}$, so is their limit $(x \,,\, i_o\, Y)$. But $(x \,,\, i_o\, y) \in \mathcal{V}$. Therefore $Y \neq y$.

Beginning with the assumption $i_o < i_y$, we arrived at (8.4.6). It says that there are two distinct unit speed geodesics $\exp_x(ty)$ and $\exp_x(tY)$, which emanate from x and meet at time $t = i_o$. This contradicts part (3) of Proposition 8.2.1, which assures us that since $i_o < i_y$, the geodesic $\exp_x(ty)$, $0 \leqslant t \leqslant i_0$ is the *unique* minimizer of arc length among *all* piecewise C^∞ curves which share its endpoints. Thus we must have $i_y \leqslant i_o$ in the first place, and the proof of (2) is complete. □

Exercises

Exercise 8.4.1: Using the fact that an accumulation point of conjugate points is again a conjugate point, give an alternate argument for the statement $i_y \leqslant i_o$ in Proposition 8.4.1.

Exercise 8.4.2:

(a) Explain intuitively why the map $(x, y) \mapsto c_y$ is typically not (fully) continuous.

(b) Can it ever be (fully) continuous?

8.5 Generic Properties of the Cut Locus Cut_x

In this section, we enumerate some generic features of the cut locus. These are valid *without* having to assume that M is compact. Forward geodesic completeness, however, is presumed.

First we list those that follow from our work in §8.2.

Lemma 8.5.1. *Fix a point x in a forward geodesically complete, connected Finsler manifold (M, F). Let $\tilde{x}$ be any point that is* **not** *in the cut locus of x. Then there is exactly one minimizing geodesic from x to $\tilde{x}$.*

Define the following star-shaped "domain" in T_xM:

$$D_x := \{ ty : F(x,y) = 1 \text{ and } 0 \leqslant t < i_y \} . \tag{8.5.1}$$

Keep in mind that i_y can equal ∞. Correspondingly, in M we define

$$\mathcal{D}_x := \exp_x(D_x) . \tag{8.5.2}$$

It is reasonable to expect that $\mathcal{D}_x$ is also star-shaped. This would certainly be the case if the exponential map $\exp_x$, when restricted to D_x, is a diffeomorphism. Such is the thesis of the following proposition.

Proposition 8.5.2. *Fix a point x in a forward geodesically complete Finsler manifold (M, F) and let D_x, $\mathcal{D}_x$ be as defined above. Let Cut_x denote the cut locus of x. Then:*

(1) *$\exp_x$ maps the connected open set D_x diffeomorphically onto the connected open set $\mathcal{D}_x$. Both D_x and $\mathcal{D}_x$ are star-shaped.*

(2) *The boundary of $\mathcal{D}_x$ is $\mathcal{C}ut_x$.*

(3) *Whenever M is connected, it is the disjoint union of $\mathcal{D}_x$ with $\mathcal{C}ut_x$.*

Proof.

For (1), it suffices to check that $\exp_x$ is injective on D_x. Suppose not; then we have

$$\text{(8.5.3a)} \qquad \exp_x(t_1\, y_1) \;=\; \exp_x(t_2\, y_2) \;=: \text{ some } \tilde{x} \text{ , say,}$$

for certain unit vectors y_1, y_2 and $t_1 < i_{y_1}$, $t_2 < i_{y_2}$. In fact, we must have

$$\text{(8.5.3b)} \qquad t_1 \;=\; t_2 \;=\; \operatorname{dist}(x, \tilde{x}) \ .$$

This is because, of the two geodesics $\exp_x(t\, y_1)$, $\exp_x(t\, y_2)$ in question, neither has reached its cut point when they meet at $\tilde{x}$, and hence must still be minimizing at that moment. Nevertheless, according to part (3) of Proposition 8.2.1, the situation we have just depicted in (8.5.3) is a contradiction. Thus $\exp_x$ is a diffeomorphism from D_x onto $\mathcal{D}_x$. Since D_x is connected and star-shaped, so is $\mathcal{D}_x$.

For (2), first note that $\mathcal{C}ut_x \subset \partial\mathcal{D}_x$. Indeed, any neighborhood about a cut point $\hat{x}_y := \exp_x(i_y\, y)$ intersects both $\mathcal{D}_x$ (say at $\exp_x[(i_y - \epsilon)\, y]$) and its complement (say at $\hat{x}_y$).

Conversely, take any $\tilde{x} \in \partial\mathcal{D}_x$; it must lie in the connected component of M that contains $\mathcal{D}_x$. Proposition 6.5.1 says that there exists a minimal unit speed geodesic, say $\exp_x(ty)$, $0 \leqslant t \leqslant r$, from x to $\tilde{x}$. Neither $i_y = \infty$ nor $r < i_y$ is possible; for otherwise $\tilde{x}$ and hence a neighborhood of $\tilde{x}$ lies entirely inside the open set $\mathcal{D}_x$. Also, since the said unit speed geodesic is minimizing, we cannot have $r > i_y$ either. Thus $r = i_y$, and $\tilde{x}$ is a cut point of x.

For (3), suppose M is connected. Given any $\tilde{x} \in M$, take as before a minimal unit speed geodesic $\exp_x(ty)$, $0 \leqslant t \leqslant r$ from x to $\tilde{x}$. If $i_y = \infty$, then of course $r < i_y$ and $\tilde{x}$ must be in $\mathcal{D}_x$. If i_y is finite, then $r \leqslant i_y$ because our geodesic is minimizing; hence $\tilde{x}$ is either in $\mathcal{D}_x$ (if $r < i_y$) or in $\mathcal{C}ut_x$ (if $r = i_y$). ☐

Since the boundary of $\mathcal{D}_x$ is $\mathcal{C}ut_x$, it follows from general point set topology that the cut locus is a closed subset of M. In this trend of thought, the hard work is buried in Proposition 6.5.1. As an alternative, we can invoke the continuity of the function $y \mapsto i_y$ instead. Such is carried out in Exercise 8.5.3. Of course, hard work is also implicit in this second approach, as we can see from the proof of Proposition 8.4.1. Let us now state some consequences of that proposition. The proofs are relegated to the exercises at the end of this section.

Corollary 8.5.3. *Let (M, F) be a forward geodesically complete Finsler manifold. Fix $x \in M$. Let S_x denote the indicatrix at x. For any $y \in S_x$ with $i_y < \infty$, use $\hat{x}_y$ to abbreviate the cut point $\exp_x(i_y\, y)$.*

(1) *The function $y \to dist(x\,,\, \hat{x}_y)$, with its value set equal to ∞ whenever i_y is infinite, is continuous from S_x into $(0, \infty]$.*

(2) *The map $y \mapsto \hat{x}_y$ from S_x onto the cut locus Cut_x of x is continuous wherever it is defined.*

Lemma 8.5.4. *Let (M, F) be a forward geodesically complete Finsler manifold. For each $x \in M$:*

- *The cut locus Cut_x is a closed subset of M.*
- *Cut_x has null Lebesgue (hence null Hausdorff) measure.*

Exercises

Exercise 8.5.1: Prove Lemma 8.5.1. Hints: you will need to use Proposition 6.5.1, the definition of cut points, and part (3) of Proposition 8.2.1.

Exercise 8.5.2: Establish Corollary 8.5.3. For part (2) in that corollary, recall that in order to define the cut point of x in the direction y, one must have $i_y < \infty$.

Exercise 8.5.3: Prove Lemma 8.5.4. Here are some suggestions.

(a) For the part about the Lebesgue measure, use the fact that given any continuous function into the one point compactified interval $[0, \infty]$, its graph must have null measure.

(b) As for the closure of the cut locus, fill in the details of the following two arguments.

* One way is to use $Cut_x = \partial \mathcal{D}_x$ to represent Cut_x as the intersection of two closed subsets of M. This was the point set topology argument we alluded to before.

* Alternatively, one can use the sequential compactness of the indicatrix S_x, the fact that every point in the cut locus has the form $\exp_x(i_y\, y)$ for some $y \in \mathrm{S}_x$, together with the continuity of the function $y \mapsto i_y$.

Exercise 8.5.4: Let (M, F) be a forward geodesically complete Finsler manifold.

(a) Show that the function which sends x to the injectivity radius at x, namely,

$$x \quad \mapsto \quad \mathrm{i}_x := \inf_{y \in \mathrm{S}_x} i_y \ ,$$

is lower semicontinuous from M into $(0, \infty]$. Explain why this function measures the metric distance from x to its cut locus Cut_x.

(b) Show that given any compact subset $K \subset M$, there is a point $\tilde{x} \in K$ such that

$$0 < i_{\tilde{x}} = \inf_{x \in K} \mathfrak{i}_x =: \mathfrak{i}_K .$$

(c) Using this $\mathfrak{i}_K$, Proposition 6.5.1, and part (3) of Proposition 8.2.1, conclude that:

Given any two points p, q in a compact subset K such that $d(p,q) < \mathfrak{i}_K$, there is one and only one globally minimizing geodesic that connects them.

(d) Must that special geodesic lie entirely in the compact set K?

Exercise 8.5.5: Let x be any point in a forward geodesically complete Finsler manifold (M, F). Show that:

- $\mathrm{B}_x(\mathfrak{i}_x) := \{ty : F(x,y) = 1 \text{ and } 0 \leqslant t < \mathfrak{i}_x\}$ is the largest open ball on which the exponential map $\exp_x$ is injective.
- $\exp_x$ maps $\mathrm{B}_x(\mathfrak{i}_x)$ diffeomorphically onto the forward metric ball $\mathcal{B}_x^+(\mathfrak{i}_x) := \{\hat{x} : d(x, \hat{x}) < \mathfrak{i}_x\}$.

8.6 Additional Properties of Cut_x When M Is Compact

We begin this section with a simple but interesting statement from doCarmo [doC3].

Lemma 8.6.1. *Let (M, F) be a forward geodesically complete Finsler manifold.*

(1) *If M is compact, every geodesic contains a cut point.*

(2) *If M is connected and if every geodesic emanating from some particular x contains a cut point, then M is compact.*

Proof. Actually, (1) has already been addressed in Exercise 8.1.2. If M is compact, it must have finite diameter D. No geodesic longer than D can remain minimizing. Thus (1) holds.

Consider (2). Our special x here guarantees that each i_y is finite, hence $\mathcal{D}_x$ [see (8.5.2)] is forward bounded and so is M (because all points can be reached from x). Of course M is closed in its own topology. The Hopf–Rinow theorem (Theorem 6.6.1) now tells us that M is compact. □

When M is compact, the significance of any specific cut locus Cut_x is encapsulated in the following proposition.

Proposition 8.6.2. *Fix a point x in a compact Finsler manifold (M, F). Let Cut_x denote its cut locus. Then:*

(1) *Cut_x is a compact and connected subset of M.*

(2) *The distance $d(x\,, Cut_x)$ is attained at some $\exp_x(i_y\, y)$ in Cut_x.*

(3) *Whenever M is connected, Cut_x is a deformation retract of the punctured space $M \smallsetminus x$.*

Proof. Let $\hat{x}_y$ abbreviate the cut point $\exp_x(i_y\, y)$ of x in the direction of y. Since M is compact, the map $y \mapsto \hat{x}_y$ is defined everywhere on the indicatrix S_x. This map is continuous in view of Corollary 8.5.3. Since S_x is both compact and connected, (1) follows. The continuity (again, see Corollary 8.5.3) of the function $y \mapsto d(x\,, \hat{x}_y)$ gives (2).

Now suppose M is connected. Given any point $\tilde{x}$ in $M \smallsetminus x$, we express it as $\exp_x(t_{\tilde{x}}\, y)$ and define $r(\tilde{x}) := \exp_x(i_y\, y)$. The map $r : M \smallsetminus x \to Cut_x$ is a retraction because it is the identity on Cut_x.

Let $i : Cut_x \to M \smallsetminus x$ be the inclusion. The map $i \circ r : M \smallsetminus x \to M \smallsetminus x$ is homotopic to the identity through

$$h(\tilde{x}\,, \epsilon) \ := \ \exp_x(\,[\,t_{\tilde{x}} \ + \ \epsilon(i_y - t_{\tilde{x}})\,]\; y\,)\ ,$$

with $0 \leqslant \epsilon \leqslant 1$. So r is a deformation retraction of $M \smallsetminus x$ onto the cut locus Cut_x. $\square$

This proposition implies that Cut_x has the same homotopy type, hence the same cohomology, as the punctured space $M \smallsetminus x$.

Exercises

Exercise 8.6.1: Suppose M is compact, connected, and has dimension n at least 3.

(a) Use a Mayer–Vietoris sequence [Spl] to relate the cohomology of Cut_x with that of M.

(b) Consult Kobayashi [Ko] for the argument that relates the homotopy groups of Cut_x to that of M.

Specifically, one has:

$$H^i(\,Cut_x\,) \ \cong \ H^i(M) \quad \text{for} \ \ i \leqslant n-2\ ,$$
$$\pi_i(\,Cut_x\,) \ \cong \ \pi_i(M) \quad \text{for} \ \ i \leqslant n-2\ ,$$

together with a homomorphism of $\pi_{n-1}(\,Cut_x\,)$ onto $\pi_{n-1}(M)$.

Exercise 8.6.2: For the example depicted in Figure 8.2, draw the retraction from $M \smallsetminus x$ onto the cut locus Cut_x.

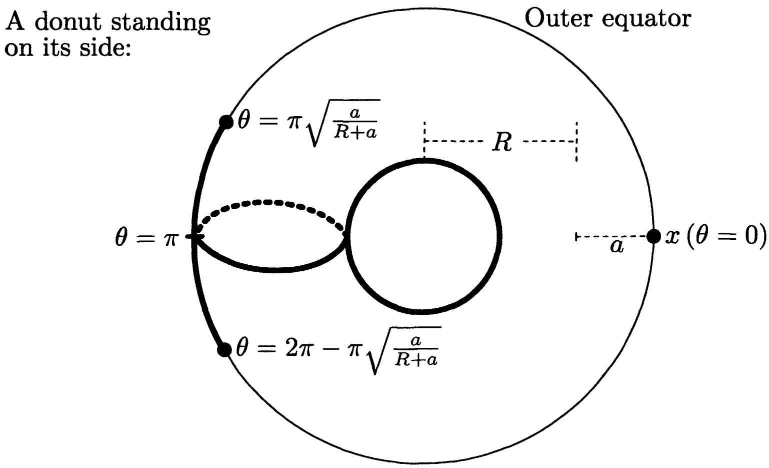

Figure 8.2

Whenever M is connected, Cut_x is a deformation retract of the punctured space $M \smallsetminus x$. Take for example the situation on the torus of revolution, as discussed in Berger–Gostiaux [BG]. Here, x is a point on the outer equator. The cut locus Cut_x is comprised of one arc (endpoints included) and two circles, all *thickened.*

8.7 Shortest Geodesics within Homotopy Classes

In this section, we present a theorem whose proof contains some ideas borrowed from the exposition in Spivak [Sp4]. The underlying trend of thought was, according to Spivak, first used by Hilbert to answer existence questions posed by some calculus of variations problems. The proof also relies on the notion of the injectivity radius of a compact set. This was treated in detail by Exercise 8.5.4.

Theorem 8.7.1. *Let (M, F) be a connected Finsler manifold.*

(1) *Suppose (M, F) is forward geodesically complete.*

- *Fix any two points p and q in M. Then every homotopy class of paths from p to q contains a shortest (but perhaps not uniquely so) smooth geodesic within that class.*

- *In particular, fix any $p \in M$. Then every homotopy class of loops based at p contains a shortest (but perhaps not uniquely so) smooth closed geodesic within that class.*

(2) *Suppose M is compact. Then, in addition to the above, every free homotopy class of loops in M contains a shortest smooth closed geodesic within that class.*

Remarks:

* Free homotopy classes of loops are elements of the space $[\mathbb{S}^1; M]$, which is a groupoid. Homotopy classes of loops based at a fixed p are elements of the **fundamental group** $\pi(M,p)$, which is a bonafide group. Only for some very special manifolds M are the two collections of homotopy classes bijective to each other. This is the case, for example, when M is a path-connected topological group.
* In Chapter 9, we show that when the flag curvature of (M, F) is nonpositive, then those shortest smooth geodesics with designated endpoints are actually unique. So, within each homotopy class of paths from p to q, there is a *unique* shortest smooth geodesic. And, every homotopy class of loops based at p contains a *unique* shortest smooth closed geodesic within that class.
* It is also interesting to compare Theorem 8.7.1, with or without the refinement afforded by nonpositive curvature, with the **Hodge decomposition theorem**. This concerns Finsler manifolds that are smooth, compact, boundaryless, and orientable. One of the conclusions of the Hodge theorem [BL2] says that each cohomology class of M contains a unique harmonic representative. Note also that, just as geodesics are extremals of the arc length functional on the space of paths, harmonic forms are extremals of certain energy functionals on the space of differential forms. The Euler–Lagrange equation for geodesics is a system of second-order quasilinear ODEs, while that for harmonic forms is a system of second-order linear elliptic PDEs. In this light, Theorem 8.7.1 may be viewed as a key component in the homotopy analogue of Hodge theory.

Proof.

Consider a specific homotopy class, say α, of piecewise smooth curves. [In case (1), these would be piecewise smooth curves from the fixed p to the fixed q.] The *geometric length* of this class is defined as

$$|\alpha| \ := \ \inf_{c \in \alpha} \ L(c) \ ,$$

where the infimum is taken over all piecewise smooth curves c in the class. A piecewise smooth path σ in the homotopy class α is said to be shortest among curves in α if $L(\sigma) = |\alpha|$. The existence of such a σ is demonstrated momentarily. Let us now digress to check that if it exists, it must be a smooth geodesic.

- **Regularity:**

Without loss of generality, let us suppose that $\sigma(t)$, $0 \leqslant t \leqslant r$ has been parametrized to have unit speed. Denote its velocity field by $T(t)$. *A priori*, σ is only piecewise smooth, so there is an associated partition $0 =: t_o < t_1 < \cdots < t_k := r$ such that σ is smooth on each closed subinterval $[t_{i-1}\,,\, t_i]$.

Let $D_T T$ be defined with reference vector T. We construct a variation of σ with variation vector field $U(t) = f(t)\, D_T T$, where $f(t)$ is zero at $t_o, \ldots, t_k$ and is positive on each $(t_{i-1}\,,\, t_i)$. The t-curves of this variation are then necessarily homotopic to σ. Substituting this $U(t)$ into the first variation (see Exercise 5.2.4) and using the fact that σ is a shortest curve in the class α, we have

$$- \int_0^r f(t)\, g_T(D_T T\,,\, D_T T)\, dt \;=\; L'(0) \;=\; 0\,.$$

So $D_T T$ (with reference vector T) must vanish identically. Therefore σ is a unit speed geodesic.

Given this, the first variation (Exercise 5.2.4 again) reads

$$L'(0) \;=\; - \sum_{i=1}^{k-1} g_T\Big(\, U(t_i)\,,\; T(t_i^+) - T(t_i^-)\,\Big)$$

for any variation of σ. Next, construct a variation such that the variation vector field satisfies $U(t_i) = T(t_i^+) - T(t_i^-)$ for $i = 1, \ldots, k-1$. The t-curves here are again homotopic to σ. Since σ is shortest, we see that

$$- \sum_{i=1}^{k-1} g_T\Big(\, U(t_i)\,,\; T(t_i^+) - T(t_i^-)\,\Big) \;=\; L'(0) \;=\; 0\,.$$

This implies that the "jumps" $T(t_i^+) - T(t_i^-)$ must all vanish. Hence the geodesic σ is C^1. By the discussion in Exercise 5.3.1, it must in fact be C^∞ (provided that our Finsler structure is C^∞ away from the zero section in TM).

- **Existence:**

We next turn to the existence of a shortest σ in the given homotopy class α. Let c_i be a sequence of paths in the class α such that $L(c_i) \longrightarrow |\alpha|$. Let us enumerate some information about these paths:

(a) The number $L := \sup L(c_i)$ is finite because the lengths $L(c_i)$ form a convergent sequence.

(b) In the case that M is forward geodesically complete, the paths c_i must all lie inside some sufficiently large "forward" metric ball $\mathcal{B}_p^+(R)$. Otherwise, some c_i always wanders outside any given $\mathcal{B}_p^+(r)$, no matter how large r may be; but then the lengths $L(c_i)$ won't even be uniformly bounded, let alone convergent. The closure of $\mathcal{B}_p^+(R)$

is forward bounded, hence a compact set by the Hopf–Rinow theorem (Theorem 6.6.1). Call that compact set K. Thus, the fact that the paths in question have fixed endpoints enables us to confine the dynamics to a compact ball K.

(c) In the case that M is compact, the analogue of the above K is simply M itself. Note that in this case, the homotopy classes in question are not assumed to have a fixed base point. Since free loops are not pinned down, there is *a priori* no builtin mechanism to prevent them from escaping to infinity. For instance, on the bugle surface (which is complete), the circular parallels are all homotopic to each other, and there is always a shorter one farther out. Replacing completeness by the considerably more stringent compactness rules out this phenomenon, and restores our analytical control of the situation. This is why compactness is needed when dealing with free homotopy groups.

Exercise 8.5.4 assures us that

$$\mathfrak{i}_K \ := \ \inf_{x \in K} \mathfrak{i}_x \ > \ 0 \ .$$

Part (c) of the same exercise says that:

> Given any two points in (the compact) K that are less than $\mathfrak{i}_K$ units apart (as measured by the metric distance function), there is one (by completeness) and only one [by part (3) of Proposition 8.2.1] globally minimizing geodesic connecting them.

Getting things set up:

Using this observation, we cut up every c_i into arcs (say N_i of them) with lengths less than $\frac{1}{2}\mathfrak{i}_K$. In particular, the metric distance between the endpoints of each arc is less than $\frac{1}{2}\mathfrak{i}_K$. (That factor of $\frac{1}{2}$ was introduced for technical reasons which later become clear.) A *single* large enough N can be used for *all* the c_i. Indeed, any integer N that satisfies

$$N \left(\frac{1}{2} \, \mathfrak{i}_K \right) \ > \ L$$

will do.

Replace each arc, say c_i^s $(s = 1, \ldots, N)$, by the corresponding globally minimizing geodesic discussed above. This move is not absolutely necessary. It is undertaken to streamline arguments at and after (*) below. Denote the resulting modified version of c_i as σ_i. By construction, $L(\sigma_i^s) \leqslant L(c_i^s)$ for each s, so

$$L(\sigma_i) \ \leqslant \ L(c_i) \ .$$

Each of the newly constructed σ_i is a piecewise smooth curve made up of N geodesic segments σ_i^s, $s = 1, \ldots, N$. Let x_i^{s-1} and x_i^s denote the two

For each fixed $s \in \{1, \ldots, N\}$:

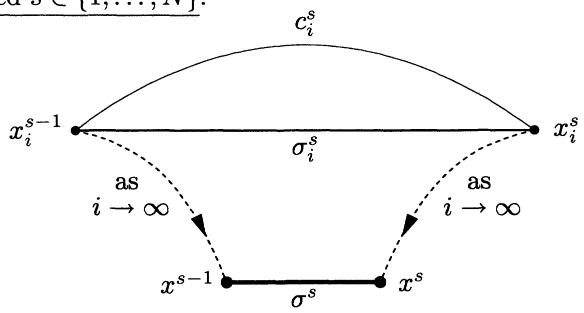

Figure 8.3

A sequence of paths c_i in the homotopy class α. They lie inside some common compact set K, and $L(c_i) \to |\alpha|$. We cut each c_i up into a large number N of short arcs, each with length $< \frac{1}{2}\mathfrak{i}_K$. Each short arc c_i^s $(s = 1, \ldots, N)$ is homotopic to a minimizing geodesic σ_i^s. The homotopy exists because c_i^s and σ_i^s both lie inside some small forward metric ball, which is in turn diffeomorphic to a simply connected tangent Finsler ball. For each s, the endpoints of the arcs converge to a pair of points x^{s-1} and x^s. These are joined by a minimizing geodesic σ^s.

endpoints of each such segment. We digress to check that σ_i^s and c_i^s are homotopic, through a homotopy that keeps their common endpoints fixed. This would tell us that σ_i can be deformed to c_i by a piecewise smooth homotopy.

Indeed, the statement

$$L(\sigma_i^s) \leqslant L(c_i^s) < \frac{1}{2}\mathfrak{i}_K < \mathfrak{i}_K$$

implies that for any point z along c_i^s or σ_i^s, we have $d(x_i^{s-1}, z) < \mathfrak{i}_K$. In order to avoid clutter, let us temporarily abbreviate x_i^{s-1} as x. Thus the two segments c_i^s and σ_i^s are both contained in the forward metric ball $\mathcal{B}_x^+(\mathfrak{i}_K)$. Using $\mathfrak{i}_K \leqslant \mathfrak{i}_x$, together with the first part of Proposition 8.5.2, we conclude that $\exp_x$ maps $\mathrm{B}_x(\mathfrak{i}_K) := \{ty : F(x, y) = 1 \text{ and } 0 \leqslant t < \mathfrak{i}_K\}$ diffeomorphically onto the above metric ball. The sought homotopy now follows readily from the fact that the tangent Finsler ball $\mathrm{B}_x(\mathfrak{i}_K)$ is simply connected.

Since σ_i is homotopic to c_i, the two belong to the same homotopy class α. In particular, $|\alpha| \leqslant L(\sigma_i)$. Combining this with our previous observation $L(\sigma_i) \leqslant L(c_i)$, we get

$$|\alpha| \leqslant L(\sigma_i) \leqslant L(c_i) .$$

Given that $L(c_i) \to |\alpha|$, one must have

$$L(\sigma_i) \;\to\; |\alpha|$$

as well.

The construction of a shortest geodesic:

Let us resume the discussion proper. For each fixed s, the sequence of points $\{x_i^s : i = 1, \ldots, \infty\}$ lies in the compact set K, and hence contains a convergent subsequence. There are only finitely many values for the superscript s on x_i^s. Namely, $0, \ldots, N$. So, by passing to a subsequence (in the index i) if necessary, we can assume without loss of generality that for every value s, there exists some point $x^s \in K$ such that

$$\lim_{i \to \infty} d(x^s\,,\, x_i^s) \;=\; 0\,.$$

Also,

$$\begin{aligned} d(x_i^{s-1}\,,\, x_i^s) \;-\; d(x^{s-1}\,,\, x^s) \;=\; & \;[\; d(x_i^{s-1}\,,\, x_i^s) \;-\; d(x_i^{s-1}\,,\, x^s)\;] \\ & + [\; d(x_i^{s-1}\,,\, x^s) \;-\; d(x^{s-1}\,,\, x^s)\;]\,. \end{aligned}$$

Thus the *equicontinuity* of the distance function, detailed in Exercise 6.2.9, can be used to deduce that

$$\lim_{i \to \infty} d(x_i^{s-1}\,,\, x_i^s) \;=\; d(x^{s-1}\,,\, x^s)\,.$$

Now

$$d(x_i^{s-1}\,,\, x_i^s) \;<\; \frac{1}{2}\,\mathrm{i}_K$$

which, upon taking the limit, gives

$$d(x^{s-1}\,,\, x^s) \;\leqslant\; \frac{1}{2}\,\mathrm{i}_K \;<\; \mathrm{i}_K\,.$$

Therefore there exists a (unique) globally minimizing geodesic segment σ^s from x^{s-1} to x^s. Denote by σ the union of the segments $\sigma^1, \ldots, \sigma^N$. It is *a priori* a broken geodesic. Note however that

$$\begin{aligned} L(\sigma) \;&=\; \sum_{s=1}^{N} L(\sigma^s) \;=\; \sum_{s=1}^{N} d(x^{s-1}\,,\, x^s) \\ &=\; \sum_{s=1}^{N} \lim_{i \to \infty} d(x_i^{s-1}\,,\, x_i^s) \;=\; \sum_{s=1}^{N} \lim_{i \to \infty} L(\sigma_i^s) \\ &=\; \lim_{i \to \infty} L(\sigma_i) \\ (*) \qquad &=\; |\alpha|\,. \end{aligned}$$

(We will soon show that σ belongs to the class α.) Hence σ is shortest and, by the established regularity, it is actually a smooth geodesic.

Showing that our minimizer is indeed in the homotopy class α: We must check that this σ really lies in the homotopy class α. It suffices to show that σ is homotopic to σ_i (for some sufficiently large i) because we already know that σ_i is homotopic to c_i.

To this end, fix a large i such that for every $s = 0, \dots, N$, we have $d(x^s, x_i^s) < \frac{1}{2}\mathfrak{i}_K$. This is possible because for every s (and there are only finitely many s), the sequence $\{x_i^s : i = 1, \dots, \infty\}$ converges to x^s. Since the exponential map at x^s is a diffeomorphism from the tangent Finsler ball $\mathrm{B}_{x^s}(\frac{1}{2}\mathfrak{i}_K)$ onto the forward metric ball $\mathcal{B}^+_{x^s}(\frac{1}{2}\mathfrak{i}_K)$, we can connect x^s to x_i^s by a unique globally minimizing radial geodesic γ^s that lies entirely inside $\mathcal{B}^+_{x^s}(\frac{1}{2}\mathfrak{i}_K)$. We construct our deformation of σ to σ_i by "sliding" along these "guideposts" γ^s.

Fix a large i.

For each fixed $s \in \{1, \dots, N\}$:

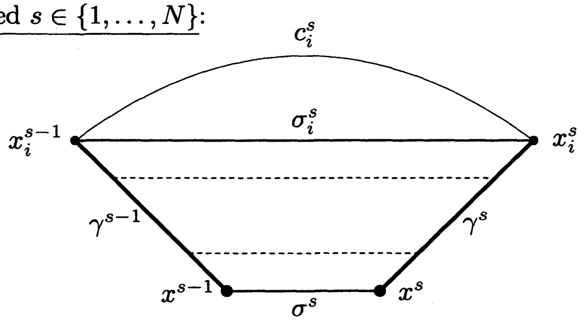

Figure 8.4

Our shortest geodesic σ indeed belongs to the same homotopy class α as the intermediate σ_i, each of which is comprised of geodesic arcs. This is so because σ is homotopic to some σ_i, for i sufficiently large. The intuition behind this fact is fairly simple. In the forward metric ball $\mathcal{B}^+_{x^{s-1}}(\mathfrak{i}_K)$, we have a curvilinear rectangle with geodesic edges. Here, the latter are drawn as "straight" segments to emphasize geodecity. The top and bottom edges are, respectively, given by σ_i^s and σ^s. The left and the right edges are, respectively, the "guideposts" γ^{s-1} and γ^s. The latter are minimizing geodesics that join x_i^{s-1} and x_i^s to x^{s-1} and x^s, respectively. A homotopy (see those dotted lines) between σ_i^s and σ^s is obtained by "sliding" the endpoints x^{s-1} and x^s along the said guideposts. This is made possible by the fact that the forward metric ball $\mathcal{B}^+_{x^{s-1}}(\mathfrak{i}_K)$ is small, and is therefore diffeomorphic to a simply connected tangent Finsler ball.

Take any point x along σ_i^s, which has left endpoint x_i^{s-1} and right endpoint x_i^s. The triangle inequality tells us that

$$d(x^{s-1}, x) \leqslant d(x^{s-1}, x_i^{s-1}) + d(x_i^{s-1}, x) < \frac{\mathfrak{i}_K}{2} + \frac{\mathfrak{i}_K}{2} = \mathfrak{i}_K .$$

Also, $d(x^{s-1}, x^s) \leqslant \frac{1}{2} i_K$ implies that every point x along σ^s satisfies the criterion $d(x^{s-1}, x) \leqslant \frac{1}{2}\mathfrak{i}_K$. Thus both segments σ_i^s and σ^s are contained in the forward metric ball $\mathcal{B}^+_{x^{s-1}}(\mathfrak{i}_K)$.

The left endpoints of σ^s and σ_i^s, namely, x^{s-1} and x_i^{s-1}, are connected by the path γ^{s-1} which lies in $\mathcal{B}^+_{x^{s-1}}(\frac{1}{2}\mathfrak{i}_K)$. Likewise, their right endpoints x^s and x_i^s are connected by γ^s, which lies in $\mathcal{B}^+_{x^s}(\frac{1}{2}\mathfrak{i}_K)$. Note that

$$\mathcal{B}^+_{x^s}\left(\frac{\mathfrak{i}_K}{2}\right) \subset \mathcal{B}^+_{x^{s-1}}(\mathfrak{i}_K) .$$

Indeed, let x be any point in $\mathcal{B}^+_{x^s}(\frac{1}{2}\mathfrak{i}_K)$; then $d(x^{s-1},x) \leqslant d(x^{s-1},x^s) + d(x^s,x)$, which is $< \mathfrak{i}_K$ because the first term on the right is at most $\frac{1}{2}\mathfrak{i}_K$, while the second is strictly less than $\frac{1}{2}\mathfrak{i}_K$.

Thus, in the forward metric ball $\mathcal{B}^+_{x^{s-1}}(\mathfrak{i}_K)$, we have the two segments σ^s and σ_i^s, together with the geodesic γ^{s-1} which connects their left endpoints, and the geodesic γ^s connecting their right endpoints. Since this metric ball is diffeomorphic to the (simply connected) tangent Finsler ball $\mathrm{B}_{x^{s-1}}(\mathfrak{i}_K)$, a homotopy h^s between σ^s and σ_i^s can be constructed, with the prescribed "boundary data" γ^{s-1} and γ^s. All the intermediate curves in this deformation are contained in $\mathcal{B}^+_{x^{s-1}}(\mathfrak{i}_K)$.

We digress to elaborate on case (1), wherein (M,F) is forward geodesically complete and the homotopy classes concern paths with fixed endpoints p and q. In that case, the x_i^0 are equal to p and the x_i^N are equal to q, for $i = 1, \dots, \infty$. So, the very first guidepost γ^0 and the very last guidepost γ^N are both constant curves; they are equal to p and q, respectively.

Let's return to the proof proper. The homotopies h^s, with $s = 1, \dots, N$, collectively produce a piecewise smooth homotopy from our (possibly non-unique!) shortest geodesic σ to the broken geodesic σ_i. Furthermore, σ_i is by construction homotopic to the path c_i, which belongs to the homotopy class α. This shows that our shortest smooth geodesic σ indeed belongs to the class α, and completes the proof of the theorem. ☐

Exercises

Exercise 8.7.1: In the proof of Theorem 8.7.1, we have the sequence of broken geodesics $\{\sigma_i\}$, together with the geodesic σ which is a shortest curve in the homotopy class α. Without loss of generality, let us suppose that the curves σ_i, and σ, have all been parametrized to have constant speed and time domain $[0,1]$. By passing to a subsequence if necessary, show that as maps from $[0,1]$ into the compact set K, the σ_i converge uniformly pointwise to σ.

Exercise 8.7.2: When setting things up to prove existence, we cut up each path c_i into N pieces of length $< \frac{1}{2}\mathfrak{i}_K$. We mentioned that our use of that $\frac{1}{2}$ was due to a technical reason. Pinpoint exactly where and how that factor of $\frac{1}{2}$ was needed in the proof of Theorem 8.7.1.

Exercise 8.7.3: Would the conclusions of Theorem 8.7.1 remain the same if (M,F) were *backward* geodesically complete (see §6.6) instead?

8.8 Synge's Theorem

Here's a result that makes pivotal use of part (2) of Theorem 8.7.1. It was first derived for Finsler manifolds by Auslander [Au]. See also [Daz].

Theorem 8.8.1 (Synge). *Let (M,F) be a Finsler manifold. Suppose:*

- *M is even-dimensional, oriented, and connected.*
- *(M,F) is forward geodesically complete.*
- *All flag curvatures are bounded below uniformly by some positive constant λ.*

Then:

(a) *Every free homotopy class of loops in M is trivial. That is, every loop is freely homotopic in M to a point.*

(b) *M is simply connected. That is, for any fixed $x \in M$, every loop based at x can be deformed to the point x through a family of loops that are all based at x.*

Proof.

Checking that conclusions (a) and (b) are equivalent:

* Suppose (a) holds. Take any loop $c(t)$, $1 \leqslant t \leqslant 2$ based at x. By ignoring the base point x, we use (a) to obtain a free homotopy from c to a single point, say p. This free homotopy consists of a 1-parameter family of loops $c_\kappa(t)$, $1 \leqslant t \leqslant 2$. Here, $\kappa \in [0,1]$ and, $c_0(t) = c(t)$, $c_1(t) = p$.

As κ increases from 0 to 1, $c_\kappa(1)$ [equivalently $c_\kappa(2)$] traces out an arc from x to the point p. This arc will help us define a family of loops that are all based at x, and which serves as a deformation of c to the point x.

To this end, for each $\kappa \in [0,1]$, define

$$h_\kappa(t) \;:=\; \begin{cases} c_{t\kappa}(1) & \text{for } 0 \leqslant t \leqslant 1 \\ c_\kappa(t) & \text{for } 1 \leqslant t \leqslant 2 \\ c_{3\kappa - t\kappa}(2) & \text{for } 2 \leqslant t \leqslant 3\,. \end{cases}$$

This h_κ heads from x to the starting point of c_κ, goes around the loop c_κ once, and then returns from the final point of c_κ back to x. When $\kappa = 0$, h_κ is a reparametrized form of c. When $\kappa = 1$, h_κ goes from x to p along some arc during the first second, sits at p for a full second, and then returns from p back to x using the reverse of the said arc. Call this loop γ. A reparametrized version of c is thus homotopic to γ through loops that are all based at x.

It is straightforward to deform γ to x using loops that are all based at x. Therefore we have demonstrated that (a) implies (b).

* Suppose (b) holds. Take any loop $c(t)$ and declare its starting point as our base point x. Then (a) immediately follows.

Deducing (a) from the hypotheses of the theorem:

It therefore suffices to establish (a). In view of the second and the third hypotheses, we see from the Bonnet–Myers theorem (Theorem 7.7.1) that M must be compact.

Suppose, for the sake of argument, that there is a nontrivial free homotopy class α. We derive a contradiction.

By part (2) of Theorem 8.7.1, α contains a shortest smooth closed geodesic σ. Without loss of generality, we may assume that $\sigma(t)$, $0 \leqslant t \leqslant L$ has already been parametrized to have unit speed. Since the loops in α are, by the above supposition, not homotopic to a point, σ cannot have zero length. In other words, L must be positive.

Abbreviate $\sigma(0)$ as x, and the initial velocity of σ as T. The unit velocity field $T(t)$ of σ is occasionally abbreviated as T also, when there is no danger of confusion.

Parallel transport once around the geodesic loop σ, with reference vector $T(t)$, is an orientation preserving isomorphism $P_\sigma : T_xM \to T_xM$. Let $T^\perp$ denote the g_T-orthogonal complement of the unit vector T in the tangent space T_xM. Recall that g_T refers to the inner product $g_{(\sigma,T)}$. By Exercise 5.2.3, the map P_σ preserves g_T lengths and g_T angles. So its restriction to $T^\perp$, say $\mathcal{P}_\sigma : T^\perp \to T^\perp$, is well defined. We know that:

* $T^\perp$ is odd-dimensional because M has even dimension.
* $\mathcal{P}_\sigma$ is an orthogonal transformation with positive determinant (equal to 1) because its predecessor P_σ is orientation preserving, and $P_\sigma(T) = T$.

Observe that:

- The characteristic polynomial of $\mathcal{P}_\sigma$ has real coefficients. So its complex eigenvalues must come in complex conjugate pairs.
- That polynomial also has odd degree; hence there is necessarily an odd number of real eigenvalues. One of those real eigenvalues must be positive because $\det \mathcal{P}_\sigma = 1$.
- Since $\mathcal{P}_\sigma$ preserves lengths of vectors, all its eigenvalues have norm one. Hence that real positive eigenvalue must be equal to 1.

Consequently, there is a vector $u \in T_xM$ that is g_T-orthogonal to T and kept invariant by P_σ. By rescaling u if necessary, we may assume that it has g_T length one.

Follow the evolution of u as it gets parallel transported (with reference vector T) once around σ. It generates a parallel vector field $U(t)$ along σ before returning to itself. Produce a variation of σ by piecewise smooth loops such that the resulting variation vector field is this $U(t)$. Keep in mind that:

* The base geodesic σ in this variation is a closed curve of unit speed.
* The variation vector field $U(t)$ is g_T-orthogonal to σ at all times.
* $D_T U = 0$ all along σ.
* $g_T(U(t), U(t)) = 1$ for all $t \in [0, L]$.

Here, the second and the fourth items follow from Exercise 5.2.3. These observations reduce the second variation of arc length (see Exercises 5.2.7 and 5.2.6) to

$$L''(0) = \int_0^L -K(T,U)\, dt .$$

By hypothesis, all flag curvatures of (M, F) are bounded below by the positive constant λ. Thus

$$L''(0) \leqslant -\lambda L < 0 .$$

This implies that many loops (near σ) in this variation are strictly shorter than σ, contradicting the defining property of σ as a shortest loop. So our original supposition must be wrong; there is no nontrivial free homotopy class after all. □

Similar ideas are involved in the proof of an estimate (due to Klingenberg) on the injectivity radius of certain Riemannian manifolds. See [CE].

Exercises

Exercise 8.8.1: For Synge's theorem, verify that

(a) Real projective space of dimension two is a counterexample if we drop the orientability hypothesis.

(b) Real projective space of dimension three is a counterexample if we drop the hypothesis that M is even-dimensional.

See [Sp4] for more discussions.

Exercise 8.8.2: Contemplate the following questions which complement Synge's theorem. See [Sp4], [GHL], and also [Daz] for insights.

> Let (M, F) be a forward geodesically complete Finsler manifold whose flag curvatures are uniformly bounded below by a positive constant.
>
> - If M is even-dimensional and nonorientable, must its fundamental group be $\mathbb{Z}_2$?
> - If M is odd-dimensional, must it be orientable?

References

[Au] L. Auslander, *On curvature in Finsler geometry*, Trans. AMS **79** (1955), 378–388.

[BG] M. Berger and B. Gostiaux, *Differential Geometry: Manifolds, Curves, and Surfaces*, Graduate Texts in Mathematics **115**, Springer-Verlag, 1988.

[BL2] D. Bao and B. Lackey, *A Hodge decomposition theorem for Finsler spaces*, C.R. Acad. Sci. Paris **323** (1996), 51–56.

[CE] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, North Holland/American Elsevier, 1975.

[Daz] P. Dazord, *Propriétés globales des géodésiques des Espaces de Finsler*, Theses, Université de Lyon, 1969.

[doC3] M. P. do Carmo, *Riemannian Geometry*, Mathematics: Theory and Applications, Birkhäuser, 1992.

[GHL] S. Gallot, D. Hulin and J. Lafontaine, *Riemannian Geometry*, Universitext, 2nd ed., Springer-Verlag, 1990.

[Ko] S. Kobayashi, *On conjugate and cut loci*, Global Differential Geometry, S. S. Chern, ed., Math. Assoc. America, 1989, pp. 140–169.

[Sp1] M. Spivak, *Differential Geometry*, vol. I, Publish or Perish, 1975.

[Sp4] M. Spivak, *Differential Geometry*, vol. IV, Publish or Perish, 1975.

Chapter 9

The Cartan–Hadamard Theorem and Rauch's First Theorem

9.1 Estimating the Growth of Jacobi Fields

In §5.5, we estimated the growth of certain Jacobi fields using the first few terms of a power series. That was valid only for a short time interval. In the present section, we use a more delicate approach—known as a comparison argument. The resulting estimate holds for long time intervals.

The hypotheses here are as follows:

- $\sigma(t)$, $0 \leqslant t \leqslant L$ is a *unit speed* geodesic with velocity field T.
- J is a nonzero Jacobi field along σ such that $J(0) = 0$.

We do **not** assume *a priori* that our Finsler manifold is forward geodesically complete.

Our goal is to estimate the growth of

$$\|J\| \ := \ \sqrt{g_T(J,J)} \ .$$

We seek a *lower* bound on $\|J(t)\|$ when the flag curvature $K(T, J)$ is uniformly bounded *above* by a constant λ. The result is used to help deduce the Cartan–Hadamard theorem in §9.4. It turns out that in order to derive this lower bound on $\|J(t)\|$, we need to first estimate the index $I(J, J)$ from below.

> The case wherein K is uniformly bounded *below* is not treated in the section proper. Instead, we relegate it to a much guided Exercise 9.1.3. In this case, we seek an *upper* bound (rather than a lower bound) on $\|J(t)\|$. As we show in part (b) of Exercise 9.1.3, a key step calls for an upper bound on the index $I(J, J)$. Thus the basic index lemma is happily relevant, thereby simplifying a technical portion of the argument.

Note that J can be resolved into two components, one tangential to σ and one which is g_T-orthogonal to it. In view of Exercises 5.4.3 and 5.4.5, the growth rate of the tangential component is well understood. As a matter of fact, it is linear in t. For this reason, there is no conceptual loss in assuming that:

- J is everywhere g_T-orthogonal to σ.

We first dispense with two technical issues:

* Since J vanishes at $t = 0$ but is nonzero, there must exist a positive $l \leqslant L$ such that J is nowhere zero on $(0, l]$. Our analysis is carried out on $[0, l]$. Such caution is necessary because some of the intermediate formulas below involve division by $\|J\|$. Later on, we explain why the estimate actually holds on $[0, L]$, provided that

$$L \leqslant \frac{\pi}{\sqrt{\lambda}} \quad \text{in the case of positive } \lambda \, .$$

* For each fixed r in $(0, l]$, let I_r denote the index form in which the integration is carried out from 0 to r.

Estimating the index form:

According to Exercise 7.2.3,

$$\frac{1}{2} \frac{d}{dt}_{|t=r} \, g_T(J, J) \;=\; I_r(J, J) \, .$$

Let us estimate $I_r(J, J)$ under the curvature assumption

$$K(T, J) \;\leqslant\; \lambda \, .$$

As a preliminary step, we have:

$$\begin{aligned} I_r(J, J) \;&=\; \int_0^r \left[\, g_T(J', J') \;-\; K(T, J)\, \|J\|^2 \,\right] dt \\ &\geqslant\; \int_0^r \left[\, (\,\|J\|'\,)^2 \;-\; \lambda\, \|J\|^2 \,\right] dt \, . \end{aligned}$$

Here, J' abbreviates the covariant derivative $D_T J$, taken with reference vector T. Also, a Schwarz inequality has been applied without mention:

$$(\|J\|')^2 = \left[\frac{g_T(J', J)}{\|J\|} \right]^2 \leqslant g_T(J', J') .$$

Anyway, we have just shown that

$$\frac{1}{2} (\|J\|^2)'(r) \geqslant \int_0^r [(\|J\|')^2 - \lambda \|J\|^2] \, dt . \tag{9.1.1}$$

A comparison argument on functions:

In order to estimate that integral, we appeal to Exercise 7.3.3. Define

$$\mathcal{F} := \left\{ \begin{array}{c} \frac{\|J(r)\|}{\sin(\sqrt{\lambda}\, r)} \sin(\sqrt{\lambda}\, t) \\ \frac{\|J(r)\|}{r} \, t \\ \frac{\|J(r)\|}{\sinh(\sqrt{-\lambda}\, r)} \sinh(\sqrt{-\lambda}\, t) \end{array} \right\} \text{ resp., for } \left\{ \begin{array}{c} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \end{array} \right\} .$$

Since our fixed r can be any number in $(0, l]$, we must impose the constraint:

$$0 < r \leqslant l < \frac{\pi}{\sqrt{\lambda}} \quad \text{in the case of } \lambda > 0 .$$

The exercise in question says that:

$$\int_0^r [(\|J\|')^2 - \lambda \|J\|^2] \, dt \geqslant \mathcal{F}'(r) \, \mathcal{F}(r) . \tag{9.1.2}$$

Some calculus:

Let us combine (9.1.1) with (9.1.2), and divide the resulting statement by $\|J(r)\|^2$. Since $\mathcal{F}(r) = \|J(r)\|$ by construction, we have

$$\frac{1}{2 \, \|J(r)\|^2} (\|J\|^2)'(r) \geqslant \frac{\mathcal{F}'(r)}{\mathcal{F}(r)} = \frac{\mathfrak{s}_\lambda'(r)}{\mathfrak{s}_\lambda(r)} ,$$

where

$$\boxed{\mathfrak{s}_\lambda(t) := \left\{ \begin{array}{c} \frac{1}{\sqrt{\lambda}} \sin(\sqrt{\lambda}\, t) \\ t \\ \frac{1}{\sqrt{-\lambda}} \sinh(\sqrt{-\lambda}\, t) \end{array} \right\} \text{ resp., for } \left\{ \begin{array}{c} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \end{array} \right\}} . \tag{9.1.3}$$

Thus

$$(\log \|J\|)'(r) \geqslant (\log \mathfrak{s}_\lambda)'(r) \quad \text{at all } r \in (0, l] .$$

Equivalently:

$$\left(\log \frac{\|J\|}{\mathfrak{s}_\lambda}\right)' \geqslant 0 \quad \text{on} \quad (0, l] \ .$$

In particular, the quotient $\|J\|/\mathfrak{s}_\lambda$ is nondecreasing and

$$\frac{\|J\|}{\mathfrak{s}_\lambda}(t) \geqslant \lim_{r \to 0^+} \frac{\|J\|}{\mathfrak{s}_\lambda}(r) \ .$$

It turns out that the limit of $\|J\|^2/\mathfrak{s}_\lambda^2$ is considerably easier to compute than that of $\|J\|/\mathfrak{s}_\lambda$. Carrying out that calculation using L'Hôpital's rule (see Exercise 9.1.1), we update the above inequality to

$$\|J(t)\| \geqslant \mathfrak{s}_\lambda(t) \ \|J'(0)\| \ . \tag{9.1.4}$$

Since $J(0) = 0$, we must have $J'(0) \neq 0$, otherwise our Jacobi field J would have to vanish identically.

Showing that our estimate is actually valid on $[0, L]$:

As of now, (9.1.4) is valid on all intervals $[0, l]$ such that $J(t), 0 < t \leqslant l$ is nowhere zero.

* Because of (9.1.4), the first positive zero of $J(t)$ cannot possibly come before the first positive zero of $\mathfrak{s}_\lambda(t)$. (*Prove* it.)
* For the $\lambda < 0$ and $\lambda = 0$ cases, $\mathfrak{s}_\lambda(t)$ has no positive zero. Therefore neither does $J(t)$. This means we could have chosen l to be L. That is, (9.1.4) holds on $[0, L]$.
* For the $\lambda > 0$ case, the first positive zero of $\mathfrak{s}_\lambda(t)$ occurs at

  $$t = \frac{\pi}{\sqrt{\lambda}} \ .$$

  In this case, let us impose the restriction

  $$L \leqslant \frac{\pi}{\sqrt{\lambda}} \ .$$

  Our proof tells us that (9.1.4) is valid on $[0, L]$ if $L < \pi/\sqrt{\lambda}$. An inspection assures us that the validity persists even if $L = \pi/\sqrt{\lambda}$.

Let us summarize:

Theorem 9.1.1. *Suppose:*

- *(M, F) is a Finsler manifold whose flag curvature is bounded above by a constant λ.*
- *$\sigma(t)$, $0 \leqslant t \leqslant L$ is a unit speed geodesic with velocity field T, and $L \leqslant \frac{\pi}{\sqrt{\lambda}}$ in the case of positive λ.*
- *J is a Jacobi field that is g_T-orthogonal to σ, and we have $J(0) = 0$, $J'(0) \neq 0$.*

Then the norm $\|J\| := \sqrt{g_T(J,J)}$ *satisfies the inequality*

$$\boxed{\|J(t)\| \;\geqslant\; \mathfrak{s}_\lambda(t)\; \|J'(0)\|} \quad \textit{on} \;\; [0, L] \; .$$

Remarks:

* Here, $\mathfrak{s}_\lambda$ is given explicitly by (9.1.3).
* The said norm $\| \quad \|$ is induced by g_T, which is an inner product. It is *not* the Finsler norm F. Put another way, $g_T(V, V)$ and $F^2(V)$ are typically different, unless V is a positive multiple of T. For this reason, the Jacobi field estimate is lamentably not a sharp one. Nevertheless, it is still useful. This echoes similar sentiments expressed by Egloff [E].
* Exercise 9.1.3 deals with an analogue of this theorem when the flag curvature is bounded below. In that case, we seek an upper bound on $\|J(t)\|$. It reads:

  $$\|J(t)\| \;\leqslant\; \mathfrak{s}_\lambda(t)\; \|J'(0)\| \quad \text{on} \;\; (0, l) \; .$$

  The number l is made precise in that exercise.

Proposition 9.1.2. *On any Finsler manifold* (M, F) *of nonpositive flag curvature, no geodesic can contain any conjugate points.*

Proof. This follows from Theorem 9.1.1 by setting $\lambda = 0$ there. Indeed, let $\sigma(t)$, $0 \leqslant t \leqslant L$ be any geodesic with velocity T. Suppose there is a nonzero Jacobi field J along σ such that $J(0) = 0 = J(r)$, with $r \in [0, L]$. According to Exercise 5.4.4, J must be g_T-orthogonal to σ at all times. Furthermore, since $J(0) = 0$ but J is nonzero, we must have $J'(0) \neq 0$. By Theorem 9.1.1, the following inequality holds on $[0, L]$:

$$\|J(t)\| \;\geqslant\; t\; \|J'(0)\| \; .$$

In particular, $J(r)$ cannot possibly be zero. Thus we have arrived at a contradiction, and there are no conjugate points after all. $\square$

Exercises

Exercise 9.1.1: Apply L'Hôpital's rule twice to deduce that

$$\begin{aligned}\lim_{r\to 0^+} \frac{\|J\|^2}{\mathfrak{s}_\lambda^2} \;&=\; \lim_{r\to 0^+} \frac{g_T(J', J)}{\mathfrak{s}_\lambda\; \mathfrak{s}_\lambda'} \;=\; \lim_{r\to 0^+} \frac{g_T(J', J)}{\mathfrak{s}_\lambda} \\ &=\; \lim_{r\to 0^+} \frac{g_T(J'', J) \;+\; g_T(J', J')}{\mathfrak{s}_\lambda'} \\ &=\; g_T\Big(J'(0)\, ,\, J'(0) \Big) \; .\end{aligned}$$

Try to compute $\lim \frac{\|J\|}{\mathfrak{s}_\lambda}$ (namely, without first squaring) using L'Hôpital's rule and see what happens.

Exercise 9.1.2: Deduce the following conclusion from Theorem 9.1.1. Suppose the flag curvature of our Finsler manifold is bounded above by a positive constant λ. Then, given any unit speed geodesic $\sigma(t)$, the first conjugate point of $\sigma(0)$ cannot occur before $t = \frac{\pi}{\sqrt{\lambda}}$. Put another way:

> **Given a constant positive upper bound λ on the flag curvature, successive conjugate points along a geodesic must be at least $\frac{\pi}{\sqrt{\lambda}}$ units apart.**

Exercise 9.1.3: Suppose

- (M, F) is a Finsler manifold whose flag curvature is bounded below by a constant λ.
- $\sigma(t)$, $0 \leqslant t \leqslant c$ is a unit speed geodesic with velocity field T. Here, c is chosen such that $\sigma(c)$ is the first conjugate point of $\sigma(0)$. If no conjugate point exists, set c equal to $+\infty$. In that case, replace $0 \leqslant t \leqslant c$ by $0 \leqslant t < \infty$.
- J is a Jacobi field that is g_T-orthogonal to σ, and $J(0) = 0$, $J'(0) \neq 0$. This J is not identically zero. By the definition of conjugate points, J must be nowhere zero on the open interval $(0, c)$.

Our exercise deals with an estimate of $\|J\| := \sqrt{g_T(J, J)}$ from above. It is the other side of the coin, so to speak, of Theorem 9.1.1.

(a) Fix any $r \in (0, c)$. Explain why there exists a parallel vector field E along σ such that $E(r) = J(r)$.

(b) Set

$$l := \begin{cases} \min\left\{ c \,,\, \dfrac{\pi}{\sqrt{\lambda}} \right\} & \text{for } \lambda > 0 \,, \\ \qquad\qquad c & \text{for } \lambda \leqslant 0 \,. \end{cases}$$

Fix $r \in (0, l)$. Consider the vector field

$$W(t) := \frac{\mathfrak{s}_\lambda(t)}{\mathfrak{s}_\lambda(r)} \, E(t)$$

along $\sigma(t)$, $0 \leqslant t \leqslant r$. Check that

$$\frac{1}{2} \left(\|J\|^2 \right)'(r) = I_r(J, J) \leqslant I_r(W, W) \,.$$

(c) Explain why $\|E(t)\| = \|E(r)\| = \|J(r)\|$.

(d) Prove that under the said curvature assumption, we have

$$I_r(W, W) \leqslant \|J(r)\|^2 \left(\log \mathfrak{s}_\lambda \right)'(r) \,.$$

(e) Put parts (b) and (d) together, and manipulate the result into the form

$$\left(\log \frac{\mathfrak{s}_\lambda}{\|J\|} \right)' \geqslant 0 \quad \text{on} \quad (0, l) .$$

(f) Carry out a computation much like that in Exercise 9.1.1. Then show that

$$\boxed{\|J(t)\| \leqslant \mathfrak{s}_\lambda(t) \, \|J'(0)\| \quad \text{on} \quad (0, l)} .$$

(g) Explain how you would deduce from this inequality that $c \leqslant \frac{\pi}{\sqrt{\lambda}}$ in the case of positive λ. Note: somewhere in the arguments, one needs to acknowledge the conclusion of Exercise 5.4.4.

So:

> **Whenever the flag curvature has a constant positive lower bound, conjugate points must exist along every geodesic, and these are spaced at most $\frac{\pi}{\sqrt{\lambda}}$ units apart.**

We have therefore derived, from a different perspective, conclusion (1) of the Bonnet–Myers theorem (Theorem 7.7.1). See also [M2]. No definite conclusion can be made however, if the uniform lower bound is either zero or negative.

9.2 When Do Local Diffeomorphisms Become Covering Maps?

Fix any point p in a Finsler manifold (M, F).

- In Proposition 9.1.2, we saw that if (M, F) has nonpositive flag curvature, then no geodesic can contain any conjugate points. In that case, Proposition 7.1.1 assures us that the exponential map $\exp_p$ is a local diffeomorphism wherever it is defined.
- Now suppose that, in addition to being nonpositively curved, our Finsler manifold is also connected and forward geodesically complete (see §6.6). Proposition 6.5.1 then assures us that we can go from p to any point $q \in M$ by way of a globally minimizing geodesic $\sigma(t) := \exp_p(tv)$. In other words, $\exp_p$ maps T_pM onto M.

Thus, for any p in a forward geodesically complete, connected Finsler manifold of nonpositive flag curvature, the exponential map $\exp_p : T_pM \to M$ is a globally defined surjective local diffeomorphism. It is C^1 at the origin and C^∞ elsewhere. Our goal in this section is to show that, under the stated circumstances, $\exp_p$ is also a covering projection.

A map $\varphi : \tilde{M} \to M$ is said to be a **covering projection** if the following holds. Take any fixed $x \in M$ together with the points $\tilde{x}_\alpha$ in its inverse image $\varphi^{-1}(x)$. We must find a neighborhood $\mathcal{O}$ of x such that:

- $\varphi^{-1}\mathcal{O}$ is a *disjoint* union of open neighborhoods $\tilde{\mathcal{O}}_\alpha$, one for each point $\tilde{x}_\alpha \in \varphi^{-1}(x)$. This $\mathcal{O}$ is then said to be **evenly covered**.
- φ is a homeomorphism from each $\tilde{\mathcal{O}}_\alpha$ onto $\mathcal{O}$.

Our basic tool is the following elegant result in doCarmo [doC2]:

> Let us be given a local homeomorphism $\varphi : \tilde{M} \to M$ that lifts continuous arcs. Suppose $\tilde{M}$ is locally arcwise connected and M is locally simply connected. Then φ is a covering projection.

Manifolds are *locally* both arcwise and simply connected. Our manifolds are smooth, and the φ that we consider are at least C^1. A study of the proof in [doC2] shows that the following variant holds:

Proposition 9.2.0.

- *Let $\varphi : \tilde{M} \to M$ be a local C^1 diffeomorphism between manifolds.*
- *Suppose every piecewise C^1 curve in M, defined on a closed interval $[a, b]$, can be lifted to one in $\tilde{M}$.*

Then φ must be a covering projection.

This proposition lets us generalize a theorem in [KN1] to the Finslerian category. Its original Riemannian version is useful in §13.4, when we discuss Hopf's classification of Riemannian space forms.

Theorem 9.2.1. *Let $(\tilde{M}, \tilde{F})$ and (M, F) be two Finsler manifolds of the same dimension. Here, the Finsler structures are positively (but perhaps not absolutely) homogeneous of degree one. Suppose:*

* *Both Finsler manifolds are connected, and the domain $(\tilde{M}, \tilde{F})$ is forward geodesically complete.*

* *$\varphi : (\tilde{M}, \tilde{F}) \to (M, F)$ is a smooth local isometry. That is, φ is a smooth local diffeomorphism and $\tilde{F} = F \circ \varphi_* = \varphi^* F$.*

Then:

- *(M, F) must in fact be forward geodesically complete as well.*
- *φ must necessarily be surjective.*
- *φ is a covering projection.*

The above Finsler structures are presumed smooth and strongly convex away from the zero section in the respective tangent bundles.

Remarks:

* According to the first conclusion, the forward geodesic completeness of the domain $(\tilde{M}, \tilde{F})$ automatically implies that of the target (M, F). A certain converse [KN1] holds, but it is not as elegant.

* Suppose we do not care whether (M, F) is complete or not, and we hypothesize that φ is onto. Then the assumption on φ being a smooth local isometry can be weakened, and it would still be a covering projection. See [doC3] and Exercise 9.2.2.

Proof. The completeness of the target and the surjectivity of φ both follow from a precise understanding of how short geodesics in the two spaces are related to each other. We relegate the proof to Exercise 9.2.1.

It remains to deduce that φ is a covering projection. We accomplish this by checking that φ lifts piecewise C^1 curves. Proposition 9.2.0 then delivers the said conclusion.

Take a piecewise C^1 curve in M. A typical C^1 segment of it can be reparametrized as $c(u)$, $0 \leqslant u \leqslant 1$. At the endpoints, it has the one-sided velocities $\dot{c}(0^+)$ and $\dot{c}(1^-)$. We want to show that c, together with the one-sided velocities, admits a C^1 lift $\tilde{c}(u)$, $0 \leqslant u \leqslant 1$.

Whenever the lift $\tilde{c}$ exists, it must satisfy $\varphi \circ \tilde{c} = c$. Since φ is a local diffeomorphism and both it and c are at least C^1, so is the lift. Furthermore, the derivative φ_* is continuous and $\dot{\tilde{c}} = (\varphi_*)^{-1}\, \dot{c}$. So $\tilde{c}$ has one-sided endpoint velocities, and they correspond properly to those of c.

It suffices to show that the lift exists. Since φ is a surjective local diffeomorphism, we can at least lift a small beginning portion of c. Take the maximal subinterval $\mathcal{L}$ (containing 0) on which c can be lifted. *A priori*, there are three possibilities for $\mathcal{L}$. It is either $[0, \mu]$ or $[0, \mu)$, with $0 < \mu < 1$, or $[0, 1]$. The first is not viable because the local diffeomorphism at $\tilde{c}(\mu)$ always enables the lift to be extended past $u = \mu$, so $[0, \mu]$ cannot be maximal. The third is what we want. Let us rule out the second scenario.

Let $\{u_i\}$ be a positive increasing sequence that converges to μ. Note that we have

$$\tilde{d}(\, \tilde{c}(0)\, ,\, \tilde{c}(u_i)\,) \;\leqslant\; L\left(\tilde{c}_{|[0,u_i]}\right) \;=\; L\left(c_{|[0,u_i]}\right) \;\leqslant\; L(c) \;<\; \infty \,,$$

where the equality comes from $\tilde{F} = F \circ \varphi_*$. Thus the sequence of points $\{\tilde{c}(u_i)\}$ is forward bounded in some forward $\tilde{d}$ metric ball of $\tilde{M}$. By the Hopf–Rinow theorem (§6.6), the closure of this ball is compact because $\tilde{F}$ is assumed to be forward geodesically complete. Hence our sequence $\{\tilde{c}(u_i)\}$ has an accumulation point $\tilde{x}$.

By passing to a subsequence if necessary, we may suppose that $\tilde{c}(u_i) \to \tilde{x}$. Applying φ to this statement, we see that $c(u_i) \to x =: \varphi(\tilde{x})$. On the other hand, $c(u_i) \to c(\mu)$. So $x = c(\mu)$.

Since φ is a local diffeomorphism, it maps an open neighborhood $\tilde{U}$ of $\tilde{x}$ diffeomorphically onto an open neighborhood U of $\varphi(\tilde{x}) =: x = c(\mu)$. Also, the lifted portion on $[0, \mu)$ "heads towards" $\tilde{x}$ because $\tilde{c}(u_i) \to \tilde{x}$. It is now apparent that c can be lifted on an interval larger than $[0, \mu)$. Thus the latter cannot be maximal, thereby leaving $[0, 1]$ as the only possibility for $\mathcal{L}$, as desired. $\square$

Proposition 9.2.2. *Let (M, F) be any forward geodesically complete, connected Finsler manifold of nonpositive flag curvature. For any $p \in M$, the exponential map $\exp_p : T_pM \to M$ is a covering projection.*

Proof. Given the hypotheses, our $\exp_p : T_pM \to M$ is a surjective C^1 local diffeomorphism. If we had set $\tilde{M} := T_pM$, $\varphi := \exp_p$, $\tilde{F} := \varphi^* F = F \circ \varphi_*$, and then tried to apply Theorem 9.2.1, we would be wrong. Finsler metrics are supposed to be smooth at *all* positions and in all nonzero directions. This $\tilde{F}$ is typically only continuous at the origin (which is a position!) of $\tilde{M}$ because φ is generically only C^1 (hence φ_* is only C^0) there.

What saves the day is the following general fact. Let

$$c(u) \ := \ \exp_p[\, t(u)\, v(u)\,] \ , \quad \text{with} \ \ F_p[\, v(u)\,] = 1 \ ,$$

be a curve of the "liftable" form, defined on $0 \leqslant u \leqslant u_o$. The scalar t and the unit vector v are both C^1 in the parameter u. The endpoints of c are

$$c(0) \ = \ \exp_p[\, t(0)\, v(0)\,] \quad \text{and} \quad c(u_o) \ = \ \exp_p[\, t(u_o)\, v(u_o)\,] \ .$$

We claim that the length of c satisfies the estimate

$$\boxed{L(c) \ \geqslant \ t(u_o) - t(0) \ = \ F_p[\, t(u_o)\, v(u_o)\,] \ - \ F_p[\, t(0)\, v(0)\,]} \ .$$

The proof uses the Gauss Lemma (§6.1) and the fundamental inequality (1.2.3). It begins with the paragraph after (6.3.4) and is word-for-word the same, up to and including the quantity $t(1) - t(0)$ in the statement (*) in our proof of Theorem 6.3.1. After that we simply replace $t(1)$ by $t(u_o)$.

In view of Proposition 9.2.0, our local C^1 diffeomorphism $\exp_p$ is a covering projection if it lifts piecewise C^1 curves. We explained in the proof of Theorem 9.2.1 why one only needs to show that each C^1 segment can be lifted. The issue of one-sided endpoint velocities always resolves itself.

A typical C^1 segment can be reparametrized as $c(u)$, $0 \leqslant u \leqslant 1$. The surjectivity of our local C^1 diffeomorphism $\exp_p$ implies that a small beginning portion of c admits a lifting. Take the maximal subinterval $\mathcal{L}$ (containing 0) on which c can be lifted. Namely, there exists a C^1 curve in T_pM, given by the "position vectors"

$$\tilde{c}(u) \ := \ t(u)\, v(u) \ , \quad u \in \mathcal{L} \ , \quad \text{with} \ \ F_p[\, v(u)\,] = 1 \ ,$$

such that $\exp_p[\tilde{c}(u)] = c(u)$. As in the proof of Theorem 9.2.1, $\mathcal{L}$ is *a priori* either $[0, \mu)$, with $0 < \mu < 1$, or $[0, 1]$. We now rule out the former.

Let $\{u_i\}$ be a positive increasing sequence that converges to μ. The general estimate stated above helps us conclude that

$$F_p[\, \tilde{c}(u_i)\,] \ - \ F_p[\, \tilde{c}(0)\,] \ \leqslant \ L\left(c_{|[0,u_i]}\right) \ \leqslant \ L(c) \ < \ \infty \ .$$

Therefore the sequence of points $\{\tilde{c}(u_i)\}$ in T_pM must be bounded in the Minkowski norm F_p, or else we would have a contradiction.

By choosing a basis for the vector space T_pM, it inherits the usual Euclidean norm $|\ |$ of $\mathbb{R}^n$. The second conclusion of Lemma 6.2.1 says that F_p

and $| \ |$ are equivalent. The completeness of Euclidean space then implies that our bounded sequence $\{\tilde{c}(u_i)\}$ has an accumulation point $\tilde{x} \in T_pM$. This occurs in the topology defined (by $| \ |$, hence that defined) by F_p, which is what we use to make sense of "local" on T_pM. The rest of the argument is identical to the last two paragraphs in our proof of Theorem 9.2.1. □

Exercises

Exercise 9.2.1: Let $\varphi : (\tilde{M}, \tilde{F}) \to (M, F)$ be a local isometry between Finsler manifolds. Short geodesics in $\tilde{M}$ and M then correspond to each other. This fact can be stated as a **commutation relation**:

$$\varphi \circ \widetilde{\exp} = \exp \circ \varphi_* .$$

Now suppose $(\tilde{M}, \tilde{F})$ is forward geodesically complete. Use that commutation relation to help in the following tasks. (See [KN1] only if necessary.)

(a) Show that the target (M, F) must *necessarily* be forward geodesically complete.

(b) Suppose, in addition to $(\tilde{M}, \tilde{F})$ being complete, we also assume that M is connected. Prove that φ must *necessarily* be surjective.

Exercise 9.2.2: Use Proposition 9.2.0 to help prove a result in [doC3].

> Let φ be a local C^1 diffeomorphism from the Finsler manifold $(\tilde{M}, \tilde{F})$ **onto** the Finsler manifold (M, F). If $(\tilde{M}, \tilde{F})$ is forward geodesically complete, and $F \circ \varphi_* \geqslant \tilde{F}$, then φ is a covering map.

Exercise 9.2.3: If (M, F) is a Berwald space, can Proposition 9.2.2 be deduced from Theorem 9.2.1? Hint: see §5.3.

Exercise 9.2.4: Unlike [KN1], our proof of Theorem 9.2.1 does not begin with a geometrically chosen $\mathcal{O}$. Study the Riemannian case in [KN1], then reprove Theorem 9.2.1 for absolutely homogeneous Finsler manifolds. After that, modify (using Lemma 6.2.1) your proof to handle the positively homogeneous case. In both cases, make clear where completeness gets used.

9.3 Some Consequences of the Covering Homotopy Theorem

Fix any point p in a forward geodesically complete Finsler manifold (M, F). Recall from Theorem 8.7.1 that for each homotopy class of paths on M from our p to another fixed q, there is a shortest smooth geodesic within the same class.

Now suppose that in addition to being forward geodesically complete, the Finsler manifold is also connected and has nonpositive flag curvature. Our goal in this section is to show that under these circumstances, there is only one shortest smooth geodesic within each homotopy class of paths from p to q.

We saw in §9.1 and §9.2 that given the hypotheses on (M, F), especially the fact that its flag curvature is nonpositive, the exponential map $\exp_p : T_pM \to M$ is a covering projection. A rather useful fact about any covering projection $\varphi : \tilde{M} \to M$ is the **Covering Homotopy theorem** described below. It is precisely the tool we need to accomplish the stated goal.

Theorem 9.3.1.

* *Let $\varphi : \tilde{M} \to M$ be a covering projection.*
* *Let $\sigma(t)$, $t \in [a, b]$ be a curve in M. Let $\tilde{\sigma}(t)$, $t \in [a, b]$ be a lift of σ. That is, $\varphi \circ \tilde{\sigma} = \sigma$.*
* *Suppose $h : [a, b] \times [0, 1] \to M$ is a homotopy with $h(t, 0) = \sigma(t)$ for all $t \in [a, b]$.*

Then:

- *There exists a unique lift $\tilde{h} : [a, b] \times [0, 1] \to \tilde{M}$ of h such that $\tilde{h}(t, 0) = \tilde{\sigma}(t)$. Here, $\varphi \circ \tilde{h} = h$.*
- *Furthermore, if for some $t_o \in [a, b]$, $h(t_o, u)$ happens to be constant for all u, then $\tilde{h}(t_o, u)$ will also be constant in u.*

Remarks: There are two immediate corollaries of this theorem.

* Let γ be any curve in M that emanates from p. By performing a reparametrization and a relabeling if necessary, we may write γ as $\gamma(u)$, with $u \in [0, 1]$. View γ as a homotopy $h(t, u) := \gamma(u)$ in which all the intermediate t-curves are constant curves. Fix any $\tilde{p} \in \tilde{M}$ such that $\varphi(\tilde{p}) = p$. Then the first conclusion of Theorem 9.3.1 ensures that γ admits a unique lift $\tilde{\gamma}$ into $\tilde{M}$ with $\tilde{\gamma}(0) = \tilde{p}$.
* Let $h : [a, b] \times [0, 1] \to M$ be a homotopy among paths with fixed endpoints p and q. That is, $h(a, u) = p$ and $h(b, u) = q$ for all $u \in [0, 1]$. Then the second conclusion of Theorem 9.3.1 tells us that its lift $\tilde{h}$ will also be a homotopy among paths with fixed endpoints.

For a more general statement of the Covering Homotopy theorem and a detailed proof, see [ST].

We are ready to establish the following interesting fact.

Lemma 9.3.2. *Let (M, F) be a Finsler manifold.*

- *Suppose that at some $p \in M$, the exponential map $\exp_p : T_pM \to M$ is a covering projection.*
- *Let $\sigma_0(t) := \exp_p(tT_0)$ and $\sigma_1(t) := \exp_p(tT_1)$, $0 \leqslant t \leqslant L$ be any two (smooth) geodesics emanating from p and terminating at some common $q \in M$.*

The following conclusions hold:

- * *If σ_0 is homotopic to σ_1 through a homotopy with fixed endpoints p and q, then $T_0 = T_1$ (equivalently, $\sigma_0 = \sigma_1$).*
- * *In particular, if σ_0 and σ_1 are not reparametrizations of each other, then they cannot be deformed to each other through a homotopy with fixed endpoints p and q.*

Proof. The contrapositive of the first conclusion encompasses the second conclusion. So it suffices to establish the first one.

Suppose σ_0 is homotopic to σ_1, through a homotopy $h(t, u)$, $0 \leqslant t \leqslant L$, $0 \leqslant u \leqslant 1$ with fixed endpoints p and q. Using Theorem 9.3.1, we lift this h to a homotopy $\tilde{h} : [0, L] \times [0, 1] \to T_pM$ with $\tilde{h}(t, 0) = tT_0$.

By hypothesis, every t-curve of the homotopy h begins at p and ends at q. Theorem 9.3.1 assures us that correspondingly, every t-curve of the lifted homotopy $\tilde{h}$ begins at the origin of T_pM and ends at the tip of $LT_0 \in T_pM$.

Note that both $\tilde{h}(t, 1)$ and tT_1 are lifts of σ_1 which emanate from the origin of T_pM. So, by a corollary of Theorem 9.3.1, they must be the same. Consequently, $\tilde{h}$ is a homotopy between the rays tT_0 and tT_1, and all the intermediate t-curves share the same endpoints. However, the only way for the two rays tT_0 and tT_1, $0 \leqslant t \leqslant L$, to have the same endpoints would be $T_0 = T_1$. This is equivalent to saying that σ_0 is actually identical to σ_1. □

Let us now fulfill the goal stated at the beginning of this section.

Theorem 9.3.3. *Let (M, F) be any forward geodesically complete, connected Finsler manifold of nonpositive flag curvature.*

- *Fix p, $q \in M$. Then, within each homotopy class of paths from p to q, there exists a unique shortest smooth geodesic within that class.*
- *In particular, fix $p \in M$. Then, within every homotopy class of loops based at p, there exists a unique shortest smooth closed geodesic within that class.*

Proof. The existence has been ascertained in Theorem 8.7.1; it only requires forward geodesic completeness and connectedness.

We establish uniqueness here. Since (M, F) has, by hypothesis, nonpositive flag curvature, the exponential map $\exp_p : T_pM \to M$ is a covering projection. This was what we found in §9.2.

Let α be any homotopy class of paths from p to q. Suppose $\sigma_0(t) := \exp_p(tT_0)$ and $\sigma_1(t) := \exp_p(tT_1)$, $0 \leqslant t \leqslant L$ are any two shortest geodesics in the class α. Being members of the same homotopy class, they are homotopic to each other through intermediate paths with endpoints p and q. By Lemma 9.3.2, we must have $\sigma_0 = \sigma_1$. □

Exercises

Exercise 9.3.1:

(a) Give a proof of the Covering Homotopy theorem (Theorem 9.3.1). This is a standard result covered in every algebraic topology text. It is also treated in some geometry texts; see, for example, [ST]. However, consult an external reference only if absolutely necessary.

(b) What is the intuitive message behind that theorem?

Exercise 9.3.2:

(a) In case the flag curvature of our compact connected Finsler manifold is nonpositive, do you suppose that every *free* homotopy class of loops in M contains a *unique* shortest geodesic loop?

(b) Also, are there any compact, simply connected Finsler manifolds with nonpositive flag curvature?

9.4 The Cartan–Hadamard Theorem

Let us give another application of the Covering Homotopy theorem discussed in §9.3.

Theorem 9.4.1 (Cartan–Hadamard). *Let (M, F) be any forward geodesically complete, connected Finsler manifold of nonpositive flag curvature. Then:*

(1) *Geodesics in (M, F) do not contain conjugate points.*

(2) *For any fixed $p \in M$, the exponential map $\exp_p : T_pM \to M$ is a globally defined C^1 local diffeomorphism from T_pM onto M. Furthermore, this surjection is in fact a covering projection.*

(3) *In case M happens to be simply connected, that exponential map $\exp_p$ is actually a C^1 diffeomorphism from the tangent space T_pM onto the manifold M.*

Remark: Recall from §5.3 that, in the general Finsler setting, the exponential map is only C^1 at the origin of T_pM, although it is smooth away from the origin. Exercise 5.3.5 tells us that the exponential map is smooth on the entire T_pM if and only if the Finsler structure is of Berwald type.

Proof. The first two conclusions have already been established in Propositions 9.1.2 and 9.2.2.

It remains to check that the covering projection $\exp_p$ is injective whenever the manifold M is simply connected. We give two separate and independent arguments. The first one depends on the Covering Homotopy theorem discussed in §9.3. The second argument completely avoids the material in that section, but uses the concept of deck transformations instead.

* Suppose $\exp_p(v_0) = q = \exp_p(v_1)$. Then $\sigma_0(t) := \exp_p(tv_0)$ and $\sigma_1(t) := \exp_p(tv_1)$, $0 \leqslant t \leqslant 1$ are two geodesics in M from p to q. Since M is simply connected, it can be shown that σ_0 is homotopic to σ_1 through a homotopy with fixed endpoints. See Exercise 9.4.1. By Lemma 9.3.2, which follows from the Covering Homotopy theorem, we must have $v_0 = v_1$. Thus $\exp_p$ is injective.
* Alternatively, we can apply a standard result about covering projections to $\exp_p$. Note that its domain T_pM is simply connected and, its range M, being a manifold, is always *locally* simply connected. Therefore the group of deck transformations at any $x \in M$ is isomorphic to the fundamental group $\pi(M, x)$. However, M is by hypothesis simply connected. Thus $\pi(M, x)$ contains only one element and consequently there is only one deck. That means every x has exactly one preimage under $\exp_p$. This again proves the injectivity of $\exp_p$. $\square$

Compare the treatment here with those in [Au] and [Daz].

Exercise

Exercise 9.4.1: Let $\supset$ and $\subset$ be any two curves from p to q, in a simply connected manifold. Prove that they are homotopic, through a homotopy with fixed endpoints p and q. Hints: visualize the following intuition.

* Go from p to q along $\supset$, then back to p along the reverse of $\subset$. This defines a loop based at p. Call it $\bigcirc$.
* Since M is simply connected, $\bigcirc$ can be shrunken down to the point p, using a 1-parameter family (indexed by $u \in [0, 1]$) of loops $\bigcirc_u$ in M based at p. Here, $\bigcirc_0 = \bigcirc$ and $\bigcirc_1$ is the constant curve at p.
* During this shrinking, the point q gives rise to a curve $\xi = \xi(u)$, where $\xi(0) = q$ and $\xi(1) = p$.
* Each intermediate loop $\bigcirc_u$ can be described in a convenient way as follows. Go from p to $\xi(u)$ along some $\supset_u$, then back to p along the reverse of some $\subset_u$.
* For each u, we travel along $\supset_u$ from p to $\xi(u)$, then to q along the reverse of a portion of ξ. This defines a curve from p to q. The resulting u-indexed family of curves from p to q represents a homotopy, with fixed endpoints, between $\supset$ and the reverse of ξ. Likewise, the use of $\subset_u$ gives a homotopy, again with fixed endpoints, between $\subset$ and the reverse of ξ.

* Combine the two homotopies described above.

Exercise 9.4.2: In the proof of the Cartan–Hadamard theorem, we gave two arguments for the injectivity of $\exp_p$ when M is simply connected. The second one invokes the fact that the group of deck transformations is isomorphic to $\pi(M, x)$. Prove this fact without consulting [ST].

9.5 Prelude to Rauch's Theorem

We now prepare to extend the comparison arguments in §9.1 and Exercise 9.1.3 to a broader setting. The preparation involves two technical ingredients that are interesting and important in their own right.

9.5 A. Transplanting Vector Fields

Begin with an n-dimensional Finsler manifold (M, F). Denote the induced inner product on π^*TM by g. Let $\sigma(t)$, $a \leqslant t \leqslant b$ be a unit speed geodesic in M, with velocity field $T(t)$.

We compare (M, F) with a Finsler space $(M_o\,,\ F_o)$ of dimension $\geqslant n$, say $= n + k$. Denote the induced inner product on $\pi_o^*TM_o$ by g_o . Let $\sigma_o(t)$, $a \leqslant t \leqslant b$ be a unit speed geodesic in M_o, with velocity field $T_o(t)$.

Let $W(t)$ be a piecewise C^∞ vector field along $\sigma(t)$. In order to compare W with something in M_o, we construct its **transplant** $\tilde{W}(t)$, which is a certain piecewise C^∞ vector field along the geodesic $\sigma_o(t)$. To avoid clutter, introduce the abbreviations

$$W' := D_T W \ , \qquad \tilde{W}' := D_{T_o} \tilde{W} \ ;$$

$$g_T := g_{(\sigma,T)} \ , \qquad g_{T_o} := g_{o\,(\sigma_o,T_o)} \ .$$

We would like our $\tilde{W}$ to have the following properties:

(1) $\tilde{W}'(t)$ and $W'(t)$ have discontinuities (if any) at the same t values.

(2) $g_{T_o}(\tilde{W}, \tilde{W}) = g_T(W, W)$.

(3) $g_{T_o}(\tilde{W}, T_o) = g_T(W, T)$.

(4) $g_{T_o}(\tilde{W}', \tilde{W}') = g_T(W', W')$.

(5) If $W(b)$ is nonzero and g_T-orthogonal to $T(b)$, then $\tilde{W}(b)$ can be arranged to be equal to ζ times $\sqrt{g_T(\,W(b)\,,\ W(b)\,)}$, where ζ is any unit vector that is g_{T_o}-orthogonal to $T_o(b)$.

Remark: Admittedly, that bit of freedom in (5) is unmotivated; but it becomes useful in §9.6, during the proof of the Rauch comparison theorem.

To construct the $\tilde{W}$ described, we proceed as in Spivak [Sp4].

Choose a g_T-orthonormal basis $\{E_i : i = 1, \ldots, n\}$ for $T_{\sigma(a)}M$, with $E_n := T(a)$. Parallel transport (with reference vector field T) the basis $\{E_i\}$ along σ to obtain $\{E_i(t)\}$. In view of Exercise 5.2.3, $\{E_i(t)\}$ is a g_T-orthonormal frame field along σ. Being a geodesic with constant speed, σ is an autoparallel, therefore

$$E_n(t) \;=\; T(t) \ .$$

Likewise, we carry out parallel transport (with reference vector field T_o) along σ_o to find $\{F_I(t) : I = 1, \ldots, n, n+1, \ldots, n+k\}$, a g_{T_o}-orthonormal frame field with

$$F_n(t) \;=\; T_o(t) \ .$$

Whenever $W(b)$ is nonzero and g_T-orthogonal to $T(b)$, let us choose the initial data for $E_1(t)$ appropriately in order to effect

$$E_1(b) \;=\; \frac{W(b)}{\sqrt{g_T(\,W(b)\,,\,W(b)\,)}} \ .$$

This is always possible because parallel transport (with reference vector T) along σ generates an isomorphism from $T_{\sigma(a)}M$ onto $T_{\sigma(b)}M$. By the same token, we can arrange to have

$$F_1(b) \;=\; \zeta \ ,$$

for any (freely specifiable) unit vector ζ [by that we mean $g_{T_o}(\zeta, \zeta) = 1$] which is g_{T_o}-orthogonal to $T_o(b)$.

In any case, expand $W(t)$ as

$$W(t) \;=\; \varphi^i(t)\; E_i(t)$$

and then use the coefficients φ^i to define

$$\tilde{W}(t) \;:=\; \varphi^i(t)\; F_i(t) \;+ 0\, F_{n+1}(t) + \cdots + 0\, F_{n+k}(t) \ . \tag{9.5.1}$$

It can be checked that this "transplant" $\tilde{W}$ has the five desired properties mentioned earlier.

9.5 B. A Second Basic Property of the Index Form

For (M, F), the index form was defined in (7.2.1):

$$I(V, W) \;:=\; \int_a^b \frac{1}{F(T)}\, [\, g_T(D_T V\,,\, D_T W) \;-\; g_T(\, R(V,T)T\,,\, W\,)\,]\; dt \ .$$

Here, our geodesics are unit speed, so $F(T) \;=\; 1$. Furthermore, we are concerned with the associated quadratic form

$$I(W, W) \;:=\; \int_a^b [\, g_T(D_T W\,,\, D_T W) \;-\; g_T(\, R(W,T)T\,,\, W\,)\,]\; dt \ . \tag{9.5.2}$$

All D_T are calculated with reference vector T, and R is evaluated along the canonical lift (σ, T).

The flag curvature $K(T,W)$ was described in §3.9. It is related to the quantity $g_T(\,R(W,T)T\,,\,W\,)$ as follows:

$$K(T,W) \;=\; \frac{g_T(\,R(W,T)T\,,\,W\,)}{g_T(W,W) \;-\; [g_T(W,T)]^2}\;. \tag{9.5.3}$$

Substituting (9.5.3) into (9.5.2) gives

$$\begin{aligned} I(W,W) \;=\; \int_a^b \Big[\; & g_T(D_T W\,,\,D_T W) \\ & - \Big(g_T(W,W) \;-\; [g_T(W,T)]^2\Big)\; K(T,W)\;\Big]\,dt\;. \end{aligned} \tag{9.5.4}$$

A similar discussion holds for the unit speed geodesic σ_o (with velocity field T_o) in the comparison Finsler space $(M_o\,,\,F_o)$. The quadratic form associated with the index form is

$$\begin{aligned} I(W_o,W_o) \;=\; \int_a^b \Big[\; & g_{T_o}(D_{T_o}W_o\,,\,D_{T_o}W_o) \\ & - \Big(g_{T_o}(W_o,W_o) \;-\; [g_{T_o}(W_o,T_o)]^2\Big)\; K(T_o,W_o)\;\Big]\,dt\,, \end{aligned} \tag{9.5.5}$$

where D_{T_o} is calculated with reference vector T_o.

One important property of the index form is given in Lemma 7.3.2, also known as the **basic index lemma**. It says that:

> **In the absence of conjugate points, the index is typically decreased if we replace our vector field by the Jacobi field that shares its endpoint values.**

The following lemma (Lemma 9.5.1) brings out a **second basic property**:

> **The index is typically decreased whenever we transplant our vector field into a space of higher flag curvature.**

This is quite evident if we compare the integrand of (9.5.4) with its counterpart in (9.5.5). The following lemma makes precise that intuitive statement.

Lemma 9.5.1. *Let (M,F) be an n-dimensional Finsler manifold. Let $\sigma(t)$, $a \leqslant t \leqslant b$ be a unit speed geodesic in M, with velocity field $T(t)$. Similarly, let $(M_o\,,\,F_o)$ be a comparison Finsler space of dimension $n+k$, and let $\sigma_o(t)$, $a \leqslant t \leqslant b$ be a unit speed geodesic in M_o, with velocity field $T_o(t)$. Take any piecewise C^∞ vector field W along σ. Construct its "transplant" $\tilde{W}$ onto σ_o, in such a way that:*

(1) $\tilde{W}'(t)$ and $W'(t)$ have discontinuities (if any) at the same t values.

(2) $g_{T_o}(\tilde{W},\tilde{W}) = g_T(W,W)$.

(3) $g_{T_o}(\tilde{W},T_o) = g_T(W,T)$.

(4) $g_{T_o}(\tilde{W}', \tilde{W}') = g_T(W', W')$, *with* $W' := D_T W$, $\tilde{W}' := D_{T_o}\tilde{W}$.

If $K(T,W) \leqslant K(T_o, W_o)$ *for any* $W \in T_{\sigma(t)}M$ *and* $W_o \in T_{\sigma_o(t)}M_o$,

$$\text{then} \quad \boxed{I(W,W) \geqslant I(\tilde{W}, \tilde{W})} .$$

9.5 C. Flag Curvature Versus Conjugate Points

Let us give some applications of this new property.

Theorem 9.5.2. *Let* $\sigma(t)$, $a \leqslant t \leqslant b$ *be a unit speed geodesic, with velocity field* T, *in a Finsler manifold of dimension* n. *Let* K *abbreviate the collection of flag curvatures* $\{ K(T,W) : W \in T_{\sigma(t)}M , \ a \leqslant t \leqslant b \}$.

(1) *If* $K \leqslant 0$, *then* $\sigma(t)$, $a \leqslant t < b$ *contains no conjugate point of* $\sigma(a)$.

(2) *If* $K \leqslant \frac{1}{r^2}$ *and* $L(\sigma) < \pi r$, *then* $\sigma(t)$, $a \leqslant t < b$ *contains no conjugate point of* $\sigma(a)$.

(3) *If* $K \geqslant \frac{1}{r^2}$ *and* $L(\sigma) > \pi r$, *then* $\sigma(t)$, $a < t \leqslant b$ *contains at least one point that is conjugate to* $\sigma(a)$.

Remarks: According to [Sp4],

* In the Riemannian case, items (2) and (3) are collectively known as the **Morse–Schoenberg theorem**.
* However, when $\dim M = 2$, they are considered parts of **Bonnet's theorem**.

Proof. In the following proof, when we speak of piecewise C^∞ vector fields along a geodesic, we mean those that vanish at the endpoints of that geodesic. This omission effects a less cumbersome prose. Note that W vanishes at the endpoints of σ if and only if its transplant $\tilde{W}$ vanishes at the endpoints of σ_o. This is due to the second property of $\tilde{W}$.

For (1), let M be the given Finsler manifold and let M_o be Euclidean $\mathbb{R}^n$. The flatness of M_o implies that its index quadratic form is nonnegative. This, together with Lemma 9.5.1, tell us that $I(W,W) \geqslant 0$ for all piecewise C^∞ vector fields along σ. Thus, in view of Proposition 7.4.1, σ cannot contain any conjugate point in its "interior."

Next consider (2). As above, let M be the Finsler manifold in question. We choose M_o to be the Euclidean n-sphere $\mathbb{S}^n(r)$ of radius r in $\mathbb{R}^{n+1}$, and specify $\sigma_o(t)$, $a \leqslant t \leqslant b$ to be a unit speed arc along (but strictly less than half of) some great circle. Since σ_o has length $b - a < \pi r$, it contains no conjugate points. By Proposition 7.3.1, we must then have $I(\tilde{W}, \tilde{W}) \geqslant 0$.

Our hypothesis says that K is bounded above by $\frac{1}{r^2}$, which is the constant flag (sectional) curvature of M_o. Substituting this and the conclusion of the above paragraph into Lemma 9.5.1, we see that $I(W, W) \geqslant 0$ for all piecewise C^∞ vector fields along σ. Therefore (2) follows from Proposition 7.4.1.

As for (3), we let the M in Lemma 9.5.1 be the standard $\mathbb{S}^n(r)$. Our given Finsler manifold now plays the role of M_o (!) in that lemma. The hypothesis $K \geqslant \frac{1}{r^2}$ says that the flag curvatures of M_o dominate those of M. Hence Lemma 9.5.1 assures us that the index quadratic form (with argument $\tilde{W}$) of M_o is bounded above by that (with argument W) of M.

On our given Finsler manifold M_o, we rename the geodesic (denoted by σ in the statement of the theorem) as σ_o. On $M := \mathbb{S}^n(r)$, let us take a unit speed arc $\sigma(t)$, $a \leqslant t \leqslant b$ along (and covering definitely more than half of) some great circle. Since σ has length $b - a > \pi\, r$, it must contain an "interior" point conjugate to $\sigma(a)$. As a result, by Proposition 7.4.1, there exists a piecewise C^∞ vector field W along σ satisfying $I(W, W) < 0$.

Putting these last two paragraphs together, we get $I(\tilde{W}, \tilde{W}) < 0$ for some piecewise C^∞ vector field $\tilde{W}$ [albeit one transplanted from $\mathbb{S}^n(r)$]. By Proposition 7.3.1, the geodesic on our given Finsler manifold must contain some point that is conjugate to $\sigma(a)$. □

Exercises

Exercise 9.5.1:

(a) Verify that the constructed $\tilde{W}$ in (9.5.1) does indeed satisfy the five intended properties.

(b) Deduce Lemma 9.5.1.

Exercise 9.5.2:

(a) Compare the third conclusion of Theorem 9.5.2 with the Bonnet–Myers theorem (Theorem 7.7.1).

(b) What is the common ingredient in the proofs of these two results?

9.6 Rauch's First Comparison Theorem

We are now ready to extend *and* unify the comparison arguments used in §9.1 and Exercise 9.1.3. Our goal here is the very basic **first Rauch theorem.** It relates the growth of Jacobi fields to flag curvature bounds. The geometrical setup is exactly as in Lemma 9.5.1. But instead of stopping at the conclusion $I(W, W) \geqslant I(\tilde{W}, \tilde{W})$, we probe deeper with a bit of calculus. Loosely speaking, we find that:

> **Raising (more positive or less negative) the flag curvature slows down the growth of Jacobi fields. In particular, Jacobi fields grow fastest on negatively curved spaces, less so on flat spaces, and considerably slower on positively curved spaces.**

This is consistent with what we learned in §5.5. The method we use borrows from that in Spivak [Sp4] and Cheeger–Ebin [CE].

Theorem 9.6.1 (Rauch).

- **Geometrical setup.** *Let (M, F) be an n-dimensional Finsler manifold. Let $\sigma(t)$, $0 \leqslant t \leqslant L$ be a unit speed geodesic in M, with velocity field $T(t)$. Similarly, let $(M_o\,,\, F_o)$ be a comparison Finsler space of dimension $n + k$, and let $\sigma_o(t)$, $0 \leqslant t \leqslant L$ be a unit speed geodesic in M_o, with velocity field $T_o(t)$.*
- **Bound on curvature.** *Suppose the flag curvatures of M and M_o satisfy*

$$K(T, W) \;\leqslant\; K(T_o, W_o) \tag{9.6.1}$$

for any $W \in T_{\sigma(t)}M$ and $W_o \in T_{\sigma_o(t)}M_o$.

- **Data on Jacobi fields.** *Let $J := (\alpha t + \beta)\, T + J^\perp$ and $J_o := (\alpha t + \beta)\, T_o + J_o^\perp$ be two Jacobi fields, respectively, along σ and σ_o. Note that it is the same α, β for J and J_o. These Jacobi fields are to satisfy*

$$J^\perp(0) \;=\; 0\;, \quad J_o^\perp(0) \;=\; 0 \tag{9.6.2}$$

and

$$g_T\Big(\, J'(0)\,,\, J'(0)\Big) \;=\; g_{T_o}\Big(\, J_o'(0)\,,\, J_o'(0)\Big)\;. \tag{9.6.3a}$$

Equivalently,

$$g_T\Big(\,(J^\perp)'(0)\,,\,(J^\perp)'(0)\Big) \;=\; g_{T_o}\Big(\,(J_o^\perp)'(0)\,,\,(J_o^\perp)'(0)\Big)\;. \tag{9.6.3b}$$

Statement proper. *Assume the above setup, bound, and data.*

If $\sigma_o(t)$, $0 \leqslant t \leqslant L$ contains no conjugate point of $\sigma_o(0)$,

then:

(1) $\dfrac{g_T\big(\,(J^\perp)'(t)\,,\,J^\perp(t)\big)}{g_T\big(\,J^\perp(t)\,,\,J^\perp(t)\big)} \geqslant \dfrac{g_{T_o}\big(\,(J_o^\perp)'(t)\,,\,J_o^\perp(t)\big)}{g_{T_o}\big(\,J_o^\perp(t)\,,\,J_o^\perp(t)\big)}$ *for all* $t \in (0, L]$.

(2) *Hence* $g_T\big(J(t), J(t)\big) \geqslant g_{T_o}\big(J_o(t), J_o(t)\big)$ *for* $0 \leqslant t \leqslant L$. *That is, the Jacobi field J grows no slower than J_o.*

(3) *In particular,* $\sigma(t)$, $0 \leqslant t \leqslant L$ *will not contain any conjugate points either.*

Notations: We have continued to use $(\ )'$ to abbreviate D_T or D_{T_o}, depending on whether one is dealing with vector fields on M or M_o. Other abbreviations include $g_T := g_{(\sigma,T)}$ and $g_{T_o} := g_{o\ (\sigma_o,T_o)}$, as well as the following that are used:

$$\| W \| := \sqrt{g_T(W,W)} \quad \text{for } W \in T_{\sigma(t)}M ,$$

$$\| W_o \|_o := \sqrt{g_{T_o}(W_o,W_o)} \quad \text{for } W_o \in T_{\sigma_o(t)}M_o .$$

Remarks:

- It is somewhat lamentable that the norms $\| J(t) \|$ and $\| J_o(t) \|_o$ involved in **(2)** are typically *not* equal to the squares of the Finsler norms $F(J)$ and $F_o(J_o)$.
- Quite often one is given a Jacobi field $J := (\alpha t + \beta)T + J^\perp$ on the Finsler space we are trying to study. In order to apply Rauch's theorem, we must build a Jacobi field J_o on the comparison space such that hypotheses (9.6.2), (9.6.3) are both satisfied. This can *always* be done. In fact, one first constructs $J_o^\perp$ and then adds on $(\alpha t + \beta)T_o$. See Exercise 9.6.2.

Proof of Rauch's theorem.

(A) Some technical concerns:

Since

(9.6.4a) $$\| J \|^2 = (\alpha t + \beta)^2 + \| J^\perp \|^2 ,$$

(9.6.4b) $$\| J_o \|_o^2 = (\alpha t + \beta)^2 + \| J_o^\perp \|_o^2 ,$$

the inequality

(9.6.5) $$\| J^\perp(t) \|^2 \geqslant \| J_o^\perp(t) \|_o^2 \quad \text{on } [0,L]$$

is equivalent to **(2)**. Thus we work with $J^\perp$, $J_o^\perp$ instead of J, J_o .

If (9.6.3b) reads $0 = 0$, then in view of (9.6.2) both Jacobi fields $J^\perp$, $J_o^\perp$ are identically zero. In that case (9.6.5) is trivially true. So let us assume that neither $(J^\perp)'(0)$ nor $(J_o^\perp)'(0)$ vanishes; then $J^\perp$ and $J_o^\perp$ are nonzero Jacobi fields.

Since σ_o contains no conjugate points by assumption, and since $J_o^\perp$ already vanishes at $\sigma_o(0)$, we know that $J_o^\perp(t)$ must be nowhere zero on $(0,L]$. The same cannot *yet* be said about $J^\perp(t)$ on $(0,L]$. *A priori*, this nowhere vanishing criterion is needed in order to make sense of **(1)**. Nevertheless, since one must travel along σ at least for a short while before encountering the first conjugate point, and since $J^\perp$ already vanished once [namely, at the initial point $\sigma(0)$], we are sure that there exists a positive $l \leqslant L$ such that $J^\perp(t)$ is nowhere zero on $(0, l]$.

Our next two steps are carried out on $[0, l]$. After that, we use a continuity argument to extend our conclusions from $[0, l]$ to $[0, L]$.

(B) Proving that $\frac{g_T\left((J^\perp)'(t)\,,\,J^\perp(t)\right)}{\|\,J^\perp(t)\,\|^2} \geqslant \frac{g_{T_o}\left((J_o^\perp)'(t)\,,\,J_o^\perp(t)\right)}{\|\,J_o^\perp(t)\,\|_o^2}$ **on** $(0, l]$ **:**

For each fixed ϵ in $(0, l]$, let I_ϵ denote the index form in which the integration is carried out from 0 to ϵ. That is,

$$I_\epsilon(W, W) \;:=\; \int_0^\epsilon \Big[\; g_T(D_T W\,,\, D_T W) - \Big(g_T(W, W) \;-\; [g_T(W, T)]^2\Big)\; K(T, W) \;\Big]\; dt$$

and

$$I_\epsilon(W_o, W_o) \;:=\; \int_0^\epsilon \Big[\; g_{T_o}(D_{T_o} W_o\,,\, D_{T_o} W_o) - \Big(g_{T_o}(W_o, W_o) \;-\; [g_{T_o}(W_o, T_o)]^2\Big)\; K(T_o, W_o) \;\Big]\; dt\;.$$

Let $\widetilde{J^\perp}$ denote the transplant of $J^\perp$, with the five properties listed in §9.5.A. For the case at hand, these read:

- $\widetilde{J^\perp}(t)$ is as smooth as $J^\perp(t)$.
- $g_{T_o}\left(\widetilde{J^\perp}\,,\,\widetilde{J^\perp}\right) \;=\; g_T(J^\perp, J^\perp)$.
- $g_{T_o}\left(\widetilde{J^\perp}, T_o\right) \;=\; g_T(J^\perp, T)$.
- $g_{T_o}\left((\widetilde{J^\perp})'\,,\,(\widetilde{J^\perp})'\right) \;=\; g_T\left((J^\perp)'\,,\,(J^\perp)'\right)$.
- $\widetilde{J^\perp}(\epsilon) \;=\; \|\,J^\perp(\epsilon)\,\|\;\frac{J_o^\perp(\epsilon)}{\|\,J_o^\perp(\epsilon)\,\|_o}$. Namely, set $\zeta := \frac{J_o^\perp(\epsilon)}{\|\,J_o^\perp(\epsilon)\,\|_o}$.

For the last property, recall that one has the freedom to specify $\widetilde{J^\perp}(\epsilon)$ to be $\|\,J^\perp(\epsilon)\,\|$ times any unit vector $\zeta \in T_{\sigma_o(\epsilon)}M_o$, as long as the latter is g_{T_o}-orthogonal to σ_o. Here, we have chosen that ζ as indicated.

The following calculation holds:

$$\begin{aligned}
&\frac{1}{\|\,J^\perp(\epsilon)\,\|^2}\; I_\epsilon(J^\perp\,,\,J^\perp) \\
\geqslant\;& \frac{1}{\|\,J^\perp(\epsilon)\,\|^2}\; I_\epsilon\left(\widetilde{J^\perp}\,,\,\widetilde{J^\perp}\right) \\
\geqslant\;& \frac{1}{\|\,J^\perp(\epsilon)\,\|^2}\; I_\epsilon\left(\frac{\|\,J^\perp(\epsilon)\,\|}{\|\,J_o^\perp(\epsilon)\,\|_o}\; J_o^\perp\,,\; \frac{\|\,J^\perp(\epsilon)\,\|}{\|\,J_o^\perp(\epsilon)\,\|_o}\; J_o^\perp\right) \\
=\;& \frac{1}{\|\,J_o^\perp(\epsilon)\,\|_o^2}\; I_\epsilon(J_o^\perp\,,\,J_o^\perp)\;.
\end{aligned}$$

The first inequality comes from Lemma 9.5.1, which is a second important (albeit straightforward) fact about the index form. Its validity requires the assumed curvature bound and the first four properties of the transplant $\widetilde{J^\perp}$. The second inequality is our so-called basic index lemma (Lemma 7.3.2). Here, we have used again the hypothesis that the geodesic σ_o contains no conjugate points. Also, we have invoked the fifth property of the transplant. It says that on the interval $[0, \epsilon]$, the vector field $\widetilde{J^\perp}$ has the same endpoint values as the Jacobi field $\frac{\| J^\perp(\epsilon) \|}{\| J_o^\perp(\epsilon) \|_o} J_o^\perp$.

Since $J^\perp$ and $J_o^\perp$ are Jacobi fields, formula (7.2.4) applies and we get

$$\begin{aligned} I_\epsilon(J^\perp, J^\perp) &= g_T\Big((J^\perp)'(\epsilon) \, , \, J^\perp(\epsilon) \Big) \, , \\ I_\epsilon(J_o^\perp, J_o^\perp) &= g_{T_o}\Big((J_o^\perp)'(\epsilon) \, , \, J_o^\perp(\epsilon) \Big) \, . \end{aligned}$$

The hypothesis that σ and σ_o are unit speed geodesics, as well as (9.6.2), have all been used without mention.

Putting the last two paragraphs together, and relabeling ϵ as t, we see that

$$\text{(9.6.6)} \qquad \frac{g_T\Big((J^\perp)'(t) \, , \, J^\perp(t) \Big)}{\| J^\perp(t) \|^2} \geqslant \frac{g_{T_o}\Big((J_o^\perp)'(t) \, , \, J_o^\perp(t) \Big)}{\| J_o^\perp(t) \|_o^2} \quad \text{on} \quad (0, l] \, ,$$

as desired.

(C) Deducing $\| J(t) \|^2 \geqslant \| J_o(t) \|_o^2$ on $[0, l]$:

Multiplying (9.6.6) by 2, we can re-express it as

$$\frac{d}{dt} \left[\log \| J^\perp(t) \|^2 \right] \geqslant \frac{d}{dt} \left[\log \| J_o^\perp(t) \|_o^2 \right] \quad \text{on } (0, l] \, .$$

That is,

$$\text{(9.6.7)} \qquad \frac{d}{dt} \left[\log \left(\frac{\| J^\perp(t) \|^2}{\| J_o^\perp(t) \|_o^2} \right) \right] \geqslant 0 \quad \text{on } (0, l] \, .$$

This implies that on $(0, l]$, the function $\frac{\| J^\perp(t) \|^2}{\| J_o^\perp(t) \|_o^2}$ is nondecreasing. In particular,

$$\text{(9.6.8)} \qquad \text{for } 0 < t \leqslant l \, , \quad \frac{\| J^\perp(t) \|^2}{\| J_o^\perp(t) \|_o^2} \geqslant \lim_{t \to 0^+} \frac{\| J^\perp(t) \|^2}{\| J_o^\perp(t) \|_o^2} \, .$$

We calculate that limit by using Taylor's formula.

Since $J^\perp(0) = 0 = J_o^\perp(0)$, our remark right after the proof of Lemma 5.4.1 tells us that

$$\begin{aligned} J^\perp(t) &= (\exp_x)_{*\,[t\,T(0)]} \, \{ t \, (J^\perp)'(0) \} \, , \quad && \text{with } x := \sigma(0) \, , \\ J_o^\perp(t) &= (\exp_{x_o})_{*\,[t\,T_o(0)]} \, \{ t \, (J_o^\perp)'(0) \} \, , \quad && \text{with } x_o := \sigma_o(0) \, . \end{aligned}$$

With the help of these representations, formula (5.5.2) reduces to

$$\| J^{\perp}(t) \|^2 = \| (J^{\perp})'(0) \|^2 \left[t^2 - \frac{t^4}{3} K\left(T(0), (J^{\perp})'(0)\right) \right] + O(t^5) ,$$

$$\| J_o^{\perp}(t) \|_o^2 = \| (J_o^{\perp})'(0) \|_o^2 \left[t^2 - \frac{t^4}{3} K\left(T_o(0), (J_o^{\perp})'(0)\right) \right] + O(t^5) .$$

Only the t^2 terms matter in evaluating $\lim_{t\to 0^+} \frac{\| J^{\perp}(t) \|^2}{\| J_o^{\perp}(t) \|_o^2}$. By hypothesis we have

$$\| (J^{\perp})'(0) \|^2 = \| (J_o^{\perp})'(0) \|_o^2 ,$$

so

$$\text{(9.6.9)} \qquad \lim_{t\to 0^+} \frac{\| J^{\perp}(t) \|^2}{\| J_o^{\perp}(t) \|_o^2} = 1 .$$

Thus (9.6.8) says that,

$$\text{(9.6.10)} \qquad \text{for } 0 < t \leqslant l , \quad \frac{\| J^{\perp}(t) \|^2}{\| J_o^{\perp}(t) \|_o^2} \geqslant 1 .$$

Rearranging and using (9.6.2), followed by (9.6.4), we get

$$\text{(9.6.11)} \qquad \| J(t) \|^2 \geqslant \| J_o(t) \|_o^2 \quad \text{on } [0,l] .$$

(D) Passing (bootstrapping) from $[0,l]$ to $[0,L]$:

Here, we could use an argument similar to the one we gave just before stating Theorem 9.1.1. But, for fun, let us try a totally abstract perspective.

Consider the following subset of $(0,L]$,

$$\mathcal{L} := \{ l \in (0,L] : J^{\perp}(t) \text{ is nowhere zero on } (0,l] \} .$$

What we said in **(A)** shows that $\mathcal{L}$ contains an interval of the form $(0,l]$, for some positive l. In particular, $\mathcal{L}$ is nonempty. Next, if $l \in \mathcal{L}$, then the continuity of $J^{\perp}(t)$ implies that numbers near l again belong to $\mathcal{L}$. In other words, $\mathcal{L}$ is an open set.

It is also closed. To see this, let $\{l_i\}$ be a sequence in $\mathcal{L}$, converging to some $l \in [0,L]$. Inequality (9.6.10) tells us that

$$\| J^{\perp}(l_i) \|^2 \geqslant \| J_o^{\perp}(l_i) \|_o^2 .$$

Letting $l_i \to l$, we get

$$\| J^{\perp}(l) \|^2 \geqslant \| J_o^{\perp}(l) \|_o^2 > 0 ,$$

where the strict inequality follows from the assumption that σ_o contains no conjugate points [see the discussion in **(A)**]. By continuity, $\| J^{\perp}(l) \|^2$ must remain positive in a neighborhood of l, which in turn must capture some l_i. It is now clear that $J^{\perp}(t)$ is nowhere zero on $(0,l]$; hence $l \in \mathcal{L}$.

We have demonstrated that our $\mathcal{L}$ is a nonempty subset which is both open and closed in the connected $(0,L]$. Thus $\mathcal{L} = (0,L]$. Consequently, (9.6.6) and (9.6.11) are in fact respectively valid on $(0, L]$ and $[0, L]$. These are precisely the first two conclusions in the statement of the theorem.

(E) Why σ cannot contain any conjugate points either:
Assume the contrary. Equivalently (see Proposition 7.1.1), suppose there exists some nonzero Jacobi field along the unit speed geodesic $\sigma(t)$, $0 \leqslant t \leqslant L$ that vanishes at t equals 0 and again at some positive $l \leqslant L$. We derive a contradiction.

According to Exercise 5.4.4, this nonzero Jacobi field and its derivative must both be g_T-orthogonal to σ at all times. So, let us denote it by $J^\perp$. By Exercise 9.6.2, one can construct a Jacobi field $J_o^\perp$ along σ_o such that:

(i) $J_o^\perp$ is g_{T_o}-orthogonal to σ_o at all times.
(ii) $J_o^\perp(0) = 0$.
(iii) $\| (J^\perp)'(0) \|^2 = \| (J_o^\perp)'(0) \|_o^2$.
(iv) $J_o^\perp$ is nonzero.

The Jacobi fields $J^\perp$ and $J_o^\perp$ satisfy all the hypotheses of the Rauch theorem; hence

$$\| J^\perp(t) \|^2 \ \geqslant\ \| J_o^\perp(t) \|_o^2 \quad \text{on} \quad [0, L] \ .$$

Since σ_o contains no conjugate points, the nonzero Jacobi field $J_o^\perp$ (which vanishes at $t = 0$) must be nowhere zero on $(0, L]$. At the beginning of this discussion, we assumed that $J^\perp$ vanishes at t equals 0 and again at some positive $l \leqslant L$. But this contradicts the above inequality. So our supposition was wrong and σ contains no conjugate points after all. This proves the third and last conclusion of the theorem. □

Exercises

Exercise 9.6.1: Show that (9.6.3a) and (9.6.3b) are equivalent by first establishing the following formulas:

$$\begin{aligned} J' &= \alpha T + (J^\perp)' , \\ \| J' \|^2 &= \alpha^2 + \| (J^\perp)' \|^2 ; \\ J_o' &= \alpha T_o + (J_o^\perp)' , \\ \| J_o' \|_o^2 &= \alpha^2 + \| (J_o^\perp)' \|_o^2 . \end{aligned}$$

Exercise 9.6.2: This exercise concerns our proposed data on the Jacobi fields $J^\perp$ and $J_o^\perp$. Let the geometrical setup be as stated in Rauch's theorem. Given any Jacobi field $J^\perp$ that is g_T-orthogonal to σ, we claim that a Jacobi field $J_o^\perp$ along σ_o can always be constructed such that:

(i) $J_o^\perp$ is g_{T_o}-orthogonal to σ_o .
(ii) $\| J^\perp(0) \|^2 = \| J_o^\perp(0) \|_o^2$.

(iii) $\| (J^\perp)'(0) \|^2 = \| (J_o^\perp)'(0) \|_o^2$.
(iv) $J_o^\perp$ is nonzero if and only if $J^\perp$ is nonzero.

Use the following guidelines. Choose a g_T-orthonormal basis $\{E_i : i = 1, \ldots, n\}$ for $T_{\sigma(0)}M$, with $E_n := T(0)$. Parallel transport (with reference vector field T) the basis $\{E_i\}$ along σ to obtain $\{E_i(t)\}$. In view of Exercise 5.2.3, $\{E_i(t)\}$ is a g_T-orthonormal frame field along σ. Since σ is an autoparallel, we get $E_n(t) = T(t)$.

(a) Show that

$$\begin{aligned} J^\perp(t) &= \varphi^\alpha(t)\ E_\alpha(t) \ , \\ (J^\perp)'(t) &= (\varphi^\alpha)'(t)\ E_\alpha(t) \ , \\ \| J^\perp(t) \|^2 &= \delta_{\alpha\beta}\ \varphi^\alpha(t)\ \varphi^\beta(t) \ , \\ \| (J^\perp)'(t) \|^2 &= \delta_{\alpha\beta}\ (\varphi^\alpha)'(t)\ (\varphi^\beta)'(t) \ , \end{aligned}$$

where the Greek summation indices α, β run from 1 to $n-1$.

(b) Analogous to the definition of $E_i(t)$, carry out parallel transport (with reference vector T_o) along σ_o to get a g_{T_o}-orthonormal frame field $\{F_I(t) : I = 1, \ldots, n, n+1, \ldots, n+k\}$, with $F_n(t) = T_o(t)$.

By ODE theory, there exists a unique Jacobi field $J_o^\perp$ along σ_o with the initial data

$$\begin{aligned} J_o^\perp(0) &:= \varphi^\alpha(0)\ F_\alpha(0) \ , \\ (J_o^\perp)'(0) &:= (\varphi^\alpha)'(0)\ F_\alpha(0) \ . \end{aligned}$$

The summation only runs from 1 to $n-1$, so neither $J_o(0)$ nor $J_o'(0)$ has any component along $F_n(0)$ [which is $T_o(0)$], $F_{n+1}(0)$, ..., $F_{n+k}(0)$. Use Exercise 5.4.3 to conclude that the Jacobi field $J_o^\perp$ is indeed g_{T_o}-orthogonal to σ_o at all times. This justifies the superscript $^\perp$ and gives property (i).

(c) Check that

$$\begin{aligned} \| J_o^\perp(0) \|_o^2 &= \delta_{\alpha\beta}\ \varphi^\alpha(0)\ \varphi^\beta(0) \ , \\ \| (J_o^\perp)'(0) \|_o^2 &= \delta_{\alpha\beta}\ (\varphi^\alpha)'(0)\ (\varphi^\beta)'(0) \ . \end{aligned}$$

Do these imply properties (ii) and (iii)? How about (iv)?

9.7 Jacobi Fields on Space Forms

Let (M, F) be a Finsler manifold of constant flag curvature λ. Let $\sigma(t)$, $0 \leqslant t \leqslant L$ be a unit speed geodesic in (M, F), with velocity field T. We learned from Exercises 5.4.3 and 5.4.5 that every Jacobi field J along σ splits into two individual Jacobi fields $J_\|$ and $J^\perp$, respectively, tangent to and g_T-orthogonal to σ. As before, g_T abbreviates $g_{(\sigma,T)}$, and $(J^\perp)'$ means $D_T J^\perp$.

The structure of $J_{||}$ is well understood. Indeed, Exercise 5.4.3 says that if $J_{||}(0) = 0$, then $J_{||}(t)$ must be a constant multiple of tT. In this section, let us focus our attention on the g_T-orthogonal piece $J^{\perp}$, which we suppose is nonzero. Exercise 5.4.5 tells us that $J^{\perp}$ satisfies the ODE

$$(J^{\perp})'' + \lambda J^{\perp} = 0 .$$

In order to solve this equation, we let $\{E_i(t)\}$ be a basis of parallel vector fields along σ. This is obtained through parallel transport along σ, with reference vector T. Carry out the expansion

$$J^{\perp}(t) = f^i(t)\, E_i(t) .$$

Exercise 5.4.5 assures us that the coefficient functions must satisfy the scalar Jacobi equation

$$(f^i)'' + \lambda f^i = 0 .$$

Suppose $J^{\perp}(0) = 0$. Since $J^{\perp}$ is not identically zero, we must have $(J^{\perp})'(0) \neq 0$. Let us arrange to have

$$E_n(0) = (J^{\perp})'(0) = (f^i)'(0)\, E_i(0) .$$

Then the initial data for the coefficients f^i are:

$$\begin{aligned} f^{\alpha}(0) &= 0 , \\ (f^{\alpha})'(0) &= 0 , \end{aligned}$$

where α runs from 1 to $n-1$, and

$$\begin{aligned} f^{n}(0) &= 0 , \\ (f^{n})'(0) &= 1 . \end{aligned}$$

Thus the first $n-1$ coefficient functions $f^1(t), \ldots, f^{n-1}(t)$ are all identically zero. The expansion $J^{\perp} = f^i E_i$ now reduces to

$$\boxed{J^{\perp}(t) = \mathfrak{s}_{\lambda}(t)\, E(t)} , \tag{9.7.1}$$

where we have relabeled f^n as $\mathfrak{s}_{\lambda}$ and E_n as E. Here,

- $J^{\perp}(0) = 0$.
- $E(t)$ is the parallel transport along σ (with reference vector T) of the nonzero vector $(J^{\perp})'(0)$.
- $\mathfrak{s}_{\lambda}(t)$ is the solution of the ODE $\mathfrak{s}_{\lambda}'' + \lambda \mathfrak{s}_{\lambda} = 0$, subject to the initial data $\mathfrak{s}_{\lambda}(0) = 0$, $\mathfrak{s}_{\lambda}'(0) = 1$. It has been explicitly displayed in (9.1.3). We reproduce it here for convenience and for emphasis:

$$\mathfrak{s}_{\lambda}(t) := \left\{ \begin{array}{c} \dfrac{1}{\sqrt{\lambda}} \sin(\sqrt{\lambda}\, t) \\ t \\ \dfrac{1}{\sqrt{-\lambda}} \sinh(\sqrt{-\lambda}\, t) \end{array} \right\} \quad \text{resp., for} \quad \left\{ \begin{array}{c} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \end{array} \right\} .$$

Exercises

Exercise 9.7.1: This concerns formula (9.7.1).

(a) Explain why $\| E(t) \|^2 := g_T(E(t) , E(t))$ is a positive *constant.*

(b) By identifying that constant, check that

$$\| J^\perp(t) \|^2 \ := \ g_T(J^\perp(t) , J^\perp(t)) \ = \ \| (J^\perp)'(0) \|^2 \, \mathfrak{s}_\lambda^2(t) \ .$$

Is this formula consistent with the Taylor expansion (5.5.3)?

Exercise 9.7.2: For any t at which $\mathfrak{s}_\lambda(t) \neq 0$, demonstrate that

$$\frac{g_T\Big((J^\perp)'(t) , J^\perp(t) \Big)}{\| J^\perp(t) \|^2} \ = \ \frac{1}{2} \frac{d}{dt} [\log \| J^\perp(t) \|^2] \ = \ \frac{\mathfrak{s}_\lambda'(t)}{\mathfrak{s}_\lambda(t)} \ ,$$

where

$$\left.\begin{aligned} \frac{\mathfrak{s}_\lambda'(t)}{\mathfrak{s}_\lambda(t)} &= \sqrt{\lambda}\ \cot(\sqrt{\lambda}\, t) \ , \quad 0 < t < \frac{\pi}{\sqrt{\lambda}} \\ \frac{\mathfrak{s}_\lambda'(t)}{\mathfrak{s}_\lambda(t)} &= \frac{1}{t} \ , \quad 0 < t \\ \frac{\mathfrak{s}_\lambda'(t)}{\mathfrak{s}_\lambda(t)} &= \sqrt{-\lambda}\ \coth(\sqrt{-\lambda}\, t) \ , \quad 0 < t \end{aligned}\right\} \ \text{resp., for} \ \begin{cases} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \ . \end{cases}$$

Exercise 9.7.3: Fix any point x in a Finsler manifold of constant flag curvature λ. Let $\sigma(t)$ be any unit speed geodesic that emanates from x with initial velocity T. Show that:

- If $\lambda > 0$, then the conjugate value in the direction T is $\frac{\pi}{\sqrt{\lambda}}$. In other words, $c_T = \frac{\pi}{\sqrt{\lambda}}$.
- If $\lambda \leqslant 0$, then σ contains no conjugate point. That is, $c_T = \infty$.

9.8 Applications of Rauch's Theorem

In this section, we specialize the Rauch theorem to the case in which one of the two spaces (M, F), (M_o , F_o) has constant flag curvature λ.

Using information we have just gathered about Jacobi fields on space forms, one can recover (not surprisingly!) the estimates obtained in Theorem 9.1.1 and Exercise 9.1.3. However, this time they are obtained much more systematically.

In order to state the results in question, let us use (as before)

$$\left.\begin{aligned} & g_T \\ & (\ \ )' \\ & \| W \|^2 \end{aligned}\right\} \quad \text{to abbreviate} \quad \begin{cases} g_{(\sigma,T)} \\ D_T \ \ \text{(with reference vector } T) \\ g_T(W , W) \ . \end{cases}$$

Also as before, let

$$\mathfrak{s}_\lambda(t) := \left\{ \begin{array}{c} \frac{1}{\sqrt{\lambda}} \sin(\sqrt{\lambda}\, t) \\ t \\ \frac{1}{\sqrt{-\lambda}} \sinh(\sqrt{-\lambda}\, t) \end{array} \right\} \text{ resp., for } \left\{ \begin{array}{c} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \end{array} \right\} ,$$

so that

$$\left. \begin{array}{l} \frac{\mathfrak{s}_\lambda'(t)}{\mathfrak{s}_\lambda(t)} = \sqrt{\lambda}\, \cot(\sqrt{\lambda}\, t) , \quad 0 < t < \frac{\pi}{\sqrt{\lambda}} \\ \frac{\mathfrak{s}_\lambda'(t)}{\mathfrak{s}_\lambda(t)} = \frac{1}{t} , \quad 0 < t \\ \frac{\mathfrak{s}_\lambda'(t)}{\mathfrak{s}_\lambda(t)} = \sqrt{-\lambda}\, \coth(\sqrt{-\lambda}\, t) , \quad 0 < t \end{array} \right\} \text{ resp., for } \left\{ \begin{array}{l} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \end{array} \right. .$$

Judicious use of Rauch's theorem (Theorem 9.6.1) and the discussion in §9.7 give the following two results.

Corollary 9.8.1. *Let (M,F) be an n-dimensional Finsler manifold. Let $\sigma(t)$, $0 \leqslant t \leqslant L$ be a unit speed geodesic in M, with velocity field $T(t)$. Suppose:*

- *The flag curvature $K(T,W) \leqslant \lambda$ for any $W \in T_{\sigma(t)}M$.*
- *$J^\perp$ is a Jacobi field along σ that is g_T-orthogonal to σ.*
- *$J^\perp(0) = 0$.*
- *For simplicity, $L \geqslant \frac{\pi}{\sqrt{\lambda}}$ in case λ is positive.*

Then:

$$\frac{g_T\Big((J^\perp)'(t) \,,\, J^\perp(t) \Big)}{\| J^\perp(t) \|^2} \geqslant \frac{\mathfrak{s}_\lambda'(t)}{\mathfrak{s}_\lambda(t)} \quad \textit{for} \quad \left\{ \begin{array}{ll} 0 < t < \frac{\pi}{\sqrt{\lambda}} & \textit{if } \lambda > 0 \\ 0 < t \leqslant L & \textit{if } \lambda \leqslant 0 \end{array} \right.$$

$$\| J^\perp(t) \|^2 \geqslant \| (J^\perp)'(0) \|^2 \, \mathfrak{s}_\lambda^2(t) \quad \textit{for} \quad \left\{ \begin{array}{ll} 0 \leqslant t \leqslant \frac{\pi}{\sqrt{\lambda}} & \textit{if } \lambda > 0 \\ 0 \leqslant t \leqslant L & \textit{if } \lambda \leqslant 0 \end{array} \right. .$$

Corollary 9.8.2. *Let (M,F) be an n-dimensional Finsler manifold. Let $\sigma(t)$, $0 \leqslant t \leqslant L$ be a unit speed geodesic in M, with velocity field $T(t)$. Suppose:*

- *The flag curvature $K(T,W) \geqslant \lambda$ for any $W \in T_{\sigma(t)}M$.*
- *$J^\perp$ is a Jacobi field along σ that is g_T-orthogonal to σ.*
- *$J^\perp(0) = 0$.*

- *For simplicity,* $L \geqslant c_T$*, where* $\sigma(c_T)$ *is the first conjugate point of* $\sigma(0)$ *along* σ*.*

Then:

$$\frac{g_T\Big((J^\perp)'(t) \, , \, J^\perp(t) \Big)}{\| \, J^\perp(t) \, \|^2} \;\leqslant\; \frac{\mathfrak{s}_\lambda'(t)}{\mathfrak{s}_\lambda(t)} \quad \textit{for } 0 < t < c_T \, .$$

$$\| \, J^\perp(t) \, \|^2 \;\leqslant\; \| \, (J^\perp)'(0) \, \|^2 \; \mathfrak{s}_\lambda^2(t) \quad \textit{for } 0 \leqslant t \leqslant c_T \, .$$

Exercises

Exercise 9.8.1: Establish Corollaries 9.8.1 and 9.8.2 using the Rauch comparison theorem (Theorem 9.6.1), together with §9.7.

(a) For Corollary 9.8.1: Let the comparison space $(M_o \, , \, F_o)$ be an n-dimensional complete Riemannian manifold of constant sectional curvature λ. Fix any unit speed geodesic $\sigma_o(t)$, $t \geqslant 0$ in that M_o. If $0 < \lambda = \frac{1}{r^2}$, let $(M_o \, , \, F_o)$ be the standard n-sphere of radius r, whose geodesics are great circles. In this case, $\sigma_o(t)$ contains no conjugate points as long as $0 \leqslant t < \pi \, r = \frac{\pi}{\sqrt{\lambda}}$. If $\lambda \leqslant 0$, let $(M_o \, , \, F_o)$ be either Euclidean $\mathbb{R}^n$ or hyperbolic space; hence σ_o contains no conjugate points.

(b) For Corollary 9.8.2: Let $(M_o \, , \, F_o)$ be an n-dimensional complete Riemannian manifold of constant sectional curvature λ. Fix any unit speed geodesic $\sigma_o(t)$, $t \geqslant 0$ in M_o . When we apply the Rauch theorem, be sure to let our given Finsler manifold (M, F) play the role of the "$(M_o \, , \, F_o)$" in the statement of Rauch's theorem. This means the space form we just defined now takes the place of the "(M, F)" in that theorem. One final hypothesis needs to be satisfied. Namely, our present geodesic σ should contain no conjugate points. This explains the appearance of the parameter value c_T.

In both cases, one also needs to use Exercise 9.6.2 to construct a Jacobi field $J_o^\perp$ along σ_o such that $J_o^\perp$ is g_{T_o}-orthogonal to σ_o, $J_o^\perp(0) = 0$, and $\| \, (J_o^\perp)'(0) \, \|_o^2 = \| \, (J^\perp)'(0) \, \|^2$.

Exercise 9.8.2: Derive the following consequence of Corollaries 9.8.1 and 9.8.2. Let (M, F) be an n-dimensional Finsler manifold. Let $\sigma(t)$, $0 \leqslant t \leqslant L$ be a unit speed geodesic in M, with velocity field $T(t)$. Suppose the flag curvature has the following uniform bounds

$$0 \;\leqslant\; \lambda \;\leqslant\; K(T, W) \;\leqslant\; \Lambda$$

for any $W \in T_{\sigma(t)}M$. For simplicity, suppose $L \geqslant \text{Max}\{\,c_T,\ \frac{\pi}{\sqrt{\Lambda}}\,\}$, where $\sigma(c_T)$ is the first conjugate point of $\sigma(0)$ along σ. Then

$$\frac{\pi}{\sqrt{\Lambda}} \leqslant c_T \leqslant \frac{\pi}{\sqrt{\lambda}}\,.$$

In other words, the distance between any two consecutive conjugate points along σ is at least $\frac{\pi}{\sqrt{\Lambda}}$ but no more than $\frac{\pi}{\sqrt{\lambda}}$. (For a simpler statement, one can set $\Lambda = \infty$.) How does this result compare with Theorem 9.5.2?

References

[Au] L. Auslander, *On curvature in Finsler geometry*, Trans. AMS **79** (1955), 378–388.

[CE] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, North Holland/American Elsevier, 1975.

[Daz] P. Dazord, *Propriétés globales des géodésiques des Espaces de Finsler*, Theses, Université de Lyon, 1969.

[doC2] M.P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, 1976.

[doC3] M.P. do Carmo, *Riemannian Geometry*, Mathematics: Theory and Applications, Birkhäuser, 1992.

[E] D. Egloff, *Uniform Finsler Hadamard manifolds*, Ann. Inst. Henri Poincaré **66** (1997), 323–357.

[KN1] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, vol. I, Wiley-Interscience, 1963 (1996).

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

[ST] I. M. Singer and J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry*, Undergraduate Texts in Mathematics, Springer-Verlag, 1976.

[Sp4] M. Spivak, *Differential Geometry*, vol. IV, Publish or Perish, 1975.

Chapter 10

Berwald Spaces and Szabó's Theorem for Berwald Surfaces

10.0 Prologue

In this chapter, we study Berwald spaces in some detail. Here are several reasons why such spaces are so important. These reasons are elaborated upon as the chapter unfolds.

* Berwald spaces are just a bit more general than Riemannian and locally Minkowskian spaces. They provide examples that are more properly Finslerian, but only slightly so.

* For Berwald spaces, the Chern connection (in natural coordinates) actually defines a linear connection directly on the underlying manifold M. (For that matter, so does the Berwald connection.) A theorem of Szabó's tells us that this linear connection also happens to be the Levi-Civita (Christoffel) connection of a (nonunique) Riemannian metric on M.

* Given a Berwald space, all its tangent spaces are linearly isometric to a common Minkowski space. One might say that the Berwald space in question is modeled on a single Minkowski space.

Since we have an eye toward global theorems, we focus on Finsler structures F that are smooth and strongly convex on $TM \setminus 0$. These F are y-**global**. Surprisingly, it takes some work to explicitly locate a y-global Berwald space that is neither Riemannian nor locally Minkowskian. In fact, according to a rigidity result (see §10.6) of Szabó's, such a creature does not even exist in dimension two. Happily, examples of the desired vintage do exist in dimension three or higher. But they won't materialize until we reach §11.6, after an excursion into the territory of Randers spaces.

10.1 Berwald Spaces

A Finsler structure F is said to be of **Berwald** type if the Chern connection coefficients $\Gamma^i{}_{jk}$ in natural coordinates have no y dependence. We see from part (a) of Exercise 2.4.8 that this is a well-defined concept. For a Berwald space, the coefficients $\Gamma^i{}_{jk}$ define a *linear* covariant derivative D directly on the underlying manifold M.

** The derivative of a vector field $W := W^i \frac{\partial}{\partial x^i}$ in the direction of $v := v^k \frac{\partial}{\partial x^k}$ is

$$D_v W \;:=\; v^k \left(\frac{\partial W^i}{\partial x^k} \;+\; W^j \, \Gamma^i{}_{jk} \right) \frac{\partial}{\partial x^i} \;.$$

** Let $\sigma(t)$ be any curve in M with velocity field $T(t) := \frac{d\sigma}{dt}$. Suppose $W(t) := W^i(t) \frac{\partial}{\partial x^i}$ is a vector field along σ. The covariant derivative of W along σ is

$$D_T W \;:=\; \left[\frac{dW^i}{dt} \;+\; W^j \, T^k \, \Gamma^i{}_{jk} \right] \frac{\partial}{\partial x^i}_{|\sigma(t)} \;.$$

The vector field $W(t)$ is said to be **parallel** along σ if $D_T W = 0$.

Since we are dealing with a Finsler metric of Berwald type, there is by assumption no y dependence in $\Gamma^i{}_{jk}$. For this reason, the concept of a *reference vector*, first introduced amidst the exercises of §5.2 and used thereafter, is happily irrelevant here.

Recall that each tangent space $T_x M$ of a Finsler manifold is a Minkowski normed linear space $(T_x M, F_x)$. In this context, a result of Ichijyō's [I] says the following.

Proposition 10.1.1 (Ichijyō). *Let* (M, F) *be a* **Berwald** *space. Then:*

- *Given any parallel vector field* W *along a curve* σ *in* M*, its Finslerian norm* $F(W)$ *is necessarily constant along* σ*.*

- *Whenever M is connected, its Minkowski normed linear spaces, namely, (T_xM, F_x), are all* **linearly** *isometric to each other.*

Proof. By (1.2.5), we can express the Finslerian norm of W as

$$F(W) = \sqrt{g_W(W,W)} \ ,$$

where $g_W := g_{ij\,(\sigma,W)}\, dx^i \otimes dx^j$. Note that even though $\Gamma^i{}_{jk}$ has no y dependence, the same cannot be said of g_{ij}.

Proving the first assertion:

We proceed analogous to Exercise 5.2.2. There, one worked with the canonical lift (σ, T), where T is the velocity field of σ. But here let us use the lift (σ, W) instead. The almost g-compatibility criterion (2.4.6) of the Chern connection reads

$$dg_{ij} - g_{kj}\, \omega_i{}^k - g_{ik}\, \omega_j{}^k = 2\, A_{ijs}\, \frac{\delta y^s}{F} \ ,$$

where $\omega_j{}^i = \Gamma^i{}_{jk}\, dx^k$. Restrict it to the lift $(\sigma, W)(t)$. Then contract with W^i and W^j. The A term promptly drops out because of (1.4.6). So we are left with

$$W^i\, W^j\, [\, dg_{ij} - g_{kj}\, \Gamma^k{}_{is}\, dx^s - g_{ik}\, \Gamma^k{}_{js}\, dx^s\,] = 0 \ .$$

Next, contract this 1-form equation with the velocity of the lift (σ, W). The result is

$$W^i\, W^j \left(\frac{dg_{ij}}{dt} - g_{kj}\, \Gamma^k{}_{is}\, T^s - g_{ik}\, \Gamma^k{}_{js}\, T^s \right) = 0 \ .$$

Here, we have used the fact that in natural coordinates, the Chern connection forms $\omega_j{}^i$ contain only dx terms. After some straightforward manipulations and relabeling, the above becomes

$$\frac{d}{dt}\, g_W(W,W) = 2\, g_W(\, D_T W\, ,\, W\,) \ .$$

The first assertion of the proposition is now immediate.

Proving the second assertion:

To that end, we take any two points p, q in the connected M and a curve σ (with velocity field T) joining them. In an arbitrary Finsler space, the differential equation $D_T W = 0$ is linear in W if the reference vector is T, and typically nonlinear in W if the reference vector is W. But presently we are dealing with a Berwald space, so reference vectors are irrelevant, and that ODE is always linear in W. It can therefore be used to define a linear map from T_pM into T_qM. Furthermore, we have ascertained above that this parallel transport preserves the Finsler norm of the vector being transported. Hence it must be an isomorphism as well. □

Each tangent space T_xM of an n-dimensional Riemannian manifold M is linearly isometric to $\mathbb{R}^n$ equipped with the dot product. This can be seen by taking an orthonormal basis for T_xM. In this sense, one says that *all* n-dimensional Riemannian manifolds are modeled on a *canonical* inner product space, namely, Euclidean $\mathbb{R}^n$. By contrast, the tangent Minkowski spaces (T_xM, F_x) of an arbitrary Finsler manifold (M, F) are typically not isometric to each other. So, generic Finsler manifolds are *not* modeled on any single Minkowski space, let alone a canonical one.

The result of Ichijyō's that we have just described puts Berwald spaces somewhere in between Riemannian manifolds and generic Finsler manifolds. In essence, it says that each *connected* Berwald space is modeled on precisely *one* Minkowski space, although the exact identity of the latter does vary from one Berwald space to another. So, while Euclidean $\mathbb{R}^n$ gives rise to the entire category of Riemannian manifolds, every single Minkowski norm (among an inexhaustible supply!) gives rise to a whole family of connected Berwald spaces. Of course, when the Minkowski norm in question happens to be the norm associated with the usual dot product, the family it generates is comprised of all Riemannian manifolds. Viewed from this perspective, the category of Berwald spaces extends and unquestionably dwarfs that of Riemannian manifolds.

Exercises

Exercise 10.1.1: Let (M, F) be a Berwald space. Let σ be a curve in M with velocity field T. Suppose V and W are vector fields defined along σ.

(a) Imitate the proof of Proposition 10.1.1 to show that

$$\frac{d}{dt}\, g_W(V, W) \;=\; g_W(D_T V\,,\, W) \;+\; g_W(V\,,\, D_T W)\;.$$

(b) How does this result compare with that in Exercise 5.2.3?

Exercise 10.1.2: Let us now work in the setting of an *arbitrary* Finsler manifold. Then the Chern connection coefficients in natural coordinates do (generically) have a directional y-dependence. By extrapolating slightly from the exercise portion of §5.2, one could define $D_T V$ and $D_T W$, *both* with reference vector W. Explicitly:

$$D_T W \;:=\; \left[\frac{dW^i}{dt} \;+\; W^j\, T^k\, ({\Gamma^i}_{jk})_{(\sigma, W)}\right] \frac{\partial}{\partial x^i}_{|\sigma(t)}\;,$$

$$D_T V \;:=\; \left[\frac{dV^i}{dt} \;+\; V^j\, T^k\, ({\Gamma^i}_{jk})_{(\sigma, W)}\right] \frac{\partial}{\partial x^i}_{|\sigma(t)}\;.$$

(a) Show that we again have

$$\frac{d}{dt}\, g_W(V, W) \;=\; g_W(D_T V\,,\, W) \;+\; g_W(V\,,\, D_T W)\;.$$

Compare your arguments with that in Exercise 5.2.3.

(b) Explain why the quantity $D_T W$, with reference vector W, is nonlinear in W.

(c) Define a nonlinear notion of parallel transport by solving $D_T W = 0$ (with reference vector W). Check that this process, though nonlinear, still preserves the Finsler norm $F(W)$ of the vector being transported.

Exercise 10.1.3: Given any Finsler manifold (M, F), one computes as usual the components $g_{ij} := (\frac{1}{2}F^2)_{y^i y^j}$ of the fundamental tensor

$$g \ := \ g_{ij\,(x,y)} \ dx^i \otimes dx^j \ .$$

This tensor is a section of the pulled-back tensor bundle $\pi^*T^*M \otimes \pi^*T^*M$, which sits over the slit tangent bundle $TM \setminus 0$.

On the other hand, we can also define the object

$$\hat{g} \ := \ g_{ij\,(x,y)} \ dy^i \otimes dy^j \ .$$

For each fixed $x \in M$, this represents a Riemannian metric $\hat{g}_x$ on the punctured space $T_xM \setminus 0$. The indicatrix S_xM therefore inherits a Riemannian metric, say $\dot{g}_x$. We briefly pursued this type of thinking in §1.4, and then revisited it in §4.1, 4.5.

(a) Suppose σ is a curve in M from p to q. The nonlinear parallel transport (along σ) discussed in part (c) of Exercise 10.1.2 uses, as reference vector, the one that's being transported. Explain why it gives rise to a map from $(T_pM,\ \hat{g}_p)$ to $(T_qM,\ \hat{g}_q)$, and also one from $(\mathrm{S}_pM,\ \dot{g}_p)$ to $(\mathrm{S}_qM,\ \dot{g}_q)$.

(b) Prove that these maps are actually diffeomorphisms. Hints: let's say that a tangent vector W_p (at p) got parallel transported along σ to a tangent vector W_q (at q). Explain why parallel transporting W_q along the reverse of σ should give us back W_p. Equivalently, suppose $W(t)$ is parallel along $\sigma(t)$, $0 \leqslant t \leqslant r$. (Keep in mind that W is also the reference vector here !) Check that $W(r - t)$ is parallel along the reverse of σ, namely, $\sigma(r - t)$.

(c) In case (M, F) is a Berwald space, the said diffeomorphisms come from restricting the linear isometry between the Minkowski spaces (T_pM, F_p) and (T_qM, F_q). Prove that in this case, the diffeomorphisms in question are also isometries between Riemannian manifolds. A map $\phi : (M, g) \to (N, h)$ between Riemannian manifolds is a **Riemannian isometry** if $h(\phi_* u, \phi_* v) = g(u, v)$.

(d) Other than the Berwald case, can you think of any other setting in which the above diffeomorphisms are Riemannian isometries? How about Finsler spaces for which $\dot{A}_{ijk} = 0$?

Exercise 10.1.4: Let (M, F) be a Berwald space. The Chern connection coefficients ${\Gamma^i}_{jk}$ in natural coordinates are then functions on M. They are always symmetric in j and k. It is therefore natural to wonder whether they coincide with some (*nonunique*) Riemannian metric's Christoffel symbols of the second kind. As Szabó [Sz] demonstrated, this is indeed the case. Note that one's obvious guess—the fundamental tensor—does not work. This is because the g_{ij} for Berwald spaces typically depend on y, and hence do not live on M.

Recall the covariant derivative D that we discussed in this section. Relative to a coordinate basis $\{\frac{\partial}{\partial x^i}\}$, it is described by the Chern connection coefficients ${\Gamma^i}_{jk}$, which have the symmetry ${\Gamma^i}_{kj} = {\Gamma^i}_{jk}$. As a result, the operator D is torsion-free:

$$D_X Y \; - \; D_Y X \; = \; [X,Y] \, .$$

To effect Szabó's claim, it suffices to construct a Riemannian metric h on M that satisfies

$$Dh \; = \; 0 \, .$$

In other words, D is compatible with h. We can then "deduce" from Theorem 2.4.1 (or quote directly from §13.1) that D is the unique torsion-free connection which is compatible with the constructed h. So, in natural coordinates, its connection coefficients must coincide with the Christoffel symbols (of the second kind) of h.

Let us go through Szabó's construction below. Fix any $x_o \in M$. Denote by G the group of all linear isomorphisms of $T_{x_o}M$ that preserve the indicatrix $S_{x_o}M$. Since the indicatrix is compact, this G is a compact Lie group. The strategy here is to define an appropriate inner product at x_o by G-averaging, and then use parallel translation to propagate it to all points of M that lie in the same connected component as x_o.

(a) Begin with any inner product, say $\mathcal{I}$, on $T_{x_o}M$. Since G acts on this tangent space, each $g \in G$ induces a new inner product which we denote by $g^*\mathcal{I}$. Define

$$h_{x_o} \; := \; \frac{1}{\mathrm{Vol}(G)} \int_G g^*\mathcal{I} \;\, \mu_G \, ,$$

where μ_G denotes the Haar measure on the compact group G. Verify that the inner product h_{x_o} is G-invariant. That is,

$$g^* \, h_{x_o} \; = \; h_{x_o}$$

for all $g \in G$.

(b) Consider any $x \in M$ that is in the same connected component as x_o. Connect x to x_o by a curve σ. Let P_σ denote the parallel translation operator from T_xM to $T_{x_o}M$, along σ. Define

$$h_x \; := \; P_\sigma^* \, h_{x_o} \; := \; h_{x_o}(\, P_\sigma \,\cdot \,, \, P_\sigma \,\cdot \,) \, .$$

It must be checked that if a different path $\tilde{\sigma}$ from x to x_o were chosen, one would still get the same answer for h_x. To this end, let

σ_- denote the reverse of σ. By considering the loop $\tilde{\sigma} \circ \sigma_-$ based at x_o, check that

$$P^*_{\tilde{\sigma}}\, h_{x_o} \;=\; P^*_{\sigma}\, h_{x_o} \;.$$

(c) Show that h satisfies all the axioms of a Riemannian metric. How does its differentiability class depend on that of D?

(d) It remains to show that D is compatible with h. Take any tangent vector $v \in T_xM$; it suffices to check that

$$(D_v h)(\, W_i(x) \,,\, W_j(x)\,) \;=\; 0 \;,$$

where $\{W_i(x)\}$ is a basis for T_xM. Let us realize v as the initial velocity of a curve $\gamma(t)$ and parallel translate $\{W_i(x)\}$ along γ to form a field of bases $\{W_i\}$. Explain why the formula

$$\left(D_{\frac{d\gamma}{dt}} h\right)(U,V) \;=\; \frac{d}{dt}[\,h(U,V)\,] \\ \qquad - h\Big(D_{\frac{d\gamma}{dt}} U \,,\, V\Big) \;-\; h\Big(U \,,\, D_{\frac{d\gamma}{dt}} V\Big)$$

reduces in our setting to

$$(D_v h)(\, W_i(x) \,,\, W_j(x)\,) \;=\; \frac{d}{dt}_{|t=0}[\,h(W_i \,,\, W_j)\,] \;.$$

Calculate $h(W_i, W_j)$ using a curve that runs from the point $\gamma(t)$ to x, and then from x to x_o. Conclude that our $h(W_i, W_j)$ does not depend on t. Thus $Dh = 0$.

With the conclusion of this exercise, one can deduce a **Gauss–Bonnet–Chern theorem for Berwald spaces** of arbitrary dimension. This was first observed by Shen. For more details, see the last section of [BCS1], and then [Ch3], [Ch4].

10.2 Various Characterizations of Berwald Spaces

In this section, we produce several characterizations of Berwald spaces. Some of these are particularly useful to us later.

Proposition 10.2.1. *Let (M, F) be a Finsler manifold. Then the following five criteria are equivalent:*

(a) *The hv-part of the Chern curvature vanishes identically: $P_j{}^i{}_{kl} = 0$.*

(b) *The Cartan tensor is covariantly constant along all horizontal directions on the slit tangent bundle $TM \setminus 0$. Namely: $A_{ijk|l} = 0$.*

(c) *The Chern connection coefficients $\Gamma^i{}_{jk}$ in natural coordinates do not depend on the directional variable y. In other words, the Finsler structure F is of Berwald type.*

(d) *The quantities* $(\Gamma^i{}_{jk}\, y^j y^k)_{y^p y^q}$ *do not depend on* y.

(e) *The quantities* $(G^i)_{y^p y^q}$*, with* $G^i := \gamma^i{}_{jk}\, y^j y^k$*, do not depend on* y.

Proof. Let us make a few preliminary observations:

- The equivalence between (a) and (b) was established near the end of §3.4. See (3.4.13). The key consists of constitutive formulas that come from first Bianchi identities. These formulas express the curvature P entirely in terms of the horizontal covariant derivatives of A, and vice versa.
- Formula (3.3.3) says that in natural coordinates,

$$P_j{}^i{}_{kl} \;=\; -\,F\,\frac{\partial \Gamma^i{}_{jk}}{\partial y^l}\ .$$

  This immediately gives the equivalence between (a) and (c).
- In (e), the $\gamma^i{}_{jk}$ are the fundamental tensor's formal Christoffel symbols of the second kind. See (2.3.1). Using the explicit formula (2.4.9) of the Chern connection Γ, one can check that $\Gamma^i{}_{jk}\, y^j y^k$ reduces to $\gamma^i{}_{jk}\, y^j y^k$. The equivalence between (d) and (e) is now transparent.
- Given (c), straightforward differentiation shows that $(\Gamma^i{}_{jk}\, y^j y^k)_{y^p y^q}$ is equal to $2\Gamma^i{}_{pq}$. Thus (c) indeed implies (d).

In view of these remarks, it remains to prove that (d) $\Rightarrow$ (c). To this end, the first step is to establish that

$$(\,\Gamma^i{}_{jk}\; y^j\, y^k\,)_{y^p y^q} \;=\; 2\;(\,\dot{A}^i{}_{pq} \;+\; \Gamma^i{}_{pq}\,)\ .$$

This is accomplished by applying (3.3.3) and (3.4.9) twice, and (3.2.3) once. More detailed guidance can be found in Exercise 10.2.1.

Suppose (d) holds. Namely, the quantities $(\Gamma^i{}_{jk}\, y^j y^k)_{y^p y^q}$ do not depend on y. The above identity then implies that, in natural coordinates, the Berwald connection $\dot{A}^i{}_{pq} + \Gamma^i{}_{pq}$ is independent of y. So

$$\dot{A}^i{}_{pq;s} \;+\; \Gamma^i{}_{pq;s} \;=\; 0\ ,$$

where each semicolon abbreviates the operation $F\frac{\partial}{\partial y}$. The use of (3.3.3) converts this to

$$(*)\qquad P_p{}^i{}_{qs} \;=\; \dot{A}^i{}_{pq;s}\ .$$

Formulas (3.4.8), (3.4.9) together say that

$$P_{ijkl} \;+\; P_{jikl} \;=\; 2\;A_{iju}\;\dot{A}^u{}_{kl} \;-\; 2\;A_{ijl|k}\ .$$

Thus

$$\ell^j\;P_{ijkl} \;=\; -\;\ell^j\;P_{jikl} \;=\; \dot{A}_{ikl}\ .$$

In view of this, contraction of (*) with ℓ_i gives

$$\dot{A}_{pqs} = \ell_i \, \dot{A}^i{}_{pq;s} = - \, \ell_{i;s} \, \dot{A}^i{}_{pq} \ .$$

But (2.5.17) tells us that $\ell_{i;s} = g_{is} - \ell_i \ell_s$. The above then simplifies to

$$2 \, \dot{A}_{pqs} = 0 \ ,$$

showing that the Berwald and Chern connections in fact coincide. We already know that the Berwald connection in natural coordinates is independent of y, so the same holds for the Chern connection. □

We close by highlighting a few facts:

- For a Berwald space, the Chern connection coefficients in natural coordinates have no y-dependence. This is the definition of a Berwald space.
- Equivalent to this definition is the vanishing of the *hv* Chern curvature. Namely, $P_j{}^i{}_{kl} = 0$. Also equivalent is the vanishing of the *hv* Berwald curvature ${}^bP_j{}^i{}_{kl}$ introduced in Exercise 3.8.4. This can be seen by putting part (b) of that exercise together with criteria (a) and (e) of Proposition 10.2.1.
- The *hh*-Chern curvature of a Berwald space is given by

$$\boxed{R_j{}^i{}_{kl} = \frac{\partial \Gamma^i{}_{jl}}{\partial x^k} - \frac{\partial \Gamma^i{}_{jk}}{\partial x^l} + \Gamma^i{}_{hk} \, \Gamma^h{}_{jl} - \Gamma^i{}_{hl} \, \Gamma^h{}_{jk}} \ .$$

  Note that these are ordinary partial derivatives with respect to x.

Exercises

Exercise 10.2.1: Let $\Gamma^i{}_{jk}$ be the Chern connection of an arbitrary Finsler manifold. In this exercise, compute with $\Gamma^i{}_{jk} \, y^j y^k$ as is. That is, do not reduce it first to $\gamma^i{}_{jk} \, y^j y^k$.

(a) Show that in spite of the fact that $\Gamma^i{}_{jk}$ generically depends on y, one always has

$$\boxed{(\Gamma^i{}_{jk} \, y^j \, y^k)_{y^p} = 2 \, \Gamma^i{}_{pk} \, y^k} \ .$$

You will need to use the fact that $(\Gamma^i{}_{jk})_{y^l} = (-1/F) \, P_j{}^i{}_{kl}$, namely (3.3.3). You will also need the Bianchi identity (3.4.9), which says that $\ell^j \, P_j{}^i{}_{kl} = -\dot{A}^i{}_{kl}$.

(b) In the same spirit, deduce the statement

$$\boxed{(\Gamma^i{}_{jk} \, y^j \, y^k)_{y^p y^q} = 2 \, (\dot{A}^i{}_{pq} + \Gamma^i{}_{pq})} \ .$$

Note that the right-hand side is twice the Berwald connection. As in part (a), you will need to use (3.3.3). But the symmetry property

$$P_j{}^i{}_{kl} \;=\; P_k{}^i{}_{jl} \;,$$

namely (3.2.3), will have to be invoked before (3.4.9) is relevant again.

(c) Now suppose $\Gamma^i{}_{jk}$ has no y dependence. The computational steps in parts (a) and (b) then simplify considerably. Check that they lead directly to the statement $(\Gamma^i{}_{jk}\, y^j y^k)_{y^p y^q} = 2\Gamma^i{}_{pq}$ instead.

Exercise 10.2.2: Show that for a Berwald space, the *hh* part of the Chern curvature has the simple formula

$$R_j{}^i{}_{kl} \;=\; \frac{\partial \Gamma^i{}_{jl}}{\partial x^k} \;-\; \frac{\partial \Gamma^i{}_{jk}}{\partial x^l} \;+\; \Gamma^i{}_{hk}\,\Gamma^h{}_{jl} \;-\; \Gamma^i{}_{hl}\,\Gamma^h{}_{jk} \;.$$

Hint: you may want to review (3.3.2). Note that the right-hand side consists entirely of quantities defined on the underlying manifold M. This formula is symbolically the same as that for the curvature tensor of a Riemannian manifold.

Exercise 10.2.3: Verify that on a Berwald space, the Chern connection $\Gamma^i{}_{jk}$ and the Berwald connection $\Gamma^i{}_{jk} + \dot{A}^i{}_{jk}$ are identical.

10.3 Examples of Berwald Spaces

At the end of §2.4, we determined the Chern connection, in natural coordinates, for Riemannian manifolds and locally Minkowski spaces. Let us review our findings:

* *Riemannian manifolds.* In natural coordinates, the Chern connection coefficients $\Gamma^i{}_{jk}$ coincide with the underlying Riemannian metric's Christoffel symbols of the second kind. In particular, they are independent of y.
* *Locally Minkowski spaces.* In certain natural coordinate charts, the Chern connection coefficients $\Gamma^i{}_{jk}$ vanish identically. Hence, in arbitrary natural coordinates, they can have at most an x dependence.

Thus all Riemannian manifolds and locally Minkowski spaces are examples of Berwald spaces.

The rest of this section concerns an example described in Rund's book [R]. This example had its genesis in a work of Berwald's. We refer to it as the **Berwald–Rund example**. It is a Finsler surface with:

- constant Cartan scalar $I = \frac{3}{\sqrt{2}}$,
- vanishing Landsberg scalar $J = 0$, but
- nonconstant Gaussian curvature K.

As we show below, this is indeed a Berwald space that is *neither* Riemannian *nor* locally Minkowskian, albeit one with a lamentable fault. Namely, its Finsler function F is only positive and strongly convex on the upper half (and with a ray excluded) of each tangent plane T_xM.

In general, if F is only smooth and strongly convex on a *proper* subset of $TM \setminus 0$, it is said to be **y-local**. For example, the interesting 3-dimensional conformally flat Berwald spaces given by Matsumoto [M9] are y-local because they do *not* have strong convexity on all of $TM \setminus 0$.

We digress to recall something. Let (M, F) be an orientable Finsler surface. Its sphere bundle SM was introduced in §4.3. When equipped with the Sasaki (type) metric, SM is a 3-dimensional Riemannian manifold. In §4.3, a globally defined orthonormal frame field $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ was constructed on the sphere bundle SM. The first two vectors are horizontal, and the third one is vertical. Given any scalar f on SM, the quantities f_1, f_2, f_3 denote, respectively, its directional derivatives along $\hat{e}_1$, $\hat{e}_2$, $\hat{e}_3$.

Lemma 10.3.1. *Given any Finsler surface (M, F) which is either y-local or y-global, the following two statements are equivalent:*

(a) *The Finsler structure F is of Berwald type.*

(b) *Its Cartan scalar I is horizontally constant; that is, $I_1 = 0 = I_2$.*

Proof. In Exercise 4.4.7, we enumerated the only two *a priori* nonvanishing components of the hv-curvature P, relative to the Berwald frame $\{e_1, e_2\}$ of the pulled-back bundle p^*TM (which sits over SM). They are P_{2111} and P_{1111}. The same exercise, together with the Bianchi identity (4.4.7), also shows that

$$\begin{aligned} P_{2111} &= - I_2 \,, \\ P_{1111} &= I\, I_2 - I_1 \,. \end{aligned}$$

Thus $P = 0$ (so that F is Berwald by Proposition 10.2.1) if and only if $I_1 = 0 = I_2$. □

Remarks:

* In the Berwald–Rund example that we present, I is equal to the constant $3/\sqrt{2}$. Thus the Finsler surface in question is a Berwald space. Its Finsler structure is non-Riemannian since $I \neq 0$. And it is not locally Minkowskian because we show that K, one of the two surviving components of R (see Exercise 4.4.7), is nonzero.
* The y-local feature that we lamented about is inevitable whenever I is a *nonzero* constant. The reason for that is Corollary 4.1.2. It says that if the Finsler function is smooth and strongly convex on any given $T_xM \setminus 0$, then I must have zero average on the indicatrix in that T_xM. Hence, for Finsler structures that are smooth and

strongly convex on $TM \setminus 0$ (that is, y-global), the only way I can stay constant is by vanishing.

* As we show in §10.6, Szabó's rigidity theorem says that every connected y-global Berwald surface is necessarily Riemannian or locally Minkowskian. Such a theorem puts the mandatory y-locality of the Berwald–Rund example in a broader context. In contradistinction to this, we show in §11.6 that for higher dimensions, there are indeed y-global Berwald spaces that are neither Riemannian nor locally Minkowskian.

We now give some details on the Berwald–Rund example. The underlying manifold is $M := \mathbb{R}^2$, with coordinates (x^1, x^2). As usual, the coordinate basis $\{\frac{\partial}{\partial x^i}\}$ induces global coordinates (y^1, y^2) in each tangent plane of M. To avoid clutter, let us adopt the following abbreviations:

$$a := x^1 \ , \quad b := x^2$$
$$p := y^1 \ , \quad q := y^2 \ , \quad r := \frac{p}{q} \ .$$

The Finsler metric he gave involves a function

$$\xi = \xi(a, b)$$

which is a *non*constant solution of the PDE

$$\xi \, \frac{\partial \xi}{\partial a} - \frac{\partial \xi}{\partial b} = 0 \ .$$

Berwald [Ber2] found (and one can verify) that the solutions ξ of this PDE are implicitly given by

$$a + \xi \, b = \psi(\xi) \ ,$$

where ψ is an arbitrary analytic function with

$$\psi'' \neq 0 \ .$$

With this background, the Finsler function in question is defined as

$$F(x,y) := q \, (\xi + r)^2 \ .$$

Note that $F \leqslant 0$ if $q \leqslant 0$. Also, in the open upper half (where $q > 0$) plane, $F = 0$ along the ray $r = -\xi$. By Theorem 1.2.2, this F can only be y-local.

Straightforward calculations give

$$\ell^1 = \frac{r}{(\xi + r)^2} \ ,$$
$$\ell^2 = \frac{1}{(\xi + r)^2} \ ,$$

and

$$\ell_1 = 2 \, (\xi + r) \ ,$$
$$\ell_2 = (\xi - r) \, (\xi + r) \ .$$

$$\begin{pmatrix} F_{pp} & F_{pq} \\ F_{qp} & F_{qq} \end{pmatrix} = \frac{2}{q} \begin{pmatrix} 1 & -r \\ -r & r^2 \end{pmatrix} .$$

$$\begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = (\xi + r)^2 \begin{pmatrix} 6 & 2[\xi - 2r] \\ 2[\xi - 2r] & 2r^2 + [\xi - r]^2 \end{pmatrix} .$$

Hence

$$g = 2(\xi + r)^6 , \quad \sqrt{g} = \sqrt{2}(\xi + r)^3 ,$$

$$\begin{pmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{pmatrix} = \frac{1}{2(\xi + r)^4} \begin{pmatrix} 2r^2 + [\xi - r]^2 & -2[\xi - 2r] \\ -2[\xi - 2r] & 6 \end{pmatrix} .$$

The Cartan tensor A_{ijk} has the following components:

$$\begin{aligned} A_{111} &= 6(\xi + r)^3 \\ A_{112} &= (-r)\,6(\xi + r)^3 \\ A_{122} &= (r^2)\,6(\xi + r)^3 \\ A_{222} &= (-r^3)\,6(\xi + r)^3 . \end{aligned}$$

So far, all the formulas are in natural coordinates. But the scalars of interest, namely, I, J, K, are more readily obtained through orthonormal frames. Specifically:

* For I we only need the Berwald frame $\{e_1, e_2\}$ for the pulled-back tangent bundle p^*TM, and the formula $I = A(e_1, e_1, e_1)$.

For the remaining invariants, we need to work with $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$. This is a vector basis (orthonormal with respect to a Sasaki type metric) on the sphere bundle SM, over which the pulled-back tangent bundle sits.

* In that context, $J = I_2$.
* And K can be obtained by exterior differentiation of ω^3, the dual 1-form of $\hat{e}_3$.

This formalism was detailed in §4.3.

We find that the e_1 [see (4.3.1)] in the Berwald frame is given by

$$e_1 = \frac{1}{\sqrt{2}(\xi + r)^2} \left[(\xi - r)\frac{\partial}{\partial a} - 2\frac{\partial}{\partial b} \right] .$$

Using this, a straightforward computation gives

$$I = \frac{3}{\sqrt{2}}$$

as claimed. In particular,

$$J = 0 .$$

We now turn to the Gaussian curvature K. To this end, let us first prolong our computations in natural coordinates. The fundamental tensor's formal Christoffel symbols of the first kind are:

$$\begin{aligned}
\gamma_{111} &= \frac{2\,(\xi + r)}{\psi'(\xi) - b}\,[3] \\
\gamma_{211} &= \frac{2\,(\xi + r)}{\psi'(\xi) - b}\,[-3r] \\
\gamma_{112} &= \frac{2\,(\xi + r)}{\psi'(\xi) - b}\,[3\xi] \\
\gamma_{212} &= \frac{2\,(\xi + r)}{\psi'(\xi) - b}\,[\xi^2 - \xi r + r^2] \\
\gamma_{122} &= \frac{2\,(\xi + r)}{\psi'(\xi) - b}\,[2\xi^2 - 2\xi r - r^2] \\
\gamma_{222} &= \frac{2\,(\xi + r)}{\psi'(\xi) - b}\,[\xi(\xi^2 - \xi r + r^2)]\,.
\end{aligned}$$

The quantity $\psi'(\xi) - b$ is never zero. Indeed, differentiating $a + b\xi = \psi(\xi)$ with respect to a and rearranging, we get $1 = [\psi'(\xi) - b]\,\xi_a$.

In the following, a subscript n on γ_{ijk} signifies contraction with ℓ. Thus,

$$\begin{aligned}
\gamma_{ijn} &:= \gamma_{ijk}\,\ell^k\,, \\
\gamma_{inn} &:= \gamma_{ijk}\,\ell^j\,\ell^k\,.
\end{aligned}$$

We have

$$\begin{aligned}
\gamma_{11n} &= \frac{2}{\psi'(\xi) - b}\,[3] \\
\gamma_{12n} &= \frac{2}{\psi'(\xi) - b}\,[2\xi - r] \\
\gamma_{21n} &= \frac{2}{\psi'(\xi) - b}\,[\xi - 2r] \\
\gamma_{22n} &= \frac{2}{\psi'(\xi) - b}\,[\xi^2 - \xi r + r^2]\,.
\end{aligned}$$

Hence

$$\begin{aligned}
\gamma_{1nn} &= \frac{2}{(\psi'(\xi) - b)\,(\xi + r)}\,[2] \\
\gamma_{2nn} &= \frac{2}{(\psi'(\xi) - b)\,(\xi + r)}\,[\xi - r]\,.
\end{aligned}$$

The formula for the nonlinear connection is

$$\frac{N_{ij}}{F} = \gamma_{ijn} - A_{ij}{}^{k}\,\gamma_{knn}\,.$$

It requires us to first calculate

$$\begin{aligned} A_{11}{}^1 &= 3\,(\xi - r) \\ A_{11}{}^2 &= -\,6 \\ A_{12}{}^1 &= 3\,(\xi - r)\,(-r) \\ A_{12}{}^2 &= -\,6\,(-\,r) \\ A_{22}{}^1 &= 3\,(\xi - r)\,(r^2) \\ A_{22}{}^2 &= -\,6\,(r^2)\,. \end{aligned}$$

Using these, one finds that

$$\frac{N_{ij}}{F} = \gamma_{ijn}\ .$$

Consequently,

$$\begin{aligned} \frac{N^1{}_1}{F} &= \frac{1}{(\psi'(\xi) - b)}\,\frac{1}{(\xi + r)^2} \\ \frac{N^1{}_2}{F} &= \frac{1}{(\psi'(\xi) - b)}\,\frac{r}{(\xi + r)^2} \\ \frac{N^2{}_1}{F} &= 0 \\ \frac{N^2{}_2}{F} &= \frac{1}{(\psi'(\xi) - b)}\,\frac{2}{(\xi + r)^2}\ . \end{aligned}$$

The above computations lead to

$$\begin{aligned} \frac{\delta p}{F} &= \frac{1}{q\,(\xi + r)^2}\,dp + \frac{1}{(\psi'(\xi) - b)}\,\frac{1}{(\xi + r)^2}\,[da + r\,db] \\ \frac{\delta q}{F} &= \frac{1}{q\,(\xi + r)^2}\,dq + \frac{1}{(\psi'(\xi) - b)}\,\frac{1}{(\xi + r)^2}\,[2\,db]\ . \end{aligned}$$

At the beginning of this section, we recalled (from §4.3) the global orthonormal frame field $\{\hat{e}_1, \hat{e}_2, \hat{e}_3\}$ for the Sasaki (type) metric on SM. The corresponding coframe $\{\omega^1, \omega^2, \omega^3\}$ has also been given explicit formulas in that section. In particular,

$$\omega^3 := \sqrt{g}\,\left(\ell^2\,\frac{\delta y^1}{F} - \ell^1\,\frac{\delta y^2}{F}\right)\ .$$

For the example at hand, and with $p := y^1$, $q := y^2$, we find that

$$\omega^3 = \frac{\sqrt{2}}{(\xi + r)}\,\left[\frac{1}{q}\,(dp - r\,dq) + \frac{1}{\psi'(\xi) - b}\,(da - r\,db)\right]\ .$$

In general, (4.4.5) says that $d\omega^3 = K\,\omega^1 \wedge \omega^2 - J\,\omega^1 \wedge \omega^3$. But $J = I_2$ is zero here because our I is constant. Thus

$$d\omega^3 = K\,\omega^1 \wedge \omega^2 = K\,\sqrt{g}\,da \wedge db\ .$$

A somewhat tedious calculation indeed gives

$$d\omega^3 = \frac{\psi''(\xi)}{(\psi'(\xi) - b)^3} \sqrt{2}\, da \wedge db \,.$$

This and our formula for $\sqrt{g}$ then imply that

$$K = \frac{\psi''(\xi)}{(\xi + r)^3 \, (\psi'(\xi) - b)^3} \,.$$

By hypothesis, $\psi''(\xi) \neq 0$, hence K is nonzero.

Exercises

Exercise 10.3.1: Denote partial differentiation by subscripts.

(a) Verify that all functions ξ defined implicitly by the relation $a + b\xi = \psi(\xi)$ are indeed solutions of $\xi\, \xi_a - \xi_b = 0$.

(b) Prove that every solution of the above PDE can be characterized that way.

(c) Check that the functions $\psi'(\xi) - b$ and ξ_a are never zero. Hint: apply $\frac{\partial}{\partial a}$ to the implicit relation.

(d) Explain why the zeroes of ξ and ξ_b must coincide. Hint: apply $\frac{\partial}{\partial b}$ to that relation.

Exercise 10.3.2: For the Berwald–Rund example, derive the stated formula of the Gaussian curvature K.

10.4 A Fact about Flat Linear Connections

In this section, let us describe a result about flat linear connections D on finite-dimensional manifolds M. It so happens that for the present purpose, the formalism based on differential forms is less intuitive than that which uses vector fields and covariant differentiation. Thus we adopt the latter approach. In that case, the torsion and curvature of D are, respectively, given by the operators

$$T(X,Y) := D_X Y - D_Y X - [X,Y] \,,$$
$$R(X,Y)Z := \Big(D_X D_Y - D_Y D_X - D_{[X,Y]} \Big) Z \,.$$

Proposition 10.4.1. *Let D be a torsion-free linear connection on a finite dimensional manifold M. Let p be any point in M. If the curvature of D vanishes in a neighborhood of p, then there is a local coordinate system (x^i) about p in which all the connection coefficients ${\Gamma^i}_{jk}$ are zero.*

Proof. We sketch an argument from Volume II of Spivak's book [Sp2]. Specifically, we focus on his "Test Case III".

Let (u^i) be any local coordinate system defined on some open set U that contains our point p. Without loss of generality, we may assume that

(a) the coordinates of p are all zero,

(b) the open set U is path-connected, and

(c) the curvature of D vanishes on U.

In order to avoid cumbersome prose, let us not distinguish the points in U from their coordinate representation (u^i).

Our first step is to demonstrate that any vector $X(p) \in T_pM$ can be extended to a covariantly constant vector field X (that is, $DX = 0$) on U. Let $\gamma(u^1) := (u^1, 0, \ldots, 0)$ be the u^1-coordinate curve that passes through p. Parallel translate $X(p)$ along γ. At each point along γ, we now have a vector $X(u^1, 0, \ldots, 0)$ and a u^2-coordinate curve σ. Parallel translate this vector along σ. We have thus extended $X(p)$ to a vector field $X(u^1, u^2, 0, \ldots, 0)$ on the u^1u^2-coordinate surface Σ that passes through p.

Since the curvature of D vanishes on U and hence on Σ, we have

$$(*) \qquad D_{\frac{\partial}{\partial u^1}} D_{\frac{\partial}{\partial u^2}} X - D_{\frac{\partial}{\partial u^2}} D_{\frac{\partial}{\partial u^1}} X = 0 \quad \text{on } \Sigma .$$

But by construction, $D_{\frac{\partial}{\partial u^2}} X = 0$ on Σ. So (*) reduces to a statement which says that the vector field $D_{\frac{\partial}{\partial u^1}} X$ is parallel along the u^2-coordinate curve σ passing through the point $\gamma(u^1)$. By construction, the said vector field vanishes at $\gamma(u^1)$. Therefore the linearity of parallel transport (which comes from the linearity of D) implies that it must be identically zero along σ. This holds for every u^2-coordinate curve that traverses γ; thus $D_{\frac{\partial}{\partial u^1}} X = 0$ on Σ. This process of extension can be continued until one obtains a vector field X on U satisfying $D_{\frac{\partial}{\partial u^i}} X = 0$ for all i. And this last property is equivalent to

$$D_Y X = 0 \quad \text{for all vector fields } Y .$$

Next, take any basis $\{X_i(p)\}$ of T_pM and, through the above procedure, extend it to a collection of covariantly constant vector fields $\{X_i\}$ on U. It is a consequence of the linearity of parallel transport by D that these vector fields form a basis at every tangent space in U. Also, the torsion-freeness of D implies that

$$[X_i , X_j] = 0 .$$

Using this information, one can construct local coordinates (x^i) on U such that

$$X_i = \frac{\partial}{\partial x^i} .$$

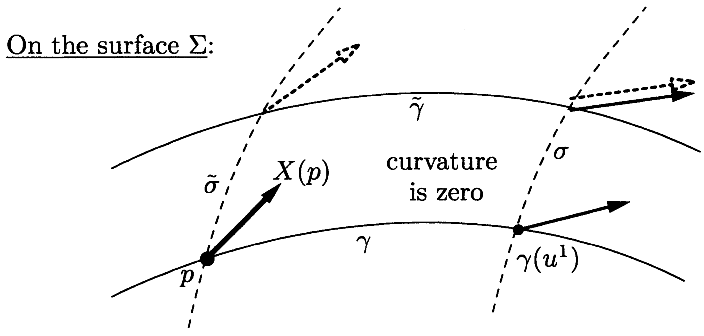

Figure 10.1

Start with a single vector $X(p)$. We parallel translate it along the u^1-coordinate curve γ through p. At each point $\gamma(u^1)$ along γ, we now have a vector X. Parallel translate that along the u^2-coordinate curve σ through $\gamma(u^1)$. This procedure generates a vector field X (solid arrows) defined on the u^1u^2-coordinate surface Σ through p. As we explained in the proof of Proposition 10.4.1, the vanishing of the curvature implies that this X is covariantly constant on the entire coordinate surface Σ. Now, we could have first parallel translated $X(p)$ along the u^2-coordinate curve $\tilde{\sigma}$ through p, and then along the u^1-coordinate curve $\tilde{\gamma}$. The vanishing of the curvature would again imply that the resulting vector field $\tilde{X}$ (dotted arrows) is covariantly constant on Σ. Actually, $\tilde{X}$ is identical to X! Note that their difference is covariantly constant on Σ, and is zero at p. A moment's thought shows that it must remain zero everywhere.

The statement $D_{X_k} X_j = 0$ then implies that

$$0 \;=\; D_{\frac{\partial}{\partial x^k}} \left(\frac{\partial}{\partial x^j} \right) \;=\; \Gamma^i{}_{jk} \, \frac{\partial}{\partial x^i} \; .$$

Therefore $\Gamma^i{}_{jk} = 0$. □

Exercises

Exercise 10.4.1:

(a) Continue the process of extension mentioned in the proof, until one obtains a vector field on U satisfying $D_{\frac{\partial}{\partial u^i}} X = 0$ for all i.

(b) Explain why the above property is equivalent to $D_Y X = 0$ for all vector fields Y.

Exercise 10.4.2:

(a) Show that it is a consequence of the linearity of parallel transport by D that the vector fields $\{X_i\}$ form a basis at every tangent space in U.

(b) Deduce from the torsion-freeness of D that $[X_i , X_j] = 0$.

(c) Explain why there must exist local coordinates (x^i) on U such that $X_i = \frac{\partial}{\partial x^i}$.

10.5 Characterizing Locally Minkowski Spaces by Curvature

Let us review our nomenclature. A Finsler manifold (M, F) is called a **locally Minkowski space** if there exist certain privileged local coordinates (x^i) on M which, together with coordinates on TM induced by $y = y^i \frac{\partial}{\partial x^i}$, render F dependent only on y and not on x. On the other hand, a **Minkowski space** consists of a vector space V and a Minkowski norm F (see §1.2), the latter inducing a Finsler structure on V by translation (see §1.3). A very little amount of definition-chasing will show that Minkowski spaces are always locally Minkowskian.

Proposition 10.5.1. *Let (M, F) be a Finsler manifold. Let $R_j{}^i{}_{kl}$ and $P_j{}^i{}_{kl}$ be, respectively, the hh- and hv-curvatures of the Chern connection. Then the following three conditions are equivalent:*

(a) *(M, F) is locally Minkowskian.*

(b) *$R_j{}^i{}_{kl} = 0$ and $P_j{}^i{}_{kl} = 0$.*

(c) *$R_{ik} := \ell^j R_{jikl} \ell^l = 0$ and $P_{jikl} = 0$.*

Remark: In view of Proposition 10.2.1 and §3.10, criterion (c) describes Berwald spaces with zero flag curvature.

Proof. Let us make some preliminary observations.

* Our discussions at the end of §2.4 and §3.3 show that (a) $\Rightarrow$ (b).
* It is apparent that (b) $\Rightarrow$ (c).

Thus it remains to check that (c) $\Rightarrow$ (a).

Suppose (c) holds. By Proposition 10.2.1, the vanishing of P implies that in natural coordinates, the Chern connection coefficients $\Gamma^i{}_{jk}$ have no y-dependence. As described in §10.1, they then define a linear connection D directly on the underlying manifold M. These $\Gamma^i{}_{jk}$ are the components of $D_{\frac{\partial}{\partial x^k}} \frac{\partial}{\partial x^j}$ with respect to $\{\frac{\partial}{\partial x^i}\}$.

Since P vanishes, so does $\dot{A}$. This follows from the first Bianchi identity (3.4.9). Using $\dot{A} = 0$ and $R^i{}_k = 0$ in the constitutive relation (3.5.6), which in turn comes from a second Bianchi identity, we see that

$$R_j{}^i{}_{kl} = 0 .$$

For the left-hand side of this equation, we substitute (3.3.2) for R, but with one important simplification. Namely, we use $\frac{\partial}{\partial x}$ instead of $\frac{\delta}{\delta x}$ because the $\Gamma^i{}_{jk}$ here have no y-dependence. The end result is

$$\frac{\partial \Gamma^i{}_{jl}}{\partial x^k} - \frac{\partial \Gamma^i{}_{jk}}{\partial x^l} + \Gamma^i{}_{hk}\,\Gamma^h{}_{jl} - \Gamma^i{}_{hl}\,\Gamma^h{}_{jk} = 0 .$$

It says that the torsion-free connection D on M is flat.

Proposition 10.4.1 now tells us that, by changing to a different coordinate system (on M) if necessary, we may assume without any loss of generality that the $\Gamma^i{}_{jk}$ are zero. Hence $N^i{}_k = 0$ by part (a) of Exercise 2.4.6. These then give

$$\frac{\partial g_{ij}}{\partial x^k} = 0$$

by (2.4.10). So the g_{ij} have no x-dependence and, by (1.2.5), neither does the Finsler function F. We have finally obtained criterion (a). ☐

The above theorem shows that the definition of locally Minkowski spaces translates into a curvature criterion. Typically, there are topological obstructions to the fulfillment of any curvature condition. Here, one such obstruction is the vanishing of the Euler characteristic of the underlying manifold M. A derivation of this fact is through the use of a generalized Gauss–Bonnet Theorem; see [BC2].

Exercises

Exercise 10.5.1: Verify that Minkowski spaces are locally Minkowskian.

Exercise 10.5.2: Give all the details in our proof of Proposition 10.5.1.

10.6 Szabó's Rigidity Theorem for Berwald Surfaces

10.6 A. The Theorem and Its Proof

The Berwald–Rund example (of a Berwald surface) was discussed in §10.3. It has the merit of being neither Riemannian nor locally Minkowskian. Unfortunately, it also has a fairly serious fault. Namely, its Finsler structure F is only strongly convex on *part* of each *punctured* tangent plane.

One might wonder if there is a Berwald *surface* in which the Finsler structure is smooth and strongly convex on $TM \setminus 0$. Szabó [Sz] addressed this question and found a surprisingly rigid picture. Namely, any such surface must necessarily be of the Riemannian or locally Minkowskian variety.

Given this fact, if we want an example of a Berwald space (M, F) that is neither Riemannian nor locally Minkowskian, *and* such that F is smooth and strongly convex on all of $TM \setminus 0$, we must search for it in dimension three or higher. Our chapter on Randers metrics shows that this goal can be realized by searching instead for Riemannian manifolds which admit globally defined parallel vector fields. This is so thanks to the works of Matsumoto [M4]; Hashiguchi and Ichijyō [HI]; Shibata, Shimada, Azuma, and Yasuda [SSAY]; and Kikuchi [Ki]. See §11.5 and then §11.6.

We now give a proof of Szabó's rigidity theorem. To this end, let us first establish a fact about the Gaussian curvature K of Landsberg surfaces. These are the ones for which the Landsberg scalar $J = I_2$ vanishes. We considered them in §4.6. The Landsberg family includes Riemannian surfaces, locally Minkowskian surfaces, and more generally, Berwald surfaces. The reason is that Berwald spaces have $P = 0$. This holds in natural coordinates and, since P is a tensor, remains so in the Berwald frame $\{e_1, e_2\}$. By part (d) of Exercise 4.4.7, J must therefore vanish.

Consider any fixed indicatrix $\mathrm{S}_x M$, which is a simple closed convex curve in our formalism.

- As we discussed in §4.1, this curve has Riemannian arc length L. Give it a unit speed parametrization $y(t)$, $0 \leqslant t < L$ such that the velocity field is precisely the $\hat{e}_3$ defined in (4.3.9). This is done in order to effect

$$f_3[\,y(t)\,] \;=\; \frac{d}{dt} f[\,y(t)\,] \;=:\; \dot{f}$$

  for any function f. Part (c) of Exercise 4.3.2 is of interest here, though not directly relevant.
- To avoid clutter, let us also adopt the abbreviations

$$I(t) \;:=\; I[\,y(t)\,]\,, \qquad K(t) \;:=\; K[\,y(t)\,]\,.$$

Proposition 10.6.1. *Let (M, F) be a Landsberg surface for which the Finsler structure F is smooth on $TM \setminus 0$. Then the value of K at any point $y(t)$ of the indicatrix $S_x M$ is determined by the Cartan scalar I according to the following formula*

$$K(t) \;=\; K(0)\; e^{\,[\int_0^t I(\tau)\, d\tau\,]}\,. \tag{10.6.1}$$

Remarks:

- * As a consistency check, set $t = L$ in (10.6.1). Since $K(L) = K(0)$, we get $\int_0^L I(t)\, dt = 0$ whenever $K(0) \neq 0$. This is consistent with (4.1.13), which says that the average of I over any indicatrix is unconditionally zero.
- * Formula (10.6.1) says that for a Landsberg surface, the way in which K varies on any given indicatrix is completely controlled by the

Cartan scalar in an explicit manner. In particular, on any S_xM, K is either nowhere zero or vanishes identically.

Proof. Since $J = 0$ on a Landsberg surface, (4.4.8) reduces to

$$K_3 + I K = 0 .$$

Restricting this to S_xM gives $\dot{K}(t) + I(t)\, K(t) = 0$. Our hypothesis on F implies that $I(t)$ is continuous; hence the solution is as claimed. □

As an application of Proposition 10.6.1, we now give a short proof of Szabó's rigidity theorem about Berwald surfaces. Recall from Lemma 10.3.1 that a Berwald surface is one for which $I_1 = 0 = I_2$. That is, its Cartan scalar I does not vary in the horizontal directions. Since $J = I_2$, we see that Berwald surfaces are of Landsberg type.

Theorem 10.6.2 (Szabó) [Sz]. *Let (M, F) be a connected Berwald surface for which the Finsler structure F is smooth and strongly convex on all of $TM \setminus 0$.*

- *If K vanishes identically, then F is locally Minkowskian everywhere.*
- *If K is not identically zero, then F is Riemannian everywhere.*

Proof. According to Exercise 4.4.7, K vanishing identically implies the same of the hh-Chern curvature $R_j{}^i{}_{kl}$. And being Berwald is synonymous with $P_j{}^i{}_{kl} = 0$. Proposition 10.5.1 now tells us that (M, F) must be locally Minkowskian. This takes care of the first scenario.

Next, suppose K is not identically zero on SM. Then there exists an indicatrix S_pM on which K is nonzero at some point, and hence at all points (of that S_pM) on account of Proposition 10.6.1. Applying a Ricci identity (from Exercise 4.4.4) to the Cartan scalar I, we get

$$I_{12} - I_{21} = - K\, I_3 .$$

Since our Finsler structure is of Berwald type, namely, $I_1 = I_2 = 0$, the left-hand side of this Ricci identity vanishes. The resulting equation, when restricted to S_pM, reads

$$K\, \dot{I} = 0 .$$

But K is nowhere vanishing on the connected set S_pM, so $\dot{I} = 0$. This means that I must be constant along the indicatrix S_pM. That constant is necessarily zero because I has zero average (see Corollary 4.1.2) on the indicatrix S_pM.

Since I vanishes on S_pM, $F(p, \tilde{y})$ must have the form

$$F(p, \tilde{y}) = \sqrt{g_{ij}(p)\ \tilde{y}^i \tilde{y}^j}$$

for all $\tilde{y}$ in T_pM. Now, on a connected Berwald surface, every Minkowskian plane (T_xM, F_x) is *linearly* isometric to (T_pM, F_p). This is the essence of Proposition 10.1.1. Thus $F(x, y)$ is equal to $F(p, \tilde{y})$, where $\tilde{y}$ is related to y by a linear transformation that depends only on x and a path from x to p. It is then straightforward to check that $F(x, \cdot)$ inherits from $F(p, \cdot)$ the form

$$F(x,y) \;=\; \sqrt{g_{ij}(x)\; y^i y^j}\;.$$

In other words, the Finsler structure is Riemannian everywhere. □

10.6 B. Distinguishing between y-local and y-global

We hasten to point out that the above theorem applies only to Finsler structures which are y-global. That is, those which are smooth and strongly convex on $TM \setminus 0$. It is perfectly consistent with the *local* treatment of Finsler surfaces by Matsumoto in the references [M2] and [AIM].

* In [M2] and [AIM], one reads that Berwald surfaces which are not locally Minkowskian are characterized by I being constant. In these references, the Finsler functions are typically only smooth and strongly convex on some open cone in each tangent plane. Thus the indicatrices are not necessarily closed curves, and our Corollary 4.1.2 *cannot* be recklessly applied to conclude that the I in question is zero. As a result, the y-local Finsler structures in question are spared the fate of having to be Riemannian.
* The same references also give Berwald's classification of y-local Finsler surfaces whose I is a function of the position x alone.

Exercises

In the two exercises below, we consider the **Okubo metrics** [Oku]

$$F(x,y) \;=\; \left\{ [\,\lambda\,(y^1)^2 \;+\; \mu\,(y^2)^2\,]\,[\,\mu\,(y^1)^2 \;+\; \lambda\,(y^2)^2\,] \right\}^{1/4} ,$$

where λ and μ are positive functions of position $x = (x^1, x^2) \in M := \mathbb{R}^2$.

Exercise 10.6.1: Prove that the ratio λ/μ has the uniform bounds

$$\frac{1}{3+2\sqrt{2}} \;=\; 3-2\sqrt{2} \;<\; \frac{\lambda}{\mu} \;<\; 3+2\sqrt{2}$$

if and only if F is smooth and strongly convex on $T\mathbb{R}^2 \setminus 0$.

Exercise 10.6.2: Can λ and μ be chosen to produce a *y-global* Berwald structure (on any connected open domain of $\mathbb{R}^2$) that is neither Riemannian nor locally Minkowskian? Make a theoretical decision, then give some direct computational verifications. For the latter, recall the quantities G^i, $i = 1, 2$

given by (3.8.1). Formulas for their machine computation are given in §12.5. According to (3.8.4), taking three y derivatives of G^i yields a multiple of the Berwald hv-curvature ${}^bP$. As remarked after the proof of Proposition 10.2.1, the vanishing of ${}^bP$ also characterizes Berwald spaces.

References

[AIM] P. L. Antonelli, R. S. Ingarden, and M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology*, FTPH 58, Kluwer Academic Publishers, 1993.

[BC2] D. Bao and S. S. Chern, *A note on the Gauss–Bonnet theorem for Finsler spaces*, Ann. Math. **143** (1996), 233–252.

[BCS1] D. Bao, S. S. Chern, and Z. Shen, *On the Gauss–Bonnet integrand for 4-dimensional Landsberg spaces*, Cont. Math. **196** (1996), 15–25.

[Ber2] L. Berwald, *Two-dimensional Finsler spaces with rectilinear extremals*, Ann. Math. **42** (1941), 84–112.

[Ch3] S. S. Chern, *A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds*, Ann. Math. **45(4)** (1944), 747–752.

[Ch4] S. S. Chern, *On the curvatura integra in a Riemannian manifold*, Ann. Math. **46(4)** (1945), 674–684.

[HI] M. Hashiguchi and Y. Ichijyō, *On some special (α, β)-metrics*, Rep. Fac. Sci. Kagoshima Univ. **8** (1975), 39–46.

[I] Y. Ichijyō, *Finsler manifolds modeled on a Minkowski space*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **16-3** (1976), 639–652.

[Ki] S. Kikuchi, *On the condition that a space with (α, β)-metric be locally Minkowskian*, Tensor, N.S. **33** (1979), 242–246.

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

[M4] M. Matsumoto, *On Finsler spaces with Randers' metric and special forms of important tensors*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **14** (1974), 477–498.

[M9] M. Matsumoto, *Theory of Finsler spaces with m-th root metric II*, Publ. Math. Debr. **49** (1996), 135–155.

[Oku] K. Okubo, Lecture at the Symposium on Finsler Geometry, 1977, unpublished (communicated to us by M. Matsumoto).

[R] H. Rund, *The Differential Geometry of Finsler Spaces*, Springer-Verlag, 1959.

[SSAY] C. Shibata, H. Shimada, M. Azuma, and H. Yasuda, *On Finsler spaces with Randers' metric*, Tensor, N.S. **31** (1977), 219–226.

[Sp2] M. Spivak, *Differential Geometry*, vol. II, Publish or Perish, 1970.

[Sz] Z. Szabó, *Positive definite Berwald spaces (structure theorems on Berwald spaces)*, Tensor, N.S. **35** (1981), 25–39.

Chapter 11

Randers Spaces and an Elegant Theorem

11.0 The Importance of Randers Spaces

In 1941, G. Randers [Ra] studied a very interesting type of Finsler structures. These are called **Randers metrics**, and we first encountered them in §1.3. Randers metrics are important for six reasons.

- They occur naturally in physical applications, most notably in electron optics. According to Ingarden's account in [AIM], the Lagrangian of relativistic electrons gives rise to the following Finsler function $F(x, y)$ of Randers type:

  $$\sqrt{\Phi(x) + \frac{1}{4}\Phi^2(x)}\ \sqrt{(y^1)^2 + (y^2)^2 + (y^3)^2} + \mathcal{A}_i(x)\, y^i .$$

  Here, Φ and $\mathcal{A}$ are respectively normalized versions of the electric ("scalar") and magnetic ("vector") potentials. The normalization involves the physical constants of the theory.

- They provide a rich source of y-global Berwald spaces, particularly those that are neither Riemannian nor locally Minkowskian. Recall (from §10.6) that in dimension two, every *y-global* (smooth and strongly convex on $TM \setminus 0$) Berwald structure is necessarily Riemannian or locally Minkowskian. This is Szabó's theorem. Thus, to locate the examples described, we must look in dimensions three or higher. As we show in §11.5 and §11.6, an elegant result demonstrates that Randers metrics are most relevant to our quest.
- Thanks to the Yasuda–Shimada theorem [YS], Randers metrics provide intriguing examples of non-Riemannian Finsler spaces with constant negative flag curvature. See §12.6 for the construction and the detailed analysis of one such example—the Finslerian Poincaré disc. Also, we explain carefully in §12.6 why the said example is perfectly consistent with Akbar-Zadeh's rigidity theorem (Theorem 12.4.1). Since we do not prove the Yasuda–Shimada theorem in this book, all assertions about that example are verified explicitly, either by hand or by symbolic computation on the computer.
- More generally, Randers spaces represent a medium in which Riemannian geometry interfaces with Finsler geometry proper. Studying them induces a mind set that opens up many new possibilities. A bewildering plethora of Finsler spaces can be constructed with Riemannian metrics satisfying certain Ricci curvature criteria, together with 1-forms of geometrical or topological significance. For instance, using parallel 1-forms gives Berwald spaces (see §11.5, §11.6), while using closed 1-forms gives Douglas spaces (see Exercise 11.5.4 and especially Exercise 11.3.4).
- Many new geometrical invariants are first explicitly computed for Randers spaces. This is the case for the geometric ratio introduced by Bao–Lackey [BL1], and for the S-curvature given by Shen [Sh]. Also, Randers spaces (with drift term strictly less than 1) are in some sense invariant under the $\mathcal{L}$-duality proposed by Miron. This involves the Legendre transformation that we study in §14.8. See Hrimiuc–Shimada ([HS1], [HS2]) and Sabau–Shimada [SaS] for detailed treatments.
- Finally, from an axiomatic standpoint, Randers spaces form a self-contained category. We say this because of the following. Every submanifold of a Randers space is itself a Randers space. This is explicitly illustrated in Exercise 11.1.4; see also Exercise 11.1.5. Furthermore, any properly weighted Cartesian product of two Randers spaces is again a Randers space. This is made precise in Exercise 11.1.6. Loosely speaking, the alleged self-containment comes about because Riemannian metrics and 1-forms, the data that constitute Randers spaces, are both well behaved under pull-backs and direct sums.

11.1 Randers Spaces, Positivity, and Strong Convexity

The given data of a Randers space consist of:

- a Riemannian metric $\tilde{a} := \tilde{a}_{ij}\, dx^i \otimes dx^j$ on a smooth n-dimensional manifold M, and
- a 1-form $\tilde{b} := \tilde{b}_i\, dx^i$ on M.

Together they define a Finsler structure F in a deceptively simple way:

$$\boxed{F(x,y) \;:=\; \alpha(x,y) \;+\; \beta(x,y)}\,,$$

where

$$\boxed{\begin{aligned} \alpha(x,y) \;&:=\; \sqrt{\tilde{a}_{ij\,(x)}\; y^i\, y^j} \\ \beta(x,y) \;&:=\; \tilde{b}_{i\,(x)}\; y^i \end{aligned}}\,.$$

For some explicit mathematical examples, see §1.3 and §11.6.

By inspection, one sees that a Randers metric F is absolutely homogeneous of degree 1 [that is, $F(x, \lambda y) = |\lambda|\, F(x,y)$ for all real λ] if and only if the 1-form $\tilde{b}$ vanishes identically. This in turn is equivalent to F being Riemannian. Therefore, by excluding the case $\tilde{b} \equiv 0$, the remaining Randers metrics are non-Riemannian and only positively homogeneous in y.

As in §1.3, the convention described below solves a good amount of bookkeeping problems:

> **The indices on certain objects are lowered and raised by $(\tilde{a}_{ij})$ and its inverse matrix $(\tilde{a}^{ij})$. Such objects are decorated with a tilde.**

Since $\beta(x,y)$ is linear in y, it cannot possibly have a fixed sign. Exercise 11.1.1 explains how the size of $\tilde{b}$ should be controlled, in order for F to be positive on $TM \setminus 0$. Namely, the said positivity holds if and only if

$$\boxed{\|\tilde{b}\| \;:=\; \sqrt{\tilde{b}_i\, \tilde{b}^i} \;<\; 1}\,,$$

where

$$\tilde{b}^i \;:=\; \tilde{a}^{ij}\, \tilde{b}_j\,.$$

Let

$$\tilde{\ell}_i \;:=\; \alpha_{y^i} \;=\; \frac{\tilde{a}_{ij}\; y^j}{\alpha}\,. \tag{11.1.1}$$

One can check that

$$\boxed{\ell_i \;:=\; F_{y^i} \;=\; \tilde{\ell}_i \;+\; \tilde{b}_i}\,. \tag{11.1.2}$$

The fundamental tensor can then be expressed as

$$g_{ij} = \frac{F}{\alpha}\,\tilde{a}_{ij} - \frac{\beta}{\alpha}\,\tilde{\ell}_i\,\tilde{\ell}_j + \tilde{\ell}_i\,\tilde{b}_j + \tilde{\ell}_j\,\tilde{b}_i + \tilde{b}_i\,\tilde{b}_j\ . \tag{11.1.3}$$

Equivalently,

$$\boxed{g_{ij} = \frac{F}{\alpha}\,(\tilde{a}_{ij} - \tilde{\ell}_i\,\tilde{\ell}_j) + \ell_i\,\ell_j}\ . \tag{11.1.4}$$

The issue of strong convexity needs to be addressed. This pertains to the positive-definiteness of the fundamental tensor g_{ij}. It turns out that the criterion $\|\tilde{b}\| < 1$, which guarantees the positivity of F, also ensures strong convexity! The crux of the argument involves the following computational fact:

$$\det(g_{ij}) = \left(\frac{F}{\alpha}\right)^{n+1} \det(\tilde{a}_{ij})\ . \tag{*}$$

Its derivation can be found in [M2], albeit in the more general context of $(\alpha\,,\,\beta)$ metrics. A self-contained exposition is given in §11.2.

Anyway, here's how we can establish strong convexity. Consider

$$F_\epsilon := \sqrt{\tilde{a}_{ij}\ y^i\,y^j} + \epsilon\,\tilde{b}_i\ y^i\ ,$$

where $\|\tilde{b}\| < 1$ and $0 \leqslant \epsilon \leqslant 1$. Note that:

* F_ϵ is positive on $TM \setminus 0$. Let g_ϵ abbreviate the fundamental tensor of F_ϵ. Its determinant is given by the right-hand side of (*), with F replaced by F_ϵ. So $\det(g_\epsilon)$ is positive. In particular, none of the eigenvalues of g_ϵ can vanish.
* The eigenvalues of g_ϵ depend continuously on ϵ. At $\epsilon = 0$, they are simply those of $(\tilde{a}_{ij})$, and hence are all positive. As ϵ changes from 0 to 1, none of these eigenvalues can become negative. This is because doing so would necessitate crossing zero first, which is forbidden. Thus they stay positive.

Setting ϵ equal to 1 tells us that the fundamental tensor (g_{ij}) of F is positive-definite whenever $\|\tilde{b}\| < 1$. This is the conclusion we seek. See Exercise 11.1.4 for an explicit example, to which is applied what we have just established.

We conclude now with two more formulas. A simple one for the **angular metric**

$$h_{ij} := g_{ij} - \ell_i\,\ell_j\ ,$$

and a preliminary one for the Cartan tensor

$$A_{ijk} := \frac{F}{4}\,(F^2)_{y^i y^j y^k}\ .$$

They read:

(11.1.5)
$$h_{ij} \;=\; \frac{F}{\alpha}\,(\,\tilde a_{ij} \;-\; \tilde\ell_i\,\tilde\ell_j\,) \;=\; \frac{F}{\alpha}\,\tilde h_{ij}\ .$$

(11.1.6)
$$A_{ijk} \;=\; \frac{1}{2}\left[h_{ij}\left(\tilde b_k - \frac{\beta}{\alpha}\,\tilde\ell_k\right) + h_{jk}\left(\tilde b_i - \frac{\beta}{\alpha}\,\tilde\ell_i\right) + h_{ki}\left(\tilde b_j - \frac{\beta}{\alpha}\,\tilde\ell_j\right)\right].$$

Exercises

Exercise 11.1.1: Fix $x \in M$. The positivity of the Randers metric F on $T_xM \smallsetminus 0$ means that

(**)
$$\sqrt{\tilde a_{ij}\; y^i\, y^j} \;>\; -\,\tilde b_i\; y^i \quad \text{for all} \quad y \neq 0\ .$$

(a) Suppose positivity holds. If $\tilde b \neq 0$, substitute
$$y^i \;=\; -\,\tilde b^i \;:=\; -\,\tilde a^{ij}\,\tilde b_j$$
into (**). Show that one obtains the criterion $\|\,\tilde b\,\| < 1$. Of course, this same criterion is also trivially satisfied by the case $\tilde b = 0$.

(b) Conversely, suppose $\|\,\tilde b\,\| < 1$ holds. Use a Cauchy–Schwarz inequality to show that if $y \neq 0$, then
$$|\,\tilde b_i\, y^i\,| \;<\; \sqrt{\tilde a_{ij}\; y^i\, y^j}\ .$$
Hence (**) follows.

Exercise 11.1.2: Explain why the positive-definiteness of (g_{ij}) is obvious if the pointwise Riemannian norm $\|\,\tilde b\,\|$ is *much* smaller than 1.

Exercise 11.1.3: Derive (11.1.6).

Exercise 11.1.4: Let M be the manifold $\mathbb{R}^3$, with Cartesian coordinates x^1, x^2, x^3. Denote arbitrary tangent vectors by $y^i\partial_{x^i}$. Define
$$F(x,y) \;:=\; \sqrt{(\,y^1\,)^2 \;+\; (\,y^2\,)^2 \;+\; (\,y^3\,)^2} \;+\; \lambda\, y^3\ ,$$
where $\lambda < 1$ is some positive constant. (M, F) is a Randers space with Riemannian metric
$$\tilde a \;:=\; \delta_{ij}\; dx^i \otimes dx^j \;=\; dx^1 \otimes dx^1 \;+\; dx^2 \otimes dx^2 \;+\; dx^3 \otimes dx^3$$
and 1-form
$$\tilde b \;:=\; \lambda\; dx^3\ .$$
The assumption $\lambda < 1$ is equivalent to $\|\,\tilde b\,\| < 1$.

Now consider a surface of the form $x^3 = f(\,x^1\,,\,x^2\,)$ in $M := \mathbb{R}^3$. Call that surface $\hat M$, and parametrize it by
$$(\,u\,,\,v\,) \;\mapsto\; \Big(u\,,\,v\,,\,f(\,u\,,\,v\,)\Big)\ .$$

Denote partial derivatives of f by f_u and f_v.

(a) Since u, v are local coordinates for $\hat{M}$, arbitrary tangent vectors of $\hat{M}$ have the form $p\,\partial_u + q\,\partial_v$. Check that

$$\partial_u = \partial_{x^1} + f_u\,\partial_{x^3}\,, \qquad \partial_v = \partial_{x^2} + f_v\,\partial_{x^3}\,.$$

Hence

$$v = p\,\partial_{x^1} + q\,\partial_{x^2} + (p\,f_u + q\,f_v)\,\partial_{x^3}\,.$$

(b) Show that when $\tilde{a}$ is pulled back to the surface $\hat{M}$, it yields a Riemannian metric $\hat{a}$ whose explicit formula is

$$\begin{aligned} &[\,1 + (f_u)^2\,]\; du \otimes du \\ &+ f_u\, f_v\,(\,du \otimes dv + dv \otimes du\,) \\ &+ [\,1 + (f_v)^2\,]\; dv \otimes dv\,. \end{aligned}$$

Call this Riemannian metric $\hat{a}$. Likewise, show that the pull-back of $\tilde{b}$ to $\hat{M}$ gives the 1-form

$$\hat{b} := \lambda\,(\,f_u\,du + f_v\,dv\,)\,.$$

(c) Prove that the *square* of the Riemannian length of $\hat{b}$ with respect to $\hat{a}$ is

$$\lambda^2\,\frac{(f_u)^2 + (f_v)^2}{1 + (f_u)^2 + (f_v)^2}\,.$$

Explain why we have $\|\,\hat{b}\,\| < 1$ everywhere on $\hat{M}$.

(d) On each tangent plane of $\hat{M}$, the Finsler norm F induces a function $\hat{F}$. Check that $\hat{F}$ sends the tangent vector $p\,\partial_u + q\,\partial_v$ to the number

$$\sqrt{p^2 + q^2 + (p\,f_u + q\,f_v)^2} + \lambda\,(p\,f_u + q\,f_v)\,.$$

Verify that $\hat{F}$ is positive homogeneous of degree one.

(e) Using parts (b), (c), and what we discussed in the section proper, explain why $\hat{F}$ is positive and strongly convex. Equivalently, verify that $(\hat{M}, \hat{F})$ is a Randers surface whose underlying Riemannian metric and 1-form are, respectively, $\hat{a}$ and $\hat{b}$.

Exercise 11.1.5: Let (M, F) be a Randers space, with Riemannian metric $\tilde{a}$ and 1-form $\tilde{b}$ such that $\|\,\tilde{b}\,\| < 1$. Let $\hat{M}$ be any submanifold of M.

(a) Let $\hat{a}$ denote the pull-back of $\tilde{a}$, from M to $\hat{M}$. Check that $\hat{a}$ is a Riemannian metric on the submanifold $\hat{M}$.

(b) Let $\hat{b}$ denote the pull-back of $\tilde{b}$, from M to $\hat{M}$. Prove that the Riemannian length of $\hat{b}$ with respect to $\hat{a}$ is strictly less than 1.

(c) On each tangent space of $\hat{M}$, F induces a function $\hat{F}$. Check that $\hat{F}$ is positive homogeneous of degree one. Express $\hat{F}$ in terms of $\hat{a}$

and $\hat{b}$. Then use part (b) and the discussion proper of this section to explain why $\hat{F}$ must be positive and strongly convex.

(d) Conclude that $(\hat{M}, \hat{F})$ is a Randers space with underlying Riemannian metric $\hat{a}$ and 1-form $\hat{b}$.

Thus, every submanifold of a Randers space is itself a Randers space.

Exercise 11.1.6: Let us be given two Randers spaces $(\tilde{M}, \tilde{F})$ and $(\check{M}, \check{F})$: one with Riemannian metric $\tilde{a}$ and 1-form $\tilde{b}$; the other with Riemannian metric $\check{a}$ and 1-form $\check{b}$. Recall that we must have $\| \tilde{b} \| < 1$ and $\| \check{b} \| < 1$. This is to ensure that $\tilde{F}$ and $\check{F}$ are both positive and strongly convex.

(a) Check that $\tilde{a} \oplus \check{a}$ is a Riemannian metric on the Cartesian product $\tilde{M} \times \check{M}$.

(b) Check that $(1/\sqrt{2})\,(\tilde{b} \oplus \check{b})$ is a 1-form on $\tilde{M} \times \check{M}$. Prove that its pointwise norm with respect to $\tilde{a} \oplus \check{a}$ is strictly less than 1 on $\tilde{M} \times \check{M}$.

(c) On each tangent space of $\tilde{M} \times \check{M}$, show that $\sqrt{\tilde{F}^2 \oplus \check{F}^2}$ is always positively homogeneous of degree one, positive, and strongly convex. On the other hand, explain why $\tilde{F} \oplus \check{F}$ may not be strongly convex.

(d) There is *no* canonical way of inducing a Finsler structure on the Cartesian product $\tilde{M} \times \check{M}$. Justify this statement.

(e) Show that on $\tilde{M} \times \check{M}$, if we *declare* the underlying Riemannian metric and 1-form to be

$$\tilde{a} \oplus \check{a} \quad \text{and} \quad \frac{1}{\sqrt{2}}\,(\tilde{b} \oplus \check{b}) \ ,$$

respectively, then it is indeed a Randers space. Express its Finsler function $\mathcal{F}$ explicitly in terms of $\tilde{F}$ and $\check{F}$. How does $\mathcal{F}^2$ differ from $\tilde{F}^2 \oplus \check{F}^2$?

11.2 A Matrix Result and Its Consequences

Our goal in this section is to present formulas for the determinant and the inverse of the fundamental tensor. To that end, the following standard matrix identity plays a central role.

Proposition 11.2.1. *Suppose:*

* (Q_{ij}) *is a nonsingular* $n \times n$ *complex matrix with inverse* (Q^{ij}).
* C_j, *with* $j = 1, \ldots, n$, *are* n *complex numbers.*

Let us define $C^s := Q^{sj} C_j$. *Then:*

- $\det(\, Q_{ij} + C_i C_j \,) = (\, 1 + C^s C_s \,) \det(Q_{ij})$.

- *Whenever* $1 + C^s C_s \neq 0$, *the matrix* $(Q_{ij} + C_i C_j)$ *is invertible. In that case, its inverse is*

$$\left(Q^{ij} - \frac{1}{1 + C^s C_s} C^i C^j \right) .$$

Remarks:

* In [M2], the above proposition was stated for symmetric nonsingular matrices (Q_{ij}). This symmetry hypothesis on Q is removable.
* Q certainly need not be positive-definite.

Proof. The asserted formula for the inverse is correct, because multiplying that on the right by $(Q_{jk} + C_j C_k)$ indeed gives $\delta^i{}_k$. Thus it remains to compute the determinant of $(Q_{ij} + C_i C_j)$.

Let C (resp., C^t) stand for the column (resp., row) vector whose entries are C_s, $s = 1, \ldots, n$. Then

$$(Q_{ij} + C_i C_j) = Q + CC^t .$$

Hence

$$\begin{aligned} &\det\{ Q + CC^t \} \\ &= \det\{ Q [I + (Q^{-1}C) C^t] \} \\ &= \det Q \det\{ I + (Q^{-1}C) C^t \} . \end{aligned}$$

Now we invoke a matrix fact which is perhaps better known. It says that if v and w are column vectors, then

$$(**) \qquad \det\{ I + wv^t \} = 1 + v^t w .$$

Suppose neither v nor w is identically zero, or else there is nothing to check. Given that, there are two cases to be analyzed.

* $v^t w \neq 0$: One can find $n - 1$ linearly independent complex column vectors w_α such that $v^t w_\alpha = 0$. The vectors w_α, together with w, form a basis for $\mathbb{C}^n$. Consider the matrix $I + wv^t$. It has the w_α as eigenvectors, of eigenvalue 1. The vector w is also an eigenvector, but with eigenvalue $1 + v^t w$. This observation gives (**).
* $v^t w = 0$: As agreed, we can suppose that both v and w are nonzero. Thus there is a complex column vector u such that $u^t w \neq 0$. Abbreviate $v + \epsilon u$ as v_ϵ. Note that $v_\epsilon^t w = \epsilon u^t w \neq 0$ for all nonzero ϵ. And in that case, we already know that $\det\{I + wv_\epsilon^t\} = 1 + v_\epsilon^t w$. Letting $\epsilon \longrightarrow 0$ gives (**).

Let us substitute, into the above, $Q^{-1}C$ for w and C for v. The said matrix fact then tells us that

$$\det\{ I + (Q^{-1}C) C^t \} = 1 + C^t (Q^{-1}C) = 1 + C_s C^s ,$$

which finishes the proof. □

Two applications of Proposition 11.2.1 will produce formulas for $\det(g_{ij})$ and g^{ij}. These read:

$$\boxed{\det(g_{ij}) = \left(\frac{F}{\alpha}\right)^{n+1} \det(\tilde{a}_{ij})} , \tag{11.2.1}$$

$$\boxed{g^{ij} = \frac{\alpha}{F}\,\tilde{a}^{ij} + \frac{\alpha^2}{F^2}\,\frac{\beta + \alpha\,\|\tilde{b}\|^2}{F}\,\tilde{\ell}^i\,\tilde{\ell}^j - \frac{\alpha^2}{F^2}\,(\tilde{\ell}^i\,\tilde{b}^j + \tilde{\ell}^j\,\tilde{b}^i)} . \tag{11.2.2}$$

Here:

$$(\tilde{a}^{ij}) = (\tilde{a}_{ij})^{-1} ,$$

$$\tilde{b}^i = \tilde{a}^{ij}\,\tilde{b}_j \quad \text{and} \quad \|\tilde{b}\|^2 = \tilde{b}^i\,\tilde{b}_i ,$$

$$\tilde{\ell}^i = \frac{y^i}{\alpha} .$$

Recall from (11.1.4) that

$$g_{ij} = \frac{F}{\alpha}\,(\tilde{a}_{ij} - \tilde{\ell}_i\tilde{\ell}_j) + \ell_i\ell_j .$$

However, this is not amenable for Proposition 11.2.1 because the matrix $(\tilde{a}_{ij} - \tilde{\ell}_i\tilde{\ell}_j)$ is singular. In fact, its null space is the line generated by $\tilde{\ell}^j$. Happily, the rearrangement

$$g_{ij} = \left(\frac{F}{\alpha}\,\tilde{a}_{ij} + \ell_i\,\ell_j\right) - \frac{F}{\alpha}\,\tilde{\ell}_i\,\tilde{\ell}_j$$

remedies the problem. Also, do keep in mind that we must work on the slit tangent bundle $TM \setminus 0$, or else the g_{ij} would not make sense.

In the first application, we set

$$Q_{ij} := \frac{F}{\alpha}\,\tilde{a}_{ij} ,$$

$$C_i := \ell_i .$$

Note that

$$C^s\,C_s = \frac{\alpha}{F}\,(\tilde{\ell}^s + \tilde{b}^s)(\tilde{\ell}_s + \tilde{b}_s) \geqslant 0$$

because both F and $\tilde{a}$ are positive-definite. In particular, the quantity $1 + C^sC_s$ is strictly positive. Further manipulations give

$$0 < 1 + C^sC_s = 2 + \frac{\beta + \alpha\,\|\tilde{b}\|^2}{F} . \tag{11.2.3}$$

Proposition 11.2.1 then assures us that

$$\det\left(\frac{F}{\alpha}\,\tilde{a}_{ij} \,+\, \ell_i\,\ell_j\right) = \left(\frac{F}{\alpha}\right)^n \det(\tilde{a}_{ij}) \left[2 \,+\, \frac{\beta+\alpha\,\|\tilde{b}\|^2}{F}\right] \tag{11.2.4}$$

and

$$\left(\frac{F}{\alpha}\,\tilde{a}_{ij} \,+\, \ell_i\,\ell_j\right)^{-1} = \left(\frac{\alpha}{F}\,\tilde{a}^{ij} \,-\, \frac{F}{[\,2F+\beta+\alpha\,\|\tilde{b}\|^2\,]}\,\frac{\alpha^2}{F^2}\,[\,\tilde{\ell}^i+\tilde{b}^i\,]\,[\,\tilde{\ell}^j+\tilde{b}^j\,]\right) . \tag{11.2.5}$$

Now we apply Proposition 11.2.1 again. This time, set

$$\mathcal{Q}_{ij} \;:=\; \frac{F}{\alpha}\,\tilde{a}_{ij} \,+\, \ell_i\,\ell_j \;,$$

$$\mathcal{C}_i \;:=\; \sqrt{-1}\,\sqrt{\frac{F}{\alpha}}\;\tilde{\ell}_i \;.$$

The inverse $(\mathcal{Q}^{ij})$ of $(\mathcal{Q}_{ij})$ exists, and is given by the right-hand side of formula (11.2.5). Using that, one can compute

$$\mathcal{C}^i \;:=\; \mathcal{Q}^{ij}\,\mathcal{C}_j \quad \text{and} \quad 1 \,+\, \mathcal{C}^s\,\mathcal{C}_s \;.$$

The answers are:

$$\mathcal{C}^i \;=\; \sqrt{-1}\,\sqrt{\frac{\alpha}{F}}\;\frac{\left(1+\frac{\beta+\alpha\,\|\tilde{b}\|^2}{F}\right)\tilde{\ell}^i \,-\, \tilde{b}^i}{2 \,+\, \frac{\beta+\alpha\,\|\tilde{b}\|^2}{F}} \;, \tag{11.2.6}$$

$$1 \,+\, \mathcal{C}^s\mathcal{C}_s \;=\; \left(\frac{F}{\alpha}\right)\frac{1}{2 \,+\, \frac{\beta+\alpha\,\|\tilde{b}\|^2}{F}} \;. \tag{11.2.7}$$

In view of (11.2.3), the quantity $1+\mathcal{C}^s\mathcal{C}_s$ here is strictly positive on $TM \setminus 0$. Thus Proposition 11.2.1 can indeed be applied. Straightforward calculations, with intermediate steps involving (11.2.4) and (11.2.5), will then give (11.2.1) and (11.2.2).

The formulas for the determinant of (g_{ij}) and its inverse (g^{ij}) are rather useful.

- Having formula (11.2.1) for $\det(g_{ij})$, an argument given in §11.1 tells us that Randers metrics satisfying $\|\tilde{b}\| < 1$ are strongly convex. More precisely, if

  $$\sqrt{\tilde{a}^{ij}\,\tilde{b}_i\,\tilde{b}_j} \;<\; 1$$

  at $x \in M$, then $F(x,y)$ is positive and $(g_{ij\,(x,y)})$ is positive-definite at each nonzero y in T_xM.

- Using (11.2.2) for (g^{ij}), one can check that

$$A_k \; := \; g^{ij} \, A_{ijk} \; = \; \frac{n+1}{2} \left(\tilde{b}_k \; - \; \frac{\beta}{\alpha} \, \tilde{\ell}_k \right) . \tag{11.2.8}$$

This converts the preliminary formula (11.1.6) to

$$\boxed{A_{ijk} \; = \; \frac{1}{n+1} \, (h_{ij} \, A_k \; + \; h_{jk} \, A_i \; + \; h_{ki} \, A_j)} \; . \tag{11.2.9}$$

In the language of [M2], **every Randers space thus has reducible Cartan tensor**. For $n > 2$, only Randers and Kropina spaces (the latter not treated in our book) have this property. See [M7H]. Incidentally, every Finsler surface, whether it is of Randers type or not, has reducible Cartan tensor. See Exercise 11.2.4.

Exercises

Exercise 11.2.1: Deduce (11.2.3)–(11.2.5) from Proposition 11.2.1.

Exercise 11.2.2: Verify (11.2.6) and (11.2.7). Then use them, together with Proposition 11.2.1, to deduce (11.2.1) and (11.2.2).

Exercise 11.2.3: Verify (11.2.8), which enables us to say that every Randers space has reducible Cartan tensor.

Exercise 11.2.4:

(a) Prove that **every Finsler** ***surface*** **has reducible Cartan tensor**,

$$A_{ijk} \; = \; \frac{1}{3} \, (h_{ij} \, A_k \; + \; h_{jk} \, A_i \; + \; h_{ki} \, A_j) \; ,$$

whether it is of Randers type or not. Here, $h_{ij} := g_{ij} - \ell_i \, \ell_j$ is the angular metric.

(b) Explain in detail how the A_i, A_j, A_k in the above formula are related to the Cartan scalar I.

Suggestion: what does the above equation look like in the Berwald frame we studied in §4.3?

Exercise 11.2.5: This exercise concerns the Cartan scalar I of a Randers surface. Our goal is to derive a specific formula in [AIM].

(a) Recall the definition of I from (4.4.1). Verify that it can be re-expressed as $A_k \, (e_1)^k$, where e_1 is one of the vectors in the Berwald frame, introduced in (4.3.1).

(b) With the help of formulas (11.2.8) and (11.2.1), show that

$$I \; = \; \frac{3}{2} \left(\frac{\alpha}{F} \right)^{\frac{1}{2}} \frac{1}{\sqrt{\tilde{a}}} \, (\tilde{b}_1 \, \tilde{\ell}_2 \; - \; \tilde{b}_2 \, \tilde{\ell}_1) \; .$$

(c) Square what we have just derived. For now, focus on the resulting $1/\tilde{a}$ and just *one* factor of $(\tilde{b}_1 \tilde{\ell}_2 - \tilde{b}_2 \tilde{\ell}_1)$. Express the $\tilde{\ell}_i$ here as $\tilde{a}_{ij} \tilde{\ell}^j$. Show that these maneuvers lead to

$$\frac{1}{\tilde{a}} \, (\tilde{b}_1 \, \tilde{\ell}_2 \; - \; \tilde{b}_2 \, \tilde{\ell}_1) \;=\; (\tilde{b}^1 \, \tilde{\ell}^2 \; - \; \tilde{b}^2 \, \tilde{\ell}^1) \, .$$

(d) Prove that

$$\boxed{ I^2 \;=\; \frac{9}{4} \left(\frac{\alpha}{F} \right) \left(\| \tilde{b} \|^2 \; - \; \frac{\beta^2}{\alpha^2} \right) } \, .$$

Exercise 11.2.6: (Suggested by Brad Lackey)

(a) Consider part (d) of Exercise 11.2.5. Using the Riemannian metric $\tilde{a}_{ij}$, we can introduce an angle φ between the arbitrary y and the fixed $\tilde{b}$ (converted to a vector). Show that this gives

$$I^2 \;=\; \frac{9}{4} \left(\frac{\| \tilde{b} \|^2 \, \sin^2 \varphi}{1 \, + \, \| \tilde{b} \| \, \cos \varphi} \right) .$$

(b) As the continuous and periodic I^2 varies over the compact interval $0 \leqslant \varphi \leqslant 2\pi$, it attains its absolute maximum. Use calculus to verify that the value of this absolute maximum is

$$\frac{9}{2} \left(1 \; - \; \sqrt{1 \; - \; \| \tilde{b} \|^2} \right) .$$

(c) Conclude that at any point x on a Randers surface, one must have

$$\boxed{ | \, I_{(x,y)} \, | \;\leqslant\; \frac{3}{\sqrt{2}} \, \sqrt{1 \; - \; \sqrt{1 \; - \; \| \tilde{b} \|^2}} }$$

for all $y \in T_x M \setminus 0$. Is this upper bound sharp?

(d) Explain why

$$| \, I_{(x,y)} \, | \;\leqslant\; \frac{3}{\sqrt{2}}$$

at all $(x, y) \in TM \setminus 0$.

Exercise 11.2.7: Let (M, F) be an arbitrary Finsler manifold. Let A denote its Cartan tensor. At each $x \in M$, define

$$\| \, A \, \|_x \;:=\; \max_{y \in \mathrm{S}_x} \sqrt{A_{\alpha_1 \alpha_2 \alpha_3} \; A^{\alpha_1 \alpha_2 \alpha_3}} \, .$$

(a) Consider an arbitrary Finsler surface. Use the Berwald frame of §4.3 to express A. Check that

$$\|A\|_x \;=\; \max_{y \in \mathrm{S}_x} |I| \, .$$

(b) Identify $\|A\|_x$ explicitly for every Randers surface.

(c) Make precise the statement that:

> For every Randers surface, the norm of the Cartan tensor has $\frac{3}{\sqrt{2}}$ as a universal uniform upper bound.

(d) In that regard, what can you say about the Cartan tensor of higher-dimensional Randers spaces?

11.3 The Geodesic Spray Coefficients of a Randers Metric

In §5.3, we first encountered the equation for a Finslerian geodesic σ, when it is parametrized to have constant speed. The equation reads $D_T T = 0$, where T is the velocity field of σ. In local coordinates, this translates into the system of equations

$$\frac{d^2\sigma^i}{dt^2} + \frac{d\sigma^j}{dt}\,\frac{d\sigma^k}{dt}\,(\gamma^i{}_{jk})_{(\sigma,T)} \;=\; 0\,.$$

Here, the $\gamma^i{}_{jk}$ are the formal Christoffel symbols (of the second kind) of the fundamental tensor g_{ij}.

The main goal of this section is to write down the geodesic spray coefficients G^i of a Randers metric. We accomplish this by computing the following quantities in succession:

* $g_{ij,x^k} := \frac{\partial g_{ij}}{\partial x^k}$;
* $\gamma_{ijk} := \frac{1}{2}\,(g_{ij,x^k} - g_{jk,x^i} + g_{ki,x^j})$;
* $\gamma_{ijk}\,\ell^j\ell^k$;
* $\gamma^i{}_{jk}\,\ell^j\ell^k$;
* $G^i := \gamma^i{}_{jk}\,y^j y^k$, which is **twice** that in [AIM].

Before the task proper, here are two preliminaries:

- The Riemannian metric $\tilde{a}_{ij}$ has its own Christoffel symbols of the first and second kind. These are respectively denoted by $\tilde{\gamma}_{ijk}$ and $\tilde{\gamma}^i{}_{jk}$. In order to minimize clutter, let us use the following abbreviation scheme in the ensuing calculations.

  > **Whenever an index on $\tilde{\gamma}_{ijk}$ is contracted with the quantity $\tilde{\ell}^s := \frac{y^s}{\alpha}$ (not to be confused with ℓ^s, which is $\frac{y^s}{F}$), we replace it by the letter n.**

  For example:

$$\tilde{\gamma}_{njk} \;:=\; \tilde{\ell}^i\,\tilde{\gamma}_{ijk}\,,$$
$$\tilde{\gamma}_{ink} \;:=\; \tilde{\ell}^j\,\tilde{\gamma}_{ijk}\,,$$

  etc.

- We use formula (11.1.4) for g_{ij}. Namely,

$$g_{ij} = \frac{F}{\alpha} \, (\tilde{a}_{ij} - \tilde{\ell}_i \tilde{\ell}_j) + \ell_i \, \ell_j \, .$$

However, from time to time it would be advantageous to re-express the factor $\frac{F}{\alpha}$ as

$$\frac{F}{\alpha} = 1 + \frac{\beta}{\alpha} \, .$$

Begin with the fact that the Riemannian metric $\tilde{a}_{ij}$ is covariantly constant with respect to its Levi-Civita (Christoffel) connection. Equivalently,

$$\tilde{a}_{ij,x^k} = \tilde{\gamma}_{ijk} + \tilde{\gamma}_{jik} \, , \tag{11.3.1}$$

where

$$\tilde{\gamma}_{ijk} := \frac{1}{2} \, (\tilde{a}_{ij,x^k} - \tilde{a}_{jk,x^i} + \tilde{a}_{ki,x^j}) \, .$$

Substituting this into the last step of the following computation

$$\alpha_{x^k} = \frac{\partial}{\partial x^k} \sqrt{\tilde{a}_{pq} \, y^p \, y^q} = \frac{1}{2\alpha} \, \tilde{a}_{pq,x^k} \, y^p \, y^q = \frac{\alpha}{2} \, \tilde{a}_{pq,x^k} \, \tilde{\ell}^p \, \tilde{\ell}^q \, ,$$

we get

$$\alpha_{x^k} = \alpha \, \tilde{\gamma}_{nnk} \, . \tag{11.3.2}$$

Also,

$$\beta_{x^k} = \alpha \, \tilde{b}_{p,x^k} \, \tilde{\ell}^p \, . \tag{11.3.3}$$

These two statements then imply that

$$\left(\frac{\beta}{\alpha} \right)_{x^k} = \tilde{b}_{p,x^k} \, \tilde{\ell}^p - \frac{\beta}{\alpha} \, \tilde{\gamma}_{nnk} \, . \tag{11.3.4}$$

Similarly, from

$$\tilde{\ell}_i = \frac{\tilde{a}_{ip} \, y^p}{\alpha} \, ,$$

we obtain

$$\tilde{\ell}_{i,x^k} = \tilde{\gamma}_{ink} + \tilde{\gamma}_{nik} - \tilde{\ell}_i \, \tilde{\gamma}_{nnk} \, . \tag{11.3.5}$$

That and (11.1.2) then gives

$$\ell_{i,x^k} = \tilde{\gamma}_{ink} + \tilde{\gamma}_{nik} - \tilde{\ell}_i \, \tilde{\gamma}_{nnk} + \tilde{b}_{i,x^k} \, . \tag{11.3.6}$$

Throughout, the subscript x^k signifies partial differentiation. On tensors, it is preceded by a comma for clarity.

Using what we have just established, it is straightforward to deduce the formulas for g_{ij,x^k} and γ_{ijk}. For the sake of reducing clutter, we introduce the abbreviation

$$\boxed{\xi_i := \tilde{b}_i - \frac{\beta}{\alpha} \, \tilde{\ell}_i = \ell_i - \frac{F}{\alpha} \, \tilde{\ell}_i} \, . \tag{11.3.7}$$

Using $\ell^i = \frac{\alpha}{F}\tilde{\ell}^i$, we can check that

$$\ell^i\,\xi_i \;=\; 0 \;=\; \tilde{\ell}^i\,\xi_i\,. \tag{11.3.8}$$

The said formulas now read:

$$\begin{aligned} g_{ij,x^k} \;=\;& \frac{F}{\alpha}\,\tilde{a}_{ij,x^k} \\ &+ \left(\tilde{b}_{p,x^k}\,\tilde{\ell}^p \;-\; \frac{\beta}{\alpha}\,\tilde{\gamma}_{nnk}\right)\tilde{h}_{ij} \\ &+ \tilde{b}_{i,x^k}\,\ell_j \;+\; \tilde{b}_{j,x^k}\,\ell_i \\ &+ \xi_i\,(\tilde{\gamma}_{jnk} + \tilde{\gamma}_{njk}) \;+\; \xi_j\,(\tilde{\gamma}_{ink} + \tilde{\gamma}_{nik}) \\ &- \xi_i\,\tilde{\ell}_j\,\tilde{\gamma}_{nnk} \;-\; \xi_j\,\tilde{\ell}_i\,\tilde{\gamma}_{nnk}\,. \end{aligned} \tag{11.3.9}$$

And

$$\begin{aligned} \gamma_{ijk} \;=\;& \frac{F}{\alpha}\,\tilde{\gamma}_{ijk} \\ &- \frac{1}{2}\left(\tilde{b}_{p,x^i}\,\tilde{\ell}^p \;-\; \frac{\beta}{\alpha}\,\tilde{\gamma}_{nni}\right)\tilde{h}_{jk} \\ &+ \frac{1}{2}\,(\clubsuit_{ijk} \;+\; \clubsuit_{ikj})\,, \end{aligned} \tag{11.3.10}$$

where

$$\begin{aligned} \clubsuit_{ijk} \;:=\;& \tilde{h}_{ij}\left(\tilde{b}_{p,x^k}\,\tilde{\ell}^p \;-\; \frac{\beta}{\alpha}\,\tilde{\gamma}_{nnk}\right) \\ &+ \ell_i\,\tilde{b}_{j,x^k} \;-\; \ell_j\,\tilde{b}_{k,x^i} \;-\; \ell_k\,\tilde{b}_{i,x^j} \\ &+ \xi_i\,(\tilde{\gamma}_{jnk} + \tilde{\gamma}_{njk}) \;-\; \xi_j\,(\tilde{\gamma}_{kni} + \tilde{\gamma}_{nki}) \;+\; \xi_k\,(\tilde{\gamma}_{inj} + \tilde{\gamma}_{nij}) \\ &- \xi_i\,\tilde{\ell}_j\,\tilde{\gamma}_{nnk} \;+\; \xi_j\,\tilde{\ell}_k\,\tilde{\gamma}_{nni} \;-\; \xi_k\,\tilde{\ell}_i\,\tilde{\gamma}_{nnj}\,. \end{aligned}$$

With the help of (11.3.8) and (11.3.10), we find that

$$\begin{aligned} \gamma_{ijk}\,\ell^j\,\ell^k \;=\;& \frac{\alpha}{F}\,\tilde{\gamma}_{inn} \;+\; \frac{\alpha^2}{F^2}\,\xi_i\,\tilde{\gamma}_{nnn} \\ &+ \ell_i\,\tilde{b}_{j,x^k}\,\ell^j\,\ell^k \;+\; (\tilde{b}_{i,x^j} \;-\; \tilde{b}_{j,x^i})\,\ell^j\,. \end{aligned}$$

The index i is now raised using formula (11.2.2) for g^{**}. After a straightforward computation and some relabeling, one gets

$$\begin{aligned} \gamma^i{}_{jk}\,\ell^j\,\ell^k \;=\;& \frac{\alpha^2}{F^2}\left(\tilde{\gamma}^i{}_{nn} \;+\; [\,\tilde{b}_{j,x^k} - \tilde{b}_{k,x^j}\,]\,\tilde{a}^{ij}\,\tilde{\ell}^k\right) \\ &+ \frac{\alpha^3}{F^3}\,\tilde{\ell}^i\left(\tilde{b}_{j|k}\,\tilde{\ell}^j\,\tilde{\ell}^k \;+\; [\,\tilde{b}_{j,x^k} - \tilde{b}_{k,x^j}\,]\,\tilde{\ell}^j\,\tilde{b}^k\right). \end{aligned}$$

Here,

$$\tilde{b}_{j|k} \;:=\; \tilde{b}_{j,x^k} \;-\; \tilde{b}_s\,\tilde{\gamma}^s{}_{jk}$$

is the covariant derivative of the 1-form $\tilde{b}$ with respect to the Levi-Civita (Christoffel) connection $\tilde{\gamma}$ of the Riemannian metric $\tilde{a}$. By inspection,

$$\tilde{b}_{j,x^k} - \tilde{b}_{k,x^j} = \tilde{b}_{j|k} - \tilde{b}_{k|j} .$$

Thus the above, after some manipulation, becomes

$$\begin{aligned}\frac{F^2}{\alpha^2}\,\gamma^i{}_{jk}\,\ell^j\,\ell^k \;=\;& \tilde{\gamma}^i{}_{nn} \\ &+ \tilde{b}_{j|k}\,(\tilde{a}^{ij}\,\tilde{\ell}^k - \tilde{a}^{ik}\,\tilde{\ell}^j) \\ &+ \tilde{b}_{j|k}\,\ell^i\,(\tilde{\ell}^j\,\tilde{\ell}^k + \tilde{\ell}^j\,\tilde{b}^k - \tilde{\ell}^k\,\tilde{b}^j) .\end{aligned}$$

In other words, our G^i (twice that of [AIM]'s) is given by:

$$\begin{aligned}G^i \;:=\;& \gamma^i{}_{jk}\,y^j\,y^k \\ =\;& \tilde{\gamma}^i{}_{jk}\,y^j\,y^k \\ &+ \tilde{b}_{j|k}\,[\tilde{a}^{ij}\,y^k - \tilde{a}^{ik}\,y^j]\,\alpha \\ &+ \tilde{b}_{j|k}\,\ell^i\,(\,y^j\,y^k + [y^j\,\tilde{b}^k - y^k\,\tilde{b}^j]\,\alpha\,) .\end{aligned} \tag{11.3.11}$$

Sometimes the following equivalent version (given in [M10]) is more useful:

$$\begin{aligned}G^i \;:=\;& \gamma^i{}_{jk}\,y^j\,y^k \\ =\;& (\tilde{\gamma}^i{}_{jk} + \ell^i\,\tilde{b}_{j|k})\,y^j\,y^k \\ &+ (\tilde{a}^{ij} - \ell^i\,\tilde{b}^j)\,(\tilde{b}_{j|k} - \tilde{b}_{k|j})\,\alpha\,y^k .\end{aligned} \tag{11.3.12}$$

That ℓ^i is not to be mistaken as $\tilde{\ell}^i$. Also, as a reminder,

$$\tilde{\gamma}^i{}_{jk} := \tilde{a}^{is}\,\tilde{\gamma}_{sjk} = \frac{\tilde{a}^{is}}{2}\,(\tilde{a}_{sj,x^k} - \tilde{a}_{jk,x^s} + \tilde{a}_{ks,x^j}) .$$

By contrast,

$$\gamma^i{}_{jk} := g^{is}\,\gamma_{sjk} = \frac{g^{is}}{2}\,(g_{sj,x^k} - g_{jk,x^s} + g_{ks,x^j}) .$$

Formula (11.3.12) immediately leads to the equation $\ddot{\sigma}^i + G^i = 0$ for autoparallels in a Randers space. Namely, a curve $(\sigma^i(t))$ in M is a **constant** *Finslerian* **speed** geodesic of the Randers metric $F := \alpha + \beta$ if

$$\boxed{\begin{aligned}&\frac{d^2\sigma^i}{dt^2} + (\tilde{\gamma}^i{}_{jk} + \ell^i\,\tilde{b}_{j|k})\,\frac{d\sigma^j}{dt}\,\frac{d\sigma^k}{dt} \\ &\quad + (\tilde{a}^{ij} - \ell^i\,\tilde{b}^j)\,(\tilde{b}_{j|k} - \tilde{b}_{k|j})\,\alpha(T)\,\frac{d\sigma^k}{dt} = 0\end{aligned}}\;,$$

where $T := \frac{d\sigma}{dt}$. Unlike its counterpart in [AIM], our geodesic equation gives rise to curves that are parametrized to have constant *Finslerian* (rather than Riemannian) speed. Put another way, $F(\frac{d\sigma}{dt})$ is constant while $\alpha(\frac{d\sigma}{dt})$ is typically not. For a treatment in which the Riemannian speed is kept constant, see the guided Exercises 11.3.3 and 11.3.4.

Exercises

Exercise 11.3.1: Verify (11.3.8).

Exercise 11.3.2:

(a) Derive (11.3.11) in detail.

(b) Re-express (11.3.11) as (11.3.12).

Exercise 11.3.3: Let us suppose that the geodesics of Randers spaces were reparametrized to have **constant** *Riemannian* **speed** instead. Namely, $\alpha(T)$ is constant, where $T := \frac{d\sigma}{dt}$. It is the objective of this exercise to show that the geodesic equation would then read

$$\boxed{\frac{d^2\sigma^i}{dt^2} + \tilde{\gamma}^i{}_{jk}\,\frac{d\sigma^j}{dt}\,\frac{d\sigma^k}{dt} + \tilde{a}^{ij}\,(\tilde{b}_{j|k} - \tilde{b}_{k|j})\,\alpha(T)\,\frac{d\sigma^k}{dt} = 0}\ .$$

In order to facilitate some later discussions in this exercise, let us abbreviate the left-hand side of this equation as $\bigstar^i$.

Recall from Exercise 5.2.4 that Finslerian geodesics, without any assumption about their speed, are described by the equation

$$D_T\left[\frac{T}{F(T)}\right] = 0 \quad \text{with reference vector } T\ .$$

In Exercise 5.3.1, we showed that this represents the following system of differential equations:

$$(*) \qquad \frac{d^2\sigma^i}{dt^2} + \frac{d\sigma^j}{dt}\,\frac{d\sigma^k}{dt}\,(\gamma^i{}_{jk})_{(\sigma,T)} = \frac{d}{dt}[\log F(T)]\,\frac{d\sigma^i}{dt}\ .$$

Note that on the left-hand side, it is $\gamma^i{}_{jk}$ (the Randers metric's formal Christoffel symbols of the second kind) rather than $\tilde{\gamma}^i{}_{jk}$ (that of the underlying Riemannian metric).

(a) Suppose $\sigma(t)$ has been parametrized to have constant Riemannian speed. Show that

$$\frac{d}{dt}[\log F(T)] = \frac{1}{F(T)}\left[\tilde{b}_{j,x^k}\,\frac{d\sigma^j}{dt}\,\frac{d\sigma^k}{dt} + \tilde{b}_j\,\frac{d^2\sigma^j}{dt^2}\right]\ .$$

(b) Substitute part (a) into the right-hand side of (*), and use (11.3.12) on the left-hand side. After a cancellation, check that (*) can be rearranged to read

$$\bigstar^i = \frac{1}{F(T)}\,\frac{d\sigma^i}{dt}\,\tilde{b}_s\,\bigstar^s\ .$$

(c) By contracting part (b) with $\tilde{b}_i$, show that one obtains the statement

$$\frac{\alpha(T)}{F(T)} \; \tilde{b}_s \bigstar^s = 0 \,.$$

(d) How will part (c) lead to the desired geodesic equation?

Exercise 11.3.4: Let (M, F) be a Randers space, where F is determined by a Riemannian metric $\tilde{a}$ and a 1-form $\tilde{b}$, both globally defined on M.

(a) Show that

> **If $\tilde{b}$ is a *closed* 1-form, then the Finslerian geodesics have the same trajectories as the geodesics of the underlying Riemannian metric $\tilde{a}$.**

Hints: see Exercise 11.3.3; also, how are the quantities $\tilde{b}_{j,x^k} - \tilde{b}_{k,x^j}$ and $\tilde{b}_{j|k} - \tilde{b}_{k|j}$ related to each other?

(b) Is the converse to part (a) true? If not, can you provide a counter-example?

Exercise 11.3.5: Recall the Randers spaces (M, F) and $(\hat{M}, \hat{F})$ that we studied in Exercise 11.1.4. Briefly:

* $M := \mathbb{R}^3$, with Cartesian coordinates x^1, x^2, x^3. Its underlying Riemannian metric $\tilde{a}$ is the Euclidean inner product, and the 1-form in question is $\tilde{b} := \lambda \, dx^3$. Here, $\lambda < 1$ is a positive constant.
* $\hat{M}$ is the graph of a function $f(x^1, x^2)$. The data $\hat{a}$ and $\hat{b}$ for $(\hat{M}, \hat{F})$ are simply the pull-backs of $\tilde{a}$ and $\tilde{b}$.

(a) Explain why the geodesics of (M, F) have the same trajectories as those of its underlying Riemannian metric $\tilde{a}$.

(b) Explain why the geodesics of $(\hat{M}, \hat{F})$ have the same trajectories as those of its underlying Riemannian metric $\hat{a}$.

(c) What do the geodesics of $\tilde{a}$ in M look like?

(d) Write down the system of two coupled ODEs that describe the geodesics of $\hat{a}$ on $\hat{M}$. Can you solve this system explicitly? Can you reduce it to one single equation?

11.4 The Nonlinear Connection for Randers Spaces

The nonlinear connection $N^i{}_j$ was first defined in (2.3.2). But we found in Exercise 2.3.3 that it can be computed through $N^i{}_j = \frac{1}{2} \, (G^i)_{y^j}$. In (11.3.11), we have successfully obtained the formula for the geodesic spray coefficients G^i of Randers spaces. Using it, one can show that the nonlinear

connection for Randers spaces has the following structure:
(11.4.1)

$$\begin{aligned}
\frac{N^i{}_j}{F} \;=\;& \frac{\alpha}{F}\,\tilde{\gamma}^i{}_{jn} \;+\; \frac{1}{2}\,\tilde{b}_{r|s} \text{ contracted with the following terms:} \\
& \frac{\alpha}{F}\left[\tilde{a}^{ir}\,(\delta^s{}_j + \tilde{\ell}^s\,\tilde{\ell}_j) \;-\; \tilde{a}^{is}\,(\delta^r{}_j + \tilde{\ell}^r\,\tilde{\ell}_j)\right] \\
& + \frac{\alpha^2}{F^2}\Big\{\delta^i{}_j\,\tilde{\ell}^r\,\tilde{\ell}^s + (\delta^i{}_j + \tilde{\ell}^i\,\tilde{\ell}_j)\,(\tilde{\ell}^r\,\tilde{b}^s - \tilde{\ell}^s\,\tilde{b}^r) \\
& \qquad + \tilde{\ell}^i\,\delta^r{}_j\,(\tilde{\ell}^s + \tilde{b}^s) + \tilde{\ell}^i\,\delta^s{}_j\,(\tilde{\ell}^r - \tilde{b}^r)\Big\} \\
& - \frac{\alpha^3}{F^3}\,\tilde{\ell}^i\,(\tilde{\ell}_j + \tilde{b}_j)\left[\tilde{\ell}^r\,\tilde{\ell}^s + (\tilde{\ell}^r\,\tilde{b}^s - \tilde{\ell}^s\,\tilde{b}^r)\right].
\end{aligned}$$

One explicit formula of the Chern connection in natural coordinates is given by (2.4.9). Namely,

$$\Gamma^i{}_{jk} \;=\; \gamma^i{}_{jk} \;-\; g^{il}\,(A_{ljs}\,N^s{}_k \;-\; A_{jks}\,N^s{}_l \;+\; A_{kls}\,N^s{}_j)\,.$$

For Randers spaces, we have already computed the following quantities:

* g^{ij} in (11.2.2);
* A_{ijk} in (11.2.8), (11.2.9);
* γ_{ijk} in (11.3.10);
* $N^i{}_j$ in (11.4.1).

In principle, the Chern connection $\Gamma^i{}_{jk}$ can then be written down, as soon as one raises the index i on γ_{ijk} . But the resulting expression is not of manageable size.

Alternatively, one can begin with the geodesic spray coefficients

$$G^i \;:=\; \gamma^i{}_{jk}\,y^j\,y^k$$

and the nonlinear connection

$$N^i{}_j \;=\; \frac{1}{2}\,(G^i)_{y^j}\,.$$

For Randers spaces, G^i is explicitly given by (11.3.11), and $N^i{}_j$ is F **times** the right-hand side of (11.4.1). Their successive y derivatives produce the Berwald connection and the tensor $\dot{A}$. These can then be used to obtain the Chern connection. Specifically:

** Formula (3.8.3) says that the Berwald connection $\Gamma^i{}_{jk} + \dot{A}^i{}_{jk}$ is equal to $\frac{1}{2}(G^i)_{y^j y^k}$, namely, $(N^i{}_j)_{y^k}$.

** Formula (3.8.5) tells us that $\dot{A}_{jkl} = -\frac{1}{4}\,y_i\,(G^i)_{y^j y^k y^l}$, which in turn is equal to $-\frac{1}{2}\,y_i\,(N^i{}_j)_{y^k y^l}$.

Raise an index on $\dot{A}$ and subtract that from the Berwald connection. The resulting formula gives the Chern connection:

$$\Gamma^i{}_{jk} = (N^i{}_j)_{y^k} + \frac{1}{2} g^{it} y_s (N^s{}_t)_{y^j y^k} . \tag{11.4.2}$$

Now we turn to the curvatures. The hv-Chern curvature is given by (3.3.3), which says that

$$P_j{}^i{}_{kl} = - F (\Gamma^i{}_{jk})_{y^l} .$$

This is computationally reasonable. On the other hand, the full hh-Chern curvature $R_j{}^i{}_{kl}$, though given by (3.3.2), that is,

$$R_j{}^i{}_{kl} = \frac{\delta \Gamma^i{}_{jl}}{\delta x^k} - \frac{\delta \Gamma^i{}_{jk}}{\delta x^l} + \Gamma^i{}_{hk} \Gamma^h{}_{jl} - \Gamma^i{}_{hl} \Gamma^h{}_{jk} ,$$

does not have a manageable expression in terms of the data $\tilde{a}_{ij}$ and $\tilde{b}_i$.

In view of this, one might want to concentrate on the predecessor (see §3.9)

$$R^i{}_k := \ell^j R_j{}^i{}_{kl} \ell^l$$

of the flag curvature. After all, there is a constitutive relation (3.5.6) that expresses $R_j{}^i{}_{kl}$ in terms of $R^i{}_k$, $\dot{A}_{ijk}$, and their covariant derivatives. Furthermore, Exercise 3.3.4 gives an elegant statement:

$$R^i{}_k = \ell^j \left(\frac{\delta}{\delta x^k} \frac{N^i{}_j}{F} - \frac{\delta}{\delta x^j} \frac{N^i{}_k}{F} \right) .$$

The following references are also useful: Yasuda and Shimada [YS]; Matsumoto [M5]; Shibata and Kitayama [SK]. See especially formula (1.10) in [M5] and formula (1.9) in [YS].

In this section, we have not been able to write down anything explicit, except for the nonlinear connection. This is because the requisite calculations are quite formidable. The references cited above have made substantial strides in detecting useful structure among tedium. It is a worthwhile endeavor to understand them. Incidentally, Miron has also openly lamented about the computational hurdles presented by Randers metrics. See his treatment in [Mir1].

Exercises

Exercise 11.4.1: Derive formula (11.4.1) for the nonlinear connection of Randers spaces.

Exercise 11.4.2: Let (M, F) be a Randers space. Suppose the 1-form $\tilde{b}$ on M is parallel (or covariantly constant) with respect to the Riemannian metric $\tilde{a}$. That is, $\tilde{b}_{j|k} = 0$.

(a) Prove that $\tilde{b}_{j|k} = 0$ implies $\Gamma^i{}_{jk} = \tilde{\gamma}^i{}_{jk}$. In other words, whenever $\tilde{b}$ is parallel with respect to $\tilde{a}$, the Chern connection of F and the Levi-Civita connection of $\tilde{a}$ share the same Christoffel symbols of the *second* kind. Hint: use (11.3.11) and the discussion in the second half of this section.

(b) Explain carefully why the same does *not* hold for Christoffel symbols of the first kind.

Exercise 11.4.3:

(a) Verify that

$$\Gamma^i{}_{jk}\, \ell^k \;=\; \frac{1}{F}\, N^i{}_j$$

holds for every Finsler space.

(b) Now suppose we have a Randers space with $\tilde{b}_{j|k} = 0$. Part (a) here, together with that of Exercise 11.4.2, would then imply that

$$\frac{1}{F}\, N^i{}_j \;=\; \tilde{\gamma}^i{}_{jk}\, \ell^k \;.$$

We emphasize that it is ℓ^k and *not* $\tilde{\ell}^k$. On the other hand, our formula (11.4.1) says that

$$\frac{1}{F}\, N^i{}_j \;=\; \frac{\alpha}{F}\, \tilde{\gamma}^i{}_{jn} \;.$$

How are these two conclusions in agreement with each other?

11.5 A Useful and Elegant Theorem

In this section, we describe a simple criterion that characterizes Berwald spaces among Randers spaces. This is later used in §11.6 to construct a 3-dimensional y-global Berwald space that is neither Riemannian nor locally Minkowskian.

Theorem 11.5.1. *Let (M, F) be a Randers space. Denote*

* *the underlying Riemannian metric by $\tilde{a} := \tilde{a}_{ij}\, dx^i \otimes dx^j$,*
* *the Levi-Civita (Christoffel) connection of $\tilde{a}$ by $\tilde{\gamma}^i{}_{jk}$,*
* *the underlying 1-form by $\tilde{b} := \tilde{b}_i\, dx^i$, with $\|\, \tilde{b}\, \| < 1$ everywhere.*

Abbreviate the covariant derivative of $\tilde{b}$ with respect to $\tilde{a}$ as

$$\tilde{b}_{j|k} \;:=\; \tilde{b}_{j,x^k} \;-\; \tilde{b}_s\, \tilde{\gamma}^s{}_{jk} \;.$$

- [$\tilde{b}_{j|k} = 0$ **as a sufficient condition**] *If $\tilde{b}_{j|k} = 0$, namely, the 1-form $\tilde{b}$ is parallel with respect to $\tilde{a}$, then our Randers space (M, F) is of Berwald type.*

- [$\tilde{b}_{j|k} = 0$ **as a necessary condition**] *Conversely, if our Randers space is of Berwald type, then we must have* $\tilde{b}_{j|k} = 0$.

Remarks: Let us mention some history behind this elegant result.

* In a 1974 paper [M4], Matsumoto showed that a Randers space is of Landsberg type (that is, $\dot{A}_{ijk} = 0$) if and only if $\tilde{b}_{j|k} = 0$. We recall from (11.2.9) that the Cartan tensor of a Randers space is reducible. Whenever one has such reducibility in dimensions $n \geqslant 3$, the Finsler structure is of Landsberg type if and only if it is of Berwald type. See [M7H], then [M6]. A more recent account is given in [M10].

* In a 1975 paper [HI], Hashiguchi and Ichijyō gave direct arguments to show that if $\tilde{b}_{j|k} = 0$, then the Randers space in question will be of Berwald type.

* In a 1977 paper [SSAY], Shibata, Shimada, Azuma, and Yasuda provided more concrete formulas for the connection coefficients and the curvatures in [M4]. They again proved the fact that a Randers space is of Landsberg type if and only if $\tilde{b}_{j|k} = 0$. Also, they acknowledged that for Randers spaces of all dimensions, the notions of being Landsberg and being Berwald are equivalent.

* Finally, in a 1979 paper [Ki], Kikuchi proved directly that given a Randers space of Berwald type, one must have $\tilde{b}_{j|k} = 0$.

Proof of the theorem.

- $\tilde{b}_{j|k} = 0$ **as a sufficient condition:**

Given that $\tilde{b}$ is parallel with respect to $\tilde{a}$, formula (11.3.11) reduces to $G^i = \tilde{\gamma}^i{}_{jk}\, y^j y^k$. Taking two y derivatives of G^i then gives $2\tilde{\gamma}^i{}_{jk}$, which comes from $\tilde{a}$ alone and therefore has no y-dependence. As a result, $\dot{A}_{jkl} = -\frac{1}{4}(G^i)_{y^j y^k y^l}$ [see (3.8.5)] vanishes. In that case, (3.8.3) implies that the Chern connection $\Gamma^i{}_{jk}$ in natural coordinates is given by $\frac{1}{2}(G^i)_{y^j y^k}$, which is simply $\tilde{\gamma}^i{}_{jk}$. So the coefficients $\Gamma^i{}_{jk}$ have no y-dependence, and thus the Randers space in question is by definition a Berwald space.

- $\tilde{b}_{j|k} = 0$ **as a necessary condition:**

Suppose our Randers metric is of Berwald type. Proposition 10.2.1 gives several characterizations of Berwald spaces. The one most useful to us says that the y-Hessian $(G^i)_{y^j y^k}$ of G^i must be independent of y. In (11.3.11), we have explicitly calculated the quantity G^i for Randers metrics. It has the structure

$$G^i \ = \ \tilde{\gamma}^i{}_{rs}\, y^r\, y^s \ + \ \tilde{b}_{r|s}\, \psi^{irs} \ ,$$

where

$$\psi^{irs} \ := \ [\, \tilde{a}^{ir}\, y^s \ - \ \tilde{a}^{is}\, y^r \,]\, \alpha \ + \ \ell^i\, (y^r\, y^s \ + \ [\, y^r\, \tilde{b}^s \ - \ y^s\, \tilde{b}^r \,]\, \alpha) \ .$$

This gives

$$(G^i)_{y^j y^k} \;=\; \tilde{\gamma}^i{}_{jk} \;+\; \tilde{b}_{r|s} \, (\psi^{irs})_{y^j y^k} \; .$$

Thus, in order for $(G^i)_{y^j y^k}$ to be independent of y, one of the following two scenarios must hold:

* either $(\psi^{irs})_{y^j y^k}$ is independent of y,
* or $\tilde{b}_{r|s} = 0$.

We show that the first scenario is impossible unless $\tilde{b}$ is identically zero (in which case certainly $\tilde{b}_{r|s} = 0$). Thus both scenarios effect the conclusion sought; namely, the 1-form $\tilde{b}$ is parallel with respect to the Riemannian metric $\tilde{a}$.

So, let us suppose that $(\psi^{irs})_{y^j y^k}$ is independent of y, and $\tilde{b}$ is not identically zero on M. We reach a contradiction momentarily.

Since $\tilde{a}_{rs}$ and $\tilde{b}^k$ depend only on x, our supposition implies that the function $\tilde{b}^k \, (\psi^{irs} \tilde{a}_{rs})_{y^j y^k}$ has no y-dependence. A straightforward computation gives

$$\begin{aligned} \tilde{b}^k \, (\psi^{irs} \, \tilde{a}_{rs})_{y^j y^k} \;&=\; (1+n) \, \frac{1}{F^2} \left[(\alpha \;+\; 2\beta) \, \beta \;-\; \alpha^2 \, \| \tilde{b} \|^2 \right] \\ &=: \; (1+n) \, \spadesuit \; . \end{aligned}$$

By our supposition, there exists some $x \in M$ at which $\tilde{b}$ is nonzero. Fix one such x. Within this $T_x M$, the above $\spadesuit$ supposedly has no y-dependence. It is then equal to some constant C (which varies with our choice of the fixed x). That is,

$$\frac{1}{F^2} \left[(\alpha \;+\; 2\beta) \, \beta \;-\; \alpha^2 \, \| \tilde{b} \|^2 \right] \;=\; C \; .$$

Algebraic manipulations give the equation

$$(C - 2) \, \beta^2 \;+\; (2C - 1) \, \alpha \beta \;+\; (C + \| \tilde{b} \|^2) \, \alpha^2 \;=\; 0 \; .$$

After considering the two cases $C \neq 2$ and $C = 2$ separately in the above equation, we find that they lead to similar conclusions. Namely, $\beta = \lambda \, \alpha$, where the proportionality factor λ depends *not* on y but possibly on x. Hence $F = \alpha + \beta = (\lambda + 1) \, \alpha$ is Riemannian at our fixed $x \in M$. This surely cannot be, since the 1-form $\tilde{b}$ is by assumption nonzero at this x. Thus we have finally reached a contradiction. □

Exercises

Exercise 11.5.1: Derive the defining equation for the quantity $\spadesuit$. Namely,

$$\tilde{b}^k \, (\psi^{irs} \, \tilde{a}_{rs})_{y^j y^k} \;=\; (1+n) \, \frac{1}{F^2} \left[(\alpha \;+\; 2\beta) \, \beta \;-\; \alpha^2 \, \| \tilde{b} \|^2 \right] .$$

Exercise 11.5.2: We asserted that the $C \neq 2$ and $C = 2$ cases both lead to the conclusion $\beta = \lambda\,\alpha$, where λ has no y-dependence. Fill in the algebraic details that we omitted.

Exercise 11.5.3: Suppose we want to use Theorem 11.5.1 to construct a y-global Berwald space that is neither Riemannian nor locally Minkowskian. Attempt to answer the following questions without consulting §11.6.

(a) In order to guarantee success in this endeavor, what criteria must one impose on the Riemannian metric $\tilde{a}$ and the 1-form $\tilde{b}$?

(b) Wouldn't your success in part (a) contradict Szabó's rigidity theorem (see §10.6) about Berwald surfaces?

Exercise 11.5.4:

(a) Show that $\tilde{b}_{j,x^k} - \tilde{b}_{k,x^j} = \tilde{b}_{j|k} - \tilde{b}_{k|j}$.

(b) Explain why parallel 1-forms are automatically closed.

(c) Are closed 1-forms necessarily parallel?

When the given $\tilde{b}$ is a closed 1-form, the resulting Randers space is known as a **Douglas space**. See [BM]. Also, recall from Exercises 11.3.4, 11.3.5 that for any Randers space of Douglas type, the Finslerian geodesics have the same *trajectories* as the geodesics of the underlying Riemannian metric $\tilde{a}$. More precisely, when the Finslerian geodesics are parametrized to have constant *Riemannian* speed, their defining equation is exactly the geodesic equation of $\tilde{a}$.

11.6 The Construction of y-global Berwald Spaces

11.6 A. The Algorithm

Let us take stock of the situation regarding explicit examples of Berwald spaces. Naturally, we would like to focus on those that are neither Riemannian nor locally Minkowskian.

In §10.3, one such example in dimension two was discussed. However, its Finsler structure F is not strongly convex on all of $TM \setminus 0$. Surprisingly, this y-local feature turns out to be unavoidable among Berwald surfaces that are neither Riemannian nor locally Minkowskian. That is the message of Szabó's rigidity theorem, which we proved in §10.6.

So, y-**global** (F being smooth and strongly convex on $TM \setminus 0$) Berwald spaces that are neither Riemannian nor Minkowskian can only be found in dimension three or higher. This is where the sufficiency part of Theorem 11.5.1 becomes quite useful. Namely, if $\tilde{b}_{j|k} = 0$, then the Randers space in question is of Berwald type. Here's how we deduce the algorithm:

* One constructs Randers metrics using Riemannian metrics $\tilde{a}$ on M that admit globally defined parallel 1-forms $\tilde{b}$ which are not identi-

cally zero. See also the article [Fult]. We must work only with those 1-forms $\tilde{b}$ whose Riemannian norm $\| \tilde{b} \|$ is uniformly bounded. That way they can be normalized (if necessary) to satisfy $\| \tilde{b} \| < 1$, as per §11.1. Using these $\tilde{b}$, the resulting Randers spaces will automatically be of Berwald type, thanks to Theorem 11.5.1. They will also be y-global, a much desired feature.

* Since we are using nonzero 1-forms $\tilde{b}$, the said Berwald spaces are definitely not Riemannian. This is so because the Finsler structures formed using such $\tilde{b}$ are positively homogeneous but not absolutely homogeneous. Incidentally, since the $\tilde{b}$ in question here is parallel, it is nonzero (that is, not identically zero) if and only if it is nowhere zero.
* Among these Berwald spaces, we need to ferret out those that are not locally Minkowskian. One way is to check whether there is no coordinate system in which the components of $\tilde{a}$ and $\tilde{b}$ are *simultaneously* constant. In principle, such nonexistence claims are difficult to establish.
* A more direct route is through curvature. Since the spaces in question are of Berwald type, the hv-Chern curvature P already vanishes (Proposition 10.2.1). A characterization of locally Minkowski spaces by curvature is given in Proposition 10.5.1. It tells us that among our Berwald spaces, the ones that are not locally Minkowskian will be singled out by having nonzero hh-Chern curvature tensors R.
* Recall the first half of Theorem 11.5.1's proof. It says that for a Randers space with $\tilde{b}_{j|k} = 0$, the Chern connection coefficients $\Gamma^i{}_{jk}$ are equal to $\tilde{\gamma}^i{}_{jk}$. The latter are the Christoffel symbols of the second kind of the underlying Riemannian metric $\tilde{a}$. Since these have no y-dependence, formula (3.3.2) for R simplifies to

$$R_j{}^i{}_{kl} = \frac{\partial \tilde{\gamma}^i{}_{jl}}{\partial x^k} - \frac{\partial \tilde{\gamma}^i{}_{jk}}{\partial x^l} + \tilde{\gamma}^i{}_{hk} \tilde{\gamma}^h{}_{jl} - \tilde{\gamma}^i{}_{hl} \tilde{\gamma}^h{}_{jk} .$$

The right-hand side is none other than the curvature tensor of $\tilde{a}$. Therefore, among the Berwald spaces obtained above, the ones that are not locally Minkowskian arise precisely from nonflat Riemannian metrics $\tilde{a}$.

Let us summarize the algorithm we have just discussed, before using it in an explicit construction.

Proposition 11.6.1. *Let (M, F) be a Randers space constructed from a Riemannian metric $\tilde{a}$ and a 1-form $\tilde{b}$, both globally defined on M. The following two statements are then equivalent:*

- *The Riemannian metric $\tilde{a}$ is not flat, and $\tilde{b}$ is a nonzero (hence nowhere zero) parallel 1-form with $\| \tilde{b} \| < 1$ everywhere.*

- *The Randers space (M, F) is a y-global Berwald space, and it is neither Riemannian nor locally Minkowskian.*

Remark:

There is at least one topological obstruction involved when one tries to put this proposition to practical use. It can be seen by raising the index of that nonvanishing parallel $\tilde{b}$ to form a nowhere zero vector field on M. If M is compact and boundaryless, or if this vector field should happen to be everywhere transversal to the boundary ∂M, then the Euler characteristic $\chi(M)$ must vanish in the first place. This is a consequence of the Poincaré–Hopf index theorem.

Incidentally, what does Theorem 11.5.1 produce when we apply it to surfaces? Let us restrict to compact oriented Riemannian surfaces without boundary. The only such surface that admits globally defined nonvanishing parallel 1-forms is the flat torus. (See Exercise 11.6.3.) Furthermore, although the resulting Randers metrics are indeed of Berwald type, they are at the same time all locally Minkowskian. (Exercise 11.6.3 again.) Thus there is no conflict with what's being predicted by Szabó's rigidity theorem.

11.6 B. An Explicit Example in Three Dimensions

Let us conclude this section with an example of a 3-dimensional y-global Berwald space that is neither Riemannian nor locally Minkowskian. It is given by a Randers metric constructed with the following data:

- The underlying manifold is the Cartesian product

$$M := \mathbb{S}^2 \times \mathbb{S}^1 .$$

  It is compact and boundaryless. As local coordinates, one can use the usual spherical θ, ϕ on $\mathbb{S}^2$, and t for $\mathbb{S}^1$. For concreteness, we measure ϕ from the positive z axis down. Also, t is such that $(\cos t, \sin t, 0)$ parametrizes $\mathbb{S}^1$.

- The Riemannian metric $\tilde{a}$ is the product metric on $\mathbb{S}^2 \times \mathbb{S}^1$. Here, $\mathbb{S}^2$ and $\mathbb{S}^1$ are given the standard Riemannian metrics that they inherited as submanifolds of Euclidean $\mathbb{R}^3$. Explicitly, one finds that

$$\tilde{a} := (\sin^2\phi \, d\theta \otimes d\theta + d\phi \otimes d\phi) + dt \otimes dt .$$

  This metric is not flat because it has nonzero curvature tensor.

- The parallel 1-form is

$$\tilde{b} := \epsilon \, dt ,$$

  where ϵ is any (fixed) positive constant less than 1. This 1-form is globally defined on M, even though the coordinate t is not. It is nonvanishing by inspection, and has Riemannian norm $\| \tilde{b} \| = \epsilon < 1$.

A straightforward calculation shows that it is indeed parallel with respect to the Levi-Civita (Christoffel) connection of $\tilde{a}$.

We now write down the resulting Randers metric. Let x be any point on M, with coordinates (θ, ϕ, t). Let $y \in T_x M$ be expanded as

$$y = y^\theta \, \partial_\theta + y^\phi \, \partial_\phi + y^t \, \partial_t \, .$$

Then

$$\boxed{F(x,y) = \sqrt{\sin^2\phi \, (y^\theta)^2 + (y^\phi)^2 + (y^t)^2} + \epsilon \, y^t} \, .$$

According to Proposition 11.6.1, (M, F) is a y-global Berwald space, and it is neither Riemannian nor locally Minkowskian.

* In Exercise 11.6.6, we show that this example can be generalized to all higher dimensions. Specifically, we consider the Cartesian products $\mathbb{S}^n \times \mathbb{S}^1$, where $n \geqslant 2$. It is to be endowed with the product Riemannian metric. The 1-form in question again comes from the dt on $\mathbb{S}^1$.
* There is a deeper reason *why* these examples work. It is through the theory of harmonic forms and the Bochner technique, for Riemannian manifolds with nonnegative Ricci curvature. That is explained in a later chapter on Riemannian manifolds. See the second part of Theorem 13.6.1 and Exercise 13.6.5.

As remarked in §10.3, the 3-dimensional conformally flat Berwald examples given by Matsumoto [M9] are y-local. This is because they only have strong convexity on proper subsets of $TM \setminus 0$.

Exercises

Exercise 11.6.1:

(a) Let $\tilde{a}$ be a Riemannian metric on M. Suppose it admits a parallel 1-form $\tilde{b}$. Show that $\| \tilde{b} \|$ must be constant.

(b) Explain why a parallel 1-form is not identically zero if and only if it is nowhere zero.

Exercise 11.6.2: Let $\tilde{a}$ be a Riemannian metric on M.

(a) Show that $\tilde{a}$ admits a parallel 1-form $\tilde{b}$ if and only if it admits a parallel vector field $\tilde{b}^\sharp$.

(b) Check that $\tilde{b}$ is nonvanishing if and only if $\tilde{b}^\sharp$ is nowhere zero.

(c) Suppose $\tilde{b}^\sharp$ is nowhere zero. Explain why if M is compact and boundaryless, or if $\tilde{b}^\sharp$ is everywhere transversal to the boundary ∂M, then the Euler characteristic $\chi(M)$ must vanish.

Exercise 11.6.3: Let us ponder the implications of Theorem 11.5.1 for Finsler *surfaces.*

(a) Focus on compact oriented Riemannian surfaces without boundary. Explain why among such, only the torus (together with whatever Riemannian metrics we are able to put on it) can possibly admit globally defined nonvanishing parallel 1-forms.

(b) Prove that, whenever a Riemannian surface $(M, \tilde{a})$ admits a globally defined nonvanishing parallel 1-form $\tilde{b}$, its Gaussian curvature must be identically zero. Here are some hints (from Exercise 6 on p. 333 of [On]): raise the index on $\tilde{b}$ to form a nowhere zero vector field $\tilde{b}^\sharp$ on M, and use that as one of the vectors in an orthonormal frame field. The covariant derivative of $\tilde{b}^\sharp$ is zero in every direction. What does that tell us about the connection forms?

(c) Explain why the tori discussed in part (b) are all isometric to the so-called "flat torus."

(d) In view of parts (a)–(c), the "flat torus" is the only Riemannian surface that admits globally defined nonvanishing parallel 1-forms $\tilde{b}$. Check that there exists (suggestion: see §10.4) a coordinate system in which the components of the metric are simply δ_{ij}, and those of $\tilde{b}$ are constants. Then explain why the resulting Randers metric is locally Minkowskian.

Exercise 11.6.4: This exercise concerns the example we produced in §11.6B.

(a) Write down the matrix $(\tilde{a}_{ij})$ in the described coordinates. Show that the product metric $\tilde{a}$ has the formula we asserted.

(b) Explain why the 1-form dt and the vector field $\frac{\partial}{\partial t}$ are globally defined on M. Check that they each have Riemannian norm 1.

(c) Show that $\frac{\partial}{\partial t}$ is parallel with respect to the Levi-Civita (Christoffel) connection of $\tilde{a}$. Explain why the same can then be said of dt.

(d) Compute the curvature tensor of $\tilde{a}$.

Exercise 11.6.5: This again concerns that example encountered in §11.6B.

(a) Proposition 11.5.1 tells us that the example we described is a Berwald space. Double check this by directly computing the *hv*-curvature P of the given Finsler structure F.

(b) Proposition 11.6.1 assures us that this Berwald space is not locally Minkowskian. Verify this by a brute force computation of the *hh*-curvature R of F.

(c) Does our M satisfy the topological constraint $\chi(M) = 0$? (We mentioned that immediately after the statement of Proposition 11.6.1.)

(d) Within each tangent space of M, the indicatrix is a simple closed surface. Identify this surface with a detailed sketch.

Exercise 11.6.6: This exercise demonstrates that the example studied in §11.6B is actually part of a hierarchy. To this end, fix any $n \geqslant 2$, and prescribe the following data:

- The underlying manifold is the Cartesian product $M := \mathbb{S}^n \times \mathbb{S}^1$. It is compact and boundaryless. As local coordinates, one can use the usual spherical $\theta^1, \ldots, \theta^n$ on $\mathbb{S}^n$, and the parameter t for $\mathbb{S}^1$. Let us agree that $(\cos t, \sin t, 0, \ldots, 0)$ parametrizes $\mathbb{S}^1$, as a submanifold of Euclidean $\mathbb{R}^{n+1}$.
- The Riemannian metric $\tilde{a}$ is the product metric on $\mathbb{S}^n \times \mathbb{S}^1$. Here, $\mathbb{S}^n$ and $\mathbb{S}^1$ are given the standard Riemannian metrics that they inherited as submanifolds of Euclidean $\mathbb{R}^{n+1}$.
- The 1-form we have in mind is $\tilde{b} := \epsilon\, dt$, where ϵ is any (fixed) positive constant less than 1. This 1-form is globally defined on M, even though the coordinate t is not.

(a) Write down an explicit formula for the Riemannian metric $\tilde{a}$, in terms of the spherical coordinates $\theta^1, \ldots, \theta^n$, and t.

(b) Calculate the Riemann curvature tensor of $\tilde{a}$. Your answer should be nonzero!

(c) Check that our 1-form $\tilde{b}$ has Riemannian norm $\|\,\tilde{b}\,\| = \epsilon < 1$ everywhere. Is it nowhere zero on M?

(d) Show that $\tilde{b}$ is parallel with respect to the Levi-Civita (Christoffel) connection of $\tilde{a}$.

(e) Verify directly that the Euler characteristic of $\mathbb{S}^n \times \mathbb{S}^1$ is zero by tabulating its cohomology. Explain how this also follows from the fact that the 1-form $\tilde{b}$ is nowhere zero.

(f) Write down the Finsler function F for the Randers space defined by $\tilde{a}$ and $\tilde{b}$. Explain why this F is of Berwald type, but is neither Riemannian nor locally Minkowskian.

References

[AIM] P. L. Antonelli, R. S. Ingarden, and M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology*, FTPH 58, Kluwer Academic Publishers, 1993.

[BL1] D. Bao and B. Lackey, *Randers surfaces whose Laplacians have completely positive symbol*, Nonlinear Analysis **38** (1999), 27–40.

[BM] S. Bácsó and M. Matsumoto, *On Finsler spaces of Douglas type, a generalization of the notion of Berwald space*, Publ. Math. Debrecen **51** (1997), 385–406.

[Fult] C. M. Fulton, *Parallel vector fields*, Proc. Amer. Math. Soc. **16** (1965), 136–137.

[HI] M. Hashiguchi and Y. Ichijyō, *On some special* (α, β)*-metrics*, Rep. Fac. Sci. Kagoshima Univ. **8** (1975), 39–46.

[HS1] D. Hrimiuc and H. Shimada, *On the* $\mathcal{L}$*-duality between Lagrange and Hamilton manifolds*, Nonlinear World **3** (1996), 613–641.

[HS2] D. Hrimiuc and H. Shimada, *On some special problems concerning the* $\mathcal{L}$*-duality between Finsler and Cartan spaces*, Tensor, N.S. **58** (1996), 48–61.

[Ki] S. Kikuchi, *On the condition that a space with* (α, β)*-metric be locally Minkowskian*, Tensor, N.S. **33** (1979), 242–246.

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

[M4] M. Matsumoto, *On Finsler spaces with Randers' metric and special forms of important tensors*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **14** (1974), 477–498.

[M5] M. Matsumoto, *Randers spaces of constant curvature*, Rep. on Math. Phys. **28** (1989), 249–261.

[M6] M. Matsumoto, *Remarks on Berwald and Landsberg spaces*, Cont. Math. **196** (1996), 79–82.

[M7H] M. Matsumoto and Hōjō, *A conclusive theorem on C-reducible Finsler spaces*, Tensor, N.S. **32** (1978), 225–230.

[M9] M. Matsumoto, *Theory of Finsler spaces with m-th root metric II*, Publ. Math. Debr. **49** (1996), 135–155.

[M10] M. Matsumoto, *Berwald connection of a Finsler space with an* (α, β)*-metric*, Tensor, N.S. **50** (1991), 18–21.

[Mir1] R. Miron, *General Randers spaces*, Lagrange and Finsler Geometry, P. L. Antonelli and R. Miron (eds.), Kluwer Academic Publishers, 1996, pp. 123–140.

[On] B. O'Neill, *Elementary Differential Geometry*, 2nd ed., Academic Press, 1997.

[Ra] G. Randers, *On an asymmetric metric in the four-space of general relativity*, Phys. Rev. **59** (1941), 195–199.

[SK] C. Shibata and M. Kitayama, *On Finsler spaces of constant positive curvature*, Proceedings of the Romanian–Japanese Colloquium on Finsler Geometry, Braşov, 1984, pp. 139–156.

[SSAY] C. Shibata, H. Shimada, M. Azuma, and H. Yasuda, *On Finsler spaces with Randers' metric*, Tensor, N.S. **31** (1977), 219–226.

[SaS] S. Sabau and H. Shimada, *Classes of Finsler spaces with* (α, β)*-metrics*, 1999 preprint.

[Sh] Z. Shen, *Curvature, distance and volume in Finsler geometry*, unpublished.

[YS] H. Yasuda and H. Shimada, *On Randers spaces of scalar curvature*, Rep. on Math. Phys. **11** (1977), 347–360.

Chapter 12

Constant Flag Curvature Spaces and Akbar-Zadeh's Theorem

12.0 Prologue

In §3.9, we encountered the **flag curvature**. As the name suggests, this quantity (denoted K) involves a location $x \in M$, a flagpole $\ell := \frac{y}{F(y)}$ with $y \in T_xM$, and a transverse edge $V \in T_xM$. The precise formula is quite elegantly given by (3.9.3):

$$K(\ell, V) \;:=\; \frac{V^i\ (\ell^j\ R_{jikl}\ \ell^l\ )\ V^k}{g(\ell,\ell)\ g(V,V)\ -\ [g(\ell,V)]^2} \;=\; \frac{V^i\ R_{ik}\ V^k}{g(V,V)\ -\ [g(\ell,V)]^2}\ .$$

Here, the tensor R_{ik} is known as the **predecessor** of the flag curvature.

In §3.10, we proved Schur's theorem about flag curvatures. It concerns connected Finsler manifolds of dimension at least three, and deals with the scenario in which K depends neither on the flagpole nor on the transverse edge. In other words, K is supposed to be a function of x only. In that case, the theorem asserts that K must actually be constant.

The present chapter is, in its entirety, about Finsler manifolds with constant flag curvature λ. We take the somewhat arrogant attitude that these spaces are self-evidently important. As we show:

- The forward geodesically complete $\lambda = 0$ spaces are necessarily locally Minkowskian. The $\lambda < 0$ spaces that are *both* forward and backward geodesically complete must be Riemannian. **These results are established under growth constraints imposed on the Cartan tensor**, and are due to Akbar-Zadeh [AZ]. They are treated in §12.3 and §12.4.
- Tampering with the completeness hypothesis or the growth constraint revives many interesting examples. For instance, the classification theorem of Yasuda–Shimada [YS] predicts the existence of a non-Riemannian Poincaré disc with $K = -\frac{1}{4}$. This fascinating Randers surface, known to Okada [Ok], is studied in §12.6. There is no contradiction with Akbar-Zadeh's result because, as we demonstrate, this example fails to be geodesically backward complete.
- The $\lambda > 0$ case is the least understood. Bryant has produced many interesting non-Riemannian examples on S^2 with $K = 1$. In §12.7, we present one from [Br2]. It is projectively flat (Exercise 12.7.3), and all its geodesics are the familiar great circles. They even have Finslerian length 2π.

The existence of so many esoteric non-Riemannian examples renders the "space form problem" rather complicated in Finsler geometry. By contrast, the Riemannian situation is considerably simpler, thanks to Hopf's classification theorem. That result is given a detailed treatment in §13.4.

12.1 Characterizations of Constant Flag Curvature

As in §3.10, let us use two abbreviations in order to reduce clutter. They are:

$$h_{ij} \ := \ g_{ij} \ - \ \ell_i \, \ell_j \ ,$$

the so-called **angular metric**, and

$$h_{ijk} \ := \ g_{ij} \, \ell_k \ - \ g_{ik} \, \ell_j \ .$$

To regain a feel for these quantities, one could verify the following properties:

* $h_{is}\,\ell^s = 0$.
* $h_{ijs}\,\ell^s = h_{ij}$.
* $g^{ij}\,h_{ik} = \delta^j{}_k - \ell^j\,\ell_k$.
* $g^{ij}\,h_{ij} = n-1$.
* $g^{ij}\,h_{ijk} = (n-1)\,\ell_k$.
* $h_{ij|k} = 0$.
* $h_{ijk|l} = 0$.

We first encountered that list in Exercise 3.10.3.

Specializing Proposition 3.10.1 to the case of constant flag curvature λ, we get:

Proposition 12.1.1. *Let (M,F) be a Finsler manifold. Let R be the hh-curvature of the Chern connection, for the portion of π^*TM over $TM \setminus 0$. The five statements listed below are mutually equivalent:*

- *(M,F) has constant flag curvature λ .*
- $R_{ii} = \lambda\,(\,g_{ii} - \ell_i\,\ell_i\,) =: \lambda\,h_{ii}$.
- $R_{ik} = \lambda\,(\,g_{ik} - \ell_i\,\ell_k\,) =: \lambda\,h_{ik}$.
- $R_{ikl} = \lambda\,(\,g_{ik}\,\ell_l - g_{il}\,\ell_k\,) =: \lambda\,h_{ikl}$.
- *The full hh-curvature tensor has the formula*

$$\begin{aligned} R_{jikl} \;=\;& \lambda\,(\,g_{ik}\,g_{jl} \;-\; g_{il}\,g_{jk}\,) \\ & -\,(\,\dot{A}_{ijl|k} \;-\; \dot{A}_{ijk|l} \;+\; \dot{A}_{isk}\,\dot{A}^s{}_{jl} \;-\; \dot{A}_{isl}\,\dot{A}^s{}_{jk}\,) \\ =\;& \lambda\,(\,g_{ik}\,g_{jl} \;-\; g_{il}\,g_{jk}\,) \\ & -\,\lambda\,(\,A_{ijk}\,\ell_l \;-\; A_{ijl}\,\ell_k\,) \;-\; (\,\dot{A}_{isk}\,\dot{A}^s{}_{jl} \;-\; \dot{A}_{isl}\,\dot{A}^s{}_{jk}\,)\ . \end{aligned}$$

And, given any of them, we have

$$\boxed{\ddot{A}_{ijk} \;+\; \lambda\,A_{ijk} \;=\; 0}\ , \tag{12.1.1}$$

where

$$\ddot{A}_{ijk} \;:=\; \dot{A}_{ijk|s}\,\ell^s \;=\; (\,A_{ijk|r}\,\ell^r\,)_{|s}\,\ell^s \;=\; A_{ijk|r|s}\,\ell^r\,\ell^s\ .$$

The above two formulas for R_{jikl} are equivalent because, when the flag curvature is a constant λ, we have

$$\dot{A}_{ijl|k} \;-\; \dot{A}_{ijk|l} \;=\; \lambda\,(\,A_{ijk}\,\ell_l \;-\; A_{ijl}\,\ell_k\,)\ .$$

See the second half of the proof of Proposition 3.10.1.

We conclude with some remarks about Finsler manifolds that satisfy the condition $\dot{A} = 0$. These are known as **Landsberg spaces**. They include locally Minkowski spaces ($R = 0 = P$), Riemannian spaces ($A = 0$), and more generally, Berwald spaces ($P = 0$).

* For Landsberg spaces, formula (12.1.1) reduces to

$$\lambda\, A_{ijk} \;=\; 0\,.$$

Thus A must vanish whenever $\lambda \neq 0$. Consequently,

All Landsberg spaces of nonzero constant flag curvature must be Riemannian.

This is a special case of Numata's theorem [Num].

* The statement of Proposition 12.1.1 implies that for Landsberg spaces of constant flag curvature λ, the full hh-Chern curvature has the formula

$$R_{jikl} \;=\; \lambda\,(\, g_{ik}\, g_{jl} \;-\; g_{il}\, g_{jk}\,)\,.$$

In contradistinction to that, the full hh-Berwald curvature ${}^bR_{jikl}$ has that form for *every* Finsler space of constant flag curvature λ. See Exercise 12.1.1 below.

Exercises

Exercise 12.1.1: In Exercise 3.2.2, we related the Berwald's hh-curvature ${}^bR_j{}^i{}_{kl}$ to that of Chern's. Using that relationship, explain why the following two statements are equivalent:

- (M, F) has constant flag curvature λ.
- ${}^bR_{jikl} = \lambda\,(\, g_{ik}\, g_{jl} - g_{il}\, g_{jk}\,)$.

This shows that the Berwald connection is particularly suited for studying *generic* Finsler manifolds of constant flag curvature.

Exercise 12.1.2:

(a) Suppose (M, F) is a Finsler manifold of *nonzero* constant flag curvature with $\ddot{A} = 0$. Show that it must be Riemannian.

(b) We explained why Landsberg spaces of *nonzero* constant flag curvature are necessarily Riemannian. Must Landsberg spaces of zero flag curvature be locally Minkowskian?

12.2 Useful Interpretations of $\dot{E}$ and $\ddot{E}$

On a Finsler manifold of constant flag curvature λ, the equation

$$\ddot{A} \;+\; \lambda\, A \;=\; 0\,,$$

namely, (12.1.1), is one whose usefulness was noted by Akbar-Zadeh [AZ]. It would be a Jacobi equation if those dots on A were time derivatives along some curve. Since the solutions of Jacobi-type equations can usually be written down explicitly, we might expect to obtain substantial information about the Cartan tensor A. As it stands though, the dots on $\ddot{A}$ have nothing to do with time derivatives. Happily, there is a simple construction of Akbar-Zadeh's that will bring fruition to the above trend of thought.

In retrospect, it is even necessary to consider tensor fields more general than the Cartan tensor A. So, let us consider

$$E \ := \ E_I \, dx^I \ ,$$

a section of the p-fold tensor product of the pulled-back bundle $\pi^* T^* M$. Here, I is a multi-index $i_1 \cdots i_p$, with

$$dx^I \ := \ dx^{i_1} \otimes \cdots \otimes dx^{i_p} \ .$$

We work out the proper context in which the equation

$$\ddot{E} \ + \ \lambda \, E \ = \ 0$$

makes sense as a Jacobi differential equation.

We digress to discuss the issue of invariance under positive rescaling in y. This has meaning for objects of the type

$$H \ := \ H_{IJ} \, dx^I \otimes \frac{1}{F^q} \, \delta y^J \ .$$

The J here is, like the I, also a multi-index, namely, $j_1 \cdots j_q$. And

$$\frac{1}{F^q} \, \delta y^J \ := \ \frac{\delta y^{j_1}}{F} \otimes \cdots \otimes \frac{\delta y^{j_q}}{F} \ .$$

A typical H that we have in mind is

$$\nabla_{\text{vert}} A \ := \ A_{ijk;l} \, dx^i \otimes dx^j \otimes dx^k \otimes \frac{\delta y^l}{F} \ .$$

All H under consideration are presumed invariant under $y \mapsto \lambda y$, with $\lambda > 0$. Since the forms dx and $\frac{\delta y}{F}$ already have the said invariance, the same must also hold for the components of H. In other words, its components are positively homogeneous of degree 0 in y. Euler's theorem (Theorem 1.2.1) then tells us that

$$y^s \, \frac{\partial}{\partial y^s} H_{IJ} \ = \ 0 \ .$$

Using the index gymnastics notation of §2.5, specifically (2.5.6), this reads

$$H_{IJ;s} \, \ell^s \ = \ 0 \ . \tag{12.2.1}$$

Examples of such H include:

* The fundamental tensor $g_{ij} \, dx^i \otimes dx^j$.
* The Cartan tensor $A_{ijk} \, dx^i \otimes dx^j \otimes dx^k$.
* The ubiquitous $\dot{A}_{ijk} \, dx^i \otimes dx^j \otimes dx^k$, where $\dot{A}_{ijk} := A_{ijk|l} \, \ell^l$.
* The horizontal covariant derivative $A_{ijk|l} \, dx^i \otimes dx^j \otimes dx^k \otimes dx^l$.

* The vertical covariant derivative $A_{ijk;l}\, dx^i \otimes dx^j \otimes dx^k \otimes \frac{\delta y^l}{F}$.

Exercises 12.2.1 and 12.2.2 explain why the last three have the asserted invariance.

Let us now return to the tensor $E := E_I\, dx^I$, which we presume to be invariant under positive rescaling in y. By imitating the definition (Exercise 2.5.5) of $\dot{A} := \nabla_{\hat{\ell}} A$, in which $\hat{\ell} := \ell^s \frac{\delta}{\delta x^s}$, we define

$$\dot{E} := \nabla_{\hat{\ell}} E = \dot{E}_I\, dx^I , \tag{12.2.2}$$

$$\ddot{E} := \nabla_{\hat{\ell}} \dot{E} = \ddot{E}_I\, dx^I . \tag{12.2.3}$$

The components $\dot{E}_I$ and $\ddot{E}_I$ are given explicitly by

$$\dot{E}_I = E_{I|s}\, \ell^s , \tag{12.2.4}$$

$$\ddot{E}_I = (E_{I|r}\, \ell^r)_{|s}\, \ell^s = E_{I|r|s}\, \ell^r\, \ell^s , \tag{12.2.5}$$

where in the second line we have made use of (2.5.14). It can be checked that both $\dot{E}$ and $\ddot{E}$ are, like E, again invariant under positive rescaling in y. See Exercises 12.2.1 and 12.2.2.

In order to interpret $\dot{E}$ and $\ddot{E}$ as time derivatives of sorts, we must at the very least restrict them to curves. Globally defined on $TM \setminus 0$ is the distinguished horizontal vector field $\hat{\ell} := \ell^s \frac{\delta}{\delta x^s}$. Its integral curves (x_t, y_t) are all horizontal and, as we show, suit our purpose. Here, our notation is such that (x_t, y_t) denotes the integral curve which passes through the point (x, y) at time $t = 0$.

Choose any g-orthonormal basis for $\pi^* TM$ at (x, y) and parallel-translate it along (x_t, y_t). Call the resulting field of bases $\{h_a(t)\}$. The almost g-compatibility criterion (2.4.6) of the Chern connection reads

$$dg_{ij} - g_{kj}\, \omega_i^{\ k} - g_{ik}\, \omega_j^{\ k} = 2\, A_{ijs}\, \frac{\delta y^s}{F} ,$$

where

$$\omega_j^{\ i} = \Gamma^i_{\ jk}\, dx^k .$$

Restrict it to the integral curve (x_t, y_t), which has velocity $\hat{\ell}$. The A term promptly drops out because $\frac{\delta y^s}{F}$, when evaluated on the horizontal $\hat{\ell}$, gives zero. The said criterion then becomes

$$\frac{dg_{ij}}{dt} - g_{kj}\, \Gamma^k_{\ is}\, \ell^s - g_{ik}\, \Gamma^k_{\ js}\, \ell^s = 0 .$$

Contract this equation with the components of two of our basis vectors, say h_a and h_b. The result can be manipulated into the form

$$\frac{d}{dt}\, g(h_a\, , h_b) = g\left(\nabla_{\hat{\ell}}\, h_a\, , h_b \right) + g\left(h_a\, , \nabla_{\hat{\ell}}\, h_b \right) . \tag{*}$$

But the right-hand side vanishes because h_a, h_b were obtained through parallel transport along (x_t, y_t). Therefore the entire field of bases $\{h_a(t)\}$ remains orthonormal at every point along the integral curve (x_t, y_t).

Let us suppress the t-dependence of $\{h_a(t)\}$ to minimize clutter. Form the components of E with respect to this new basis:

$$E_{a_1 \cdots a_p}(t) \;:=\; E(\,h_{a_1}\,, \ldots,\, h_{a_p}\,)\;. \tag{12.2.6}$$

Using the chain rule, the Leibniz rule, and the fact that each h_a is parallel along (x_t, y_t), it can be shown that

$$\frac{d}{dt}\, E_{a_1 \cdots a_p}(t) \;=\; \dot{E}(\,h_{a_1}\,, \ldots,\, h_{a_p}\,)\;, \tag{12.2.7}$$

$$\frac{d^2}{dt^2}\, E_{a_1 \cdots a_p}(t) \;=\; \ddot{E}(\,h_{a_1}\,, \ldots,\, h_{a_p}\,)\;. \tag{12.2.8}$$

Suppose E satisfies an equation of the type

$$\ddot{E} \;+\; \lambda\, E \;=\; 0 \tag{12.2.9}$$

for some constant λ. For instance, when dealing with Finsler manifolds of constant flag curvature λ, Proposition 12.1.1 assures us that the Cartan tensor A always satisfies such an equation. Now take the tensors E, $\ddot{E}$ in the said equation and evaluate them on the parallel g-orthonormal frame field $\{h_a\}$ we have just constructed. With the help of (12.2.6) and (12.2.8), we obtain

$$\left(\frac{d^2}{dt^2} \;+\; \lambda \right)\, E_{a_1 \cdots a_p}(t) \;=\; 0\;. \tag{12.2.10}$$

This is indeed a differential equation of the Jacobi type. So the goal of this section has been accomplished. However, two issues are worth recapitulating:

- In order to obtain the Jacobi equation (12.2.10), we have restricted $\ddot{E} + \lambda\, E = 0$ to an (arbitrary) integral curve (x_t, y_t) of the vector field $\hat{\ell}$ on $TM \smallsetminus 0$.
- The unknowns $E_{a_1 \cdots a_p}$ in that Jacobi equation are typically *not* the components of E in natural coordinates. Rather, they are the components of E relative to a parallel g-orthonormal frame field along (x_t, y_t).

Exercise

Exercise 12.2.1: In this exercise, we ascertain that if $E := E_i\, dx^i$ is invariant under positive rescaling in y, then so are $\nabla_{\mathrm{horiz}} E$ and $\nabla_{\mathrm{vert}} E$. As we mentioned in the section, the said invariance is equivalent to

$$E_{i;s}\, \ell^s \;=\; 0\;,$$

and one should expect to have to use this repeatedly.

(a) The tensor field $\nabla_{\text{horiz}} E$ is defined as $E_{i|j}\, dx^i \otimes dx^j$. Select an appropriate Ricci identity from Exercise 3.5.8 and use it to show that

$$E_{i|j;s}\, \ell^s \;=\; E_{i;s|j}\, \ell^s \,.$$

Then compute the right-hand side with the help of (2.5.14).

(b) The tensor field $\nabla_{\text{vert}} E$ is defined as $E_{i;j}\, dx^i \otimes \frac{\delta y^j}{F}$. Select an appropriate Ricci identity from Exercise 3.5.8 and use it to show that

$$E_{i;j;s}\, \ell^s \;=\; E_{i;j} \;+\; E_{i;s;j}\, \ell^s \,.$$

Then compute the right-hand side with the help of (2.5.15).

Exercise 12.2.2: Generalize Exercise 12.2.1 to the case of $E := E_I\, dx^I$.

Exercise 12.2.3: Although the flow of $\hat{\ell}$ does not project onto M, any specific integral curve (x_t, y_t) does. Denote its projection to M by x_t.

(a) Show that the velocity field of x_t is $\ell_{|(x_t, y_t)} = \frac{y_t}{F(y_t)}$. Hence it has constant unit speed.

(b) Is the canonical lift of x_t equal to (x_t, y_t)?

(c) Explain why the distance $d(x_a, x_b)$ (with $b > a$) from x_a to x_b is no more than $b - a$.

(d) Prove that x_t is a geodesic by using the following approach. Abbreviate (x_t, y_t) simply as (x, y). List the components of $\hat{\ell}$ with respect to $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$. Use that to write down the equations

$$\dot{x}^j \;=\; \ell^j \text{ (note: not } y^j) \,, \qquad \dot{y}^j \;=\; -\, y^i\, \frac{1}{F}\, N^j{}_i \,.$$

Then derive the statement

$$\ddot{x}^j \;+\; \gamma^j{}_{ik}\, \dot{x}^i\, \dot{x}^k \;=\; 0 \,.$$

Hint: imitate Exercise 5.1.8.

Exercise 12.2.4: Derive equation (*) in detail.

Exercise 12.2.5: As in the section, denote $\nabla_{\hat{\ell}} E$ by $\dot{E}$, and $\nabla_{\hat{\ell}}(\nabla_{\hat{\ell}} E)$ by $\ddot{E}$, where ∇ is the Chern connection. Let $h_1 = h_1(t)$, $\ldots$, $h_p = h_p(t)$ be any p parallel sections of $\pi^* TM$. Define

$$\begin{aligned} E_t &:= E(h_1, \ldots, h_p) \,, \\ \dot{E}_t &:= \dot{E}(h_1, \ldots, h_p) \,, \\ \ddot{E}_t &:= \ddot{E}(h_1, \ldots, h_p) \,. \end{aligned}$$

Prove that

$$\dot{E}_t \;=\; \frac{d}{dt} E_t \,, \qquad \ddot{E}_t \;=\; \frac{d^2}{dt^2} E_t \,.$$

In particular, one obtains (12.2.7) and (12.2.8).

Exercise 12.2.6: In this exercise, we use *first principles* to establish the following fact.

> Suppose M is forward complete with respect to the metric distance d induced by F, in the sense that every forward Cauchy sequence converges. (For example, this is so if M is compact.) Also, suppose our M is boundaryless as always. Then every integral curve (x_t, y_t) of the horizontal vector field $\hat{\ell}$ on $TM \setminus 0$ must be indefinitely forward extendible.

Adopt the following guidelines.

(a) Use the fact that $\hat{\ell}$ is defined at every point in $TM \setminus 0$, together with an existence theorem for first order ODEs, to show that if (x_t, y_t) is defined at $t = \tau$, then it is defined on an open time interval containing τ. This is where the fact that M is boundaryless gets used. Thus every maximally extended integral curve is defined on a maximal time interval $\alpha < t < \beta$.

(b) Suppose $\beta \neq \infty$. Take a Cauchy sequence $\{t_i\}$ in $\mathbb{R}$ such that $t_i \nearrow \beta$. Check that $\{x_{t_i}\}$ is a forward Cauchy sequence (see §6.2D). By the hypothesized forward completeness, it converges to some $x \in M$. Show that

$$\lim_{t \to \beta^-} x_t = x .$$

Hence we can define x_β to be x.

(c) Next, recall from Exercise 2.3.5 that $F(y_t)$ has constant value (say c) for all t because (x_t, y_t) is a horizontal curve. Consider the collection of tangent Finsler spheres

$$\{ \mathrm{S}_{x_t}(c) : \ \beta - \epsilon \leqslant t \leqslant \beta \} ,$$

where $\epsilon > 0$ is small. They form a compact subset of $TM \setminus 0$ (but with respect to which topology?), and contain all y_{t_i} for sufficiently large i. Show that the latter admits a subsequence that converges to some $y \in \mathrm{S}_{x_\beta}(c)$. Explain why one must then have

$$\lim_{t \to \beta^-} y_t = y .$$

So we can define y_β to be y.

(d) There exists an integral curve of $\hat{\ell}$ that passes through the point (x_β, y_β). Again, the fact that M is boundaryless has been used here. Explain why this integral curve's velocity field matches with that of our supposedly maximal curve (x_t, y_t), thereby furnishing a smooth extension of the latter to $t = \beta$ and beyond. This then contradicts the assumed maximality of the time interval $\alpha < t < \beta$.

Exercise 12.2.7: Reformulate the result in Exercise 12.2.6 for the case in which (M, F) is backward complete (see §6.2D). Then prove it using *first principles*.

Exercise 12.2.8: In this section, we have worked exclusively with the integral curves of the distinguished horizontal vector field $\hat{\ell}$. Equally important is the horizontal vector field $F\,\hat{\ell}$. Indeed, according to Exercise 5.1.6, the constant speed Finslerian geodesics on M are, through the canonical lift procedure, bijective with the integral curves of $F\,\hat{\ell}$.

Let us now rederive the results of Exercises 12.2.6 and 12.2.7 from another perspective. Show that the following statements are equivalent:

(a) The integral curves of $\hat{\ell}$ are indefinitely forward (resp., backward) extendible.

(b) The integral curves of $F\,\hat{\ell}$ are indefinitely forward (resp., backward) extendible.

(c) (M, F) is forward (resp., backward) geodesically complete.

(d) Every forward (resp., backward) Cauchy sequence in (M, F) is convergent.

Hints: recall the Hopf–Rinow theorem (Theorem 6.6.1); also, how are the integral curves of $F\,\hat{\ell}$ related to those of $\hat{\ell}$?

12.3 Growth Rates of Solutions of $\ddot{E} + \lambda\, E = 0$

In order to measure the growth rate of our tensor E, we need a norm. And to put that growth rate in perspective, we must stipulate the size of the domain on which that is being measured.

First, recall from §6.2 that the positively homogeneous Finsler function F induces a metric distance $d(p, q)$ on M which is positive-definite, satisfies the triangle inequality, but which may not be symmetric. In fact, it has the property $d(q, p) = d(p, q)$ only when $F(x, y)$ is absolutely homogeneous of degree one in y.

Given any point $p \in M$ and a radius $r > 0$, we have defined in §6.2B the notion of **forward metric balls**:

$$\mathcal{B}_p^+(r) \ := \ \{ x \in M : \ d(p, x) \ < \ r \} \ .$$

Likewise, one also defines (Exercise 6.2.10) the **backward metric balls**:

$$\mathcal{B}_p^-(r) \ := \ \{ x \in M : \ d(x, p) \ < \ r \} \ .$$

Let

$$\mathcal{B}_p(r) \ := \ \mathcal{B}_p^+(r) \ \cup \ \mathcal{B}_p^-(r) \ .$$

The three sets described above are neighborhoods in M containing the point p. They are not to be confused with the tangent Finsler balls $\mathrm{B}_p(r)$

introduced in §6.1. The latter are neighborhoods in the tangent space T_pM instead of M.

The neighborhoods $\mathcal{B}_p(r)$ serve as domains on which we measure the growth rate (in terms of r) of certain tensors. In the definition of $\mathcal{B}_p(r)$, it is the *union* and not the intersection that we want. Let us indicate why. Take any integral curve (x_t, y_t) of the vector field $\hat{\ell}$ on $TM \setminus 0$. According to Exercise 12.2.3, the projection of that integral curve onto M, denoted x_t, is a unit speed geodesic. It passes through the point x at time $t = 0$. Define

$$r_o \ := \ \max\{\, d(\Theta, x)\,,\, d(x, \Theta)\,\}\ ,$$

which measures how far the point x is from a certain designated "origin" Θ. Let τ be any positive number such that the curve x_t is defined on the time interval $-\tau < t \leqslant 0$. Then it can be shown (Exercise 12.3.2) that

$$\{\, x_t :\ -\tau < t \leqslant 0\,\} \ \subset\ \mathcal{B}_x(\tau) \ \subseteq\ \mathcal{B}_\Theta(r_o + \tau)\ . \tag{12.3.1}$$

Likewise, if x_t is defined on $0 \leqslant t < \tau$, then

$$\{\, x_t :\ 0 \leqslant t < \tau\,\} \ \subset\ \mathcal{B}_x(\tau) \ \subseteq\ \mathcal{B}_\Theta(r_o + \tau)\ . \tag{12.3.2}$$

Therefore, neighborhoods of the form $\mathcal{B}_\Theta(r)$ do contain the important curves along which we measure the growth rates of the tensor field E.

Next, we turn to norms. The pointwise norm of E on $TM \setminus 0$ is simply

$$\sqrt{E_{a_1 \dots a_p}\ E^{a_1 \dots a_p}}\ .$$

This norm is independent of the choice of coframe, be it natural or g-orthonormal. Given any punctured tangent space $T_xM \setminus 0$, our hypothesis on E says that it is constant along each ray which emanates from the origin. Thus maximizing the pointwise norm over $T_xM \setminus 0$ is the same as maximizing it over the indicatrix S_x. Let us then define

$$\|\, E\,\|_x \ := \ \max_{y \in \mathrm{S}_x} \sqrt{E_{a_1 \dots a_p}\ E^{a_1 \dots a_p}}\ .$$

Since S_x is compact, $\|\, E\,\|_x$ is well defined.

* There is an advantage in using a g-orthonormal coframe to express E, for then its components satisfy the following inequality:

$$|\, E_{a_1 \dots a_p}(x, y)\,| \ \leqslant\ \|\, E\,\|_x\ . \tag{12.3.3}$$

  See Exercise 12.3.3.

* Fix a point $\Theta \in M$ once and for all. It will serve as a **reference origin**. We say that

$$\sup_{x \in \mathcal{B}_\Theta(r)} \|\, E\,\|_x \ =\ o[\, f(r)\,]$$

  if that supremum grows *more slowly* than $f(r)$ as $r \to \infty$. Note that this is the **little "oh"** notation.

We are now ready to prove a result of Akbar-Zadeh's [AZ]:

Proposition 12.3.1 (Akbar-Zadeh). *Let (M, F) be a connected boundaryless Finsler manifold. Let Θ be any designated "origin" in M. Let E be any section of the tensor products of the portion of π^*T^*M over $TM \setminus 0$. Let's suppose that E is invariant under positive rescaling in y and obeys the equation*

$$\ddot{E} + \lambda E = 0 , \quad \textit{where } \lambda \textit{ is a constant} .$$

(a) *Suppose $\lambda = 0$ and (M, F) is forward geodesically complete.*

$$\textit{If} \sup_{q \in \mathcal{B}_\Theta(r)} \|E\|_q = o[r] , \quad \textit{then } \dot{E} = 0 .$$

(b) *Suppose $\lambda < 0$ and (M, F) is both forward and backward geodesically complete.*

$$\textit{If} \sup_{q \in \mathcal{B}_\Theta(r)} \| E \textbf{ or } \dot{E} \|_q = o[e^{\sqrt{-\lambda}\, r}] , \quad \textit{then } E = 0 .$$

Remark: Forward (resp., backward) geodesic completeness is equivalent to the hypothesis that every forward (resp., backward) Cauchy sequence converges. See §6.6 or Exercise 12.2.8.

Proof.

Fix an arbitrary $(x, y) \in TM \setminus 0$. We show that the conclusions hold at that point.

Recall the distinguished horizontal vector field $\hat{\ell}$ discussed in §12.2. Let (x_t, y_t) be the integral curve of $\hat{\ell}$ that passes through the point (x, y) at $t = 0$. According to Exercise 12.2.8:

- When (M, F) is forward geodesically complete, this integral curve on $TM \setminus 0$ is defined at all $t \in [0, \infty)$.
- When (M, F) is backward geodesically complete, this integral curve on $TM \setminus 0$ is defined at all $t \in (-\infty, 0]$.

In §12.2, we constructed a field of bases $\{h_a(t)\}$ along the said integral curve. For each t, this is a g-orthonormal basis for the fiber of π^*TM over the point (x_t, y_t). Furthermore, each $h_a(t)$ is parallel with respect to the Chern connection on π^*TM. Using this field of bases, we defined in (12.2.6) the components of E:

$$E_{a_1 \cdots a_p}(t) := E(h_{a_1}, \ldots, h_{a_p}) .$$

Then we showed that the equation $\ddot{E} + \lambda E = 0$, where $\ddot{E}$ means $\nabla_{\hat{\ell}}(\nabla_{\hat{\ell}} E)$, is equivalent to

$$(*) \qquad \left(\frac{d^2}{dt^2} + \lambda \right) E_{a_1 \cdots a_p}(t) = 0 .$$

Let $r_o := \max\{ d(\Theta, x)\, ,\, d(x, \Theta) \}$. We see from (12.3.1), (12.3.2) that

$$\{ x_t :\ -\tau < t \leqslant 0 \} \subset \mathcal{B}_\Theta(r_o + \tau)\ ,$$
$$\{ x_t :\ 0 \leqslant t < \tau \} \subset \mathcal{B}_\Theta(r_o + \tau)\ .$$

Together with (12.3.3), these give

$$(^{**}) \qquad \sup_{-\tau < t \leqslant 0} |E_{a_1 \cdots a_p}(t)| \leqslant \sup_{-\tau < t \leqslant 0} \|E\|_{x_t} \leqslant \sup_{q \in \mathcal{B}_\Theta(r_o + \tau)} \|E\|_q$$

and

$$(^{***}) \qquad \sup_{0 \leqslant t < \tau} |E_{a_1 \cdots a_p}(t)| \leqslant \sup_{0 \leqslant t < \tau} \|E\|_{x_t} \leqslant \sup_{q \in \mathcal{B}_\Theta(r_o + \tau)} \|E\|_q\ .$$

Case (a) $\lambda = 0$ **:**

The solutions of the Jacobi type equation (*) are

$$E_{a_1 \cdots a_p}(t) = \alpha + \beta t\ .$$

Using this and the hypothesis

$$\sup_{q \in \mathcal{B}_\Theta(r)} \|E\|_q = o[r]\ ,$$

we infer from (***) that

$$\sup_{0 \leqslant t < \tau} |\alpha + \beta t| \leqslant o[r_o + \tau]\ .$$

Since our integral curve is defined for all $t \in [0, \infty)$, we can let $\tau \to \infty$. Now, $o[r_o + \tau]$ means *slower than* linear growth in τ, thus β must vanish. So $E_{a_1 \cdots a_p}(t)$ is constant. According to (12.2.8), the components of $\dot{E} := \nabla_{\hat{\ell}} E$ equal $\frac{d}{dt} E_{a_1 \cdots a_p}(t)$, which are then zero. We have therefore deduced that $\dot{E} = 0$ all along (x_t, y_t). In particular, $\dot{E}$ vanishes at our fixed (x, y), as desired.

Case (b) $\lambda < 0$ **:**

In this case, the solutions of the Jacobi type equation (*) are

$$E_{a_1 \cdots a_p}(t) = \alpha\, e^{\sqrt{-\lambda}\, t} + \beta\, e^{-\sqrt{-\lambda}\, t}\ .$$

By (12.2.8), the components of $\dot{E}$ are given by $\frac{d}{dt} E_{a_1 \cdots a_p}(t)$. So,

$$\dot{E}_{a_1 \cdots a_p}(t) = \sqrt{-\lambda}\ \alpha\, e^{\sqrt{-\lambda}\, t} - \sqrt{-\lambda}\ \beta\, e^{-\sqrt{-\lambda}\, t}\ .$$

- Suppose

$$\sup_{q \in \mathcal{B}_\Theta(r)} \|E\|_q = o[e^{\sqrt{-\lambda}\, r}]\ .$$

Then (**) and (***) together tell us that

$$\sup_{-\tau < t < \tau} |\alpha\, e^{\sqrt{-\lambda}\, t} + \beta\, e^{-\sqrt{-\lambda}\, t}| \leqslant o[e^{\sqrt{-\lambda}\, (r_o + \tau)}]\ .$$

Note that $e^{\sqrt{-\lambda}\,t}$ exhibits exponential growth for $t > 0$ and $e^{-\sqrt{-\lambda}\,t}$ undergoes the same for $t < 0$. In the above growth constraint, the hypothesized forward and backward geodesic completeness allows us to let $\tau \to \infty$ and $\tau \to -\infty$. Since $o[\,e^{\sqrt{-\lambda}\,(r_o+\tau)}\,]$ means *slower than* exponential growth in τ, we must have $\alpha = 0 = \beta$. Thus $E = 0$ all along (x_t, y_t), and at (x, y) in particular.

- Suppose

$$\sup_{q \in \mathcal{B}_\Theta(r)} \|\dot{E}\|_q \;=\; o[\,e^{\sqrt{-\lambda}\,r}\,] \,.$$

Applying (**) and (***) to the tensor $\dot{E} := \nabla_{\hat{\ell}} E$, we have

$$\sup_{-\tau < t < \tau} \;|\,\sqrt{-\lambda}\;\alpha\, e^{\sqrt{-\lambda}\,t} \;-\; \sqrt{-\lambda}\;\beta\, e^{-\sqrt{-\lambda}\,t}\,| \;\leqslant\; o[\,e^{\sqrt{-\lambda}\,(r_o+\tau)}\,] \,.$$

Again, letting $\tau \to \pm\infty$ forces α and β to vanish. Thus $E = 0$ all along (x_t, y_t), and at (x, y) in particular. $\square$

Exercises

Exercise 12.3.1:

(a) Produce an example in which the forward metric ball $\mathcal{B}_p^+(r)$ and the backward metric ball $\mathcal{B}_p^-(r)$ are different.

(b) Draw a typical pair of forward and backward balls.

Hint: consult §12.6 only if absolutely necessary.

Exercise 12.3.2: Let (x_t, y_t) be any integral curve of the vector field $\hat{\ell}$. Let x_t be its projection onto M. Review Exercise 12.2.3.

(a) Let r_o be as defined in the section. Check that

$$\mathcal{B}_x(\tau) \;\subseteq\; \mathcal{B}_\Theta(r_o + \tau)$$

for every positive number τ.

(b) Let τ be any positive number such that x_t is defined on the time interval $-\tau < t \leqslant 0$. Show that

$$\{\, x_t : \; -\tau < t \leqslant 0 \,\} \;\subset\; \mathcal{B}_x(\tau) \,.$$

(c) Let τ be any positive number such that x_t is defined on the time interval $0 \leqslant t < \tau$. Show that

$$\{\, x_t : \; 0 \leqslant t < \tau \,\} \;\subset\; \mathcal{B}_x(\tau) \,.$$

Exercise 12.3.3:

(a) Show that with respect to any g-orthonormal coframe, the components of E satisfy the inequality

$$|\, E_{a_1 \ldots a_p}(x, y) \,| \;\leqslant\; \|\, E \,\|_x \,.$$

(b) Exemplify how such an inequality might fail in the natural coordinate coframe $\{dx^i\}$.

Exercise 12.3.4: Let (M,F) be a complete Finsler manifold of constant flag curvature 1. Let $\sigma(t)$ be a unit speed geodesic with velocity field T. Denote by $\hat{\sigma}$ the canonical lift (σ,T) of σ. Suppose $U=U(t)$, $V=V(t)$, $W=W(t)$ are parallel sections of π^*TM along the canonical lift $\hat{\sigma}$. Define

$$A_t := A(U,V,W) \ , \quad \dot{A}_t := \dot{A}(U,V,W) \ .$$

Show that

$$\begin{aligned} A_t &= A_0 \cos t + \dot{A}_0 \sin t \ , \\ \dot{A}_t &= - A_0 \sin t + \dot{A}_0 \cos t \ . \end{aligned}$$

12.4 Akbar-Zadeh's Rigidity Theorem

When the flag curvature is a nonpositive constant, one has rigidity under additional *analytical* hypotheses. §12.6 describes an example which shows that one of the said hypotheses is sharp. Also, for the case of constant positive flag curvature, there is no rigidity (see §12.7) unless one imposes additional curvature criteria [such as the one mentioned after (12.1.1)].

Theorem 12.4.1 (Akbar-Zadeh) [AZ]. *Let (M,F) be a connected boundaryless Finsler manifold of constant flag curvature λ. Let ∇ be the Chern connection on the portion of π^*TM over $TM \setminus 0$. Let Θ be any designated "origin" in M.*

(a) *Suppose $\lambda < 0$ and (M,F) is both forward and backward geodesically complete. If*

$$\sup_{x \in \mathcal{B}_\Theta(r)} \| A \text{ or } \dot{A} \|_x = o[\, e^{\sqrt{-\lambda}\, r} \,] \ ,$$

then (M,F) must be Riemannian.

(b) *Suppose $\lambda = 0$ and (M,F) is forward geodesically complete. If*

$$\sup_{x \in \mathcal{B}_\Theta(r)} \|A\|_x = o[\, r \,] \ ,$$

then $\dot{A} = 0$ and hence F is of Landsberg type. If, in addition,

$$\sup_{x \in \mathcal{B}_\Theta(r)} \|\nabla_{vert} A\|_x = o[\, r \,] \ ,$$

then (M,F) is locally Minkowskian.

Remark: Forward (resp., backward) geodesic completeness is equivalent to the hypothesis that every forward (resp., backward) Cauchy sequence converges. See §6.6 or Exercise 12.2.8.

Proof.

According to Proposition 12.1.1, given a Finsler manifold of constant flag curvature λ, the Cartan tensor A always satisfies the equation

$$(*) \qquad \ddot{A} + \lambda A = 0 .$$

- Suppose $\lambda < 0$. Imposing the said growth constraint on A or $\dot{A}$, and using Proposition 12.3.1, we see that $A = 0$. In other words, (M, F) must be Riemannian.
- Next, suppose $\lambda = 0$. The first constraint

$$\sup_{x \in \mathcal{B}_\Theta(r)} \|A\|_x = o[r]$$

says that A has slower than linear growth. So Proposition 12.3.1 tells us that $\dot{A}$ must vanish. That is, F is of Landsberg type.

It remains to ascertain the following:

Let us be given a Landsberg space (M, F) with zero flag curvature. If the tensor $\nabla_{\text{vert}} A$ has slower than linear growth, then (M, F) must be locally Minkowskian.

This is the companion of a statement we made near the end of §12.1:

All Landsberg spaces of nonzero constant flag curvature must be Riemannian.

Note that for Landsberg spaces with $\lambda = 0$, equation (*) reduces to $0 = 0$, which gives no information. Remarkably, Akbar-Zadeh has found a new pair of equations that saves the day.

A useful identity for Landsberg spaces:

On any Landsberg space, we have the identity

$$(12.4.1) \qquad \boxed{A_{ijk;l|s}\, \ell^s = - A_{ijk|l}} .$$

This can be derived as follows:

* Apply the Ricci identity or interchange formula (3.6.2) to A_{ijk} and impose the Landsberg criterion $\dot{A} = 0$. The resulting formula reads

$$A_{ijk;l|s} - A_{ijk|s;l} = - A_{vjk}\, P_i{}^v{}_{sl} - A_{ivk}\, P_j{}^v{}_{sl} - A_{ijv}\, P_k{}^v{}_{sl} .$$

* Contract the above with ℓ^s. Note that by the symmetry condition (3.2.3) and the Bianchi identity (3.4.9), we have

$$\ell^s\, P_i{}^v{}_{sl} = \ell^s\, P_s{}^v{}_{il} = -\dot{A}^v{}_{il} = 0 .$$

Thus we get

$$A_{ijk;l|s}\, \ell^s = A_{ijk|s;l}\, \ell^s .$$

With the help of (2.5.15), and the criterion $\dot{A} = 0$, the right-hand side is found to be equal to $-A_{ijk|l}$.

Thus (12.4.1) is valid on Landsberg spaces, as claimed.

Akbar-Zadeh's technique:

For ease of exposition, let us introduce the temporary abbreviations

$$\mathcal{H} := \nabla_{\text{horiz}} A \leftrightarrow A_{ijk|l} , \tag{12.4.2}$$

$$\mathcal{V} := \nabla_{\text{vert}} A \leftrightarrow A_{ijk;l} . \tag{12.4.3}$$

Then the identity (12.4.1) looks like

$$\dot{\mathcal{V}} = -\mathcal{H} . \tag{12.4.4}$$

Observe that

$$\dot{\mathcal{H}}_{ijkl} = A_{ijk|l|s}\, \ell^s .$$

By Proposition 12.1.1, the curvature R_{jikl} vanishes on a Landsberg space with zero flag curvature. Using this and the Ricci identity (3.6.1), we see that

$$A_{ijk|l|s}\, \ell^s = A_{ijk|s|l}\, \ell^s .$$

But the right-hand side is zero because of (2.5.14) and $\dot{A} = 0$. So, in the present setting one has

$$\dot{\mathcal{H}} = 0 .$$

Formula (12.4.4) then yields

$$\ddot{\mathcal{V}} = 0 . \tag{12.4.5}$$

Our next step is to "integrate" (12.4.5). Note that $\mathcal{V}$ is invariant under positive rescaling in y. Also:

* Condition (12.4.5) is a Jacobi-type equation concerning $\mathcal{V}$.
* The abbreviation (12.4.3), together with our hypothesis about the quantity $\nabla_{\text{vert}} A$, gives us a growth constraint on $\mathcal{V}$.
* Finally, we have the assumed forward completeness of M.

Therefore all the hypotheses in Proposition 12.3.1 are met, and we can conclude that

$$\dot{\mathcal{V}} = 0 .$$

In view of (12.4.4), $\mathcal{H}$ vanishes as well. This means that

$$A_{ijk|l} = 0 .$$

Through the constitutive relation (3.4.11), which expresses the curvature P in terms of $A_{ijk|l}$, we see that P must vanish. (So our flat Landsberg space is actually of Berwald type.)

We have just shown that the hv-Chern curvature P is zero. It has been remarked above that $\lambda = 0$ effects the vanishing of the hh-Chern curvature R (Proposition 12.1.1) as well. Therefore, by Proposition 10.5.1, the Landsberg space in question must be locally Minkowskian. □

Suppose M is compact with respect to its manifold topology. By Exercises 6.2.11 and 6.6.4, (M, F) is then both forward and backward geodesically complete. Furthermore, given compactness, the stated asymptotic growth constraints are automatically satisfied. In view of this, Theorem 12.4.1 has a most elegant corollary.

Corollary 12.4.2 (Akbar-Zadeh) [AZ]. *Let (M, F) be a compact connected boundaryless Finsler manifold of constant flag curvature λ.*

(a) *If $\lambda < 0$, then (M, F) is Riemannian.*

(b) *If $\lambda = 0$, then (M, F) is locally Minkowskian.*

Exercises

Exercise 12.4.1: Prove in detail that on a Landsberg space, the identity

$$A_{ijk;l|s}\, \ell^s \;=\; -\, A_{ijk|l}$$

holds. It was used crucially in the proof of Akbar-Zadeh's theorem (Theorem 12.4.1).

Exercise 12.4.2: Let (M, F) be a boundaryless simply connected Finsler manifold of zero flag curvature. Suppose it is forward complete with respect to the metric distance function d induced by F. Let Θ be some designated "origin" in M. Prove that if

$$\lim_{r \to \infty} \; \sup_{x \in M \smallsetminus \mathcal{B}_\Theta(r)} \; \|A\|_x \;=\; 0 \,,$$

then the Finsler structure F is Riemannian.

Exercise 12.4.3:

(a) Check that Finsler manifolds with $R_j{}^i{}_{kl} = 0$ necessarily have zero flag curvature. Explain why the converse is false. That is, having zero flag curvature does *not* imply that the hh-Chern curvature has to vanish.

(b) Recall the constitutive formula (3.5.6) for the hh-Chern curvature R, and Exercise 3.8.4 about the hh-Berwald curvature ${}^bR$. Show that a Finsler manifold has zero flag curvature if and only if ${}^bR_j{}^i{}_{kl} = 0$.

(c) In view of the Akbar-Zadeh theorem, is it difficult to find a Finsler manifold with $R = 0$ but $P \neq 0$? Do such examples exist?

12.5 Formulas for Machine Computations of K

12.5 A. The Geodesic Spray Coefficients

The quantities

$$G_i = \gamma_{ijk}\, y^j\, y^k = \frac{1}{2}\left(g_{ij,x^k} - g_{jk,x^i} + g_{ki,x^j}\right) y^j\, y^k$$

have been introduced in Exercise 2.3.2. Define

$$\boxed{\mathcal{L} := \frac{1}{2}\, F^2} \,. \tag{12.5.1}$$

Note that the fundamental tensor is given by

$$g_{ij} = \mathcal{L}_{y^i y^j} \,.$$

Since $\mathcal{L}_{y^i}$ is positively homogeneous of degree 1 in y, Euler's theorem (Theorem 1.2.1) gives

$$g_{ij}\, y^j = \mathcal{L}_{y^i} \,. \tag{12.5.2}$$

With this, one can check that

$$G_i = \mathcal{L}_{y^i x^j}\, y^j - \mathcal{L}_{x^i} \,. \tag{12.5.3}$$

Raising the index i gives

$$G^s := g^{si}\, G_i \,. \tag{12.5.4}$$

Now we specialize to Finsler surfaces. In order to reduce clutter, let us relabel the natural coordinates x^1, x^2 as x, y. The induced global coordinates y^1, y^2 on each tangent plane are relabeled as u, v. That is:

$$\boxed{\begin{array}{ccc} x & \leftrightarrow & x^1 \\ y & \leftrightarrow & x^2 \\ u & \leftrightarrow & y^1 \\ v & \leftrightarrow & y^2 \end{array}} \,. \tag{12.5.5}$$

Strategic uses of Euler's theorem, together with

$$\begin{aligned} \mathcal{L}_{xu}\, u &= 2\, \mathcal{L}_x - \mathcal{L}_{xv}\, v \\ \mathcal{L}_{yv}\, v &= 2\, \mathcal{L}_y - \mathcal{L}_{yu}\, u \end{aligned}$$

(which are themselves consequences of Euler's theorem), lead to explicit formulas for our G^1 and G^2 (twice those in [AIM]). These read:

$$\begin{aligned} G^1 &= \frac{(\mathcal{L}_{vv}\, \mathcal{L}_x - \mathcal{L}_{vx}\, \mathcal{L}_v) - (\mathcal{L}_{uv}\, \mathcal{L}_y - \mathcal{L}_{uy}\, \mathcal{L}_v)}{\det(g)} \,, \\ G^2 &= \frac{(\mathcal{L}_{uu}\, \mathcal{L}_y - \mathcal{L}_{uy}\, \mathcal{L}_u) - (\mathcal{L}_{vu}\, \mathcal{L}_x - \mathcal{L}_{vx}\, \mathcal{L}_u)}{\det(g)} \,. \end{aligned}$$

Here,

$$\text{(12.5.6)} \qquad \boxed{\det(g) := \mathcal{L}_{uu}\,\mathcal{L}_{vv} - (\mathcal{L}_{uv})^2}\ .$$

It turns out that some curvature-related quantities look slightly simpler if expressed in terms of $\frac{1}{2}G^i$ rather than G^i. Thus we introduce (12.5.7a, 12.5.7b)

$$\boxed{\begin{aligned} G &:= \frac{1}{2}\,G^1 = \frac{(\mathcal{L}_{vv}\,\mathcal{L}_x - \mathcal{L}_{vx}\,\mathcal{L}_v) - (\mathcal{L}_{uv}\,\mathcal{L}_y - \mathcal{L}_{uy}\,\mathcal{L}_v)}{2\,\det(g)}\ , \\ H &:= \frac{1}{2}\,G^2 = \frac{(\mathcal{L}_{uu}\,\mathcal{L}_y - \mathcal{L}_{uy}\,\mathcal{L}_u) - (\mathcal{L}_{vu}\,\mathcal{L}_x - \mathcal{L}_{vx}\,\mathcal{L}_u)}{2\,\det(g)}\ . \end{aligned}}$$

12.5 B. The Predecessor of the Flag Curvature

In Exercise 3.3.4, we gave a strategy for deriving the elegant formula

$$\text{(12.5.8)} \qquad R^i{}_k = \ell^j \left(\frac{\delta}{\delta x^k}\frac{N^i{}_j}{F} - \frac{\delta}{\delta x^j}\frac{N^i{}_k}{F} \right)\ .$$

The $N^i{}_s$ are the coefficients of the nonlinear connection (see §2.3).

Let us manipulate the first term $\ell^j \frac{\delta}{\delta x^k}\frac{N^i{}_j}{F}$ on the right-hand side as follows:

* We use Exercise 2.5.6 to move ℓ past $\frac{\delta}{\delta x}$, thereby gaining a "correction" term
$$\frac{1}{F^2}\,N^i{}_j\,N^j{}_k\ .$$
This correction term can be expressed in terms of y derivatives of the geodesic spray coefficients G^i, using Exercise 2.3.3.
* It can be seen from (2.3.2b) that
$$\frac{1}{F}\,\ell^j\,N^i{}_j = \frac{1}{F^2}\,G^i\ .$$
Since part (b) of Exercise 2.3.5 tells us that F is horizontally constant, the factor $1/F^2$ can be moved outside the $\frac{\delta}{\delta x}$ derivative.
* Finally, we use (2.3.3) to spell out the operator $\frac{\delta}{\delta x}$.

These maneuvers give

$$\ell^j\,\frac{\delta}{\delta x^k}\frac{N^i{}_j}{F} = \frac{1}{F^2}\left[(G^i)_{x^k} - \frac{1}{4}\,(G^i)_{y^j}\,(G^j)_{y^k} \right]\ .$$

Less work is required on the second term $-\ell^j\,\frac{\delta}{\delta x^j}\frac{N^i{}_k}{F}$:

* We use the horizontal constancy of F to move the $1/F$ past the $\frac{\delta}{\delta x}$.
* Then we spell out $\frac{\delta}{\delta x}$ and use $\ell^j\,N^s{}_j = G^s/F$.

The result is

$$-\ell^j \, \frac{\delta}{\delta x^j} \frac{N^i{}_k}{F} \;=\; \frac{1}{F^2} \left[-\frac{1}{2}\, y^j \,(G^i)_{y^k x^j} \;+\; \frac{1}{2}\, G^j \,(G^i)_{y^k y^j} \right] .$$

Together, they tell us that:

(12.5.9)

$$\boxed{\begin{aligned} F^2 \, R^i{}_k \;=\; & 2\,(\bar{G}^i)_{x^k} \;-\; (\bar{G}^i)_{y^j}\,(\bar{G}^j)_{y^k} \\ & -\, y^j\,(\bar{G}^i)_{y^k x^j} \;+\; 2\,\bar{G}^j\,(\bar{G}^i)_{y^k y^j} \end{aligned}} ,$$

where

$$\boxed{\bar{G}^s \;:=\; \frac{G^s}{2}} .$$

Specializing to two dimensions, and using the relabeling that we have introduced above, (12.5.9) gives:

(12.5.10)
$$F^2\,R^1{}_1 \;=\; (G_{xv} - G_{yu})\,v + 2\,G\,G_{uu} + 2\,HG_{uv} - G_u\,G_u - G_v H_u \;,$$

(12.5.11)
$$F^2\,R^2{}_2 \;=\; (H_{yu} - H_{xv})\,u + 2\,HH_{vv} + 2\,GH_{vu} - H_v H_v - H_u G_v \;,$$

(12.5.12)
$$F^2\,R^1{}_2 \;=\; (G_{yu} - G_{xv})\,u + 2\,HG_{vv} + 2\,G\,G_{uv} - G_v H_v - G_v\,G_u \;,$$

(12.5.13)
$$F^2\,R^2{}_1 \;=\; (H_{xv} - H_{yu})\,v + 2\,GH_{uu} + 2\,HH_{vu} - H_u G_u - H_u H_v \;.$$

12.5 C. Maple Codes for the Gaussian Curvature

Part (e) of Exercise 4.4.8 shows that, in *natural coordinates*, one has:

$$\begin{pmatrix} F^2\,R^1{}_1 & F^2\,R^1{}_2 \\ F^2\,R^2{}_1 & F^2\,R^2{}_2 \end{pmatrix} \;=\; K \begin{pmatrix} y^2\,\mathcal{L}_{y^2} & -\,y^1\,\mathcal{L}_{y^2} \\ -\,y^2\,\mathcal{L}_{y^1} & y^1\,\mathcal{L}_{y^1} \end{pmatrix} .$$

These, together with (12.5.10)–(12.5.13), lead to the following computational formulas for the Gaussian curvature K:

(12.5.14)
$$\boxed{K \;=\; \frac{(G_{xv} - G_{yu})\,v + 2\,G\,G_{uu} + 2\,HG_{uv} - G_u\,G_u - G_v H_u}{v\,\mathcal{L}_v}} ,$$

(12.5.15)
$$\boxed{K \;=\; \frac{(H_{yu} - H_{xv})\,u + 2\,HH_{vv} + 2\,GH_{vu} - H_v H_v - H_u G_v}{u\,\mathcal{L}_u}} ,$$

(12.5.16)
$$\boxed{K \;=\; \frac{(G_{yu} - G_{xv})\,u + 2\,HG_{vv} + 2\,G\,G_{uv} - G_v H_v - G_v\,G_u}{-\,u\,\mathcal{L}_v}} ,$$

(12.5.17)

$$\boxed{K \;=\; \frac{(H_{xv} - H_{yu})\,v + 2\,GH_{uu} + 2\,HH_{vu} - H_uG_u - H_uH_v}{-\;v\;\mathcal{L}_u}}$$

Since

$$u\,\mathcal{L}_u \;+\; v\,\mathcal{L}_v \;=\; 2\,\mathcal{L} \;=\; F^2$$

by Euler's theorem, we also have

$$\begin{aligned} F^2\,K \;=\;& (G_{xv} - G_{yu})\,v \;+\; (H_{yu} - H_{xv})\,u \\ & + 2\,(G\,G_{uu} + HH_{vv}) \;+\; 2\,(HG_{uv} + GH_{vu}) \\ & - (G_u\,G_u + H_vH_v) \;-\; (G_vH_u + H_uG_v)\;. \end{aligned} \tag{12.5.18}$$

This is a rather symmetric looking formula for K. However, it turns out to be not as computationally friendly as we thought it would be.

We conclude this section with the Maple code for implementing (12.5.14). For concreteness, we also include a specific example. **Review** (12.5.5).

```
> K:=proc(F)
> local L,Lx,Ly,Lu,Lv,Lxv,Lyu,Luu,Luv,Lvv,D,
      G,H,Gu,Gv,Guu,Guv,Gxv,Gyu,Hu,K;
> L:=F^2/2;
> Lx:=diff(L,x);
> Ly:=diff(L,y);
> Lu:=diff(L,u);
> Lv:=diff(L,v);
> Lxv:=diff(Lx,v);
> Lyu:=diff(Ly,u);
> Luu:=diff(Lu,u);
> Luv:=diff(Lu,v);
> Lvv:=diff(Lv,v);
> D:=simplify(2*(Luu*Lvv-Luv*Luv));
> G:=simplify(((Lx*Lvv-Ly*Luv)+(Lyu-Lxv)*Lv)/D);
> H:=simplify(((Ly*Luu-Lx*Luv)+(Lxv-Lyu)*Lu)/D);
> Gu:=diff(G,u);
> Gv:=diff(G,v);
> Guu:=diff(Gu,u);
> Guv:=diff(Gu,v);
> Gxv:=diff(Gv,x);
> Gyu:=diff(Gu,y);
> Hu:=diff(H,u);
> K:=((Gxv-Gyu)*v+2*G*Guu+2*H*Guv-Gu*Gu-Gv*Hu)/(v*Lv);
> simplify(K);
> end:
> F:=sqrt(u^2+sinh(x)^2*v^2)+tanh(x)*u:
> K(F);
```

```
    -1/4
bytes used=129197548, alloc=11139080, time=125.06
```

In the above code, the Finsler function is in **arbitrary** coordinates! See the abbreviation chart (12.5.5). At the end, the specific example entered is

$$F(x,y;u,v) \;:=\; \sqrt{u^2 \;+\; \sinh^2(x)\; v^2} \;+\; u\; \tanh(x)\;. \tag{12.5.19}$$

Its Gaussian curvature happens to be a negative constant, with value $-\frac{1}{4}$. We explain the origin of this example in §12.6.

For further information about machine computations in differential geometry, see Oprea [Op] (which uses Maple) and Gray [Gr] (which uses Mathematica).

Exercises

Exercise 12.5.1: Derive formulas (12.5.7a,b) in detail.

Exercise 12.5.2:

(a) We sketched a derivation of (12.5.9). Fill in the details.

(b) By specializing (12.5.9) to specific values of i and k, establish formulas (12.5.10)–(12.5.13).

Exercise 12.5.3:

(a) Modify the given Maple code to implement (12.5.15), (12.5.16), and (12.5.17). Rerun your code with example (12.5.19) in each case. How do the four formulas (12.5.14)–(12.5.17) compare with each other in terms of computational efficiency?

(b) Investigate what happens if you try to carry out the same program with formula (12.5.18).

(c) For the Finsler function F given in (12.5.19), are you able to calculate its Gaussian curvature K by hand?

(d) Convert the given Maple code into a Mathematica code.

12.6 A Poincaré Disc That Is Only Forward Complete

Our story begins with a theorem entitled *"On Randers spaces of scalar curvature"* by Yasuda and Shimada [YS], published in 1977. This theorem is stated for Randers spaces of dimension $n \geqslant 2$. A 1989 paper by Matsumoto [M5] corroborates the results of [YS] for $n > 2$. Shimada has assured us that the theorem remains valid for $n = 2$, as originally stated. See also the treatment in [SK].

The proof of the Yasuda–Shimada theorem is difficult, and we do not give it in this book. However, we are free to use its conclusions as *inspirations* for constructing examples. This attitude is logically fine, as long as one explicitly verifies that the constructed examples have the properties expected of them. We do just that.

12.6 A. The Example and Its Yasuda–Shimada Pedigree

The example we are going to construct is of Randers type (Chapter 11), and is also known to Okada [Ok]. It resides on the Euclidean open disc with radius 2 and center $(0,0)$ in $\mathbb{R}^2$. Denote the canonical Cartesian coordinates by X, Y. And introduce polar coordinates via

$$X = r\cos\theta\,, \qquad Y = r\sin\theta\,.$$

Then our underlying manifold is

$$M := \{(X,Y) \in \mathbb{R}^2 : r^2 = X^2 + Y^2 < 4\}\,.$$

In order to stipulate a Randers space, we need two other pieces of data besides M: a Riemannian metric $\tilde{a}$ and a 1-form $\tilde{b}$, both globally defined on M. For our example, $\tilde{a}$ is the usual Poincaré disc model of the hyperbolic metric with constant Gaussian curvature -1. See [On] for detailed discussions about this Riemannian metric. Explicitly,

$$\text{(12.6.1)} \qquad \boxed{\tilde{a} := \frac{1}{\left(1-\frac{r^2}{4}\right)^2}\,\left[dr\otimes dr + r^2\, d\theta\otimes d\theta\right]}\,.$$

The 1-form $\tilde{b}$ in question is

$$\text{(12.6.2)} \qquad \boxed{\tilde{b} := d\left(\log\left[\frac{4+r^2}{4-r^2}\right]\right) = \frac{r}{\left(1+\frac{r^2}{4}\right)\left(1-\frac{r^2}{4}\right)}\,dr}\,.$$

Note that $\tilde{b}$ is exact, hence closed in particular. The relevance of this feature is shown later in this section.

We digress to describe the genesis of this 1-form $\tilde{b}$. The **Yasuda–Shimada theorem** says, among other conclusions, that:

A Randers space has constant negative flag curvature $-\dfrac{\lambda^2}{4}$

if and only if

- The underlying Riemannian metric $\tilde{a}$ has constant negative sectional curvature (see Exercise 3.10.8) $-\lambda^2$.
- The drift 1-form $\tilde{b}$ is exact (that is, has the form df) and satisfies the system of PDEs

$$\tilde{b}_{i,x^j} - \tilde{b}_k\,\tilde{\gamma}^k{}_{ij} = \lambda\,(\tilde{a}_{ij} - \tilde{b}_i\,\tilde{b}_j)\,.$$

In our example, $\lambda = 1$. By limiting our choice of f to those with only an r dependence, we reduced the above PDE system to two ODEs. They are:

$$f_r \;=\; \frac{16\, r}{(4+r^2)(4-r^2)}\;,$$

$$f_{rr} \;-\; \frac{2\, r}{4-r^2}\, f_r \;=\; \frac{16}{(4-r^2)^2} \;-\; f_r\, f_r\;.$$

Solving the first one gives

(12.6.3)
$$f(r) \;=\; \log\left[\frac{4+r^2}{4-r^2}\right] \;+\; \text{an arbitrary constant}\;.$$

It is then checked that this f satisfies the second ODE identically. The drift 1-form $\tilde{b}$ is the exterior differential of this f, as stipulated in (12.6.2).

12.6 B. The Finsler Function and Its Gaussian Curvature

Let V be any tangent vector to M. The value of the Finsler function F on V is, in the abstract,

$$F(V) \;:=\; \sqrt{\tilde{a}(V,V)} \;+\; \tilde{b}(V)\;.$$

In §11.1, we explained why a Randers F is positive and strongly convex (that is, has positive-definite fundamental tensor) on $TM \setminus 0$ if and only if the Riemannian norm of $\tilde{b}$ is uniformly less than 1 on M. For the example in question, one finds that

(12.6.4)
$$\boxed{\;\|\tilde{b}\| \;=\; \frac{4r}{4+r^2}\;}\;.$$

Straightforward calculus shows that since $0 \leqslant r < 2$ here, we indeed have

$$\|\tilde{b}\| \;<\; 1\;.$$

Thus our F is a y-global Finsler metric. It has also been obtained by Okada [Ok] through his study of Funk and Hilbert metrics.

At any point in M, with polar coordinates (r, θ), our arbitrary tangent vector V can be expanded as

(12.6.5)
$$V \;:=\; p\,\frac{\partial}{\partial r} \;+\; q\,\frac{\partial}{\partial \theta}\;.$$

The value of the Finsler function on this tangent vector is

(12.6.6)
$$\boxed{\;F(V) \;:=\; \frac{1}{1-\frac{r^2}{4}}\,\sqrt{p^2 \;+\; r^2\, q^2} \;+\; \frac{p\, r}{\left(1-\frac{r^2}{4}\right)\left(1+\frac{r^2}{4}\right)}\;}\;.$$

The Yasuda–Shimada theorem says that this F should have constant negative Gaussian curvature $-\frac{1}{4}$. Since we have not proved the said theorem in this book, we verify their conclusion by direct calculation.

Let us convert the right-hand side into a form that is friendlier towards machine computations. To this end, define

$$s \ := \ \log\left(\frac{1+\frac{r}{2}}{1-\frac{r}{2}}\right) . \tag{12.6.7}$$

Since $0 \leqslant r < 2$, we see that $0 \leqslant s < \infty$. One can check that:

$$\begin{aligned} ds &= \frac{dr}{1-\frac{r^2}{4}} , \\ \frac{\partial}{\partial s} &= \left(1-\frac{r^2}{4}\right)\frac{\partial}{\partial r} . \end{aligned} \tag{12.6.8}$$

And

$$\begin{aligned} \sinh(s) &= \frac{r}{1-\frac{r^2}{4}} , \\ \tanh(s) &= \frac{r}{1+\frac{r^2}{4}} . \end{aligned} \tag{12.6.9}$$

Our arbitrary tangent vector V then has the expansion

$$V \ = \ \frac{p}{1-\frac{r^2}{4}}\,\frac{\partial}{\partial s} + q\,\frac{\partial}{\partial \theta} \ =: \ \mathfrak{p}\,\frac{\partial}{\partial s} + q\,\frac{\partial}{\partial \theta} , \tag{12.6.10}$$

where $\mathfrak{p}$ is p normalized as shown. And the Finsler function becomes

$$\boxed{F(s,\theta\,;\,\mathfrak{p},q) \ := \ \sqrt{\mathfrak{p}^2 \ + \ \sinh^2(s)\, q^2} \ + \ \mathfrak{p}\ \tanh(s)} . \tag{12.6.11}$$

According to the Maple calculation just before (12.5.19), **this Finsler structure does have constant negative Gaussian curvature $-\frac{1}{4}$.** We have therefore explicitly verified, for the example under discussion, the prediction of the Yasuda–Shimada theorem.

12.6 C. Geodesics; Forward and Backward Metric Discs

The drift 1-form $\tilde{b}$ in our example is exact, hence closed. In that case, Exercises 11.3.3 and 11.3.4 tell us that the geodesics of our Randers space are *trajectorywise* the same as the geodesics of the underlying Riemannian metric $\tilde{a}$. Those exercises demonstrate this fact by simply parametrizing the Finslerian geodesics to have constant Riemannian speed. Now, the geodesics of $\tilde{a}$ are well understood. See, for instance, [On]. So, **the geodesics of our example have the following trajectories**:

- Euclidean circular arcs that intersect the boundary of the Poincaré disc at Euclidean right angles. In view of the Pythagoras theorem, none of these can pass through the origin. See Figure 12.1.
- Euclidean straight rays that emanate from the origin.
- Euclidean straight rays that aim towards the origin.

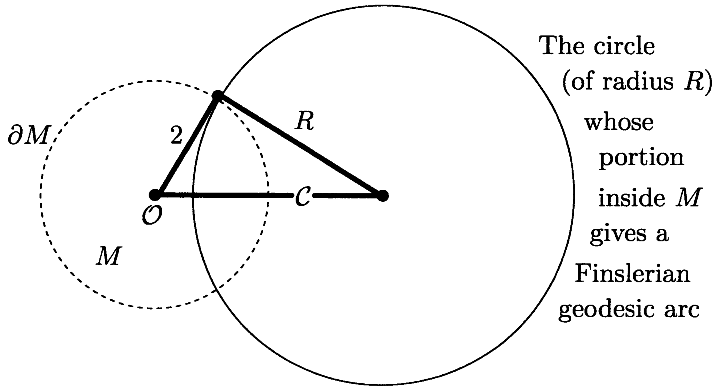

Figure 12.1

The geodesic trajectories of our Randers space (M, F). The Euclidean circle centered outside M intersects ∂M at Euclidean right angles. By Pythagoras' theorem, we have $\mathcal{C}^2 = 2^2 + R^2$. If the circular arc were to pass through the origin $\mathcal{O}$, we would have $\mathcal{C} = R$, which would then lead to a contradiction.

This is a remarkable feature. To put that in context, recall Exercise 5.3.3. There, it is shown that if the Finsler structure is absolutely homogeneous, or is of Berwald type, then the reverse of a geodesic is again a geodesic, the implication being that such conclusion does not hold for generic F. Our F under discussion is a Randers metric. It is neither absolutely homogeneous nor Berwald, so we do not *expect* the reverse of its geodesics to be geodesics. Nevertheless, this is valid on a trajectory level. Namely, **if we reverse any given geodesic (say, from P to Q) in our example, then the trajectory of the resulting curve (from Q to P) coincides with that of a geodesic.**

Next, recall the discussion in §7.5 and §8.2. Namely, each geodesic starts out by globally minimizing the Finslerian arc length until it reaches the cut point. Beyond that, it loses the status as a global minimizer. However, it will continue to minimize among nearby curves that share its endpoints, until it reaches the first conjugate point. After that, there is no minimizing ability whatsoever. Long, globally minimizing, geodesics are therefore a rare commodity. We show, in the rest of this section, that the above **Euclidean straight rays (to or from the origin) are global minimizers in their entirety!**

Fix a point P on the Euclidean circle of radius $\epsilon < 2$ and center at the origin $\mathcal{O}$. Its polar coordinates are $r = \epsilon$ and $\theta = \xi$. Both ϵ and ξ are fixed. Here is our plan:

* Take an arbitrary curve

$$c(t) \ := \ (\ r(t)\cos\theta(t)\,,\, r(t)\sin\theta(t)\ )\ , \quad 0 \leqslant t \leqslant \epsilon$$

from $\mathcal{O}$ to P. Here, $r(0) = 0$, $r(\epsilon) = \epsilon$, and $\theta(\epsilon) = \xi$. We show that its Finslerian arc length is bounded below by that of the straight ray

$$\sigma(t) \ := \ (\ t\cos\xi\,,\, t\sin\xi\ )\ , \quad 0 \leqslant t \leqslant \epsilon$$

from $\mathcal{O}$ to P.

* Take an arbitrary curve

$$c(t) \ := \ (\ r(t)\cos\theta(t)\,,\, r(t)\sin\theta(t)\ )\ , \quad 0 \leqslant t \leqslant \epsilon$$

from P to $\mathcal{O}$. Here, $r(0) = \epsilon$, $r(\epsilon) = 0$, and $\theta(0) = \xi$. We show that its Finslerian arc length is bounded below by that of the straight ray

$$\sigma(t) \ := \ (\ [\epsilon - t]\cos\xi\,,\, [\epsilon - t]\sin\xi\ )\ , \quad 0 \leqslant t \leqslant \epsilon$$

from P to $\mathcal{O}$.

Note that

$$\frac{\partial c}{\partial r} = (\ \cos\theta\,,\, \sin\theta\ ) = \frac{\partial}{\partial r}\ ,$$
$$\frac{\partial c}{\partial \theta} = r\ (\ -\sin\theta\,,\, \cos\theta\ ) = \frac{\partial}{\partial \theta}\ .$$

Thus the chain rule

$$\frac{dc}{dt} = \frac{\partial c}{\partial r}\frac{dr}{dt} + \frac{\partial c}{\partial \theta}\frac{d\theta}{dt}$$

yields the following decomposition of the velocity:

$$\dot c \ := \ \frac{dc}{dt} = \dot r\,\frac{\partial}{\partial r} + \dot\theta\,\frac{\partial}{\partial \theta}\ .$$

Hence, by (12.6.6), we have

$$\begin{aligned} F(\dot c) &= \frac{1}{1-\frac{r^2}{4}}\sqrt{\dot r^2 + r^2\,\dot\theta^2} + \frac{\dot r\ r}{\left(1-\frac{r^2}{4}\right)\left(1+\frac{r^2}{4}\right)} \\ &\geqslant \frac{1}{1-\frac{r^2}{4}}\,|\,\dot r\,| + \frac{r}{\left(1-\frac{r^2}{4}\right)\left(1+\frac{r^2}{4}\right)}\,\dot r \\ &\geqslant \left[\frac{\pm 1}{1-\frac{r^2}{4}} + \frac{r}{\left(1-\frac{r^2}{4}\right)\left(1+\frac{r^2}{4}\right)}\right]\dot r\ . \end{aligned} \tag{12.6.12}$$

Integrating this inequality, and realizing that $\dot r\,dt = dr$, gives:
(12.6.13)

$$\int_0^\epsilon F(\dot c)\ dt \ \geqslant \int_{r=r(0)}^{r=r(\epsilon)}\left[\frac{\pm 1}{1-\frac{r^2}{4}} + \frac{r}{\left(1-\frac{r^2}{4}\right)\left(1+\frac{r^2}{4}\right)}\right] dr\ .$$

The case of going from the origin $\mathcal{O}$ to P:

In the ± 1 on the right-hand side of (12.6.13), we take the plus sign. Also, since the curve c goes from $\mathcal{O}$ to P, we have $r(0) = 0$ and $r(\epsilon) = \epsilon$. Thus

$$\begin{aligned} L_F(c) &:= \int_0^{\epsilon} F(\dot{c})\, dt \\ &\geqslant \int_{r=0}^{r=\epsilon} \left[\frac{1}{1-\frac{r^2}{4}} + \frac{r}{\left(1-\frac{r^2}{4}\right)\left(1+\frac{r^2}{4}\right)} \right] dr. \end{aligned}$$

Now, the velocity of the Euclidean straight ray σ from $\mathcal{O}$ to P is $\frac{\partial}{\partial r}$. Using (12.6.6) to calculate its speed, we see that its Finslerian length is

$$L_F(\sigma) := \int_{t=0}^{t=\epsilon} \left[\frac{1}{1-\frac{t^2}{4}} + \frac{t}{\left(1-\frac{t^2}{4}\right)\left(1+\frac{t^2}{4}\right)} \right] dt .$$

This integral is identical to the lower bound of $L_F(c)$. Therefore

$$L_F(c) \geqslant L_F(\sigma) ,$$

which shows that the straight ray σ is globally shortest among all curves from $\mathcal{O}$ to P. Furthermore, the method of partial fractions can be used to compute the said integral, leading to

$$d(\mathcal{O}, P) = L_F(\sigma) = \log \left[\frac{4+\epsilon^2}{(2-\epsilon)^2} \right] . \tag{12.6.14}$$

In particular,

$$\boxed{d(\mathcal{O}, \partial M) = \lim_{\epsilon \to 2^-} \log \left[\frac{4+\epsilon^2}{(2-\epsilon)^2} \right] = +\infty} . \tag{12.6.15}$$

That is:

> **The Finslerian metric distance from the origin to the rim of the Poincaré disc is infinite.**

Using calculus, one checks that the above logarithm is a monotonically increasing (hence one-to-one) function of ϵ, for $0 \leqslant \epsilon < 2$. Based on this information, a moment's thought convinces us that

$$\boxed{\mathbb{B}_o(\epsilon) = \mathcal{B}_o^+ \left(\log \left[\frac{4+\epsilon^2}{(2-\epsilon)^2} \right] \right)} . \tag{12.6.16}$$

Here, $\mathbb{B}_o(\epsilon)$ is the Euclidean open disc of radius ϵ and center $\mathcal{O}$. Also, $\mathcal{B}_o^+$ denotes *forward* metric discs centered at $\mathcal{O}$, as defined in §6.2B. In particular,

$$\boxed{M = \mathbb{B}_o(2) = \mathcal{B}_o^+(\infty)} . \tag{12.6.17}$$

This finishes our analysis of the "outward bound" situation.

The case of going from P towards the origin $\mathcal{O}$:

In the ± 1 on the right-hand side of (12.6.13), we take the minus sign this time. Also, since the curve c goes from P to $\mathcal{O}$, we have $r(0) = \epsilon$ and $r(\epsilon) = 0$. Thus

$$\begin{aligned} L_F(c) &:= \int_0^{\epsilon} F(\dot{c})\, dt \\ &\geqslant \int_{r=\epsilon}^{r=0} \left[\frac{-1}{1-\frac{r^2}{4}} + \frac{r}{\left(1-\frac{r^2}{4}\right)\left(1+\frac{r^2}{4}\right)} \right] dr. \end{aligned}$$

Equivalently,

$$L_F(c) \geqslant \int_{r=0}^{r=\epsilon} \left[\frac{1}{1-\frac{r^2}{4}} - \frac{r}{\left(1-\frac{r^2}{4}\right)\left(1+\frac{r^2}{4}\right)} \right] dr.$$

The velocity of the Euclidean straight ray σ from P towards $\mathcal{O}$ is $-\frac{\partial}{\partial r}$. Using (12.6.6) to calculate its speed, we see that its Finslerian length is

$$L_F(\sigma) := \int_{t=0}^{t=\epsilon} \left[\frac{1}{1-\frac{(\epsilon-t)^2}{4}} - \frac{\epsilon - t}{\left(1-\frac{(\epsilon-t)^2}{4}\right)\left(1+\frac{(\epsilon-t)^2}{4}\right)} \right] dt\ .$$

Either by a change-of-variable or a graphical argument, one sees that

$$\int_0^{\epsilon} f(\epsilon - t)\, dt = \int_0^{\epsilon} f(t)\, dt\ .$$

Thus

$$L_F(\sigma) := \int_{t=0}^{t=\epsilon} \left[\frac{1}{1-\frac{t^2}{4}} - \frac{t}{\left(1-\frac{t^2}{4}\right)\left(1+\frac{t^2}{4}\right)} \right] dt\ .$$

This integral is identical to the lower bound of $L_F(c)$. Therefore

$$L_F(c) \geqslant L_F(\sigma)\ ,$$

which shows that the straight ray σ is globally shortest among all curves from P to $\mathcal{O}$. Now, the method of partial fractions can be used to calculate the said integral, giving

$$d(P, \mathcal{O}) = L_F(\sigma) = \log\left[\frac{(2+\epsilon)^2}{4+\epsilon^2} \right]\ . \tag{12.6.18}$$

In particular,

$$\boxed{d(\partial M, \mathcal{O}) = \lim_{\epsilon \to 2^-} \log\left[\frac{(2+\epsilon)^2}{4+\epsilon^2} \right] = \log 2}\ . \tag{12.6.19}$$

That is:

Coming in from the rim of the Poincaré disc to the center, the Finslerian metric distance has the finite value $\log 2$!

It can be verified that the above logarithm is a monotonically increasing (hence one-to-one) function of ϵ, for $0 \leqslant \epsilon < 2$. This information, together with a moment's thought, tells us that

$$\boxed{\mathbb{B}_o(\epsilon) \;=\; \mathcal{B}_o^-\left(\log\left[\frac{(2+\epsilon)^2}{4+\epsilon^2}\right]\right)}\,. \tag{12.6.20}$$

Again, $\mathbb{B}_o(\epsilon)$ is the Euclidean open disc of radius ϵ and center $\mathcal{O}$. Also, $\mathcal{B}_o^-$ denotes the *backward* metric discs centered at $\mathcal{O}$, as defined in Exercise 6.2.10. In particular,

$$\boxed{M \;=\; \mathbb{B}_o(2) \;=\; \mathcal{B}_o^-(\log 2)}\,. \tag{12.6.21}$$

This finishes our analysis of the "inward bound" situation.

Let us propose a model to help visualize the fact that

$$d(\mathcal{O}, \partial M) \;=\; \infty\,, \quad \text{while} \quad d(\partial M, \mathcal{O}) \;=\; \log 2\,.$$

* Suppose water in a kitchen sink is draining towards a sink-hole located at $\mathcal{O}$. Let's say that the walls of the kitchen sink are situated at ∂M. To complete the picture, imagine a tiny bug swimming either from $\mathcal{O}$ to ∂M, or from ∂M to $\mathcal{O}$, along the straight rays that we called σ.
* Let $F(\dot\sigma)\,dt = F(d\sigma)$ denote the physical time (which is distinctly *different* from the parameter increment dt) it takes the bug to traverse a short portion $d\sigma$ of its journey. Then the mathematical Finslerian "arc length" $\int F(d\sigma)$ is actually measuring the total physical time of the journey. Revisit similar discussions in §1.0.
* It is now conceivable that, swimming against a current, the bug might need an eternity to reach ∂M. On the other hand, aided by a current, it can easily reach $\mathcal{O}$ in finite physical time.

12.6 D. Consistency with Akbar-Zadeh's Rigidity Theorem

By construction, our example is of Randers type, and is manifestly non-Riemannian. It has also been verified explicitly to have constant negative Gaussian curvature $-\frac{1}{4}$. Is this then somehow a counterexample to Akbar-Zadeh's rigidity theorem (Theorem 12.4.1)?

Happily the answer is no. There are two crucial hypotheses in that theorem:

- The Cartan scalar I is supposed to have slower than exponential growth.
- The Finsler surface is supposed to be both forward and backward geodesically complete.

Let's examine them in turn. (Study Okada [Ok] in the same spirit.)

The Cartan scalar:

We see from Exercise 11.2.6 that for Randers surfaces, **the Cartan scalar has the uniform universal bound** $\frac{3}{\sqrt{2}}$. Hence our I certainly grows more slowly than exponentially!

Forward geodesic completeness:

Take any unit speed Finslerian geodesic $\sigma(t)$ that passes through the origin $\mathcal{O}$, say at time $t = 0$. In view of the discussion at the beginning of §12.6C, the trajectory of σ must be part of a Euclidean straight line through the origin. Extend $\sigma(t)$ maximally forward, so that it is defined on an interval $0 \leqslant t < b$. Since the Euclidean line reaches the boundary ∂M, we must have

$$\lim_{t \to b^-} \sigma(t) \in \partial M .$$

If b were finite, the Finslerian length of the unit speed geodesic $\sigma(t)$, $0 \leqslant t < b$ would be finite. The metric distance from $\mathcal{O}$ to ∂M would then be finite as well, contradicting (12.6.15). Therefore b has to be ∞. This means that (M, F) **is forward geodesically complete** at $\mathcal{O}$. By the Hopf–Rinow theorem (Theorem 6.6.1), it is so everywhere.

Backward geodesic completeness:

We now demonstrate that our Randers metric is *not* backward geodesically complete. To this end, extend the above $\sigma(t)$ maximally backward, so that it is defined on an interval $a < t \leqslant 0$. Since the trajectory of σ is part of a Euclidean line that reaches the boundary ∂M, we must have

$$\lim_{t \to a^+} \sigma(t) \in \partial M .$$

If a were equal to $-\infty$, the Finslerian length of the unit speed $\sigma(t)$; $a < t \leqslant 0$ would be infinite. The metric distance from ∂M to $\mathcal{O}$ would then be infinite as well, contradicting (12.6.19). This means that a cannot equal $-\infty$. Hence (M, F) **is *not* backward geodesically complete**. Given that, the negative curvature case in Akbar-Zadeh's theorem cannot be applied to our example. So all is well.

Exercises

Exercise 12.6.1:

(a) Derive the ODEs that characterize the potential function f.

(b) Solve one of them to obtain (12.6.3).

Exercise 12.6.2:

(a) Show that the Riemannian norm of the drift 1-form is indeed given by formula (12.6.4).

(b) Prove that this norm is uniformly less than 1 on the Poincaré disc.

Exercise 12.6.3:

(a) Start with (12.6.7). Derive the transformation formulas (12.6.8) and (12.6.9).

(b) Check that our Finsler function F has the two equivalent descriptions (12.6.6) and (12.6.11).

Exercise 12.6.4: In (12.6.6), take the difference instead of the sum of the two terms. Does this destroy forward completeness and restore backward completeness? Can you physically interpret this new Finsler metric?

12.7 Non-Riemannian Projectively Flat S^2 with $K = 1$

Let us turn to the case of positive flag curvatures. In [AZ], Akbar-Zadeh showed that if an n-dimensional Finsler manifold (M, F) has constant positive flag curvature, then its universal cover must be diffeomorphic to the standard sphere $\mathbb{S}^n$. Through the works of Bryant, we now know that (M, F) need not be isometric to $\mathbb{S}^n$. He has informed us that this holds for surfaces as well as for higher dimensions.

12.7 A. Bryant's 2-parameter Family of Finsler Structures

In two dimensions, Bryant [Br1, Br2] has published explicit non-Riemannian examples with constant positive Gaussian curvature $K = 1$. Here, we focus on a 2-parameter family from [Br2]. Each Finsler structure in this family has $K = 1$. In [Br2], it is explained in detail how these are related to some earlier works of Funk's [F1, F2]. See also [Br3].

To describe that 2-parameter family, let V be a 3-dimensional real vector space with basis $\{b_1, b_2, b_3\}$. Let ρ, γ be two fixed angles satisfying

$$|\gamma| \leqslant \rho < \frac{\pi}{2} \,. \tag{12.7.1}$$

Define a ρ and γ dependent, complex-valued quadratic form Q on V by

$$Q(u, v) \;:=\; e^{i\rho}\, u^1 v^1 \;+\; e^{i\gamma}\, u^2 v^2 \;+\; e^{-i\rho}\, u^3 v^3 \,. \tag{12.7.2}$$

In the above exponentials,

$$i \;:=\; \sqrt{-1} \,,$$

and

$$u \;=\; u^i\, b_i \,, \quad v \;=\; v^i\, b_i \,.$$

Let S^2 denote the set of rays in V. Equivalently, we are identifying X and X^* in V whenever $X^* = \lambda X$ for some $\lambda > 0$. Each point of S^2 can thus be denoted as an equivalence class $[X]$, with $0 \neq X \in V$. As a manifold,

S^2 is a projective sphere. A moment's thought shows that every tangent vector at the point $[X]$ on S^2 is the initial velocity to a curve of the form $[X + tY]$, for some $Y \in V$. Each such curve is half of a great circle on S^2. It makes sense to denote the said tangent vector by $[X, Y]$. Note that $[X', Y'] = [X, Y]$ if and only if $X' = \lambda X$ and $Y' = \lambda Y + \mu X$, for some $\lambda > 0$ and $\mu \in \mathbb{R}$.

For each choice of ρ and γ satisfying (12.7.1), Bryant gives an explicit formula for the corresponding Finsler function $F : TS^2 \to [0, \infty)$. It reads:
(12.7.3)

$$F([X,Y]) := \text{Re}\left[\sqrt{\frac{Q(Y,Y)\,Q(X,X) - Q^2(X,Y)}{Q^2(X,X)}} - i\,\frac{Q(X,Y)}{Q(X,X)}\right],$$

where "Re" means taking the real part. The complex square root function is taken to be branched along the negative real axis, and to satisfy $\sqrt{1} = 1$. In other words,

$$\text{Re}\,\sqrt{a + i\,b} := +\sqrt{\frac{a + \sqrt{a^2 + b^2}}{2}}.$$

For each F defined by (12.7.3), with ρ and γ constrained by (12.7.1), the method of exterior differential systems in [Br2] effects some interesting properties. We quote but do not attempt to prove them in this book.

* The F given by (12.7.3) is indeed a Finsler structure in the sense of §1.1. Unless $\rho = 0 = \gamma$, this Finsler function is non-Riemannian and is only positively homogeneous.
* Each great semicircle $[X + tY]$ is a geodesic of the Finsler structure. (Incidentally, such curves are not parametrized yet to have constant speed.) Furthermore, every Finslerian geodesic is part of a great circle, and hence traces out a curve of the form $[X + tY]$.
* More technically, there exists a privileged coordinate system in which the Finslerian geodesics are defined by linear equations. In other words, the Finsler surfaces (S^2, F) described by (12.7.3) are **projectively flat**. (See also Exercise 12.7.3.)
* The Finslerian length of each great circle is 2π. The calculation for a special case, namely, $\rho = \frac{\pi}{4} = \gamma$, is not too difficult. The reader is asked to undertake *a portion of* that task in Exercise 12.7.2.
* The Gaussian curvature K of each such Finsler surface (S^2, F) has the constant positive value 1.
* The 2-parameter (ρ and γ) family presented in (12.7.3) encompasses all projectively flat Finsler structures on S^2 with $K = 1$. The 2-dimensional moduli space is *non*-compact.

We invite the reader to read Bryant's papers, especially [Br2], and to try to verify his first few claims.

12.7 B. A Specific Finsler Metric from That Family

We next work out the manifestly real formula for one of the Finsler functions described in (12.7.3). The case we are concerned with is

$$\boxed{\rho = \frac{\pi}{4} = \gamma} \ .$$

As in (12.5.5), we use

$$\begin{aligned} x &\leftrightarrow x^1 \ , \\ y &\leftrightarrow x^2 \ , \\ u &\leftrightarrow y^1 \ , \\ v &\leftrightarrow y^2 \ . \end{aligned}$$

Take the 3-dimensional vector space to be $\mathbb{R}^3$, with Cartesian coordinates x, y, z. For concreteness, we realize S^2 as the unit sphere defined by the equation $x^2 + y^2 + z^2 = 1$. As a graph, it admits a parametrization

$$\varphi \ : \ \{ (x,y) : \ x^2 + y^2 < 1 \} \ \rightarrow \ S^2$$

given by

$$(x\,,\,y) \ \mapsto \ X \ := \ \left(x\,,\,y\,,\,s\sqrt{1 - x^2 - y^2}\, \right) \quad \text{with} \ \ s := \pm 1 \ .$$

We have omitted the equator, where $x^2 + y^2 = 1$, because some subsequent quantities involve division by positive powers of $1 - x^2 - y^2$. Here, $s = \pm 1$ corresponds to the upper and lower hemispheres, respectively. It turns out that only s^2 (which is 1) enters the crucial steps leading to the formula for F. Thus the Finsler structure on the upper hemisphere is the same as that on the lower hemisphere. We later see—from the "$2C$ term" in (12.7.4)—that F is positively but not absolutely homogeneous.

With the present *graph* parametrization, one finds that

$$\partial_x \ := \ \varphi_* \frac{\partial}{\partial x} \ = \ \left(1\,,\,0\,,\,\frac{-\,x\,s}{\sqrt{1 - x^2 - y^2}} \right) \ ,$$

$$\partial_y \ := \ \varphi_* \frac{\partial}{\partial x} \ = \ \left(0\,,\,1\,,\,\frac{-\,y\,s}{\sqrt{1 - x^2 - y^2}} \right) \ .$$

Thus an arbitrary tangent vector to S^2 at X is

$$Y \ := \ u\,\partial_x \ + \ v\,\partial_y \ = \ \left(u\,,\,v\,,\,\frac{-\,s\,[x\,u + y\,v]}{\sqrt{1 - x^2 - y^2}} \right) \ .$$

Note that by construction, our X lies on S^2 and Y is tangent to S^2. So there is no need to projectivize.

To minimize clutter, let us also introduce the following abbreviations:

$$\begin{aligned} r^2 &:= x^2 + y^2 , \\ P^2 &:= 1 - r^2 , \\ B &:= 2r^2 - 1 , \\ R^2 &:= u^2 + v^2 , \\ C &:= x\,u + y\,v , \\ a &:= (1+B^2)[(P^2 R^2 + C^2) + B\,(P^2 R^2 - C^2)] + 8\,(1+B)\,C^2 P^4 , \\ b &:= (1+B^2)[(P^2 R^2 - C^2) - B\,(P^2 R^2 + C^2)] - 8\,(0+B)\,C^2 P^2 . \end{aligned}$$

We emphasize that in b, the very last term contains $C^2\,P^2$ and not $C^2\,P^4$. Straightforward computations yield:

$$\begin{aligned} Q(X,X) &= \frac{1}{\sqrt{2}}\,(1 \,+\, i\,B)\,, \\ Q(X,Y) &= \frac{1}{\sqrt{2}}\,(i\,2\,C)\,, \\ Q(Y,Y) &= \frac{1}{\sqrt{2}}\,\frac{1}{P^2}\,[(P^2 R^2 + C^2) \,+\, i\,(P^2 R^2 - C^2)]\,. \end{aligned}$$

And

$$\mathrm{Re}\left[-\,i\,\frac{Q(X,Y)}{Q(X,X)}\right] = \frac{2\,C}{1+B^2}\,,$$

$$\frac{Q(Y,Y)}{Q(X,X)} - \left[\frac{Q(X,Y)}{Q(X,X)}\right]^2 = \frac{a \,+\, i\,b}{P^2\,(1+B^2)^2}\,.$$

Consequently, the formula for the Finsler function is

$$\boxed{F(X,Y) = \frac{1}{1+B^2}\left[\frac{1}{P}\sqrt{\frac{a+\sqrt{a^2+b^2}}{2}} \,+\, 2\,C\right]}\,. \tag{12.7.4}$$

Note that the quantity a is a quadratic, and that $a^2 + b^2$ is a *quartic*! Let us contrast (12.7.4) with the Riemannian metric induced on the graph S^2 by the Euclidean metric (dot product) of $\mathbb{R}^3$. The corresponding norm is

$$\check{F}(X,Y) = \frac{1}{P}\,\sqrt{P^2 R^2 \,+\, C^2}\,. \tag{12.7.5}$$

We tried to use Maple to demonstrate symbolically that the Gaussian curvature of the Finsler metric in (12.7.4) is 1, but failed. However, we have succeeded in obtaining *numerical* evidence of $K = 1$ by plugging (randomly selected) specific positions X (with $x^2 + y^2 \ll 1$) and tangent vectors Y into the Maple expression for K. The answer is, up to many decimal places of accuracy, always 1. Here is a sample documentation:

```
> r:=sqrt(x^2+y^2):
> P:=sqrt(1-r^2):
> B:=2*r^2-1:
> R:=sqrt(u^2+v^2):
> C:=x*u+y*v:
> a:=(1+B^2)*((P^2*R^2+C^2) +B*(P^2*R^2-C^2))
      +8*(1+B)*C^2*P^4:
> b:=(1+B^2)*((P^2*R^2-C^2) -B*(P^2*R^2+C^2))
      -8*(0+B)*C^2*P^2:
> rt:=sqrt((a+sqrt(a^2+b^2))/2):
> F:=(rt/P+2*C)/(1+B^2):
> L:=F^2/2:
> Lx:=diff(L,x):
> Ly:=diff(L,y):
> Lu:=diff(L,u):
> Lv:=diff(L,v):
> Lxv:=diff(Lx,v):
> Lyu:=diff(Ly,u):
> Luu:=diff(Lu,u):
> Luv:=diff(Lu,v):
> Lvv:=diff(Lv,v):
> DN:=2*(Luu*Lvv-Luv*Luv):
> G:=((Lx*Lvv-Ly*Luv)+(Lyu-Lxv)*Lv)/DN:
> H:=((Ly*Luu-Lx*Luv)+(Lxv-Lyu)*Lu)/DN:
> Gu:=diff(G,u):
> Gv:=diff(G,v):
> Guu:=diff(Gu,u):
> Guv:=diff(Gu,v):
> Gvx:=diff(Gv,x):
> Guy:=diff(Gu,y):
> Hu:=diff(H,u):
> K:=((Gvx-Guy)*v+2*G*Guu+2*H*Guv-Gu*Gu-Gv*Hu)/(v*Lv):
> x:=0.2:
> y:=0.2:
> u:=0.0:
> v:=0.5:
> K;

            1.000000089
```

This Maple computation took almost 20 minutes!

For choices of x, y such that $x^2 + y^2$ is close to the limiting value 1 (that is, for positions X near the equator), the computation time needed is considerably longer. And the accuracy drops off sharply too.

A much more efficient program can be had by changing to polar coordinates r, θ in parameter space. In that case, $x = r\cos\theta$, $y = r\sin\theta$, and

$$u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} = p \frac{\partial}{\partial r} + q \frac{\partial}{\partial \theta} .$$

One can check that

$$(12.7.6) \qquad u = p \cos\theta - q\, r \sin\theta \,, \quad v = p \sin\theta + q\, r \cos\theta \,.$$

Hence

$$(12.7.7) \qquad u^2 + v^2 = p^2 + r^2 q^2 \,,$$

$$(12.7.8) \qquad x\, u + y\, v = r\, p \,.$$

These are used to re-express the quantities R and C that enter into the definition of F, so that $F(x, y; u, v) = F(r, \theta; p, q)$.

Now comes a simple but conceptually important point. Recall formula (12.5.14), upon which our Maple code for K is built. In the code, beginning with the line "Lx:= ..." and ending with the line "K:= ...," the x and y are merely abbreviations of *arbitrary* coordinates x^1 and x^2 on S^2, while u and v abbreviate the corresponding coordinates y^1 and y^2 on the tangent planes. The point is, r and θ are legitimate choices for x^1 and x^2. Consequently, p and q are perfectly fine as y^1 and y^2. This freedom comes about because K is intrinsic. It does not depend on the coordinates used to calculate it.

Furthermore, the economy in using these new coordinates, as opposed to the Cartesian ones, is that our example has no θ-dependence whatsoever. Hence lines of code concerning θ derivatives can be dropped. The new and *simplified* Maple code, together with a sample run, reads:

```
> P:=sqrt(1-r^2):
> B:=2*r^2-1:
> R:=sqrt(p^2+r^2*q^2):
> C:=r*p:
> a:=(1+B^2)*((P^2*R^2+C^2) +B*(P^2*R^2-C^2))
      +8*(1+B)*C^2*P^4:
> b:=(1+B^2)*((P^2*R^2-C^2) -B*(P^2*R^2+C^2))
      -8*(0+B)*C^2*P^2:
> rt:=sqrt((a+sqrt(a^2+b^2))/2):
> F:=(rt/P+2*C)/(1+B^2):
> L:=F^2/2:
> Lr:=diff(L,r):
> Lp:=diff(L,p):
> Lq:=diff(L,q):
> Lrq:=diff(Lr,q):
> Lpp:=diff(Lp,p):
> Lpq:=diff(Lp,q):
> Lqq:=diff(Lq,q):
> DN:=2*(Lpp*Lqq-Lpq*Lpq):
```

```
> G:=(Lr*Lqq-Lq*Lrq)/DN:
> H:=(Lp*Lrq-Lr*Lpq)/DN:
> Gp:=diff(G,p):
> Gq:=diff(G,q):
> Gpp:=diff(Gp,p):
> Gpq:=diff(Gp,q):
> Gqr:=diff(Gq,r):
> Hp:=diff(H,p):
> K:=(Gqr*q+2*G*Gpp+2*H*Gpq-Gp*Gp-Gq*Hp)/(q*Lq):
> r:=0.5:
> p:=0.0:
> q:=1.0:
> K;
          .9999991812
     bytes used=89659664, alloc=80463472, time=48.15
```

The runtime and accuracy begin to suffer only when r exceeds 0.95.

Exercises

Exercise 12.7.1:

(a) Check that corresponding to the choice $\rho = 0 = \gamma$, Bryant's metric (12.7.3) is indeed Riemannian.

(b) Deduce the formula for the Riemannian $\breve{F}$ of (12.7.5). How does it compare with that in part (a)?

Exercise 12.7.2: Provide all the details in the derivation of (12.7.4). Then use it to show that every great circle which passes through the North Pole on Bryant's 2-sphere has Finslerian length 2π.

Exercise 12.7.3: Recall formulas of $G := \frac{1}{2}G^1$, $H := \frac{1}{2}G^2$ from (12.5.7).

(a) Calculate the quantities $\breve{G}$ and $\breve{H}$ for the Riemannian structure $\breve{F}$ on S^2, as described by (12.7.5). It gives the standard 2-sphere.

(b) Compare these with the G and H for the Bryant metric depicted in (12.7.4). Specifically, use Maple to demonstrate *symbolically* that $(G - \breve{G})/u = (H - \breve{H})/v$.

Two Finsler metrics (on a given surface) that satisfy the above criterion are said to be **projectively related**. The function common to the left-hand and right-hand sides is $\frac{1}{2}$ times *our* **projective factor**. Part (b) shows that the Bryant metric defined by (12.7.4) is projectively related to the standard Riemannian metric, both living on the 2-sphere.

Exercise 12.7.4: Consider the parametrization

$$(x\,,y) \;\mapsto\; \frac{1}{\sqrt{1+x^2+y^2}}\,(x\,,y\,,\pm 1)\ .$$

Use $+1$ (resp. -1) when parametrizing the upper (resp. lower) hemisphere.

(a) Draw a picture to bring out the geometry behind this map. Show that it parametrizes each of the hemispheres (excluding the equator) with the *entire* plane $\mathbb{R}^2$. Explain why the equator corresponds to points at infinity in $\mathbb{R}^2$.

(b) Carry out the analysis and Maple computations, as in §12.7B, for this parametrization.

References

[AZ] H. Akbar-Zadeh, *Sur les espaces de Finsler à courbures sectionnelles constantes*, Acad. Roy. Belg. Bull. Cl. Sci. (5) **74** (1988), 281–322.

[Br1] R. Bryant, *Finsler structures on the 2-sphere satisfying $K = 1$*, Cont. Math. **196** (1996), 27–42.

[Br2] R. Bryant, *Projectively flat Finsler 2-spheres of constant curvature*, Selecta Mathematica, N.S. **3** (1997), 161–203.

[Br3] R. Bryant, *Finsler surfaces with prescribed curvature conditions*, Aisenstadt Lectures, in preparation.

[F1] P. Funk, *Über zweidimensionale Finslersche Räume, insbesondere über solche mit geradlinigen Extremalen und positiver konstanter Krümmung*, Math. Zeitschr. **40** (1936), 86–93.

[F2] P. Funk, *Eine Kennzeichnung der zweidimensionalen elliptischen Geometrie*, Österreichische Akad. der Wiss. Math., Sitzungsberichte Abteilung II **172** (1963), 251–269.

[Gr] A. Gray, *Modern Differential Geometry of Curves and Surfaces with Mathematica*, 2nd ed., CRC Press, 1998.

[M5] M. Matsumoto, *Randers spaces of constant curvature*, Rep. on Math. Phys. **28** (1989), 249–261.

[Num] S. Numata, *On Landsberg spaces of scalar curvature*, J. Korean Math. Soc. **12** (1975), 97–100.

[Ok] T. Okada, *On models of projectively flat Finsler spaces of constant negative curvature*, Tensor, N.S. **40** (1983), 117–124.

[On] B. O'Neill, *Elementary Differential Geometry*, 2nd ed., Academic Press, 1997.

[Op] J. Oprea, *Differential Geometry and its Applications*, Prentice-Hall, 1997.

[SK] C. Shibata and M. Kitayama, *On Finsler spaces of constant positive curvature*, Proceedings of the Romanian–Japanese Colloquium on Finsler Geometry, Braşov, 1984, pp. 139–156.

[YS] H. Yasuda and H. Shimada, *On Randers spaces of scalar curvature*, Rep. on Math. Phys. **11** (1977), 347–360.

Chapter 13

Riemannian Manifolds and Two of Hopf's Theorems

13.1 The Levi-Civita (Christoffel) Connection

A **Riemannian metric** g on a manifold M is a family of inner products $\{g_x\}_{x \in M}$ such that the quantities

$$g_{ij\,(x)} \;:=\; g\left(\frac{\partial}{\partial x^i} \,, \frac{\partial}{\partial x^j} \right)$$

are smooth in local coordinates. The Finsler function $F(x,y)$ of a Riemannian manifold has the characteristic structure

$$F(x,y) \;=\; \sqrt{g_{ij\,(x)}\; y^i\, y^j} \;.$$

In that case, the fundamental tensor $(\frac{1}{2}F^2)_{y^i y^j}$ is simply $g_{ij\,(x)}$, which is independent of y.

Let us recall a construction in the generic Finsler case, from §2.1. For each fixed $x \in M$, we erected a copy of $T_x M$ over every point (x,y) in the parameter space $TM \setminus 0$. Then we assigned the inner product $g_{ij\,(x,y)}\, dx^i \otimes dx^j$ to each such $T_x M$. If we were to insist on this setup for the Riemannian case, the amount of redundancy would be embarrassing because all these inner product spaces corresponding to a fixed x are identical. For this reason, we work with one single copy of $T_x M$ over the point x, and endow it with the inner product $g_{ij\,(x)}\, dx^i \otimes dx^j$. Consequently, the vector bundle of relevance is the tangent bundle TM over M, rather than the pulled-back bundle $\pi^* TM$ over $TM \setminus 0$.

Under this simplification, the Chern connection of §2.4 reduces to the usual **Levi-Civita (Christoffel) connection**, described below. Take an arbitrary local frame $\{b_i : i = 1, \ldots, n\}$ of TM. Denote the corresponding (naturally dual) coframe of T^*M by $\{\omega^i\}$. The structural equations of the Levi-Civita (Christoffel) connection ${\omega_j}^i$ are:

* **Torsion-freeness:**

$$d\omega^i \;-\; \omega^j \wedge {\omega_j}^i \;=\; 0 \;.$$

* **Metric-compatibility:**

$$dg_{ij} \;-\; g_{kj}\, {\omega_i}^k \;-\; g_{ik}\, {\omega_j}^k \;=\; 0 \;.$$

Here,

$$g_{ij} \;:=\; g(b_i, b_j) \;.$$

If the local frame happens to be orthonormal, then g_{ab} equals the Kronecker delta δ_{ab}, and the metric-compatibility criterion reads $\omega_{ab} + \omega_{ba} = 0$. In that case, the connection forms, with both indices down, are skew-symmetric in those two indices.

There are two other ways to describe these structural equations. The first one is coordinate-free, while the second one is not. In any case, we begin by defining the covariant directional derivative of b_j along b_k:

$$\nabla_{b_k} b_j \;:=\; {\omega_j}^i(b_k)\, b_i \;=:\; {\Gamma^i}_{jk}\, b_i \;.$$

Extending this by linearity and the product rule, one can define the covariant derivative of any tensor field on M. As an example, for the rank $\binom{1}{1}$ tensor $T := {T^i}_j\, b_i \otimes \omega^j$, we have

$$\begin{aligned} \nabla_{b_k} T \;&=\; (\, d{T^i}_j \;+\; {T^s}_j\, {\omega_s}^i \;-\; {T^i}_s\, {\omega_j}^s\,)(b_k)\; b_i \otimes \omega^j \\ &=:\; {T^i}_{j|k}\; b_i \otimes \omega^j \;, \end{aligned}$$

where

$${T^i}_{j|k} \;:=\; (d{T^i}_j)(b_k) \;+\; {T^s}_j\, {\Gamma^i}_{sk} \;-\; {T^i}_s\, {\Gamma^s}_{jk} \;.$$

Using ∇, the coordinate-free description of the structural equations is:

$$\nabla_X Y - \nabla_Y X = [X, Y] \quad \text{for all vector fields } X, Y ,$$
$$\nabla g = 0 .$$

On the other hand, the description in terms of natural coordinates is

$$\Gamma^i{}_{kj} = \Gamma^i{}_{jk} ,$$

$$\frac{\partial g_{ij}}{\partial x^k} - g_{sj}\,\Gamma^s{}_{ik} - g_{is}\,\Gamma^s{}_{jk} = 0 .$$

This has its merits. To uncover that, rewrite metric-compatibility as

$$(*) \qquad \Gamma_{ijk} + \Gamma_{jik} = \frac{\partial g_{ij}}{\partial x^k}$$

and use the so-called **Christoffel's trick**. Namely, apply (*) to the combination

$$(\Gamma_{rjk} + \Gamma_{jrk}) - (\Gamma_{jkr} + \Gamma_{kjr}) + (\Gamma_{krj} + \Gamma_{rkj}) ,$$

and impose the symmetry

$$\Gamma_{sqp} = \Gamma_{spq} .$$

After much cancellation and raising the index r, one finds that

$$(13.1.1) \qquad \boxed{\Gamma^i{}_{jk} = \frac{g^{is}}{2}\left(\frac{\partial g_{sj}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^s} + \frac{\partial g_{ks}}{\partial x^j}\right) =: \gamma^i{}_{jk}} .$$

Compare with formula (2.4.10). The right-hand side of (13.1.1) consists of the Riemannian metric's **Christoffel symbols of the second kind.**

Exercises

Exercise 13.1.1:

(a) Use metric-compatibility to show directly that the following is *always* valid on a Riemannian manifold. Let V and W be vector fields along any smooth curve $\sigma(t)$ with velocity T, then:

$$\boxed{\frac{d}{dt}\, g(V, W) = g(D_T V, W) + g(V, D_T W)} .$$

(b) Explain how this could also have followed from Exercise 5.2.3.

Exercise 13.1.2: Carry out the Christoffel trick to derive (13.1.1) in detail.

13.2 Curvature

The **curvature 2-forms** of the Levi-Civita (Christoffel) connection are

$$(13.2.1) \qquad \boxed{\Omega_j^{\ i} \ := \ d\omega_j^{\ i} \ - \ \omega_j^{\ k} \wedge \omega_k^{\ i}} \ .$$

Our convention on the wedge product does *not* include any normalization factor. Thus, for example, the wedge product of two 1-forms is

$$\theta \wedge \zeta \ := \ \theta \otimes \zeta \ - \ \zeta \otimes \theta \ ,$$

without the factor of $\frac{1}{2}$.

Since the $\Omega_j^{\ i}$ are 2-forms on the manifold M, we can expand them as

$$\Omega_j^{\ i} \ := \ \frac{1}{2} \, R_j^{\ i}{}_{kl} \, \omega^k \wedge \omega^l \ .$$

The quantities $R_j^{\ i}{}_{kl}$ are the components of the **curvature tensor**. In natural coordinates, it has the formula

$$(13.2.2) \qquad \boxed{R_j^{\ i}{}_{kl} \ = \ \frac{\partial \gamma^i{}_{jl}}{\partial x^k} \ - \ \frac{\partial \gamma^i{}_{jk}}{\partial x^l} \ + \ \gamma^i{}_{hk} \, \gamma^h{}_{jl} \ - \ \gamma^i{}_{hl} \, \gamma^h{}_{jk}} \ .$$

Compare with formula (3.3.2).

13.2 A. Symmetries, Bianchi Identities, the Ricci Identity

Lowering the index i on R gives the **Riemann tensor**:

$$R_{jikl} \ := \ g_{is} \, R_j^{\ s}{}_{kl} \ .$$

The Riemann tensor has beautiful symmetry properties:

$$(13.2.3\text{–}5) \qquad \boxed{\begin{aligned} R_{jilk} \ &= \ - \ R_{jikl} \ , \\ R_{ijkl} \ &= \ - \ R_{jikl} \ , \\ R_{klji} \ &= \ + \ R_{jikl} \ . \end{aligned}}$$

These are obtained by specializing (3.1.3), (3.4.4), and (3.4.5) to the Riemannian setting.

Next, we turn to the Bianchi identities. The **first Bianchi identity** is algebraic:

$$(13.2.6) \qquad \boxed{R_j^{\ i}{}_{kl} \ + \ R_k^{\ i}{}_{lj} \ + \ R_l^{\ i}{}_{jk} \ = \ 0} \ .$$

The **second Bianchi identity** involves covariant differentiation:

$$(13.2.7) \qquad \boxed{R_j^{\ i}{}_{kl|m} \ + \ R_j^{\ i}{}_{lm|k} \ + \ R_j^{\ i}{}_{mk|l} \ = \ 0} \ .$$

Here, the covariant differentiation $_|$ is as defined in §13.1. For example, in natural coordinates, we have

$$R_{q\ rs|t}^{\ p} := \frac{\partial R_{q\ rs}^{\ p}}{\partial x^t} - R_{v\ rs}^{\ p}\,\gamma^v_{\ qt} + R_{q\ rs}^{\ v}\,\gamma^p_{\ vt} - R_{q\ vs}^{\ p}\,\gamma^v_{\ rt} - R_{q\ rv}^{\ p}\,\gamma^v_{\ st}\ .$$

The above Bianchi identities are the Riemannian versions of (3.2.4) and (3.5.3).

The **Ricci identity** measures the difference between the double covariant differentiation $_{|j|i}$ and its interchanged version, namely, $_{|i|j}$. For this reason, it is sometimes known as the **interchange formula**. The formula below comes from restricting (3.6.1) to Riemannian manifolds. When applied to a tensor field T of rank $\binom{1}{1}$, it reads:

$$\boxed{T^p_{\ q|j|i} - T^p_{\ q|i|j} = T^s_{\ q}\,R_{s\ ij}^{\ p} - T^p_{\ s}\,R_{q\ ij}^{\ s}}\ . \tag{13.2.8}$$

The above index correction pattern extends to tensors with any number of up and down indices.

13.2 B. Sectional Curvature

Restricting the concept of flag curvatures (§3.9) to Riemannian geometry, one gets sectional curvatures. As a review, let us begin with the notion of a flag on M. The act of installing a flag at $x \in M$ necessitates a nonzero $U \in T_xM$ which serves as the **flagpole**. The actual **flag** itself is described by one edge along the flagpole and another **transverse edge**, say $V := V^i\,\frac{\partial}{\partial x^i}$. It is denoted by $U \wedge V$. As we show, the actual "length" of the edge along the flagpole plays no role in the sectional curvature.

Now that we have a flag, we can associate with it a number $K(U,V)$. It is obtained by carrying out the following computation at the point $x \in M$:

$$\boxed{K(U,V) := \frac{V^i\ (U^j\ R_{jikl}\ U^l)\ V^k}{g(U,U)\ g(V,V) - [g(U,V)]^2}}\ . \tag{13.2.9}$$

The quantity $K(U,V)$ is called the **sectional curvature** of the flag $U \wedge V$ at the point x. It has certain basic properties:

- $K(U,V) = K(V,U)$.
- $K(U,V)$ remains unchanged if we multiply U or V by any nonzero number.
- More generally, if $\tilde{U} := \alpha\,U + \beta\,V$ and $\tilde{V} := \gamma\,U + \delta\,V$, where $\alpha\,\delta - \beta\,\gamma \neq 0$, then $K(\tilde{U},\tilde{V}) = K(U,V)$.

In the Riemannian case, the quantities R_{jikl} and g depend only on x. This is unlike the generic Finslerian case, in which they depend on both the position x and the direction $y \in T_xM$. Consequently, the sectional curvature $K(U,V)$ is a function of the position x and the flag $U \wedge V$, but has *no* separate dependence on the flagpole U.

* For Riemannian surfaces, there is only one sectional curvature $K = K(x)$ at each point x. This function K is called the **Gaussian curvature** function. It is computed as follows (see [On]). Take a g-orthonormal frame field ω^1, ω^2. The Levi-Civita (Christoffel) connection is comprised of a single 1-form because ${\omega_1}^1 = 0 = {\omega_2}^2$ and ${\omega_1}^2 = -{\omega_2}^1$. Finally,

$$d{\omega_2}^1 \;=\; K\,\omega^1 \wedge \omega^2 \;.$$

Compare this with our treatment of Finsler surfaces in Chapter 4, especially (4.4.5).

* In dimensions $n \geqslant 3$, there are many different 2-planes within each tangent space. Thus the sectional curvature typically depends on both the position x and the choice of the 2-plane. Nevertheless, one might wonder whether there are situations in which K has no dependence whatsoever on the 2-planes chosen. By specializing **Schur's lemma** (Lemma 3.10.2) to the Riemannian setting, we see that:

> The sectional curvature K has no dependence on the 2-planes defining the flags if and only if it is a constant on each connected component of our Riemannian manifold.

Sectional curvatures completely determine the full curvature tensor. If K has the constant value λ, such is apparent from Proposition 3.10.1, which says that

$$R_{jikl} \;=\; \lambda\,(g_{ik}\,g_{jl} \;-\; g_{il}\,g_{jk})\;. \tag{13.2.10}$$

See also Exercise 3.10.7. As for the general case, we need an abbreviation to minimize clutter. Set

$$(U,V) \;:=\; V^i\,(U^j\,R_{jikl}\,U^l)\,V^k\;,$$

so that (13.2.9) reads

$$(U,V) \;=\; K(U,V)\,\{\,g(U,U)\,g(V,V) \;-\; [g(U,V)]^2\,\}\;.$$

In other words, the symbol (U,V) abbreviates the sectional curvature times the area squared of the flag in question. According to Exercise 3.9.6,

$$6\,W^j\,V^i\,R_{jikl}\,X^k\,Y^l$$

is equal to

$$\begin{aligned}
&(X+V,Y+W) \;-\; (X+W,Y+V)\\
&+\;(X+W,Y) \;+\; (X+W,V) \;+\; (X,Y+V) \;+\; (W,Y+V)\\
&-\;(X+V,Y) \;-\; (X+V,W) \;-\; (X,Y+W) \;-\; (V,Y+W)\\
&+\;(X,W) \;+\; (V,Y) \;-\; (W,Y) \;-\; (X,V)\;.
\end{aligned}$$

This expression corrects two typographical errors in [CE], and agrees with the one given in [J].

13.2 C. Ricci Curvature and Einstein Metrics

The **Ricci tensor** is defined as

$$Ric_{ij} := R_i{}^s{}_{sj} . \tag{13.2.11}$$

By (13.2.3)–(13.2.5), Ric_{ij} is symmetric in the indices i, j. Its trace is the **scalar curvature** S:

$$S := g^{ij} Ric_{ij} . \tag{13.2.12}$$

Do not confuse the scalar curvature with the **Ricci scalar** Ric defined in Exercise 3.9.5. The Ricci scalar is a scaling invariant function on the tangent bundle TM. Equivalently, it is a function on the unit tangent bundle. At any unit vector U, the value of Ric is $U^i R_i{}^s{}_{sj} U^j$. Note that this is precisely a double contraction $U^i U^j Ric_{ij}$, and is manifestly different from the trace $g^{ij} Ric_{ij}$.

A Riemannian metric g is said to be an **Einstein metric** if

$$Ric_{ij} = \mathfrak{c}\, g_{ij} \quad \text{for some constant } \mathfrak{c} .$$

The manifold (M, g) is then called an **Einstein manifold**. By tracing on the i and k indices in (13.2.10), we see that:

> If (M, g) is an n-dimensional Riemannian manifold of constant sectional curvature λ, then it is also an Einstein manifold with constant $\mathfrak{c} = (n-1)\lambda$.

Conversely:

- If (M, g) is an Einstein *surface* with constant $\mathfrak{c}$, then it must have constant Gaussian curvature $\mathfrak{c}$. This follows from the fact that $Ric_{ij} = K\, g_{ij}$. See Exercise 13.2.1.
- If (M, g) is a 3-dimensional Einstein manifold with constant $\mathfrak{c}$, then it must have constant sectional curvature $\lambda = \frac{\mathfrak{c}}{2}$. According to [KN1], this is a **result of Schouten and Struik** [SS]. Its proof uses only the hypothesized dimension and the symmetry properties of the Riemann tensor. The details are relegated to the guided Exercises 13.2.6 and 13.2.7.
- Such a converse fails in higher dimensions. For example, in dimension four there are infinitely many (complete) Einstein metrics that do not have constant sectional curvature. See Exercise 13.2.8.

We conclude with a **Schur-type result**. Suppose, instead of the above definition for Einstein metrics, we merely require the Ricci tensor to be a *function* multiple of g. Namely, $Ric_{ij} = \mathfrak{c}(x)\, g_{ij}$. Then all Riemannian surfaces satisfy this criterion. For higher dimensions, it turns out that the function $\mathfrak{c}(x)$ must be constant on a connected Riemannian manifold. The derivation involves computations that center around the second Bianchi identity (13.2.7). See Exercise 13.2.9 for details.

Exercises

Exercise 13.2.1: Let (M, g) be a Riemannian surface with Riemann curvature tensor R_{jikl}. Denote its Gaussian, Ricci, and scalar curvatures by K, Ric_{ij}, and S, respectively.

(a) Explain why there is only one sectional curvature, and it is equal to the Gaussian curvature K.

(b) Prove directly that $R_{jikl} = K\,(g_{ik}\,g_{jl} - g_{il}\,g_{jk})$.

(c) Show that $Ric_{jl} = K\,g_{jl}$.

(d) Explain why (M, g) is an Einstein surface with constant $\mathfrak{c}$ if and only if it has constant Gaussian curvature $\mathfrak{c}$.

(e) Deduce that $S = 2\,K$.

Exercise 13.2.2:

(a) Explain why a Riemannian manifold is 1-dimensional if and only if its metric has the form $d\theta \otimes d\theta$ for some choice of coordinate θ.

(b) Show that the Gaussian curvature of the metric

$$dr \otimes dr \;+\; r^2\, d\theta \otimes d\theta$$

is identically zero. Here, $r > 0$.

(c) Let λ be a positive constant. Consider the metric

$$ds \otimes ds \;+\; \frac{1}{\lambda}\, \sin^2(\sqrt{\lambda}\, s)\, d\theta \otimes d\theta\;, \qquad 0 < s < \frac{\pi}{\sqrt{\lambda}}\;.$$

Show that this metric is Riemannian and has constant positive Gaussian curvature λ.

(d) Let λ be a negative constant. Consider the metric

$$ds \otimes ds \;-\; \frac{1}{\lambda}\, \sinh^2(\sqrt{-\lambda}\, s)\, d\theta \otimes d\theta\;,$$

where $s > 0$. Show that this metric is Riemannian and has constant negative Gaussian curvature λ.

These describe model Riemannian surfaces. We will understand their geometry better by the end of §13.3.

Exercise 13.2.3: On $M := (0, \pi) \times \mathbb{S}^1$, let us define a family of Riemannian metrics as follows:

$$g \;:=\; d\phi \otimes d\phi \;+\; \lambda\, \sin^2(\phi)\, d\theta \otimes d\theta\;.$$

This family is indexed by a positive constant λ. Show that (M, g) has constant Gaussian curvature 1 for all $\lambda > 0$.

Exercise 13.2.4: At each point x in $\mathbb{R}^4$, introduce the abbreviation $\rho := |x|$. On $\mathbb{R}^4 \setminus 0$, define the following global 1-forms:

$$\theta^1 := \frac{1}{\rho^2}\,(-\,x^4\,dx^1 \;-\; x^3\,dx^2 \;+\; x^2\,dx^3 \;+\; x^1\,dx^4)\ ,$$

$$\theta^2 := \frac{1}{\rho^2}\,(+\,x^3\,dx^1 \;-\; x^4\,dx^2 \;-\; x^1\,dx^3 \;+\; x^2\,dx^4)\ ,$$

$$\theta^3 := \frac{1}{\rho^2}\,(-\,x^2\,dx^1 \;+\; x^1\,dx^2 \;-\; x^4\,dx^3 \;+\; x^3\,dx^4)\ .$$

(a) Explain why these θ^i can be regarded as differential forms on the unit sphere $\mathbb{S}^3$.

(b) Verify that

$$d\theta^1 = 2\,\theta^2 \wedge \theta^3\ ,$$
$$d\theta^2 = 2\,\theta^3 \wedge \theta^1\ ,$$
$$d\theta^3 = 2\,\theta^1 \wedge \theta^2\ .$$

(c) Use part (b) to directly check that the metric $\sum_{i=1}^{3} \theta^i \otimes \theta^i$ on $\mathbb{S}^3$ has constant sectional curvature 1.

(d) Show that the standard flat metric on $\mathbb{R}^4$ can be re-expressed as

$$d\rho \otimes d\rho \;+\; \rho^2 \sum_{i=1}^{3} \theta^i \otimes \theta^i\ .$$

Exercise 13.2.5:

(a) Show that $K(U,V) = K(V,U)$.

(b) Suppose $\tilde{U} := \alpha\, U + \beta\, V$ and $\tilde{V} := \gamma\, U + \delta\, V$, where $\alpha\,\delta - \beta\,\gamma \neq 0$. Check that $K(\tilde{U},\tilde{V}) = K(U,V)$.

Exercise 13.2.6:

(a) At any unit vector U, the value of the Ricci scalar $Ric := U^i\, R_i{}^s{}_{sj}\, U^j$ is precisely the double contraction $U^i\, U^j\, Ric_{ij}$. Use this fact to help show that:

> Given any fixed unit vector U on an n-dimensional Riemannian manifold, the quantity $U^i\, U^j\, Ric_{ij}$ is the sum of $(n-1)$ sectional curvatures.

Hint: consider U as a member in an orthonormal basis.

(b) Let Ric_{ij} be the components of the Ricci tensor, relative to an arbitrary orthonormal basis $\{e_i : i = 1, \ldots, n\}$ at x. For any fixed index i_o, explain why

$$Ric_{i_o i_o} = \sum_{i \neq i_o} K(\,e_{i_o}\,,\,e_i\,)\ .$$

Exercise 13.2.7: Let (M, g) be an Einstein manifold. Express the Ricci tensor with respect to an orthonormal basis. Write out $Ric_{11}, \ldots, Ric_{nn}$ using part (b) of Exercise 13.2.6.

(a) Suppose $\dim M = 3$. Explain how the following equations come about:

$$\begin{aligned} \mathfrak{c} &= K(e_1, e_2) + K(e_1, e_3) , \\ \mathfrak{c} &= K(e_2, e_1) + K(e_2, e_3) , \\ \mathfrak{c} &= K(e_3, e_1) + K(e_3, e_2) . \end{aligned}$$

In view of part (a) of Exercise 13.2.5, these equations involve only three unknowns. Solve them to show that the sectional curvatures are all equal to $\frac{\mathfrak{c}}{2}$.

(b) Suppose $\dim M \geqslant 4$. In what way does the method in part (a) break down?

Exercise 13.2.8: Let $\mathbb{S}^2(3r)$ be the standard 2-sphere of radius $3r$. Set

$$M := \mathbb{S}^2(3r) \times \mathbb{S}^2(3r) .$$

(a) Define what should be meant by the product metric g on M.

(b) Show that g is an Einstein metric with constant $\frac{1}{r^2}$.

Exercise 13.2.9: Let (M, g) be an n-dimensional connected Riemannian manifold. In this exercise, we prove the Schur-type theorem discussed at the end of the section. Let us begin with the second Bianchi identity (13.2.7), namely,

$$R_j{}^i{}_{kl|m} + R_j{}^i{}_{lm|k} + R_j{}^i{}_{mk|l} = 0 .$$

(a) Contract with g^{jm}, lower the index i, and then contract with g^{ik}. Demonstrate that the resulting equation is

$$2\ Ric^s{}_{l|s} = S_{|l} = \frac{\partial S}{\partial x^l} .$$

Here, S is the scalar curvature defined in (13.2.12).

(b) Suppose $Ric_{ij} = \mathfrak{c}(x)\, g_{ij}$. Check that one must have $\mathfrak{c}(x) = \frac{S}{n}$.

(c) Substitute $Ric_{ij} = \frac{S}{n}\, g_{ij}$ into the equation we obtained in part (a). Show that the result can be manipulated into the form

$$(n-2)\ S_{|l} = 0 .$$

(d) Explain why, if $n \geqslant 3$ and $Ric_{ij} = \mathfrak{c}(x)\, g_{ij}$, then the function $\mathfrak{c}(x)$ must in fact be constant.

13.3 Warped Products and Riemannian Space Forms

In this section, we discuss some special Riemannian manifolds. One of the most basic properties of Riemannian metrics, which general Finsler metrics do not have, is the **splitting property**. This feature is illustrated by examples here, under the guise of warped products.

Let us be given two Riemannian manifolds (M_1, g_1) and (M_2, g_2). Form the Cartesian product $M = M_1 \times M_2$. The tangent space T_xM at $x = (x_1, x_2)$ has the natural decomposition

$$T_xM = T_{x_1}M_1 \oplus T_{x_2}M_2 .$$

The **canonical product metric** $g = g_1 \oplus g_2$ on $M = M_1 \times M_2$ is defined by

$$g(u,v) := g_1(u_1, v_1) + g_2(u_2, v_2) ,$$

where $u = u_1 \oplus u_2$, $v = v_1 \oplus v_2$ are elements in T_xM. More generally, suppose ϕ and ψ are two real-valued C^∞ positive functions on M. Then

$$g := \phi^2 g_1 \oplus \psi^2 g_2$$

is a Riemannian metric on M. It is called a **warped product metric**.

13.3 A. One Special Class of Warped Products

Many standard Riemannian metrics can be realized as warped products. Let us illustrate this phenomenon. Consider the Cartesian product $M = (a,b) \times \check{M}$, where $\check{M}$ is a smooth $(n-1)$-dimensional manifold. Let $\check{g}$ be any Riemannian metric on $\check{M}$, and let $\varphi : (a,b) \to (0,\infty)$ be any C^∞ positive function. Define the following warped product:

$$g := dt \otimes dt + \varphi^2(t)\, \check{g} . \tag{13.3.1}$$

Let us express the geometry of g in terms of that of $\check{g}$ and the first two derivatives of the function φ.

To this end:

* Let $\{\check{\omega}^\alpha : \alpha = 2, \ldots, n\}$ be an arbitrary coframe on the $(n-1)$-dimensional $\check{M}$. Set

$$\omega^1 := dt , \quad \omega^\alpha := \check{\omega}^\alpha \quad (\alpha = 2, \ldots, n) .$$

Then $\{\omega^a : a = 1, \ldots, n\}$ is an adapted coframe on M. In this section, the lower case Greek indices run from 2 to n instead of the usual 1 to $n-1$. Note that

$$\begin{pmatrix} g_{11} & g_{1\beta} \\ g_{\alpha 1} & g_{\alpha\beta} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \varphi^2\, \check{g}_{\alpha\beta} \end{pmatrix} ,$$

$$\begin{pmatrix} g^{11} & g^{1\beta} \\ g^{\alpha 1} & g^{\alpha\beta} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{\varphi^2}\, \check{g}^{\alpha\beta} \end{pmatrix} .$$

In particular, if we begin with an orthonormal coframe on $\breve{M}$, then the corresponding *adapted* coframe on M will only be orthogonal, but typically no longer orthonormal.

* Denote the Levi-Civita (Christoffel) connection forms on $\breve{M}$ by $\{\breve{\omega}_\beta{}^\alpha\}$, and those on M by $\{\omega_b{}^a\}$. They are characterized by torsion-freeness and metric compatibility. Review §13.1.

* The curvature forms of $(\breve{M}, \breve{g})$ are defined as

$$\breve{\Omega}_\beta{}^\alpha := d\breve{\omega}_\beta{}^\alpha - \breve{\omega}_\beta{}^\gamma \wedge \breve{\omega}_\gamma{}^\alpha ,$$

and those on (M, g) are defined as

$$\Omega_b{}^a := d\omega_b{}^a - \omega_b{}^c \wedge \omega_c{}^a .$$

We first compute the connection forms of M. This is most easily done by calculating the Christoffel symbols in natural coordinates. Set

$$\omega^1 := dt , \quad \omega^\alpha := \breve{\omega}^\alpha = dx^\alpha \quad (\alpha = 2, \ldots, n) ,$$

where the x^α are local coordinates on $\breve{M}$. Together with $x^1 := t$, they provide local coordinates on M. Straightforward computations give the connection forms:

$$\begin{aligned} \omega_1{}^1 &= 0 \\ \omega_\beta{}^1 &= -\frac{\varphi'}{\varphi}\, g_{\beta\gamma}\, \omega^\gamma \\ \omega_1{}^\alpha &= \frac{\varphi'}{\varphi}\, \omega^\alpha \\ \omega_\beta{}^\alpha &= \breve{\omega}_\beta{}^\alpha + \frac{\varphi'}{\varphi}\, \delta_\beta{}^\alpha\, dt . \end{aligned} \tag{13.3.2a–d}$$

Upon exterior differentiation, we get the curvature 2-forms:

$$\begin{aligned} \Omega_1{}^1 &= 0 \\ \Omega_\beta{}^1 &= -\frac{\varphi''}{\varphi}\, g_{\beta\gamma}\, dt \wedge \omega^\gamma \\ \Omega_1{}^\alpha &= \frac{\varphi''}{\varphi}\, dt \wedge \omega^\alpha \\ \Omega_\beta{}^\alpha &= \breve{\Omega}_\beta{}^\alpha + \frac{1}{2}\left(\frac{\varphi'}{\varphi}\right)^2 (g_{\beta\gamma}\, \delta^\alpha_\xi - g_{\beta\xi}\, \delta^\alpha_\gamma)\, \omega^\gamma \wedge \omega^\xi . \end{aligned} \tag{13.3.3a–d}$$

The components of the curvature tensors are contained in the above 2-forms.

$$\breve{\Omega}_\beta{}^\alpha = \frac{1}{2} \breve{R}_\beta{}^\alpha{}_{\gamma\xi}\, \breve{\omega}^\gamma \wedge \breve{\omega}^\xi ,$$

$$\Omega_b{}^a = \frac{1}{2} R_b{}^a{}_{cd}\, \omega^c \wedge \omega^d .$$

In view of (13.3.3a–d), we find that:

(13.3.4a–d)
$$
\begin{aligned}
R_{1\ cd}^{\ 1} &= 0 \\
R_{\beta\ cd}^{\ 1} &= \frac{\varphi''}{\varphi}\,(g_{\beta c}\,\delta^1_{\ d} - g_{\beta d}\,\delta^1_{\ c}) \\
R_{1\ cd}^{\ \alpha} &= -\,\frac{\varphi''}{\varphi}\,(\delta^\alpha_{\ c}\,\delta^1_{\ d} - \delta^\alpha_{\ d}\,\delta^1_{\ c}) \\
R_{\beta\ cd}^{\ \alpha} &= \breve{R}_{\beta\ \gamma\xi}^{\ \alpha}\,\delta^\gamma_{\ c}\,\delta^\xi_{\ d} + \left(\frac{\varphi'}{\varphi}\right)^2 (g_{\beta c}\,\delta^\alpha_{\ d} - g_{\beta d}\,\delta^\alpha_{\ c})\,.
\end{aligned}
$$

Recall that the Ricci tensors are defined by

$$\breve{R}ic_{\alpha\beta} := R_{\alpha\ \sigma\beta}^{\ \sigma}\,, \qquad Ric_{ab} := R_{a\ sb}^{\ s}\,,$$

in which the two automatic summations (one in σ and one in s) are understood. They are symmetric tensors. Formulas (13.3.4a–d) imply that:

(13.3.5a–c)
$$
\begin{aligned}
Ric_{11} &= -\,(n-1)\,\frac{\varphi''}{\varphi} \\
Ric_{\alpha 1} &= 0 = Ric_{1\beta} \\
Ric_{\alpha\beta} &= \breve{R}ic_{\alpha\beta} - \left[\frac{\varphi''}{\varphi} + (n-2)\left(\frac{\varphi'}{\varphi}\right)^2\right] g_{\alpha\beta}\,.
\end{aligned}
$$

Finally, the scalar curvature is the trace of the Ricci tensor:

$$\breve{S} := \breve{g}^{\alpha\beta}\,\breve{R}ic_{\alpha\beta}\,, \qquad S := g^{ab}\,Ric_{ab}\,.$$

Using (13.3.5a–c), one gets:

(13.3.6)
$$S = \frac{1}{\varphi^2}\,\breve{S} - (n-1)\left[2\,\frac{\varphi''}{\varphi} + (n-2)\left(\frac{\varphi'}{\varphi}\right)^2\right].$$

Formulas (13.3.3)–(13.3.6) are tensorial. They are valid in any coframe on $\breve{M}$, and in the corresponding adapted coframe on M. By specializing them to the case of

$$\varphi(t) := \mathfrak{s}_\lambda(t) := \left\{\begin{array}{c} \frac{1}{\sqrt{\lambda}}\,\sin(\sqrt{\lambda}\,t) \\ t \\ \frac{1}{\sqrt{-\lambda}}\,\sinh(\sqrt{-\lambda}\,t) \end{array}\right\} \text{ resp., for } \left\{\begin{array}{c} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \end{array}\right\},$$

one can deduce the following proposition.

Proposition 13.3.1.

* *Let $(\breve{M}, \breve{g})$ be a Riemannian manifold of dimension $n-1$.*
* *Set $M := (0,r) \times \breve{M}$, where $r := \infty$ if $\lambda \leqslant 0$ and $r := \frac{\pi}{\sqrt{\lambda}}$ if $\lambda > 0$.*

* *On the n-dimensional M, consider the warped product metric*

$$g := dt \otimes dt + \mathfrak{s}_\lambda^2(t)\, \breve{g} .$$

* *Let $\breve{K}$, $\breve{Ric}_{\alpha\beta}$, $\breve{S}$ be the sectional, Ricci, and scalar curvatures of $(\breve{M}, \breve{g})$. Denote their counterparts on (M, g) by K, Ric_{ab}, and S.*

Suppose $\dim \breve{M} = 1$. That is, $n = 2$.

Then the warped product (M, g) is a Riemannian surface with constant Gaussian curvature $K = \lambda$. Equivalently, $Ric_{ab} = \lambda\, g_{ab}$ or $S = 2\,\lambda$.

Suppose $\dim \breve{M} \geqslant 2$. That is, $n \geqslant 3$.

(1) *If $(\breve{M}, \breve{g})$ has constant sectional curvature $\breve{K} = 1$, then (M, g) has constant sectional curvature $K = \lambda$.*

(2) *If $\breve{Ric}_{\alpha\beta} = (n-2)\,\breve{g}_{\alpha\beta}$, then $Ric_{ab} = \lambda\,(n-1)\,g_{ab}$.*

(3) *If $\breve{S} = (n-1)\,(n-2)$, then $S = n\,(n-1)\,\lambda$.*

Remarks:

- In the $n = 2$ case, the statements about K, Ric_{ab}, and S are equivalent because of Exercise 13.2.1. The conclusion $K = \lambda$ comes from Exercise 13.2.2.
- As for the rest of the proposition, its proof involves direct applications of (13.3.4)–(13.3.6) and a keen awareness of $g_{\alpha\beta} = \varphi^2\, \breve{g}_{\alpha\beta}$. We omit the straightforward but tedious details.
- For the $n \geqslant 3$ cases, conclusion (1) is most relevant to us. It shows us how to use induction to construct Riemannian space forms. And, for some special Riemannian metrics, it offers an elegant indirect way of getting the sectional curvature. We illustrate these techniques below.

13.3 B. Spheres and Spaces of Constant Curvature

Here are two applications of Proposition 13.3.1. The first one concerns the sectional curvature of the standard unit sphere $\mathbb{S}^n$.

* Fix a pair of antipodal points on $\mathbb{S}^n$ and call them the north and south poles. Orient the Cartesian axes of $\mathbb{R}^{n+1}$ such that the north pole has coordinates $(0, \ldots, 0, 1)$. In other words, it is along the positive x^{n+1} axis.

* For each fixed choice of $0 < \phi < \pi$, the n-dimensional hyperplane $x^{n+1} = \cos\phi$ intersects $\mathbb{S}^n$ at a standard $(n-1)$-sphere with radius $\sin\phi$. In this light, the parameter

$$\phi := \cos^{-1}(x^{n+1})$$

gives the colatitude of the said $(n-1)$-sphere.

* This is the perspective of spherical coordinates. It gives a diffeomorphism $\mathfrak{p}$ from $\mathbb{S}^n \smallsetminus \{\text{the two poles}\}$ onto $(0,\pi)\times\mathbb{S}^{n-1}$. One then finds that

$$g_{\mathbb{S}^n} \;=\; \mathfrak{p}^*[\, d\phi\otimes d\phi \;+\; \sin^2(\phi)\; g_{\mathbb{S}^{n-1}} \,] \tag{13.3.7}$$

except at the two poles. The metric, however, is smoothly defined (hence continuous) at these poles. The singularities are only of coordinate nature.

For $n = 2$, either Proposition 13.3.1 or Exercise 13.2.2 tells us that $\mathbb{S}^2$ has Gaussian curvature 1 away from its two poles. By continuity, $\mathbb{S}^2$ must have $K = 1$ everywhere. Now apply Proposition 13.3.1 inductively to the recursion relation (13.3.7). We see that for $n \geqslant 3$, $\mathbb{S}^n \smallsetminus \{\text{the two poles}\}$ has constant sectional curvature 1. By continuity again, the same holds for the entire unit sphere $\mathbb{S}^n$.

As a second application of Proposition 13.3.1, we compute the curvature of some standard Riemannian metrics. In order to reduce clutter, let us introduce two abbreviations. At any point $x = (x^1,\ldots,x^n)\in\mathbb{R}^n$, set

$$|x| \;:=\; \sqrt{(x^1)^2 + \cdots + (x^n)^2}\;,$$
$$x_i \;:=\; \delta_{ik}\, x^k\;.$$

For each constant λ, the metric we are interested in is $g := g_{ij\,(x)}\, dx^i\otimes dx^j$, where

$$g_{ij\,(x)} \;:=\; \frac{x_i\, x_j}{|x|^2} \;+\; \left[\frac{\mathfrak{s}_\lambda(\,|x|\,)}{|x|}\right]^2 \left(\delta_{ij} \;-\; \frac{x_i\, x_j}{|x|^2}\right). \tag{13.3.8}$$

These g_{ij} can be extended smoothly to the origin $x = 0$. And, g is a Riemannian metric

$$\text{on } \mathcal{D} \;:=\; \left\{\begin{array}{l} \text{the entire } \mathbb{R}^n :\; |x| \;<\; \infty \\ \text{the open ball} :\; |x| \;<\; \dfrac{\pi}{\sqrt{\lambda}} \end{array}\right\} \text{ resp., for } \left\{\begin{array}{l} \lambda \leqslant 0 \\ \lambda > 0 \end{array}\right\}.$$

It is not difficult to deduce from (13.3.8) that, according to g, all points x at a fixed positive distance $\rho(=|x|)$ from the origin collectively take on the geometry of a standard sphere. But that sphere only has radius ρ if $\lambda = 0$. For $\lambda \neq 0$, the radius in question is $\mathfrak{s}_\lambda(\rho)$. To make this precise, consider the diffeomorphism

$$\mathfrak{p}(x) \;:=\; \left(|x|\,,\; \frac{x}{|x|}\right)$$

from $\mathcal{D}\smallsetminus 0$ onto $(0,\infty)\times\mathbb{S}^{n-1}$. One can check that

$$g \;=\; \mathfrak{p}^*[\, d\rho\otimes d\rho \;+\; \mathfrak{s}_\lambda^2(\rho)\; g_{\mathbb{S}^{n-1}} \,] \tag{13.3.9}$$

except at the origin. Note that the origin is only a coordinate singularity because the metric g is smoothly (hence continuously) defined there.

For $n = 2$, either Proposition 13.3.1 or Exercise 13.2.2 tells us that g has Gaussian curvature λ away from the origin. By continuity, it must have $K = \lambda$ everywhere. Since we have shown that the standard unit spheres have constant sectional curvature 1, Proposition 13.3.1 immediately implies that, when $n \geqslant 3$, our Riemannian metric g has constant sectional curvature λ away from the origin. By continuity, this must hold at the origin as well.

13.3 C. Standard Models of Riemannian Space Forms

Let us discuss the three cases $\lambda = 0$, $\lambda > 0$, and $\lambda < 0$ in some detail. As in (13.3.9), ρ means $|x|$. However, pull-back maps are suppressed.

- For zero sectional curvature, the manifold in question is $\mathbb{R}^n$. It is simply connected. The Riemannian metric at any point x away from the origin has the description

$$d\rho \otimes d\rho \ + \ \rho^2 \, g_{S^{n-1}} \ , \quad 0 < \rho < \infty \, .$$

This is none other than the usual Euclidean metric $\delta_{ij}\, dx^i \otimes dx^j$ expressed in spherical coordinates (which has the origin as a coordinate singularity). The resulting Riemannian manifold is complete, and is known as the **flat** or **Euclidean model**.

- For constant positive sectional curvature $\lambda > 0$, the manifold is technically the open ball of radius $\frac{\pi}{\sqrt{\lambda}}$, centered at the origin of Euclidean $\mathbb{R}^n$, $n \geqslant 2$. It is simply connected. The Riemannian metric is

$$d\rho \otimes d\rho \ + \ \frac{1}{\lambda} \, \sin^2(\sqrt{\lambda}\, \rho) \, g_{S^{n-1}} \ , \quad 0 < \rho < \frac{\pi}{\sqrt{\lambda}} \, .$$

It can be extended smoothly to the origin, where $\rho = 0$.

The above Riemannian metric is not complete. To find its completion, let us consider the standard n-sphere of radius $\frac{1}{\sqrt{\lambda}}$, centered at the origin of Euclidean $\mathbb{R}^{n+1}$. As in the discussions leading up to (13.3.7), we introduce the colatitude ϕ which ranges from 0 to π. These two values correspond to the north and the south poles, respectively. The ambient Euclidean structure induces a Riemannian metric on this sphere. Parametrizing the sphere with spherical coordinates (of which ϕ is one), one finds that the induced metric looks like

$$\frac{1}{\lambda} \, d\phi \otimes d\phi \ + \ \frac{1}{\lambda} \, \sin^2(\phi) \, g_{S^{n-1}} \ , \quad 0 < \phi < \pi \, .$$

Note that $\phi = 0, \pi$ are coordinate singularities.

Instead of the angular variable ϕ, let us use $s := \frac{1}{\sqrt{\lambda}}\, \phi$ instead. The quantity s measures arc length along a longitude, from the north pole down to the point with colatitude ϕ. The induced metric on our sphere then has the description

$$ds \otimes ds \;+\; \frac{1}{\lambda}\,\sin^2(\sqrt{\lambda}\,s)\;g_{\mathbb{S}^{n-1}}\;, \quad 0 < s < \frac{\pi}{\sqrt{\lambda}}$$

away from the two poles. This is identical in form to the one on the open ball.

Therefore the completion we seek is the n-sphere of radius $\frac{1}{\sqrt{\lambda}}$, together with its Riemannian structure induced by the Euclidean ambient space. Since $n \geqslant 2$, this standard sphere is simply connected. It has constant positive sectional curvature λ, and is known as a **spherical** or an **elliptical model**.

- For constant negative sectional curvature $\lambda < 0$, the manifold in question is again $\mathbb{R}^n$, which is simply connected. The Riemannian metric in this case is

$$d\rho \otimes d\rho \;-\; \frac{1}{\lambda}\,\sinh^2(\sqrt{-\lambda}\,\rho)\;g_{\mathbb{S}^{n-1}}\;, \quad 0 < \rho < \infty\,.$$

It can be smoothly extended to the origin (where $\rho = 0$). The resulting Riemannian manifold is complete.

The above Riemannian manifold is isometric to the **Poincaré model** of **hyperbolic space**. For simplicity, let us only demonstrate this for a special case: $n = 2$ and $\lambda = -1$.

The underlying manifold of the Poincaré disc is the open disc of radius 2, centered at the origin of $\mathbb{R}^2$. See [On]. (For higher dimensions and arbitrary negative constant λ, see [KN1].) Its Riemannian metric in polar coordinates r, θ is

$$\frac{1}{[1 - \frac{r^2}{4}]^2}\,(dr \otimes dr \;+\; r^2\,d\theta \otimes d\theta)\;, \quad 0 < r < 2\,.$$

The origin (where $r = 0$) is only a coordinate singularity. Set

$$s \;:=\; \log\left(\frac{1 + \frac{r}{2}}{1 - \frac{r}{2}}\right)\,.$$

Straightforward calculations show that this converts the metric to

$$ds \otimes ds \;+\; \sinh^2(s)\,d\theta \otimes d\theta\;, \quad 0 < s < \infty\;,$$

as claimed.

Any complete, simply connected Riemannian manifold of constant sectional curvature λ is called a **Riemannian space form**. We have examined three specific **models** of Riemannian space forms:

* Euclidean space $\mathbb{E}^n$.
* The standard spheres. For $\lambda = 1$, the standard sphere we need is simply the unit sphere $\mathbb{S}^n$.

* Hyperbolic spaces. For $\lambda = -1$, the usual notation for the hyperbolic space in question is $\mathbb{H}^n$.

It is shown in the next section that every Riemannian space form is necessarily isometric to one of these models.

Exercises

Exercise 13.3.1: Transform (13.3.2) from the adapted coordinate coframe into an arbitrary adapted coframe. Do the formulas remain unchanged in *form*?

Exercise 13.3.2: Fill in the details in the derivation of (13.3.3)–(13.3.6).

Exercise 13.3.3: Use (13.3.4)–(13.3.6) to prove Proposition 13.3.1.

Exercise 13.3.4:

(a) Show that the analogue of (13.3.7) for a sphere of radius c is

$$ds \otimes ds \; + \; c^2 \; \sin^2\Big(\frac{s}{c}\Big) \; g_{\mathbb{S}^{n-1}} \; .$$

Here, s equals c times the colatitude ϕ introduced in (13.3.7).

(b) Establish (13.3.9).

Exercise 13.3.5: We claimed that the metric defined by (13.3.8) extends smoothly to the origin o of $\mathbb{R}^n$. For each of the three cases $\lambda = 0$, $\lambda < 0$, $\lambda > 0$, identify explicitly the inner product on the tangent space $T_o\mathbb{R}^n$.

Exercise 13.3.6: Let (M, g) be a Riemannian manifold.

(a) Let $\mathfrak{c}$ be a positive constant. Define a new Riemannian metric $\breve{g} := \mathfrak{c}\, g$ on M. Show that:

$$\begin{aligned}
\breve{g}_{ij} &= \mathfrak{c}\; g_{ij} \; , \\
\breve{g}^{ij} &= \frac{1}{\mathfrak{c}}\; g^{ij} \; , \\
\breve{\gamma}^{i}{}_{jk} &= \gamma^{i}{}_{jk} \; , \\
\breve{R}_{j}{}^{i}{}_{kl} &= R_{j}{}^{i}{}_{kl} \; , \\
\breve{R}_{jikl} &= \mathfrak{c}\; R_{jikl} \; , \\
\breve{K}(U,V) &= \frac{1}{\mathfrak{c}}\; K(U,V) \; , \\
\breve{Ric}_{ij} &= Ric_{ij} \; , \\
\breve{S} &= \frac{1}{\mathfrak{c}}\; S \; .
\end{aligned}$$

(b) Suppose the positive constant $\mathfrak{c}$ is replaced by a positive smooth function $\mathfrak{c}(x)$ on M. What happens to the above transformation formulas? Hint: see pp. 183–184 of [SY] if necessary.

13.4 Hopf's Classification of Riemannian Space Forms

There are three technical ingredients that we need for proving Hopf's classification theorem. One holds only for complete Riemannian manifolds of constant sectional curvature, while the other two are valid on arbitrary Riemannian manifolds. Let us describe them in turn.

Let (M, g) be a complete Riemannian manifold of constant sectional curvature λ. Fix $p \in M$ and consider the exponential map $\exp_p : T_pM \to M$. Since we have hypothesized completeness, (6.1.7) is valid for all y. Namely, for all y and for all $0 < t \leqslant 1$, we have

$$(*) \qquad t^2\, g\Big((\exp_{p*})_{(ty)} V_1 \,,\, (\exp_{p*})_{(ty)} V_2 \Big) \;=\; \mathfrak{s}^2_{\lambda r^2}(t)\; \hat{g}_{p\,(ty)}(V_1, V_2) \;,$$

where r denotes $\sqrt{g(y,y)}$. Here, V_1, V_2 are tangent vectors on the linear manifold T_pM that emanate from the point ty and are $\hat{g}_p$-orthogonal to the straight ray in question. And

$$\mathfrak{s}_{\lambda r^2}(t) \;:=\; \left\{ \begin{array}{c} \dfrac{1}{\sqrt{\lambda}\, r} \;\sin(\sqrt{\lambda}\, r\, t) \\ t \\ \dfrac{1}{\sqrt{-\lambda}\, r} \;\sinh(\sqrt{-\lambda}\, r\, t) \end{array} \right\} \text{ resp., for } \left\{ \begin{array}{c} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \end{array} \right\} .$$

Let us simplify (*) a bit.

* On the right-hand side, $\hat{g}_{p\,(ty)}$ is simply $g_{ij\,(p)}\, dy^i \otimes dy^j$ because the g_{ij} depend only on p. Each tangent vector $V := V^i \frac{\partial}{\partial y^i}$ of the linear manifold T_pM can be regarded as a "point" (again denoted V) $V^i \frac{\partial}{\partial x^i}$ of T_pM. To keep pace with this change of perspective, we replace $g_{ij\,(p)}\, dy^i \otimes dy^j$ by $g_{ij\,(p)}\, dx^i \otimes dx^j$, which is g_p. Thus

$$\hat{g}_{p\,(ty)}(V_1, V_2) \;=\; g_p(V_1, V_2) \;.$$

* Express y as rY, where Y has unit length, and abbreviate rt as τ. Note that $\mathfrak{s}_{\lambda r^2}(t) = \frac{1}{r}\, \mathfrak{s}_\lambda(\tau)$. Then (*) becomes

$$\tau^2\, g\Big((\exp_{p*})_{(\tau Y)} V_1 \,,\, (\exp_{p*})_{(\tau Y)} V_2 \Big) \;=\; \mathfrak{s}^2_\lambda(\tau)\; g_p(V_1, V_2) \;,$$

where $0 < \tau < \infty$.

Now relabel Y as y and τ as t. We have just demonstrated that on a complete Riemannian manifold of constant sectional curvature λ, the following holds:

$$g\Big((\exp_{p*})_{(ty)} V_1 \,,\, (\exp_{p*})_{(ty)} V_2 \Big) \;=\; \frac{\mathfrak{s}_\lambda^2(t)}{t^2}\; g_p(V_1, V_2) \tag{13.4.1}$$

for all (*unit*) directions $y \in T_pM$, all V_1, V_2 that are g_p-*orthogonal* to y, and all positive t. Here,

$$\mathfrak{s}_\lambda(t) \;:=\; \left\{ \begin{array}{c} \frac{1}{\sqrt{\lambda}}\ \sin(\sqrt{\lambda}\, t) \\ t \\ \frac{1}{\sqrt{-\lambda}}\ \sinh(\sqrt{-\lambda}\, t) \end{array} \right\} \text{ resp., for } \left\{ \begin{array}{c} \lambda > 0 \\ \lambda = 0 \\ \lambda < 0 \end{array} \right\} .$$

As $t \to 0^+$, the left-hand side becomes $g_p(V_1, V_2)$. L'Hôpital's rule shows that on the right, $\frac{S_\lambda(t)}{t}$ approaches 1. Therefore we may say that (13.4.1) is valid at $t = 0$ as well.

Two more ingredients are needed in the proof of Hopf's classification theorem. Given an arbitrary Riemannian manifold:

- The Gauss lemma (Lemma 6.1.1) tells us that

$$g\Big((\exp_{p*})_{(ty)} V \,,\, (\exp_{p*})_{(ty)} \alpha y \Big) \;=\; 0 \;=\; g_p(V, \alpha y) \tag{13.4.2}$$

  for all constants α, and for all V that are g_p-*orthogonal* to y.

- Also, observe that $T(t) := (\exp_{p*})_{(ty)} y$ is the velocity field of the autoparallel $\exp_p(ty)$, and $T(0) = y$. Thus, by Exercise 5.2.3 or Exercise 13.1.1, we see that

$$g\Big((\exp_{p*})_{(ty)}\, \alpha y \,,\, (\exp_{p*})_{(ty)}\, \beta y \Big) \;=\; g_p(\alpha y, \beta y) \tag{13.4.3}$$

  for all constants α and β.

Theorem 13.4.1 (Hopf). *Let (M, g) be a complete connected Riemannian manifold of constant sectional curvature λ. Denote by $(\tilde{M}, \tilde{g})$ the standard model of a simply connected complete Riemannian space with constant sectional curvature λ. It is hyperbolic space for $\lambda < 0$, Euclidean space for $\lambda = 0$, and the standard sphere of radius $\frac{1}{\sqrt{\lambda}}$ for $\lambda > 0$. Then:*

(1) *There exists a smooth local isometry $\varphi : (\tilde{M}, \tilde{g}) \to (M, g)$. It is an onto map and a covering projection.*

(2) *If M is also simply connected, that surjection φ must necessarily be injective, hence a smooth diffeomorphism. In that case, (M, g) is isometric to $(\tilde{M}, \tilde{g})$.*

Remark: The essence of our proof comes from that given in Gallot–Hulin–Lafontaine [GHL]. But our exposition is quite a bit more leisurely.

Proof. Here, both $\tilde{M}$ and M are complete and connected. So, according to Theorem 9.2.1, every smooth local isometry $\varphi : \tilde{M} \to M$ must be an onto map and a covering projection.

> Also, (2) follows from (1). Indeed, $\tilde{M}$ is simply connected and M, being a manifold, is always *locally* simply connected. Therefore at any point $p \in M$, the group of deck transformations is isomorphic to the fundamental group $\pi(M,p)$. If M is hypothesized to be simply connected as well, then $\pi(M,p)$ is trivial and consequently there is only one deck. So the surjection φ is injective, and is thus a diffeomorphism. By (1), this diffeomorphism φ is presumed to be a smooth local isometry; therefore it is a smooth isometry.

It suffices to construct the smooth local isometry φ. As we show, there are three levels of subtlety. The $\lambda = 0$ case is the simplest. It is followed by the $\lambda < 0$ case, in which we have to come to terms with a conformal factor, albeit straightforwardly. The same issue of a conformal factor resurfaces in the $\lambda > 0$ case, but this time the peripheral arguments are complicated by a topological feature of the sphere.

The $\lambda = 0$ case:

Fix any point $p \in M$ and consider the exponential map $\exp_p$. It is smooth because we are in the Riemannian category.

Since $\lambda = 0$, we have $\mathfrak{s}_\lambda(t) = t$ in (13.4.1). This, together with (13.4.2) and (13.4.3), tells us that $\exp_p$ is a local isometry between (T_pM, g_p) and (M, g). The domain (T_pM, g_p), being a finite-dimensional inner product space, is complete. Also, both the domain and the target space are connected. By Theorem 9.2.1, the map $\exp_p$ must in fact be surjective, and is a covering projection. (Note: This last conclusion also follows from the Cartan–Hadamard theorem.)

Now (T_pM, g_p) is, upon the choice of an orthonormal basis, isometric to Euclidean $\mathbb{R}^n$, which is the $(\tilde{M}, \tilde{g})$ in the $\lambda = 0$ case. Call this linear isometry $f : (\tilde{M}, \tilde{g}) \to (T_pM, g_p)$.

Finally, define $\varphi := \exp_p \circ f$ and we are done.

The $\lambda < 0$ case:

Again, fix $p \in M$. We have the smooth exponential map $\exp_p : T_pM \to M$. Since the sectional curvatures are negative, Proposition 9.1.2 tells us that no geodesic can contain any conjugate points. Thus, by Proposition 7.1.1, the exponential map $\exp_p$ is a local diffeomorphism. (As a matter of fact, the Cartan–Hadamard theorem says that in this case, the exponential map is a covering projection. But that is considerably more than what is relevant.)

Let $\tilde{p}$ be any point in hyperbolic space $(\tilde{M}, \tilde{g})$. Since $\tilde{M}$ is simply connected, the Cartan–Hadamard theorem (Theorem 9.4.1) assures us that the exponential map $\widetilde{\exp}_{\tilde{p}} : T_{\tilde{p}}\tilde{M} \to \tilde{M}$ is a smooth diffeomorphism. Hence its inverse makes sense.

Choose an orthonormal basis for T_pM and one for $T_{\tilde p}\tilde M$. This gives rise to an isometry $f : (T_{\tilde p}\tilde M\,,\,\tilde g_{\tilde p}) \to (T_pM\,,\,g_p)$.

Consider the composition

$$\varphi\ :\ (\tilde M,\tilde g)\ \xrightarrow{\widetilde{\exp}_{\tilde p}^{-1}}\ (T_{\tilde p}\tilde M\,,\,\tilde g_{\tilde p})\ \xrightarrow{f}\ (T_pM\,,\,g_p)\ \xrightarrow{\exp_p}\ (M,g)\ .$$

Observe that at the point $\tilde p \in \tilde M$, the derivative φ_* is f_*, which is a linear isometry. So, it is the behavior of φ_* away from $\tilde p$ that we need to understand.

* Take any point $\tilde x \in \tilde M$ different from $\tilde p$. Completeness implies (see Proposition 6.5.1) that there is a unit speed geodesic $\widetilde{\exp}_{\tilde p}(t\tilde y)$ from $\tilde p$ which reaches $\tilde x$ when, say, $t = r > 0$. Take any two tangent vectors at $\tilde x$ that are $\tilde g$-orthogonal to this geodesic. Since $\widetilde{\exp}_{\tilde p}$ is invertible, we can express these vectors as $(\widetilde{\exp}_{\tilde p *})_{(r\tilde y)}\tilde V_1$ and $(\widetilde{\exp}_{\tilde p *})_{(r\tilde y)}\tilde V_2$. The Gauss lemma (Lemma 6.1.1) then implies that $\tilde V_1$ and $\tilde V_2$ are both $\tilde g_{\tilde p}$-orthogonal to $\tilde y$ in $T_{\tilde p}\tilde M$. Now use the fact that $(\widetilde{\exp}_{\tilde p}^{-1})_* = (\widetilde{\exp}_{\tilde p *})^{-1}$, together with (13.4.1) applied to $(\tilde M,\tilde g)$. After a slight rearrangement, we get

$$\tilde g_{\tilde p}(\tilde V_1,\tilde V_2)\ =\ \frac{(\sqrt{-\lambda}\,r)^2}{\sinh^2(\sqrt{-\lambda}\,r)}\ \tilde g\Big(\,(\widetilde{\exp}_{\tilde p *})_{(r\tilde y)}\tilde V_1\,,\,(\widetilde{\exp}_{\tilde p *})_{(r\tilde y)}\tilde V_2\Big)\,.$$

This shows that on tangent vectors at $\tilde x$ which are $\tilde g$-orthogonal to the said geodesic, the action of $(\widetilde{\exp}_{\tilde p}^{-1})_*$ is conformal, with conformal factor $\frac{\sqrt{-\lambda}\,r}{\sinh(\sqrt{-\lambda}\,r)}$.

* The map $f : (T_{\tilde p}\tilde M\,,\,\tilde g_{\tilde p}) \to (T_pM\,,\,g_p)$ is a linear isometry between inner product spaces. Since f is linear, its derivative f_* may be identified with f itself. Under the linear isometry f, the straight ray $t\tilde y$ gets transformed to one, say ty, in T_pM. The vectors $\tilde V_1$, $\tilde V_2$ that are $\tilde g_{\tilde p}$-orthogonal to $\tilde y$ become vectors V_1, V_2 in T_pM that are g_p-orthogonal to y. Furthermore,

$$g_p(V_1,V_2)\ =\ \tilde g_{\tilde p}(\tilde V_1,\tilde V_2)\ .$$

This last statement holds because f is an isometry. It remains valid even if the vectors V are not orthogonal to the straight rays in question.

* Since the vectors V_1, V_2 in T_pM are g_p-orthogonal to y, (13.4.1) says that

$$g\Big(\,(\exp_{p*})_{(ry)}V_1\,,\,(\exp_{p*})_{(ry)}V_2\Big)\ =\ \frac{\sinh^2(\sqrt{-\lambda}\,r)}{(\sqrt{-\lambda}\,r)^2}\ g_p(V_1,V_2)\,.$$

In other words, on vectors that are g_p-orthogonal to the straight ray ty in T_pM, the action of the derivative map $\exp_{p*}$ is conformal, with conformal factor $\frac{\sinh(\sqrt{-\lambda}\,r)}{\sqrt{-\lambda}\,r}$.

These three observations together establish that, on tangent vectors at $\tilde{x}$ which are $\tilde{g}$-orthogonal to the geodesic $\widetilde{\exp}_{\tilde{p}}(t\tilde{y})$, the derivative of φ is conformal, with conformal factor 1. It turns out that φ_* also has conformal factor 1 along other directions, primarily because of the Gauss lemma. Such can be deduced by decomposing arbitrary tangent vectors into components along and orthogonal to the geodesic in question, then following with judicious applications of (13.4.2), (13.4.3) to (M, g) and $(\tilde{M}, \tilde{g})$. Therefore φ is a smooth local isometry.

The $\lambda > 0$ case:

As before, fix $p \in M$ and consider the smooth exponential map $\exp_p$. Parts (2) and (3) of Theorem 9.5.2 say that the first conjugate point of p, in any direction y, occurs exactly at a distance of $\frac{\pi}{\sqrt{\lambda}}$ units along the geodesic $\exp_p(ty)$ emanating from p. Proposition 7.1.1, together with the Inverse Function theorem, then tells us that the map

$$\exp_p \ : \ \mathrm{B}_p\left(\frac{\pi}{\sqrt{\lambda}}\right) \ \rightarrow \ M$$

is a *local* diffeomorphism.

Now fix $\tilde{p}$ on the sphere $\tilde{M}$. The antipodal point $-\tilde{p}$ is the cut locus of $\tilde{p}$. So the injectivity radius at $\tilde{p}$ is $\frac{\pi}{\sqrt{\lambda}}$. This means that we have a diffeomorphism

$$\widetilde{\exp}_{\tilde{p}} \ : \ \tilde{\mathrm{B}}_{\tilde{p}}\left(\frac{\pi}{\sqrt{\lambda}}\right) \ \rightarrow \ \tilde{M} \setminus -\tilde{p} \, .$$

Thus, on $\tilde{M} \setminus -\tilde{p}$, the inverse $\widetilde{\exp}_{\tilde{p}}^{-1}$ makes sense.

Choose an orthonormal basis for T_pM and one for $T_{\tilde{p}}\tilde{M}$. This gives rise to an isometry $f : (T_{\tilde{p}}\tilde{M}, \tilde{g}_{\tilde{p}}) \rightarrow (T_pM, g_p)$. In particular, the linear transformation f restricts to an isometry between the open tangent balls $\tilde{\mathrm{B}}_{\tilde{p}}(\pi/\sqrt{\lambda})$ and $\mathrm{B}_p(\pi/\sqrt{\lambda})$.

Consider the composition

$$\varphi \ : \ \tilde{M} \setminus -\tilde{p} \ \overset{\widetilde{\exp}_{\tilde{p}}^{-1}}{\longrightarrow} \ \tilde{\mathrm{B}}_{\tilde{p}}\left(\frac{\pi}{\sqrt{\lambda}}\right) \ \overset{f}{\longrightarrow} \ \mathrm{B}_p\left(\frac{\pi}{\sqrt{\lambda}}\right) \ \overset{\exp_p}{\longrightarrow} \ M \, .$$

At the point $\tilde{p} \in \tilde{M}$, the derivative φ_* is simply f_*, which is a linear isometry. So, let us study instead the behavior of φ_* away from $\tilde{p}$. On arguments that are similar to those of the negative curvature case, we shall be brief.

* Take any point $\tilde{x} \in \tilde{M} \setminus -\tilde{p}$ that is different from $\tilde{p}$. Completeness implies that there is a unit speed geodesic $\widetilde{\exp}_{\tilde{p}}(t\tilde{y})$ from $\tilde{p}$ which reaches $\tilde{x}$ when $t = r > 0$. Take any two tangent vectors at $\tilde{x}$ that are $\tilde{g}$-orthogonal to this geodesic. Since $\widetilde{\exp}_{\tilde{p}}$ maps diffeomorphically onto $\tilde{M} \setminus -\tilde{p}$, we can express these vectors uniquely as $(\widetilde{\exp}_{\tilde{p}*})_{(r\tilde{y})}\tilde{V}_1$ and $(\widetilde{\exp}_{\tilde{p}*})_{(r\tilde{y})}\tilde{V}_2$. The Gauss lemma then assures us that both $\tilde{V}_1$

and $\tilde{V}_2$ are $\tilde{g}_{\tilde{p}}$-orthogonal to $\tilde{y}$ in $T_{\tilde{p}}\tilde{M}$. Now apply (13.4.1) to $(\tilde{M}, \tilde{g})$. It leads to

$$\tilde{g}_{\tilde{p}}(\tilde{V}_1, \tilde{V}_2) \;=\; \frac{(\sqrt{\lambda}\, r)^2}{\sin^2(\sqrt{\lambda}\, r)}\; \tilde{g}\Big((\widetilde{\exp}_{\tilde{p}*})_{(r\tilde{y})} \tilde{V}_1 \,,\, (\widetilde{\exp}_{\tilde{p}*})_{(r\tilde{y})} \tilde{V}_2 \Big)\,.$$

Thus, on tangent vectors at $\tilde{x}$ that are $\tilde{g}$-orthogonal to the said geodesic, the action of $(\widetilde{\exp}_{\tilde{p}}^{-1})_*$ is conformal, with factor $\frac{\sqrt{\lambda}\, r}{\sin(\sqrt{\lambda}\, r)}$.

* The map $f : (T_{\tilde{p}}\tilde{M}, \tilde{g}_{\tilde{p}}) \to (T_p M, g_p)$ is a linear isometry between two inner product spaces. It transforms the straight ray $t\tilde{y}$ of $\tilde{\mathrm{B}}_{\tilde{p}}(\pi/\sqrt{\lambda})$ into the straight ray ty of $\mathrm{B}_p(\pi/\sqrt{\lambda})$. The vectors $\tilde{V}_1$, $\tilde{V}_2$ in $T_{\tilde{p}}\tilde{M}$ that are $\tilde{g}_{\tilde{p}}$-orthogonal to $\tilde{y}$ become, under the derivative map f_* ($\cong f$), vectors V_1, V_2 in T_pM that are g_p-orthogonal to y. Furthermore,

$$g_p(V_1, V_2) \;=\; \tilde{g}_{\tilde{p}}(\tilde{V}_1, \tilde{V}_2)\,.$$

This last statement remains valid even if the vectors V are replaced by vectors not necessarily orthogonal to the straight rays.

* Since the vectors V_1, V_2 in T_pM are g_p-orthogonal to y, (13.4.1) says that

$$g\Big((\exp_{p*})_{(ry)} V_1 \,,\, (\exp_{p*})_{(ry)} V_2 \Big) \;=\; \frac{\sin^2(\sqrt{\lambda}\, r)}{(\sqrt{\lambda}\, r)^2}\; g_p(V_1, V_2)\,.$$

In other words, on vectors that are g_p-orthogonal to the straight ray ty, the action of $\exp_{p*}$ is conformal, with factor $\frac{\sin(\sqrt{\lambda}\, r)}{\sqrt{\lambda}\, r}$.

The two nonunit conformal factors are reciprocals of each other. So, on tangent vectors at $\tilde{x}$ that are $\tilde{g}$-orthogonal to the geodesic $\widetilde{\exp}_{\tilde{p}}(t\tilde{y})$, the derivative of φ is conformal, with conformal factor 1. As in the negative curvature case, judicious applications of (13.4.2), (13.4.3) to (M, g) and $(\tilde{M}, \tilde{g})$ will show that φ_* also has conformal factor 1 along other directions. Therefore φ is a smooth local isometry from the *punctured* sphere $\tilde{M} \setminus -\tilde{p}$ into M.

We now extend the domain of definition of this φ, so that it becomes a local isometry from the entire sphere $\tilde{M}$ into M. To this end:

- First express the local isometry as a commutation relation:

$$\varphi \circ \widetilde{\exp} \;=\; \exp \circ \varphi_*\,.$$

This statement is valid everywhere on the punctured sphere $\tilde{M} \setminus -\tilde{p}$. So we evaluate it at some fixed $\tilde{q}$ that is not $-\tilde{p}$, and get

$$\varphi \circ \widetilde{\exp}_{\tilde{q}} \;=\; \exp_{\varphi(\tilde{q})} \circ \varphi_{*\tilde{q}}\,.$$

It is now ready to accept elements of $T_{\tilde{q}}\tilde{M}$ as inputs.

- Restrict the above to the open tangent ball $\tilde{\mathrm{B}}_{\tilde{q}}(\pi/\sqrt{\lambda})$ in $T_{\tilde{q}}\tilde{M}$. Then the inverse of $\widetilde{\exp}_{\tilde{q}}$ exists, has domain $\tilde{M} \setminus -\tilde{q}$, and

$$\varphi = \exp_{\varphi(\tilde{q})} \circ \varphi_{*\tilde{q}} \circ \widetilde{\exp}_{\tilde{q}}^{-1} .$$

In order for *both* sides to make sense, the domain of this equation technically should not include the points $-\tilde{q}$ and $-\tilde{p}$.

- This is an interesting formula for φ. It says that for *every* fixed choice of $\tilde{q}$ which is different from $-\tilde{p}$, our map φ has the said representation on $\tilde{M} \setminus \{-\tilde{q}, -\tilde{p}\}$. Let us now agree to choose $\tilde{q}$ so that it is different from $\tilde{p}$ as well. Then the right-hand side of our formula actually makes sense at the point $-\tilde{p}$!
- We have just discovered that the *singularity* $-\tilde{p}$ in the original definition of φ is removable. And, we can extend φ to $-\tilde{p}$ by

$$\varphi(-\tilde{p}) := \left[\exp_{\varphi(\tilde{q})} \circ \varphi_{*\tilde{q}} \circ \widetilde{\exp}_{\tilde{q}}^{-1}\right](-\tilde{p}) .$$

Since the extension was carried out solely with the commutation relation, one can check that the extended φ retains its status as a local isometry.

We have finally obtained a smooth local isometry from the standard sphere $(\tilde{M}, \tilde{g})$ into (M, g). This completes the proof of (1). Since we have already explained at the beginning how (2) follows from (1), the proof of Hopf's classification theorem is complete. □

Exercises

Exercise 13.4.1: Revisit the proof of Hopf's classification theorem for the $\lambda < 0$ and $\lambda > 0$ cases. There, we claimed that what was said, together with (13.4.2) and (13.4.3), implied that the constructed map φ was a local isometry. Fill in the details.

Exercise 13.4.2: Recall our hyperbolic metric

$$d\rho \otimes d\rho - \frac{1}{\lambda} \sinh^2(\sqrt{-\lambda}\,\rho)\, g_{\mathbb{S}^{n-1}} , \quad 0 < \rho < \infty$$

on $\mathbb{R}^n$. Here ρ means $|x|$ and, as pointed out before, this metric can be smoothly extended to the origin (where $\rho = 0$).

(a) Write down the metric for the Poincaré model, in dimension n, with constant negative sectional curvature $\lambda < 0$. Without doing any computation, explain why it is necessarily isometric to the above hyperbolic space.

(b) Now calculate an explicit formula for this isometry. Hint: you might want to consult §13.3C.

13.5 The Divergence Lemma and Hopf's Theorem

We begin with the fact that $d\sqrt{g} = \frac{1}{2}\sqrt{g}\, g^{ij}\, dg_{ij}$. Equivalently,

$$\boxed{\frac{\partial\sqrt{g}}{\partial x^k} = \frac{1}{2}\sqrt{g}\, g^{ij}\, \frac{\partial g_{ij}}{\partial x^k}} \;. \tag{13.5.1}$$

For the purposes of this section, we rephrase this as

$$\boxed{\gamma^i{}_{ik} = \gamma^i{}_{ki} = \frac{\partial}{\partial x^k}\log\sqrt{g}} \;. \tag{13.5.2}$$

Using (13.5.2), a straightforward computation shows that

$$\boxed{V^i{}_{|i}\sqrt{g} =: (\nabla_i V^i)\sqrt{g} = \frac{\partial}{\partial x^i}(V^i\sqrt{g})} \;. \tag{13.5.3}$$

Here,

$$\boxed{V^i{}_{|i} := \partial_i V^i + V^k\gamma^i{}_{ki}}$$

is the **divergence** of the vector field V.

We have the following **divergence lemma**:

Lemma 13.5.1. *Let V be any globally defined vector field on a compact Riemannian manifold (M, g) without boundary. Denote the volume form in natural coordinates by $\sqrt{g}\, dx$. Then*

$$\int_M (\nabla_i V^i)\sqrt{g}\, dx = 0\,.$$

Proof. Because M is compact, we can decompose it as a finite union of closed subsets $\overline{U}_\alpha$ whose interiors U_α are mutually disjoint from each other. Furthermore, we may assume that each $\overline{U}_\alpha$ lies inside some coordinate neighborhood.

Over each $\overline{U}_\alpha$, we integrate (13.5.3) and get:

$$\int_{\overline{U}_\alpha}(\nabla_i V^i)\sqrt{g}\, dx = \int_{\overline{U}_\alpha}\frac{\partial}{\partial x^i}(V^i\sqrt{g})\, dx\,. \tag{*}$$

To the right-hand side of (*), we apply the standard divergence theorem in Euclidean space. This converts (*) to the statement

$$\int_{\overline{U}_\alpha}(\nabla_i V^i)\sqrt{g}\, dx = \text{the outward flux of } V \text{ through } \partial\overline{U}_\alpha\,. \tag{**}$$

Sum (**) over α.

- Adding the left-hand sides produces $\int_M (\nabla_i V^i)\sqrt{g}\, dx$. This is because the closed subsets have union equal to M, and their mutual intersections are either empty or have measure zero.

- Adding the right-hand sides gives zero. This is so because M is boundaryless, hence the oriented boundaries of the $\overline{U}_\alpha$ must cancel each other out algebraically. □

Let us apply the divergence lemma to deduce a theorem of Hopf's. Given any C^2 function f on (M, g), define its **Laplacian** as

$$\boxed{\Delta f \;:=\; g^{ij}\, f_{|i|j} \;=:\; \nabla^i \nabla_i f} \;. \tag{13.5.4}$$

Equivalently:

$$\boxed{\Delta f \;=\; \frac{1}{\sqrt{g}}\,\frac{\partial}{\partial x^i}\left(\sqrt{g}\, g^{ij}\, \frac{\partial f}{\partial x^j}\right)} \;. \tag{13.5.5}$$

The function f is said to be **harmonic** if $\Delta f = 0$.

- Our definition of the Laplacian follows the tradition in analysis and geometry. There is *no* minus sign in front. However, when analysts and geometers speak about **eigenvalues** of the Laplacian, they use the equation $(-\Delta)f = \lambda f$ instead. This switch from Δ to $-\Delta$ ensures that the eigenvalues λ are always nonnegative. See [SY].
- There is also the Laplace–Beltrami operator on differential forms, introduced in the next section. When that is applied to functions (0-forms), one gets $\delta(df)$. Here, the codifferential δ turns out to be the *negative* of the divergence. See (13.6.2). So, the Laplace–Beltrami operator and the Laplacian do *differ* by a sign.
- In the literature, the above distinction is not so fastidiously maintained. Some may use the word "Laplacian" to abbreviate the phrase "Laplace–Beltrami operator." See, for instance, [BL1]. Others may use the symbol Δ to denote the Laplace–Beltrami operator. See, for example, [J].

Theorem 13.5.2 (Hopf). *Let (M, g) be a compact connected Riemannian manifold without boundary. Then every globally defined function f, with $\Delta f \geqslant 0$ everywhere or $\Delta f \leqslant 0$ everywhere, must be constant. In particular, there are no nonconstant globally defined harmonic functions on such M.*

Proof. By replacing f with $-f$ if necessary, we may assume without loss of generality that $\Delta f \geqslant 0$ everywhere.

In view of the divergence lemma and (13.5.4), we have

$$\int_M \Delta f \,\sqrt{g}\, dx \;=\; 0 \;.$$

This integral makes sense because M is compact. Since $\Delta f \geqslant 0$ by assumption, and f is C^2, we conclude that $\Delta f = 0$ everywhere on M.

We now know that the said f is actually harmonic. In particular,

$$\int_M f \, \Delta f \, \sqrt{g} \, dx = 0 .$$

Integrating by parts and using the divergence lemma again, we obtain

$$- \int_M g^{ij} \, f_{|i} \, f_{|j} \, \sqrt{g} \, dx = 0 .$$

Hence $\partial_i f = f_{|i}$ vanishes identically on M. Since M is connected, f must be constant. □

Exercises

Exercise 13.5.1:

(a) In the proof of the divergence lemma, explain in detail why the outward fluxes cancel out each other.

(b) In the proof of Hopf's theorem, explain what is meant by "integrating by parts and using the divergence lemma."

Exercise 13.5.2: Derive (13.5.5) from (13.5.4).

Exercise 13.5.3: Suppose (M, g) is a compact Riemannian manifold without boundary. Let f and φ be twice differentiable functions on M. Check that

$$\int_M (f \, \Delta\varphi \; - \; \varphi \, \Delta f \,) \, \sqrt{g} \, dx = 0 .$$

This is a special case of **Green's second identity**. It shows that the Laplacian is a formally self-adjoint operator.

13.6 The Weitzenböck Formula and the Bochner Technique

In §13.5, we defined the Laplacian on functions. It is manifestly in divergence form. Here, we extend that definition to differential k-forms. We show that the result can be manipulated into a divergence form plus corrections terms involving the curvature. Such is known as a **Weitzenböck formula**.

Let (M, g) be an n-dimensional Riemannian manifold with volume element dV_g. A canonical one can be specified uniquely if M is orientable.

* In natural coordinates $x^1, \ldots, x^n$, the volume element is taken to be $\sqrt{g} \; dx^1 \wedge \cdots \wedge dx^n$.

* More generally, in an arbitrary coframe $\{\omega^i\}$, naturally dual to a frame $\{b_j\}$, our dV_g is given by $\sqrt{\det[\, g(b_i, b_j)\,]} \; \omega^1 \wedge \cdots \wedge \omega^n$.

Let us write generic k-forms θ on M as

$$(*) \qquad \theta \;=\; \frac{1}{k!}\,\theta_{i_1\ldots i_k}\,\omega^{i_1}\wedge\cdots\wedge\omega^{i_k}\ ,$$

where each coefficient function $\theta_{i_1\ldots i_k}$ is totally skew-symmetric in all its indices. Our convention for the wedge product is as stipulated in §13.2. Thus, as a covariant tensor of rank k, the object θ can be re-expressed as $\theta = \theta_{i_1\ldots i_k}\,\omega^{i_1}\otimes\cdots\otimes\omega^{i_k}$ without the $\frac{1}{k!}$. That factor is present in (*) because the spanning set $\{\omega^{i_1}\wedge\cdots\wedge\omega^{i_k}\}$ for k-forms is $k!$-fold redundant.

The metric g defines an inner product on the k-forms and on the covariant tensors of rank k. These inner products are

$$\langle\,\theta\,,\,\psi\,\rangle_{\text{form}} \;:=\; \frac{1}{k!}\int_M \theta_{i_1\ldots i_k}\,g^{i_1j_1}\cdots g^{i_kj_k}\,\psi_{j_1\ldots j_k}\,dV_g\ ,$$

$$\langle\,\theta\,,\,\psi\,\rangle_{\text{tensor}} \;:=\; \int_M \theta_{i_1\ldots i_k}\,g^{i_1j_1}\cdots g^{i_kj_k}\,\psi_{j_1\ldots j_k}\,dV_g\ .$$

They make sense, for example, on all objects with compact support. Not distinguishing these two inner products can lead to all sorts of paradoxes.

Using the skew-symmetry of $\theta_{i_1\ldots i_k}$, one can rewrite $d\theta$ covariantly as

$$(13.6.1)\quad d\theta \;=\; \frac{1}{(k+1)!}\,(\,\nabla_s\,\theta_{i_1\ldots i_k}\;+\;\text{cyclic perms.}\,)\;\omega^s\wedge\omega^{i_1}\wedge\cdots\wedge\omega^{i_k}\ .$$

Here:

- Covariant differentiation of tensor components is denoted by ∇_s instead of $_{|s}$. This choice of notation will bring out, in the simplest way, the nature of the combinatorics at hand.
- Those additional terms inside the parentheses are all generated from the first by *cyclic* permutations on $s, i_1, \ldots, i_k$. These terms all have plus signs when k is even, and alternating signs when k is odd. For example,

$$k=1:\qquad d\theta \;=\; \frac{1}{2!}\,(\,\nabla_s\,\theta_i - \nabla_i\,\theta_s\,)\;\omega^s\wedge\omega^i\ ,$$

$$k=2:\qquad d\theta \;=\; \frac{1}{3!}\,(\,\nabla_s\,\theta_{ij} + \nabla_i\,\theta_{js} + \nabla_j\,\theta_{si}\,)\;\omega^s\wedge\omega^i\wedge\omega^j\ .$$

Define the **codifferential** δ [which converts a k-form into a $(k-1)$-form] to be the *formal* adjoint of d with respect to $\langle\,,\,\rangle_{\text{form}}$, not $\langle\,,\,\rangle_{\text{tensor}}$. Suppose θ is a k-form and ϕ is a $(k-1)$-form, both compactly supported in the interior of M; then $\delta\theta$ is implicitly given by

$$\langle\,\delta\theta\,,\,\phi\,\rangle_{\text{form}} \;=\; \langle\,\theta\,,\,d\phi\,\rangle_{\text{form}}\ .$$

Using (13.6.1) and integration-by-parts, we get the explicit formula

$$(13.6.2)\qquad \delta\theta \;=\; \frac{1}{(k-1)!}\,(\,-\nabla^{i_1}\,\theta_{i_1i_2\ldots i_k}\,)\;\omega^{i_2}\wedge\cdots\wedge\omega^{i_k}\ .$$

* In this covariantized formula, ∇^i means $g^{ij}\,\nabla_j$.
* Traditionally, the codifferential is defined as $\delta := (-1)^{n(k+1)+1} * d *$, where $*$ is the **Hodge star** operator. See, for example, [Wa].

On differential k-forms θ, the analogue of the Laplacian is the **Laplace–Beltrami operator** $d\,\delta + \delta\, d$. The zeroes of this operator are known as the **harmonic k-forms**. Using (13.6.1) and (13.6.2) judiciously, one finds that $(d\,\delta + \delta\, d)\;\theta$ has the preliminary expression

$$\frac{1}{k!}\,[-\nabla^s\nabla_s\,\theta_{i_1\ldots i_k} \;+\; k\,(\nabla^s\nabla_{i_1} - \nabla_{i_1}\nabla^s\,)\,\theta_{si_2\ldots i_k}\,]\,\omega^{i_1}\wedge\cdots\wedge\omega^{i_k}\;.$$

It is understood that the coefficients have to be explicitly antisymmetrized in the indices $i_1,\ldots,i_k$.

Applying the Ricci identity (13.2.8), followed by the first Bianchi identity (13.2.6), that preliminary expression can be converted to the following **Weitzenböck formula**:

(13.6.3)

$$\boxed{\begin{aligned}(d\,\delta+\delta\,d)\,\theta \;=\; &\frac{1}{k!}\,\omega^{i_1}\wedge\cdots\wedge\omega^{i_k}\\ &\Big[-\nabla^s\nabla_s\,\theta_{i_1\ldots i_k}\\ &\quad+\;k\;Ric_{i_1}{}^{s}\;\theta_{si_2\ldots i_k}\;+\;\frac{k(k-1)}{2}\;R_r{}^{s}{}_{i_1i_2}\;\theta^r{}_{si_3\ldots i_k}\Big]\;.\end{aligned}}$$

We hasten to point out that:

- The coefficients on the right-hand side must be explicitly antisymmetrized in the indices i_1 through i_k.
- Our $R_q{}^p{}_{rs}$ is written as $R^p{}_{qrs}$ by some authors.

As an application, consider the Weitzenböck formula for 1-forms. The message in this case is particularly elegant, and reads:

$$(d\,\delta+\delta\,d)\,\theta \;=\; (\,-\nabla^s\nabla_s\,\theta_i \;+\; Ric_i{}^s\;\theta_s\,)\,\omega^i\;. \tag{13.6.4}$$

Our application exemplifies the so-called **Bochner technique**.

Theorem 13.6.1 (Bochner). *Let θ be a globally defined 1-form on a compact boundaryless Riemannian manifold (M,g).*

* *Suppose the quadratic form defined by the Ricci tensor of g is positive-definite. That is, $Ric_{ij}\,V^i\,V^j > 0$ at x whenever $V \neq 0$ at x. Then: θ is harmonic if and only if it is identically zero.*
* *Suppose the quadratic form defined by the Ricci tensor of g is non-negative. That is, $Ric_{ij}\,V^i\,V^j \geqslant 0$ for all V. Then: θ is harmonic if and only if it is parallel; that is, $\nabla\theta = 0$.*

Proof. Take the inner product of (13.6.4) with θ. All the integrals make sense because M is compact and θ is globally defined. Carry out an integration-by-parts on the term involving the double covariant derivative. Since M is boundaryless, there is no subsequent boundary integral because of the divergence lemma (Lemma 13.5.1). The result reads:

$$(13.6.5) \quad \langle \theta , (d\delta + \delta d)\theta \rangle_{\text{form}} = \int_M (\nabla_s \theta_i \nabla^s \theta^i + Ric_{ij}\, \theta^i \theta^j)\, dV_g .$$

Suppose the Ricci tensor is positive-definite:

If $\theta = 0$, then it is trivially harmonic. Conversely, suppose $(d\delta + \delta d)\theta = 0$. Then the left-hand side of (13.6.5) is zero. This means that two nonnegative integrals have a vanishing sum. Hence each must vanish. In particular, the integral of $Ric(\theta, \theta) := Ric_{ij}\, \theta^i \theta^j$ is zero. Since $Ric(\theta, \theta)$ is a continuous nonnegative function, it must vanish at each x. The hypothesized positive-definiteness of Ric_{ij} now implies that θ must be identically zero on M.

Suppose the Ricci tensor is nonnegative:

If $\nabla\theta = 0$, then according to (13.6.1) and (13.6.2), θ must be closed and coclosed, hence harmonic. Incidentally, if one had concentrated on (13.6.4) instead, then $\nabla\theta = 0$ would seem to only imply that $(d\delta + \delta d)\theta = Ric_i{}^s\, \theta_s\, \omega^i$. But that term on the right is actually expressible (see Exercise 13.6.4) as a commutator of covariant derivatives on θ. Thus it is zero after all, and we can again conclude that θ is harmonic.

Conversely, suppose $(d\delta + \delta d)\theta = 0$. Then the left-hand side of (13.6.5) is zero. Again, each integral must vanish. In particular, the integral of $\|\nabla\theta\|^2 := \nabla_s \theta_i \nabla^s \theta^i$ is zero. This implies that $\|\nabla\theta\|$ vanishes pointwise. Thus θ is parallel. $\square$

For matters such as Weitzenböck formulas in the Finslerian realm, see [BL3] and related articles in the monograph [AL].

Exercises

Exercise 13.6.1: Derive the covariantized formulas (13.6.1) and (13.6.2).

Exercise 13.6.2:

(a) Let θ be a globally defined k-form on a compact Riemannian manifold without boundary. Prove that it is harmonic if and only if it is closed and coclosed.

(b) Show that the Laplace–Beltrami operator is formally self-adjoint.

Exercise 13.6.3: Establish the Weitzenböck formula (13.6.3) for the two simplest cases, namely, for 1-forms and 2-forms.

Exercise 13.6.4:

(a) Produce the origin of the formula

$$\theta_{p|j|i} \;-\; \theta_{p|i|j} \;=\; -\,\theta_s\, R_p{}^s{}_{ij}\ .$$

(b) What do $Ric_i{}^s\,\theta_s$ and $(\nabla^p\nabla_i - \nabla_i\nabla^p)\theta_p$ have in common?

Exercise 13.6.5: Consider the Cartesian product $\mathbb{S}^n \times \mathbb{S}^1$, $n \geqslant 2$, endowed with the product metric g.

(a) Use (13.3.5) to help show that the Ricci tensor of g is nonnegative. Specifically, along directions tangent to $\mathbb{S}^n$, it agrees with that of $\mathbb{S}^n$. And, along the direction tangent to $\mathbb{S}^1$, it's zero.

(b) Let t denote the coordinate along $\mathbb{S}^1$. Is dt is a harmonic 1-form of the metric g? Is there any other harmonic 1-form of g that is not a constant multiple of dt?

References

[AL] P. L. Antonelli and B. Lackey (eds.), *The Theory of Finslerian Laplacians and Applications*, MAIA 459, Kluwer Academic Publishers, 1998.

[BL1] D. Bao and B. Lackey, *Randers surfaces whose Laplacians have completely positive symbol*, Nonlinear Analysis **38** (1999), 27–40.

[BL3] D. Bao and B. Lackey, *A geometric inequality and a Weitzenböck formula for Finsler surfaces*, The Theory of Finslerian Laplacians and Applications, P. Antonelli and B. Lackey (eds.), MAIA 459, Kluwer Academic Publishers, 1998, pp. 245–275.

[CE] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, North Holland/American Elsevier, 1975.

[GHL] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*, Universitext, 2nd ed., Springer-Verlag, 1990.

[J] J. Jost, *Riemannian Geometry and Geometric Analysis*, Universitext, Springer-Verlag, 1995.

[KN1] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, vol. I, Wiley-Interscience, 1963 (1996).

[On] B. O'Neill, *Elementary Differential Geometry*, 2nd ed., Academic Press, 1997.

[SS] J. Schouten and D. Struik, *On some properties of general manifolds relating to Einstein's theory of gravitation*, Amer. J. Math. **43** (1921), 213–216.

[SY] R. Schoen and S. T. Yau, *Lectures on Differential Geometry*, Conference Proceedings and Lecture Notes in Geometry and Topology, vol. I, International Press, 1994.

[Wa] F. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Scott-Foresman, 1971.

Chapter 14

Minkowski Spaces, the Theorems of Deicke and Brickell

14.1 Generalities and Examples

Let $y \mapsto F(y)$ be a Minkowski norm on $\mathbb{R}^n$. It is nonnegative and has the following defining properties.

* **Regularity**: F is smooth at **all** $y \neq 0$.
* **Positive homogeneity**: $F(\lambda y) = \lambda\, F(y)$ for all $\lambda > 0$.
* **Strong convexity**: The fundamental tensor

$$g_{ij}(y) \;:=\; \left[\frac{1}{2}\, F^2 \right]_{y^i y^j}$$

  is defined and positive-definite at **all** $y \neq 0$.

Given these, it follows from Theorem 1.2.2 that $F(y)$ must be positive for all nonzero y. More important, review the fundamental inequality (1.2.3) and its myriad interpretations (§1.2C).

The restriction of a Finsler structure F to any specific tangent space T_xM gives a Minkowski norm F_x. Thus, any Finsler manifold may be viewed as a smoothly varying family of Minkowski spaces $\{ (T_xM, F_x) : x \in M \}$.

Every Minkowski space $(\mathbb{R}^n, F)$ is at the same time a Finsler manifold, albeit a rather simple one. This is realized by assigning the same F to every tangent space of $\mathbb{R}^n$ (which is identifiable with $\mathbb{R}^n$ itself). Review §1.3A for explicit details. The resulting Finsler manifold is an example of a locally Minkowskian space.

* As such, both the Chern curvatures R and P must vanish (see the end of §3.3). Therefore the only geometrical invariant is the Cartan tensor and its "vertical" derivatives. For instance, Minkowski planes are classified by the Cartan scalar alone (Proposition 4.2.1).
* According to Exercise 5.3.4, all geodesics in locally Minkowskian spaces are straight lines. Hence these spaces are geodesically both forward and backward complete. This is true in spite of the fact that F may only be positively homogeneous of degree 1. See also Exercise 14.1.1.

Let us list some important examples of Minkowski planes $(\mathbb{R}^2, F)$. In order to reduce clutter, we use the following abbreviations:

$$p := y^1 , \quad q := y^2 .$$

- **Minkowski planes of Randers type.** Here, the Minkowski norm has the form

$$F(y) := \sqrt{c_{ij}\, y^i y^j} + b_i\, y^i ,$$

where b_i and c_{ij} are constants, and

$$c^{ij}\, b_i\, b_j < 1 .$$

This last criterion is needed to ensure strong convexity and positivity at all nonzero y; see §11.1. These Minkowski norms can be put into normal form as follows. Choose an orthonormal basis for the inner product defined by c_{ij}. Moreover, align the first vector of this basis with the vector whose components are $b^i := c^{ij} b_j$, so that they point in the *same* direction. In terms of the global linear coordinates defined by this carefully chosen basis, we have the **normal form**

$$F(y) = \sqrt{p^2 + q^2} + B\, p , \tag{14.1.1}$$

where $B < 1$ is a positive constant. A moment's thought tells us that B is the length, with respect to c_{ij}, of the covector b_i. Compare the above with Matsumoto's treatment in [M8]. See Exercise 14.1.2 for details about the indicatrix, and [BL1] for a further restriction on B (namely $B \lessapprox 0.9139497$) which has an analytical origin.

- **Regularized quartic metrics.** We encountered this family in §1.3A. The Minkowski norms in question have the formula

$$F(y) := \sqrt{\sqrt{p^4 + q^4} + \epsilon\,[p^2 + q^2]}\ , \tag{14.1.2}$$

where ϵ is a constant. This may be viewed as a perturbation of the quartic metric. As explained in §1.3A, strong convexity is violated at some nonzero y if $\epsilon = 0$. See also Exercise 14.1.3. If $\epsilon > 0$, our treatment in §1.3A shows that the fundamental tensor g_{ij} is positive-definite because both its eigenvalues are positive everywhere on $\mathbb{R}^2 \smallsetminus 0$. The function F is then a bonafide Minkowski norm for each choice of the positive constant ϵ. In this sense, the perturbation has regularized the quartic metric. We believe that the regularized quartic metric has considerable potential in physical applications.

- **Regularized quartics with drift term.** The simplest one we have in mind is

$$F(y) = \sqrt{\sqrt{p^4 + q^4} + \epsilon\,[p^2 + q^2]} + B\,p\ , \tag{14.1.3}$$

where ϵ and B are both positive constants. It would be an interesting calculation (Exercise 14.1.4) to determine the restrictions (if any) on the positive constants ϵ and B, so that the resulting F has all the defining attributes of a Minkowski norm. More generally, Bryant's example (§12.7) inspires us to consider

$$F(y) = \sqrt{\sqrt{P_4(p,q)} + \epsilon\, Q_2(p,q)} + (b_1\,p + b_2\,q)\ .$$

Here, P_4 is a fourth order polynomial, Q_2 is a quadratic, and ϵ is a positive constant. However, the constants b_1, b_2 need not be positive. They represent the components of the drift covector, in a Randers sense. There is much mathematics to be learned from this family.

We conclude this section with two useful tools:

* The **Okubo technique.** This was first introduced in Exercise 1.2.8, and is applicable to all dimensions. It is useful when the indicatrix S is specified by physical or geometrical concerns, and one wants to recover the candidate Minkowski norm F from the specified S. We say "candidate" because it is not guaranteed that the F obtained indeed satisfies the defining properties of Minkowski norms. For instance, in Exercise 1.2.8, the indicatrix was *a priori* stipulated to be the convex limaçon $\rho = 3 + \cos\phi$. The Okubo technique produced, rather effortlessly, the explicit formula of the candidate F, which turns out to be a bonafide Minkowski norm (see Exercise 4.1.4). On the other hand, in Exercise 1.2.9, the technique (trivially) produced the *unregularized* quartic metric, which fails to be strongly convex along the coordinate axes.

* An elegant **criterion for checking strong convexity**. This was derived in §4.1B, and is applicable only to candidate Minkowski norms on $\mathbb{R}^2$. The exact criterion is spelled out in (4.1.18). It is precisely the method used to ascertain that the candidate F corresponding to the convex limaçon is indeed a Minkowski norm. See Exercise 4.1.4. Its power lies in the fact that any parametrization of the indicatrix will suffice. Its drawback is that in some cases, such as the families of examples described above, there is no obvious way to parametrize the indicatrices in question.

Exercises

Exercise 14.1.1: Let $(\mathbb{R}^n, F)$ be a Minkowski space, where F is positively (but perhaps not absolutely) homogeneous of degree one.

(a) Show that there exists a constant $c > 1$, depending only on F and a choice of basis for $\mathbb{R}^n$, such that

$$\frac{1}{c}\,|\,y\,| \;\leqslant\; F(y) \;\leqslant\; c\,|\,y\,|$$

at every y. Hint: examine the proof of Lemma 6.2.1.

(b) The distance from y to z is $F(z-y)$, whereas that from z to y is $F(y-z)$. Review §6.4A. Explain why these two numbers are typically different.

(c) Check that every forward Cauchy sequence (see §6.2D) with respect to F is an ordinary Cauchy sequence with respect to $|\ |$. And, every convergent sequence in $|\ |$ is a convergent sequence in F. Show that all forward Cauchy sequences in $(\mathbb{R}^n, F)$ are convergent.

(d) Repeat part (c) for backward Cauchy sequences.

Thus, **Minkowski spaces are both forward and backward complete**, even though F may only be positively homogeneous of degree 1.

Exercise 14.1.2: By passing to polar coordinates ρ, ϕ, show that the normal form (14.1.1) can be re-expressed as

$$F \;=\; \rho\,(1 \;+\; B\,\cos\phi)\;.$$

Verify that the indicatrix is an ellipse with the following features:

* Its eccentricity is B.
* Its center is at the point $(p, q) = \left(-\frac{B}{1-B^2}, 0\right)$.
* The semimajor axis is horizontal and has length $\frac{1}{1-B^2}$.
* Its right focus is at the origin $(p, q) = (0, 0)$.
* Its left focus is at the point $(p, q) = \left(-\frac{2B}{1-B^2}, 0\right)$.

* The semiminor axis is vertical and has length $\frac{1}{\sqrt{1-B^2}}$.

Exercise 14.1.3: This exercise concerns the perturbed quartic metric that we defined in (14.1.2). Is there any value of $\epsilon < 0$ for which the stated F is smooth and strongly convex at every nonzero y?

Exercise 14.1.4: Consider the 2-parameter family of F defined by (14.1.3). Determine all the positive values of ϵ and B such that F is smooth and strongly convex on $\mathbb{R}^2 \setminus 0$.

Exercise 14.1.5: Consider the following family of candidate Minkowski norms,

$$F(y) \ := \ [\, p^4 \ + \ 6\,\lambda\, p^2\, q^2 \ + \ q^4 \,]^{1/4} \ ,$$

where λ is a constant. Show that F is smooth and strongly convex away from the origin if and only if $0 < \lambda < 1$.

14.2 The Riemannian Curvature of Each Minkowski Space

Let $(\mathbb{R}^n, F)$ be a Minkowski space. Since the manifold in question is actually the vector space $\mathbb{R}^n$, each choice of basis gives rise to a set of globally defined coordinates $(y^1, \ldots, y^n)$ on $\mathbb{R}^n$. Every change of basis is equivalent to a *linear* change of coordinates

$$y^i \ \longrightarrow \ c^i{}_j \ y^j \ ,$$

where $(c^i{}_j)$ is a *constant* invertible matrix. Also, at each point y of $\mathbb{R}^n$, the tangent space $T_y\mathbb{R}^n$ and the cotangent space $T_y^*\mathbb{R}^n$ have the bases $\{\frac{\partial}{\partial y^i}\}$ and $\{dy^i\}$, respectively.

We define a Riemannian metric on the punctured space $\mathbb{R}^n \setminus 0$, and compute its Levi-Civita (Christoffel) connection and curvature. In the treatment here, we define and calculate the said objects in the global coordinates y^i mentioned above. The metric and the curvature are then *required* to transform tensorially when we change from these privileged coordinates to arbitrary (possibly local) coordinates $\tilde{y}^p$, such as spherical coordinates. Likewise, the connection is required to transform like a connection under such coordinate changes.

Let us begin with the Riemannian metric. The Minkowski norm F gives rise to the fundamental tensor

$$g_{ij\,(y)} \ := \ \left(\frac{1}{2}\, F^2 \right)_{y^i y^j} \ = \ F\, F_{y^i y^j} \ + \ F_{y^i}\, F_{y^j} \ ,$$

which in turn defines a Riemannian metric

$$\boxed{\hat{g} \ := \ g_{ij}(y)\ dy^i \otimes dy^j} \ . \qquad (14.2.1)$$

Typically, the g_{ij} are not even continuous at $y = 0$. For this reason, the Riemannian metric $\hat{g}$ lives on the punctured space $\mathbb{R}^n \setminus 0$.

Let $\hat{\gamma}^i{}_{jk}$ be this metric's Christoffel symbols of the second kind. Using the definition (1.4.2) of the Cartan tensor, we see that

$$\boxed{\hat{\gamma}^i{}_{jk} := \frac{g^{is}}{2}\left(\frac{\partial g_{sj}}{\partial y^k} - \frac{\partial g_{jk}}{\partial y^s} + \frac{\partial g_{ks}}{\partial y^j}\right) = \frac{1}{F}\, A^i{}_{jk}} \,. \tag{14.2.2}$$

Denote the corresponding Levi-Civita (Christoffel) connection by $\hat{\nabla}$, then

$$\hat{\nabla}_{\frac{\partial}{\partial y^k}} \frac{\partial}{\partial y^j} = \hat{\gamma}^i{}_{jk}\, \frac{\partial}{\partial y^i} \,. \tag{14.2.3}$$

The curvature operator of $\hat{\nabla}$ is

$$\begin{aligned} \hat{R}(U,V)W &:= \left(\hat{\nabla}_U \hat{\nabla}_V - \hat{\nabla}_V \hat{\nabla}_U - \hat{\nabla}_{[U,V]}\right) W \\ &= W^j\, U^k\, V^l\, \hat{R}_j{}^i{}_{kl}\, \frac{\partial}{\partial y^i} \,, \end{aligned} \tag{14.2.4}$$

where U, V, W are vector fields on $\mathbb{R}^n \setminus 0$. That curvature tensor has components

$$\hat{R}_j{}^i{}_{kl} = \frac{\partial \hat{\gamma}^i{}_{jl}}{\partial y^k} - \frac{\partial \hat{\gamma}^i{}_{jk}}{\partial y^l} + \hat{\gamma}^i{}_{hk}\, \hat{\gamma}^h{}_{jl} - \hat{\gamma}^i{}_{hl}\, \hat{\gamma}^h{}_{jk} \,. \tag{14.2.5}$$

A simple expression for $\hat{R}_j{}^i{}_{kl}$ can be obtained as follows. Substitute (14.2.2) into (14.2.5). Now use the fact that

$$\begin{aligned} &\frac{\partial}{\partial y^s}\left(\frac{1}{F}\, A^p{}_{qr}\right) \\ &= \frac{-1}{F^2}\, \frac{\partial F}{\partial y^s}\, A^p{}_{qr} + \frac{1}{F}\, \frac{\partial A^p{}_{qr}}{\partial y^s} \\ &= \frac{1}{F^2}\, (A^p{}_{qr;s} - A^p{}_{qr}\, \ell_s) \,, \end{aligned}$$

followed by (2.5.12) and the Bianchi identity (3.4.14). After all that has been carried out, we get, as in Kikuchi [Kik],

$$\boxed{\hat{R}_j{}^i{}_{kl} = \frac{1}{F^2}\, (A_j{}^s{}_k\, A_s{}^i{}_l - A_j{}^s{}_l\, A_s{}^i{}_k)} \,. \tag{14.2.6}$$

This expression always reduces to zero whenever our Minkowski space is 2-dimensional! Exercise 14.2.3 uses an indirect way to deduce this somewhat surprising fact.

The expressions given for g_{ij}, $\hat{\gamma}^i{}_{jk}$, and $\hat{R}_j{}^i{}_{kl}$ retain their forms if we decide to use a different basis for $\mathbb{R}^n$. This is not difficult to verify. Now suppose, instead of a linear change of coordinates, we change from the global y^i to arbitrary (and typically local) coordinates $\tilde{y}^p$. The components of the Riemannian metric $\hat{g}$, the Levi-Civita (Christoffel) connection $\hat{\gamma}$, and the curvature $\hat{R}$ in these new coordinates are *defined* to be:

$$\begin{aligned}
g_{pq} &:= \frac{\partial y^i}{\partial \tilde{y}^p}\,\frac{\partial y^j}{\partial \tilde{y}^q}\, g_{ij}\ , \\
\hat{\gamma}^p{}_{qr} &:= \frac{\partial \tilde{y}^p}{\partial y^i}\,\frac{\partial y^j}{\partial \tilde{y}^q}\,\frac{\partial y^k}{\partial \tilde{y}^r}\,\hat{\gamma}^i{}_{jk} \;+\; \frac{\partial \tilde{y}^p}{\partial y^s}\,\frac{\partial^2 y^s}{\partial \tilde{y}^q \partial \tilde{y}^r}\ , \\
\hat{R}_q{}^p{}_{rs} &:= \frac{\partial y^j}{\partial \tilde{y}^q}\,\frac{\partial \tilde{y}^p}{\partial y^i}\,\frac{\partial y^k}{\partial \tilde{y}^r}\,\frac{\partial y^l}{\partial \tilde{y}^s}\,\hat{R}_j{}^i{}_{kl}\ .
\end{aligned}$$

We conclude with the sectional curvatures of $(\mathbb{R}^n \smallsetminus 0\,,\,\hat{g}\,)$. Take any two linearly independent tangent vectors U, V in $T_y(\mathbb{R}^n \smallsetminus 0)$. They define a 2-plane in that tangent space. The sectional curvature corresponding to this 2-plane is

$$\hat{K}(U,V) \;:=\; \frac{\hat{g}(\,\hat{R}(V,U)U\,,\,V\,)}{\hat{g}(U,U)\,\hat{g}(V,V) - [\hat{g}(U,V)]^2}\ . \tag{14.2.7}$$

The sectional curvature depends only on the 2-plane in question. See Exercise 14.2.4. *Intuitively,* it is the Gaussian curvature of the surface obtained by "exponentiating" the 2-plane.

Exercises

Exercise 14.2.1:

(a) What does it mean to say that under a *linear* change of coordinates $y^i \to z^i := c^i{}_j\, y^j$, an expression remains *unchanged in form*?

(b) Show that the expressions for g_{ij}, $\hat{\gamma}^i{}_{jk}$, and $\hat{R}_j{}^i{}_{kl}$ share this property.

Exercise 14.2.2:

(a) Derive formula (14.2.2) for $\hat{\gamma}^i{}_{jk}$.

(b) Derive formula (14.2.6) for $\hat{R}_j{}^i{}_{kl}$.

Exercise 14.2.3: In this exercise, we show that **for every Minkowski plane, the metric $\hat{g}$ is flat**. Fix any $y_o \in \mathbb{R}^2 \smallsetminus 0$, let us check that $\hat{R}$ is zero at y_o.

(a) In the tangent plane $T_{y_o}\mathbb{R}^2$, consider the right-handed $\hat{g}(y_o)$ orthonormal basis $\{\hat{e}_1, \hat{e}_2\}$ with

$$\hat{e}_2 \;:=\; \frac{y_o^1}{F(y_o)}\,\frac{\partial}{\partial y^1}_{|y_o} \;+\; \frac{y_o^2}{F(y_o)}\,\frac{\partial}{\partial y^2}_{|y_o}\ .$$

Write down $\hat{e}_1$ explicitly *without* first consulting the Berwald frame of §4.3.

(b) Show that $\hat{R}(U,V)W$ vanishes whenever U or V or W is proportional to $\hat{e}_2$.

(c) Explain why there exists a *linear* change of coordinates on $\mathbb{R}^2$, say from y^i to $z^i := c^i{}_j\, y_j$, such that at the point y_o we have $\frac{\partial}{\partial z^i}{}_{|y_o} = \hat{e}_i$.

(d) At the point y_o, demonstrate that the components of $\hat{R}$ with respect to the global coordinates z^i are all zero.

Exercise 14.2.4:

(a) Show that $\hat{K}(U, V) = \hat{K}(V, U)$.

(b) Check that if $\tilde{U} := \alpha\, U + \beta\, V$ and $\tilde{V} := \gamma\, U + \delta\, V$, where $\alpha\,\delta - \beta\,\gamma \neq 0$, then $\hat{K}(\tilde{U}, \tilde{V}) = \hat{K}(U, V)$.

14.3 The Riemannian Laplacian in Spherical Coordinates

We now get a technical ingredient ready for §14.4, when we prove Deicke's theorem. This concerns the Laplacian of $\hat{g}$, expressed in spherical coordinates. Recall from §14.2 that $\hat{g}$ is the Riemannian metric on the punctured Minkowski space $\mathbb{R}^n \setminus 0$. Its components are those of the fundamental tensor g_{ij}.

Consider the diffeomorphism

$$\Phi(r, u) \;:=\; r\, u$$

from $(0, \infty) \times \mathrm{S}$ onto $\mathbb{R}^n \setminus 0$. Here, the indicatrix S is regarded as a subset of $\mathbb{R}^n \setminus 0$. Let:

* y^i, $i = 1, \ldots, n$ be global coordinates on $\mathbb{R}^n$, from a choice of basis;
* θ^α, $\alpha = 1, \ldots, n-1$ be local coordinates that are intrinsic to S.

Some amount of definition-chasing will show that, at the image point ru in $\mathbb{R}^n \setminus 0$, we have

$$\Phi_* \frac{\partial}{\partial r} \;=\; u^i \frac{\partial}{\partial y^i}\,, \tag{14.3.1}$$

$$\Phi_* \frac{\partial}{\partial \theta^\alpha} \;=\; r\, \frac{\partial u}{\partial \theta^\alpha}\,. \tag{14.3.2}$$

Our interpretation of the right-hand sides is as follows:

* Take the "position vector" of the point u and slide it along the ray tu until the base reaches the point ru. The resulting tangent vector is $u^i \frac{\partial}{\partial y^i}$, which is an element of $T_{ru}\mathbb{R}^n$.

* The quantity $\frac{\partial u}{\partial \theta^\alpha}$ is a vector tangent to S. And is none other than the realization in $\mathbb{R}^n$ of the tangent vector $\frac{\partial}{\partial \theta^\alpha}$ intrinsic to S. We slide it along the ray tu until its base point is at ru. The resulting vector is tangent to $r\,\mathrm{S}$. Multiplying that by r gives the right-hand side of (14.3.2).

In view of the discussion that immediately precedes §6.1A, we know that

$$\text{(14.3.3)} \qquad \Phi_* \frac{\partial}{\partial r} \ \perp \ \Phi_* \frac{\partial}{\partial \theta^\alpha} \quad \text{with respect to } \hat{g} \,.$$

Let $\dot{g}$ denote the Riemannian metric induced by $\hat{g}$ on the indicatrix S, regarded as a submanifold of $\mathbb{R}^n \setminus 0$. Our understanding of (14.3.1) and (14.3.2), together with (14.3.3), is enough to effect the statement

$$\text{(14.3.4)} \qquad \Phi^* \hat{g} \ = \ dr \otimes dr \ + \ r^2 \, \dot{g} \,.$$

Equivalently, define the diffeomorphism $\Psi : \mathbb{R}^n \setminus 0 \to (0, \infty) \times \mathrm{S}$ by

$$\Psi(y) \ := \ \left(F(y) \,, \ \frac{y}{F(y)} \right) \,.$$

This is the inverse of Φ, and we have

$$\text{(14.3.5)} \qquad \hat{g} \ = \ \Psi^* (\, dr \otimes dr \ + \ r^2 \, \dot{g} \,) \,.$$

The message behind (14.3.4) or (14.3.5) is useful. It says that when expressed in terms of the spherical coordinates r, θ^α, the metric $\hat{g}$ takes on the form of a warped product (see §13.3), with warping factor $\varphi := r$.

Denote the Christoffel symbols (of the second kind) of $\hat{g}$ by $\hat{\gamma}^i{}_{jk}$, and those of $\dot{g}$ as $\dot{\gamma}^\alpha{}_{\beta\kappa}$. Now use (13.3.2) and keep in mind that

$$\hat{g}_{\alpha\beta} \ = \ r^2 \, \dot{g}_{\alpha\beta} \,.$$

The results are as follows:

$$\text{(14.3.6a–f)} \qquad \begin{aligned} \hat{\gamma}^1{}_{11} \ &= \ 0 \\ \hat{\gamma}^1{}_{\beta 1} \ &= \ 0 \\ \hat{\gamma}^1{}_{\beta\kappa} \ &= \ - \, r \, \dot{g}_{\beta\kappa} \\ \hat{\gamma}^\alpha{}_{11} \ &= \ 0 \\ \hat{\gamma}^\alpha{}_{\beta 1} \ &= \ r^{-1} \, \delta^\alpha{}_\beta \\ \hat{\gamma}^\alpha{}_{\beta\kappa} \ &= \ \dot{\gamma}^\alpha{}_{\beta\kappa} \,. \end{aligned}$$

Do not confuse the Greek κ with the Latin k.

As in (13.5.4), we define the Laplacian of $(\mathbb{R}^n \setminus 0 \,, \, \hat{g} \,)$ by

$$\Delta_{\hat{g}} \, f \ := \ g^{ij} \ \hat{\nabla}_i \hat{\nabla}_j \, f \,.$$

More explicitly,

$$\text{(14.3.7)} \qquad \Delta_{\hat{g}} \, f \ = \ g^{ij} \ (\, \partial_i \partial_j f \ - \ \hat{\gamma}^k{}_{ji} \; \partial_k f \,) \,.$$

Likewise, the Laplacian of $(\mathrm{S}, \dot{g})$ is given by

$$\text{(14.3.8)} \qquad \Delta_{\dot{g}} \, f \ = \ \dot{g}^{\alpha\beta} \ (\, \partial_\alpha \partial_\beta f \ - \ \dot{\gamma}^\kappa{}_{\beta\alpha} \; \partial_\kappa f \,) \,.$$

The definition of $\Delta_{\hat{g}} \, f$ is independent of the coordinate system we use. Let us then expand it out in terms of spherical coordinates, and use (14.3.6).

This straightforward calculation yields the formula

$$\boxed{\Delta_{\hat g} f \;=\; \frac{1}{r^2}\,[r^2\, f_{rr} \;+\; r\,(n-1)\, f_r] \;+\; \frac{1}{r^2}\,\Delta_{\dot g} f} \;. \tag{14.3.9}$$

Here, f_r and f_{rr} abbreviate partial differentiation with respect to r. In particular,

$$\Delta_{\hat g} f \;=\; \frac{1}{r^2}\,\Delta_{\dot g} f \quad \text{whenever } f \text{ is constant along rays.} \tag{14.3.10}$$

Exercises

Exercise 14.3.1: Derive (14.3.4).

Exercise 14.3.2:

(a) Supply all the details leading to (14.3.9).

(b) Is that formula consistent with what one finds in vector calculus texts?

Exercise 14.3.3:

(a) Deduce from (14.3.7) and (14.2.2) that

$$\boxed{\Delta_{\hat g} f \;=\; g^{ij}\,\partial_i\partial_j f \;-\; \frac{1}{F}\,A^k\,\partial_k f} \;.$$

Here, $A_k := g^{ij}\,A_{ijk}$ and $A^k := g^{ks}\,A_s$.

(b) The definition of $\Delta_{\hat g}$ is independent of coordinates. Using this fact, show that if $A_k = 0$, then the natural coordinate functions y^i are all $\hat g$ harmonic.

Exercise 14.3.4: This exercise relates formula (14.3.9) with the subject of spherical harmonics.

(a) Each $y \in \mathbb{R}^n \setminus 0$ can be expressed as ru, where $r := F(y)$ and $u \in \mathrm{S}$, the indicatrix. Explain why

$$\frac{\partial y^i}{\partial r} \;=\; \frac{y^i}{r} \;.$$

(b) Suppose f is a positively homogeneous function of degree m in the natural coordinates $y^1, \ldots, y^n$. Use the chain rule, part (a), and Euler theorem (Theorem 1.2.1) to show that

$$f_r \;=\; \frac{1}{r}\; m\, f \;,$$

$$f_{rr} \;=\; \frac{1}{r^2}\; m\,(m-1)\, f \;.$$

(c) Now, suppose in addition to being homogeneous of degree m in the coordinates y^i, our function f is also $\hat{g}$ harmonic. Restrict this f to the indicatrix S and call the resulting function $\mathcal{Y}$. Use (14.3.9) to help prove that $\mathcal{Y}$ is an eigenfunction of $\Delta_{\dot{g}}$. Specifically,

$$\begin{aligned} \Delta_{\dot{g}}\, \mathcal{Y} &= -\, m\,(m+n-2)\,\mathcal{Y}\,, \quad \text{or} \\ (-\,\Delta_{\dot{g}})\,\mathcal{Y} &= +\, m\,(m+n-2)\,\mathcal{Y}\,. \end{aligned}$$

These eigenfunctions $\mathcal{Y}$ are known as **spherical harmonics**.

14.4 Deicke's Theorem

We now give Brickell's proof [B1] of a theorem of Deicke's [D]. Recall that a Minkowski norm F is induced by an inner product on $\mathbb{R}^n$ if and only if the Cartan tensor A_{ijk} vanishes. Define

$$A_k \;:=\; g^{ij}\, A_{ijk}\;,$$

where g^{ij} is the inverse of the fundamental tensor of F. Then Deicke's theorem states that F comes from an inner product if and only if $A_k = 0$.

Theorem 14.4.1 (Deicke). *Let F be a Minkowski norm on $\mathbb{R}^n$, smooth and strongly convex at all $y \neq 0$. Let*

$$g_{ij} \;:=\; \left(\frac{1}{2}F^2\right)_{y^i y^j}\,, \qquad A_{ijk} \;:=\; \frac{F}{2}\,\left(\frac{1}{2}F^2\right)_{y^i y^j y^k}$$

be its fundamental tensor and Cartan tensor, respectively. Then the following statements are equivalent:

(1) *The components g_{ij} are constant.*

(2) $A_{ijk} = 0$.

(3) *The determinant* $\det(g_{ij})$ *is constant.*

(4) $A_k = 0$.

Proof.

* **Statements (1) and (2) are equivalent**, and is local. This can be seen directly from the definition of the Cartan tensor.

* **Statements (3) and (4) are equivalent**, and is local. Note that $\det(g_{ij})$ is constant if and only if $\sqrt{g}$ is constant. And, we have

$$(\sqrt{g}\,)_{y^k} \;=\; \frac{1}{2}\,\sqrt{g}\,g^{ij}\,(g_{ij}\,)_{y^k} \;=\; \frac{1}{F}\,\sqrt{g}\,A_k\,.$$

Since (2) $\Rightarrow$ (4) is clear, it suffices to demonstrate that (4) implies (1). This is where *we* use the y-global assumption on F, through the application of Hopf's theorem to a certain Laplacian on the indicatrix.

Some linear algebra preliminaries:

Denote by $g(y)$ the matrix (g_{ij}) evaluated at arbitrary y in $\mathbb{R}^n \setminus 0$. Denote by $g^{-1}(y_o)$ the inverse matrix (g^{ij}) evaluated at any **fixed** nonzero y_o. Consider the matrix product $g^{-1}(y_o)\, g(y)$. Our primary goal here is to show that all its eigenvalues are real and positive.

Note that

$$\det[\, g^{-1}(y_o)\; g(y) - \lambda\, I\,] \;=\; \det[\, g^{-1}(y_o)\,]\; \det[\, g(y) - \lambda\; g(y_o)\,]\;,$$

and that the inverse matrix $g^{-1}(y_o)$ is positive-definite because $g(y_o)$ is. Therefore the eigenvalues of our special matrix product are precisely the solutions of the equation

$$(*) \qquad \det[\, g(y) - \lambda\; g(y_o)\,] \;=\; 0\;.$$

Let λ be any solution of (*). Then there exists a (possibly complex) nonzero column vector v such that $[\, g(y) - \lambda\, g(y_o)\,]v = 0$. That is,

$$g(y)\; v \;=\; \lambda\; g(y_o)\; v\;.$$

Taking the complex conjugate of this statement and using the fact that $g(y)$, $g(y_o)$ are real, we have

$$g(y)\; \bar{v} \;=\; \bar{\lambda}\; g(y_o)\; \bar{v}\;.$$

So

$$\begin{aligned} &\lambda\; \bar{v}^T\, g(y_o)\; v \\ &= \bar{v}^T\, [\, g(y)\; v\,] \\ &= [\, g(y)\; v\,]^T\, \bar{v} \\ &= v^T\, g(y)\; \bar{v} \\ &= v^T\, \bar{\lambda}\; g(y_o)\; \bar{v} \\ &= \bar{\lambda}\; [\, v^T\, g(y_o)\; \bar{v}\,]^T \\ &= \bar{\lambda}\; \bar{v}^T\, g(y_o)\; v\;. \end{aligned}$$

In this calculation, the fact that $g(y)$, $g(y_o)$ are symmetric has been used without mention. Also, the symmetry of $g(y_o)$ implies that

$$\bar{v}^T\, g(y_o)\; v \;=\; \mathrm{Re}(v)^T\, g(y_o)\; \mathrm{Re}(v) \;+\; \mathrm{Im}(v)^T\, g(y_o)\; \mathrm{Im}(v)\;.$$

Since v is nonzero and $g(y_o)$ is positive-definite, we must have

$$\lambda \;=\; \bar{\lambda}\;.$$

Hence the solutions of (*) are indeed real. Furthermore, observe that the matrix $g(y) - \lambda\, g(y_o)$ would be positive-definite, and consequently would have to have a positive determinant, if λ were $\leqslant 0$. Therefore, in order for it to have a zero determinant, the real λ must necessarily be positive.

We have just ascertained that the eigenvalues of

$$\mathcal{P} \;:=\; g^{-1}(y_o)\, g(y)$$

are real and positive. Its trace and determinant are, in view of its Jordan canonical form, respectively, the sum and the product of these positive eigenvalues. Thus the arithmetic-geometric mean inequality tells us that

$$(**) \qquad \frac{1}{n}\,\mathrm{tr}\mathcal{P} \;\geqslant\; (\det\mathcal{P})^{\frac{1}{n}} \ .$$

Proving that (4) implies (1):

Let us now suppose that

$$A_k \;=\; 0 \ .$$

We want to demonstrate that the quantities g_{ij} are constant functions.

- As a function of the variable y, the quantity **tr$\mathcal{P}$ takes on its absolute minimum value n at the point $y = y_o$.** Indeed, since $A_k = 0$, the determinant of (g_{ij}) (and hence of its inverse) is independent of the point of evaluation. This is so because (4) $\Leftrightarrow$ (3). Consequently $\det\mathcal{P} = 1$ and, in view of (**), n is a lower bound of tr$\mathcal{P}$. Moreover, the value of tr$\mathcal{P}$ at $y = y_o$ is precisely n.
- Let $\Delta_{\hat{g}}$ denote the covariant Laplacian $g^{ij}\,\hat{\nabla}_i\hat{\nabla}_j$ of the metric

  $$\hat{g} \;:=\; g_{ij}(y)\;dy^i\otimes dy^j$$

  on $\mathbb{R}^n\setminus 0$. Then **the matrix $(\,\Delta_{\hat{g}}\,g_{pq}\,)$ is positive-semidefinite** at every point of $\mathbb{R}^n\setminus 0$. As we show below, this follows from the fact that tr$\mathcal{P}$ has its absolute minimum at $y = y_o$. For then its Hessian matrix at $y = y_o$ must be positive-semidefinite. Indeed,

  $$\begin{aligned} &[\,\mathrm{tr}\mathcal{P}\,]_{y^py^q} \\ =\;& [\,g^{ij}(y_o)\;g_{ji}(y)\,]_{y^py^q} \\ =\;& g^{ij}(y_o)\left[\frac{1}{2}\,F^2\right]_{y^iy^jy^py^q} \\ =\;& g^{ij}(y_o)\;\partial_{y^i}\partial_{y^j}\;g_{pq}(y) \ . \end{aligned}$$

  According to Exercise 14.3.3, the difference between the operator $g^{ij}(y_o)\,\partial_{y^i}\partial_{y^j}$ and the covariant Laplacian $\Delta_{\hat{g}}$ is a term proportional to $A_k(y_o)$, which is by hypothesis zero here. Therefore

  $$[\,\mathrm{tr}\mathcal{P}\,]_{y^py^q}(y_o) \;=\; [\,\Delta_{\hat{g}}\,g_{pq}\,](y_o) \ .$$

  Our criterion on the Hessian of tr$\mathcal{P}$ is tantamount to saying that the matrix $(\,\Delta_{\hat{g}}\,g_{pq}\,)$ is positive-semidefinite at every fixed $y_o \neq 0$.

Since $(\,\Delta_{\hat{g}}\,g_{pq}\,)$ is positive-semidefinite, each of its diagonal entries $\Delta_{\hat{g}}\,g_{pp}$ must be $\geqslant 0$. Now each g_{pp} is constant along rays through the origin of $\mathbb{R}^n$, so it can be regarded as living on the indicatrix S. Formula (14.3.10) then implies that $\Delta_{\dot{g}}\,g_{pp}\geqslant 0$, where $\Delta_{\dot{g}}$ is the Laplacian on S. By Hopf's theorem (Theorem 13.5.2), g_{pp} must be constant on S, and hence on $\mathbb{R}^n\setminus 0$.

Using (14.3.7), we then have $\Delta_{\hat{g}}\, g_{pp} = 0$. This says that the diagonal entries of our matrix $(\Delta_{\hat{g}}\, g_{pq})$ are all zero. In particular, its trace vanishes.

The matrix $(\Delta_{\hat{g}}\, g_{pq})$ is real and symmetric, hence it can be diagonalized at every fixed y_o. Since it is positive-semidefinite, all the resulting diagonal entries are nonnegative. But we have just seen that it has vanishing trace, which is a basis independent criterion. So all the diagonal entries are in fact zero. This means that the matrix $(\Delta_{\hat{g}}\, g_{pq})$ must have been identically zero in the first place, at every y_o. In other words, each g_{pq} is a $\Delta_{\hat{g}}$ harmonic function. By scale-invariance, each g_{pq} is $\Delta_{\dot{g}}$ harmonic as a function on the indicatrix S. The constancy of g_{pq} on S again follows from Hopf's Theorem. Scale-invariance then gives its constancy on $\mathbb{R}^n \setminus 0$. □

Remark: Instead of using the covariant Laplacian and Hopf's theorem, Brickell [B1] originally worked with the elliptic operator $g^{ij}\, \partial_{y^i} \partial_{y^j}$ and the **strong maximum principle** ([GT], [Smo]), which is also due to Hopf.

Exercises

Exercise 14.4.1: The **arithmetic-geometric mean inequality** states that given any n positive numbers $c_1, \ldots, c_n$, we have

$$\boxed{\frac{1}{n}\,(c_1 + \cdots + c_n) \;\geqslant\; (c_1 \cdots c_n)^{\frac{1}{n}}}\ .$$

Furthermore, equality holds if and only if $c_1 = \cdots = c_n$.

(a) The $n = 2$ case says that

$$c_1 + c_2 \;\geqslant\; 2\,\sqrt{c_1}\,\sqrt{c_2}\ ,$$

with equality if and only if $c_1 = c_2$. Demonstrate this algebraically.

(b) Now consider the function

$$f(x) \;:=\; \frac{c_1 + c_2 + x}{(c_1\, c_2\, x)^{\frac{1}{3}}}$$

on the half interval $(0, \infty)$, with positive constants c_1 and c_2. Using calculus and part (a), establish the $n = 3$ case of the inequality.

(c) Use part (a), calculus, and mathematical induction to prove the general case.

Exercise 14.4.2: The **strong maximum principle** for our operator $g^{ij}\, \partial_i \partial_j$ says the following.

> Suppose $g^{ij}\, \partial_i \partial_j\, f \geqslant 0$ (resp., $\leqslant 0$) on a domain (not necessarily bounded) Ω of $\mathbb{R}^n$. If f takes on its absolute maximum (resp., absolute minimum) at an interior point of Ω, then it must in fact be constant.

For an arbitrary elliptic operator, the boundedness of the domain is needed. We do not need that here because each g^{ij} is constant along rays emanating from the origin, and hence our operator satisfies the criterion of **uniform ellipticity**. See [GT] for details.

(a) Instead of Theorem 13.5.2, use the strong maximum principle to prove Deicke's theorem. Does we still need F to be y-global ?

(b) Can you give a proof of the strong maximum principle for our special operator $g^{ij}\,\partial_i\partial_j$?

14.5 The Extrinsic Curvature of the Level Spheres of F

Let us now turn our attention to a family of submanifolds of $\mathbb{R}^n \setminus 0$. These are the hypersurfaces

$$\mathrm{S}(r) \;:=\; \{\, y \in \mathbb{R}^n : \; F(y) = r \,\}$$

that we first introduced in §1.4. This family of level spheres is indexed by the radius $r > 0$.

The metric $\hat{g}$ of the ambient space restricts (or pulls back) to a Riemannian metric $\dot{g}$ on $\mathrm{S}(r)$, whose Levi-Civita (Christoffel) connection (characterized by torsion-freeness and $\dot{g}$-compatibility) we denote by $\dot{\nabla}$. The resulting curvature tensor and sectional curvatures are, respectively, $\dot{R}$, $\dot{K}$. They are defined just like $\hat{R}$ and $\hat{K}$.

For our purposes, it is advantageous to give an extrinsic description of $\dot{\nabla}$. This would eventually allow us to see, in §14.6, how $\dot{R}$, $\dot{K}$ are related to $\hat{R}$, $\hat{K}$. First recall from §1.4 that at each point $y \in \mathrm{S}(r)$, the radial vector

$$\nu \;:=\; \frac{y^i}{F(x,y)}\,\frac{\partial}{\partial y^i} \;=\; \frac{y^i}{r}\,\frac{\partial}{\partial y^i} \tag{14.5.1}$$

is the outward-pointing unit normal with respect to the metric $\hat{g}$. It is also the $\hat{g}$-gradient of F. In other words,

$$\nu_i \;=\; F_{y^i} \; . \tag{14.5.2}$$

Since $\mathrm{S}(r)$ has codimension 1, this ν spans its normal bundle. Thus we have the following $\hat{g}$-orthogonal direct sum decomposition:

$$T_y(\mathbb{R}^n \setminus 0) \;=\; T_y\mathrm{S}(r) \,\oplus\, \mathbb{R}\,\nu \; . \tag{14.5.3}$$

Given any two locally defined vector fields U, V of $\mathrm{S}(r)$, we can extend them to vector fields (also locally defined) on $\mathbb{R}^n \setminus 0$. In order to keep the notation simple, we continue to denote the extended objects by U, V. In view of (14.5.3), we can decompose the covariant derivative $\hat{\nabla}_U V$ into two pieces:

$$\hat{\nabla}_U V \;=\; (\hat{\nabla}_U V)_{\parallel} \,+\, (\hat{\nabla}_U V)^{\perp} \; .$$

The two terms on the right-hand side are, respectively, tangential and $\hat{g}$-perpendicular to $S(r)$.

- By going through some definition-chasing, one can show that the tangential piece $(\hat{\nabla}_U V)_{\|}$ defines a connection on $S(r)$ which is torsion-free and $\dot{g}$-compatible. Therefore, by the uniqueness of the Levi-Civita (Christoffel) connection, it must actually be $\dot{\nabla}$. Put another way, we have an extrinsic formula for $\dot{\nabla}$:

$$\dot{\nabla}_U V \;=\; (\hat{\nabla}_U V)_{\|} \,. \tag{14.5.4}$$

- As for the piece that is $\hat{g}$-orthogonal to $S^{n-1}(r)$, it is traditionally called (the value of) the **second fundamental form** Π of $S(r)$. In general, the second fundamental form of a Riemannian submanifold S takes values in the normal bundle of S. However, the normal bundle in our case is 1-dimensional, and is being spanned by ν. Thus we can write

$$(\hat{\nabla}_U V)^{\perp} \;=:\; \Pi(U,V) \;=\; \Pi^{\nu}(U,V)\,\nu \,. \tag{14.5.5}$$

  This Π^{ν} is a symmetric bilinear form on $S(r)$.

Let us calculate an explicit formula for the $\Pi^{\nu}(U,V)$ of $S(r)$. To begin, $\Pi^{\nu}(U,V) = \hat{g}(\hat{\nabla}_U V,\, \nu)$ which, by the $\hat{g}$-compatibility of $\hat{\nabla}$ and the fact that ν is $\hat{g}$-orthogonal to V, becomes $-\hat{g}(V,\, \hat{\nabla}_U\, \nu)$. For a general U, we have

$$\begin{aligned}\hat{\nabla}_U\, \nu \;&=\; d\left(\frac{y^j}{F}\right)(U)\, \frac{\partial}{\partial y^j} \;+\; \frac{y^j}{F}\, U^k\, \hat{\gamma}^i{}_{jk}\, \frac{\partial}{\partial y^i} \\ &=\; \left(\frac{-y^i}{F^2}\, \nu_k\, U^k \;+\; \frac{1}{F}\, U^i \;+\; \frac{y^j}{F}\, U^k\, \frac{A^i{}_{jk}}{F} \right) \frac{\partial}{\partial y^i} \,.\end{aligned}$$

Note that the last of the three terms (on the right-hand side) vanishes because $y^j\, A^i{}_{jk}(x,y) \;=\; 0$. Now, let us restrict our U to those that are tangential to $S(r)$, so that $\nu_k\, U^k \;=\; 0$. The above computation then says that

$$\boxed{\hat{\nabla}_U\, \nu \;=\; \frac{1}{F}\, U \;=\; \frac{U}{r}} \quad \text{for } U \text{ tangent to } S(r) \,. \tag{14.5.6}$$

Put another way,

$S(r)$ is umbilic with principal curvatures all equal to $\frac{1}{r}$.

Therefore

$$\boxed{\Pi^{\nu}(U,V) \;=\; \frac{-1}{F}\, \hat{g}(U,V) \;=\; \frac{-1}{r}\, \hat{g}(U,V)} \quad \text{on } S(r) \,. \tag{14.5.7}$$

We digress to mention an interesting observation by Xinyue Chen (and possibly others as well). He pointed out to us that in Minkowski spaces:

* The level spheres $S(r)$ do *not* look "round" intrinsically. Their volumes can be computed using the induced volume form (1.4.8). See §4.1 for some cited references about the $n = 2$ (Minkowski plane), $r = 1$ (indicatrix) case. One finds that, unlike the level spheres of Euclidean norms, these volumes are *not* equal to r^{n-1} multiplying universal constants which depend only on the dimension n.
* However, extrinsically these $S(r)$ do look "round." Indeed, as we have established above, they are umbilic with constant principal curvature, and hence constant mean curvature, equal to $\frac{1}{r}$.

Exercises

Exercise 14.5.1: Prove that the tangential piece $(\hat{\nabla}_U V)_{\|}$ defines a connection on $S(r)$ that is torsion-free and $\dot{g}$-compatible.

Exercise 14.5.2: Show, *without* reference to (14.5.7), that the Π^{ν} defined in (14.5.5) is a symmetric bilinear form on $S(r)$.

Exercise 14.5.3: Fill in the details leading to (14.5.6).

Exercise 14.5.4: With the help of the volume form (1.4.8) and a rescaling, prove that in any n-dimensional Minkowski space, we have

$$\mathrm{vol}[\,S(r)\,] \;=\; r^{n-1}\,\mathrm{vol}[\,S(1)\,]\ .$$

Note that $S(1)$ is the indicatrix S.

Exercise 14.5.5:

(a) Consider any two norms F and $\tilde{F}$ that arise from *inner products* on $\mathbb{R}^n$. Prove that the volumes of their level spheres $S(r)$, $\tilde{S}(r)$ are the same for each value of r. Hint: what does an inner product look like in an orthonormal basis?

(b) Suppose $F(y) = \sqrt{(\,y^1\,)^2 + \cdots + (\,y^n\,)^2}$. Then the level sphere $S(1)$ is the standard unit sphere $\mathbb{S}^{n-1}$ in Euclidean n-space. Show that

$$\mathrm{vol}(\mathbb{S}^{n-1}) \;=\; \begin{cases} \dfrac{2\,\pi^{n/2}}{(\frac{n}{2}-1)!} & \text{for } n \text{ even}\ , \\[2ex] \dfrac{2^{(n+1)/2}\,\pi^{(n-1)/2}}{1\cdot 3\cdots(n-2)} & \text{for } n \text{ odd}\ . \end{cases}$$

14.6 The Gauss Equations

For each fixed r, $(\,S(r)\,,\,\dot{\nabla}\,)$ is a Riemannian submanifold of the punctured Riemannian manifold $(\,\mathbb{R}^n \smallsetminus 0\,,\,\hat{g}\,)$. Let U, V be locally defined vector fields of $S(r)$. Denote their extensions to $\mathbb{R}^n \smallsetminus 0$ by the same symbols. Let $\hat{\nabla}$ be the Levi-Civita (Christoffel) connection of $\hat{g}$, and $\dot{\nabla}$ the one of $\dot{g}$. One of the main points of §14.5 is the $\hat{g}$-orthogonal decomposition

$$\hat{\nabla}_U V = \dot{\nabla}_U V + \Pi(U,V) , \tag{14.6.1a}$$

where Π is the second fundamental form of $\mathrm{S}(r)$. Equivalently,

$$\dot{\nabla}_U V = \hat{\nabla}_U V - \Pi(U,V) , \tag{14.6.1b}$$

which expresses the Levi-Civita (Christoffel) connection on $\mathrm{S}(r)$ in terms of that for the ambient space, and the second fundamental form.

Repeated use of (14.6.1a) in

$$\hat{\nabla}_U \hat{\nabla}_V W - \hat{\nabla}_V \hat{\nabla}_U W - \hat{\nabla}_{[U,V]} W$$

gives the **Gauss curvature equation**

(14.6.2a)
$$[\hat{R}(U,V)W]_{\|} = \dot{R}(U,V)W + [\hat{\nabla}_U \Pi(V,W)]_{\|} - [\hat{\nabla}_V \Pi(U,W)]_{\|} .$$

Here, U, V, W are tangent to $\mathrm{S}(r)$. A rearrangement gives

(14.6.2b)
$$\dot{R}(U,V)W = [\hat{R}(U,V)W]_{\|} + [\hat{\nabla}_V \Pi(U,W)]_{\|} - [\hat{\nabla}_U \Pi(V,W)]_{\|} .$$

Note that (14.6.2) carries information only when the dimension n of the ambient manifold is at least 3. If $n = 2$, the submanifold $\mathrm{S}(r)$ is 1-dimensional. In that case, (14.6.2) reduces to $0 = 0$.

For Z tangent to $\mathrm{S}(r)$, note that

$$\dot{g}\Big([\cdots]_{\|} , Z \Big) = \dot{g}\Big([\cdots] , Z \Big) .$$

This fact, together with (14.6.2b) and

$$\hat{\nabla} \hat{g} = 0 ,$$

will lead to

$$\begin{aligned} \dot{g}(\dot{R}(U,V)W, Z) = \;& \hat{g}(\hat{R}(U,V)W, Z) \\ & + \hat{g}(\Pi(V,W), \Pi(U,Z)) - \hat{g}(\Pi(U,W), \Pi(V,Z)) . \end{aligned} \tag{14.6.3}$$

For $n \geqslant 3$, (14.6.3) immediately yields

$$\begin{aligned} \dot{K}(U,V) = \;& \hat{K}(U,V) \\ & + \frac{\hat{g}(\Pi(U,U), \Pi(V,V)) - \hat{g}(\Pi(U,V), \Pi(U,V))}{\dot{g}(U,U)\, \dot{g}(V,V) - [\dot{g}(U,V)]^2} . \end{aligned} \tag{14.6.4}$$

Let us also refer to these as the Gauss equations.

Formulas (14.6.1)–(14.6.4) are perfectly general. They could have been derived for any submanifold of an ambient Riemannian manifold. It is only now that we specialize them to the case at hand.

* The ambient curvature $\hat{R}(U,V)W$ in our case study is, according to (14.2.6), a very special vector:

$$W^j\, U^k\, V^l\, \frac{1}{F^2}\, \left(A_j{}^s{}_k\, A_s{}^i{}_l \;-\; A_j{}^s{}_l\, A_s{}^i{}_k \right)\, \frac{\partial}{\partial y^i}\ .$$

A moment's reflection shows that it is always $\hat{g}$-orthogonal to the outward unit normal ν of $\mathrm{S}(r)$. Hence

$$\begin{aligned}[\hat{R}(U,V)W]_{\|} &= \hat{R}(U,V)W \\ &= \frac{1}{r^2}\, [A(V,\, A(W,U)) \;-\; A(U,\, A(W,V))]\ .\end{aligned}$$

* Next, (14.5.6) implies that

$$[\,\hat{\nabla}_U\, \Pi(V,W)\,]_{\|} \;=\; \frac{-1}{r^2}\; \dot{g}(V,W)\; U\ .$$

Therefore (14.6.2) becomes

$$\text{(14.6.5)} \quad \dot{R}(U,V)W \;=\; \hat{R}(U,V)W \;+\; \frac{1}{r^2}\, [\dot{g}(V,W)\,U \;-\; \dot{g}(U,W)\,V]\ .$$

Also, it follows from (14.6.4) and (14.5.6) that

$$\text{(14.6.6)} \qquad \boxed{\dot{K}(U,V) \;=\; \hat{K}(U,V) \;+\; \frac{1}{r^2}}\ .$$

Just like (14.6.4), this makes sense only when $n \geqslant 3$. Compare the treatment here with that by A. Kawaguchi [Kawa].

We are ready to deduce the following:

Proposition 14.6.1. *Let* $(\,\mathbb{R}^n,\ F\,)$ *be any Minkowski space of dimension* $n \geqslant 3$. *Here,* F *is typically only positively homogeneous of degree one. Then the following statements are equivalent:*

(a) *The metric* $\hat{g} := g_{ij}(y)\, dy^i \otimes dy^j$ *on* $\mathbb{R}^n \setminus 0$ *is flat.*

(b) *Every level sphere* $(\,S(r),\ \dot{g}\,)$ *has constant sectional curvature* $\frac{1}{r^2}$.

(c) *Some level sphere* $(\,S(r_o),\ \dot{g}\,)$ *has constant sectional curvature* $\frac{1}{r_o^2}$.

Remarks:

- For $n \geqslant 3$, whenever the level sphere $(\,\mathrm{S}(r),\ \dot{g}\,)$ has constant sectional curvature $\frac{1}{r^2}$, it is necessarily isometric to the standard sphere of radius r in Euclidean space $\mathbb{R}^n$. This follows from Hopf's classification of Riemannian space forms, discussed in §13.4.
- In that case, one might wonder if the Minkowski norm F arises from an inner product on $\mathbb{R}^n$, for then the ambient Minkowski space is actually Euclidean. For those F that are absolutely homogeneous

of degree one, the answer is surprisingly yes! This is the theorem of Brickell's that we prove in §14.9.

Proof. The Gauss equation (14.6.6) holds only when $n \geqslant 3$. It tells us that (a) $\Rightarrow$ (b). Also, (b) $\Rightarrow$ (c) is logically immediate.

It remains to show that (c) $\Rightarrow$ (a).

* Given (c), we conclude from (14.6.6) that all sectional curvatures $\hat{K}$ must vanish on the level sphere $\mathrm{S}(r_o)$. It is not difficult to deduce from (14.2.6) and (14.2.7) that $\hat{K}$ is homogeneous of degree -2 in y. Therefore it vanishes everywhere in $\mathbb{R}^n \setminus 0$.
* Now specialize Exercise 3.9.6 to the Riemannian setting (or see §13.2B). It says that every value of the full curvature tensor $\hat{R}$ is expressible as a sum of sectional curvatures $\hat{K}$ (each multiplied by the area of the corresponding "flag").

Combining these two observations gives (a). □

Exercises

Exercise 14.6.1: In (14.6.1), we have

$$\dot{\nabla}_U V \;=\; (\hat{\nabla}_U V)_{\parallel} \quad \text{and} \quad \Pi(U,V) \;=\; (\hat{\nabla}_U V)^{\perp} \,.$$

The U and V on the left-hand sides are local vector fields tangent to the level spheres $\mathrm{S}(r)$. Those on the right-hand sides are local extensions of the said U, V to $\mathbb{R}^n \setminus 0$.

(a) Show that the value of $\hat{\nabla}_U V$ **along** $\mathrm{S}(r)$ does not depend on the extensions we just described.

(b) Explain why $\dot{\nabla}_U V$ and $\Pi(U, V)$ are independent of the way by which one extends U, V from $\mathrm{S}(r)$ to $\mathbb{R}^n \setminus 0$.

Exercise 14.6.2:

(a) Derive the Gauss equation (14.6.2), as well as (14.6.3).

(b) Explain why, when $n = 2$, (14.6.2) reduces to $0 = 0$.

(c) Why can't one derive (14.6.4) when $n = 2$?

Exercise 14.6.3: Show that on $\mathrm{S}(r)$, we have

$$[\,\hat{\nabla}_U \,\Pi(V,W)\,]_{\parallel} \;=\; -\,\frac{1}{r^2}\,\dot{g}(V,W)\;U \,.$$

Exercise 14.6.4: Let $(S, \dot{g})$ be a submanifold of an ambient Riemannian manifold $(\mathcal{M}, \hat{g})$.

(a) Let ν be any smooth local section of the normal bundle NS. Let U be tangent to S. Define

$$\mathcal{D}_U \nu := (\hat{\nabla}_U \nu)^\perp .$$

Verify that $\mathcal{D}$ satisfies all the axioms for a connection on a vector bundle (see §2.4).

(b) The normal bundle NS has an induced metric h obtained by restricting $\hat{g}$. Is the normal connection $\mathcal{D}$ torsion-free in any appropriate sense? Is it h-compatible?

(c) For U, V, W all tangent to S, the Gauss equation addresses the part of $\hat{R}(U,V)W$ tangential to S. Show that the part which is $\hat{g}$-orthogonal to S is

$$[\hat{R}(U,V)W]^\perp = (\mathcal{D}_U \mathrm{II})(V,W) - (\mathcal{D}_V \mathrm{II})(U,W) ,$$

where the quantity $(\mathcal{D}_U \mathrm{II})(V,W)$ abbreviates

$$\mathcal{D}_U[\mathrm{II}(V,W)] - \mathrm{II}(\dot{\nabla}_U V, W) - \mathrm{II}(V, \dot{\nabla}_U W) .$$

This is the **Codazzi–Mainardi equation.**

In view of the Gauss and Codazzi–Mainardi equations, one expects that the most interesting Riemannian submanifolds are to be found inside geometrically simple ambient spaces.

14.7 The Blaschke–Santaló Inequality

Let K be a subset of $\mathbb{R}^n$.

* K is said to be **convex** if, given any y_1, y_2 in K, the line segment $(1-t)\,y_1 + t\,y_2$, $0 \leqslant t \leqslant 1$ joining them also lies in K.
* K is said to be **centrally symmetric** if, for each $y \in K$, $-y$ also belongs to K.
* Let $\bullet$ denote the usual dot product on $\mathbb{R}^n$. The **polar body** K^o of K is defined as follows:

$$K^o := \{ z \in \mathbb{R}^n : z \bullet y \leqslant 1 \text{ for all } y \in K \} .$$

As an example, the closed unit ball

$$\overline{\mathbb{B}} := \{ y \in \mathbb{R}^n : y \bullet y \leqslant 1 \} \tag{14.7.1}$$

is convex and centrally symmetric. It is also equal to its polar body.

Let $y^1, \ldots, y^n$ denote Cartesian coordinates of $\mathbb{R}^n$. Namely, these are globally defined coordinates induced by an orthonormal basis with respect to the dot product. Given any closed subset K, we define its **Euclidean volume** by

$$\mathrm{vol}(K) := \int_K dy^1 \cdots dy^n . \tag{14.7.2}$$

Here, $dy^1 \cdots dy^n$ is the usual Lebesgue measure on $\mathbb{R}^n$. See Royden [Roy] for a review of measure-theoretic issues.

Let K be a **closed, convex, centrally symmetric** subset of Euclidean space $\mathbb{R}^n$. Then

$$\boxed{\operatorname{vol}(K)\operatorname{vol}(K^o) \leqslant \operatorname{vol}(\overline{\mathbb{B}})\operatorname{vol}(\overline{\mathbb{B}}^o) = [\operatorname{vol}(\overline{\mathbb{B}})]^2} .$$

Equality holds if and only if K is an ellipsoid.

This is known as the **Blaschke–Santaló inequality**. It was proved by Blaschke [Bla] in dimensions 2 and 3. Santaló [San] established the inequality for higher dimensions. We defer to Schneider [Sch2] for a detailed methodical exposition, and to Meyer–Pajor [MP] for a short proof.

For application to Finsler geometry, take K to be the indicatrix S together with the interior region of which it is the boundary. Namely,

$$\text{(14.7.3)} \qquad \boxed{\overline{\mathrm{B}} := \{ y \in \mathbb{R}^n : F(y) \leqslant 1 \}} ,$$

where F is the Minkowski norm under study. Although the polar body $\overline{\mathrm{B}}^o$ still makes sense as defined above, it is nevertheless desirable to replace it by an object that is more akin to F than to the dot product $\bullet$.

To this end, let us define, as in functional analysis, the norm F^* on the dual space $\mathbb{R}^{n*}$. As we justify in the next section, it is appropriate to denote generic points in the dual space by p. Then F^* is defined as follows:

$$\text{(14.7.4)} \qquad F^*(p) := \sup_{v \in \mathrm{S}} p(v) = \sup_{0 \neq y \in \mathbb{R}^n} p\left[\frac{y}{F(y)} \right] .$$

Take any *nonzero* element $p \in \mathbb{R}^{n*}$. It is a nonzero linear functional on $\mathbb{R}^n$, so its level sets are parallel hyperplanes in $\mathbb{R}^n$. Since S is the boundary of a (strictly) convex set, the above supremum is actually achieved at a unique v on the indicatrix S. That is,

$$\text{(14.7.5)} \qquad F^*(p) = p(v) \quad \text{for some unique } v \in \mathrm{S} .$$

More geometrically, the hyperplane on which p has the constant value $F^*(p)$ is tangent to S at v. And this v depends implicitly on the nonzero linear functional p.

The F^* we have defined turns out to have all three properties (see §1.1) of Minkowski norms, but on $\mathbb{R}^{n*}$ rather than on $\mathbb{R}^n$. Of these properties, positive homogeneity is the most apparent; see Exercise 14.7.3. The remaining two, namely, smoothness away from $p = 0$ and strong convexity, are discussed at the end of §14.8. Anyway, using F^*, we introduce the following subset of $\mathbb{R}^{n*}$:

$$\text{(14.7.6)} \qquad \boxed{\overline{\mathrm{B}}^* := \{ p \in \mathbb{R}^{n*} : F^*(p) \leqslant 1 \}} .$$

Then we have the bijection

$$\overline{\mathrm{B}}^o \leftrightarrow \overline{\mathrm{B}}^* .$$

Indeed, definition-chasing shows that the map

$$z \mapsto z \bullet (\,\cdot\,)$$

is a diffeomorphism from $\overline{\mathrm{B}}^{o}$ onto $\overline{\mathrm{B}}^{*}$. See Exercise 14.7.4.

Let us digress to describe what is meant by $\mathrm{vol}(\overline{\mathrm{B}}^{*})$. Suppose $y^1, \ldots, y^n$ are Cartesian coordinates induced by some basis $\{e_i\}$ for $\mathbb{R}^n$. Let $p_1, \ldots, p_n$ be the coordinates induced by the corresponding dual basis $\{e^i\}$ on $\mathbb{R}^{n*}$. Then

$$\mathrm{vol}(\overline{\mathrm{B}}^{*}) := \int_{\overline{\mathrm{B}}^{*}} dp_1 \cdots dp_n \,. \tag{14.7.7}$$

Now, the above diffeomorphism $z \mapsto z \bullet (\,\cdot\,)$ from $\overline{\mathrm{B}}^{o}$ onto $\overline{\mathrm{B}}^{*}$ has Jacobian determinant equal to 1. Hence by the change-of-variables theorem, we have

$$\mathrm{vol}(\overline{\mathrm{B}}^{*}) = \mathrm{vol}(\overline{\mathrm{B}}^{o}) \,.$$

To conclude this section, we apply the **Blaschke–Santaló inequality** to $K := \overline{\mathrm{B}}$. Let F be an **absolutely homogeneous** Minkowski norm, so that $\overline{\mathrm{B}}$ is closed, convex, and most important, centrally symmetric. Then

$$\boxed{\mathrm{vol}(\overline{\mathrm{B}})\, \mathrm{vol}(\overline{\mathrm{B}}^{*}) \leqslant [\,\mathrm{vol}(\overline{\mathbb{B}})\,]^2} \,, \tag{14.7.8}$$

where equality holds if and only if $\overline{\mathrm{B}}$ is an ellipsoid. Equivalently, equality holds if and only if F is the norm induced by an inner product on $\mathbb{R}^n$.

Exercises

Exercise 14.7.1: Let $\overline{\mathbb{B}}$ denote the closed unit ball in $\mathbb{R}^n$. Show that:

(a) It is convex and centrally symmetric.

(b) The polar body of $\overline{\mathbb{B}}$ is itself.

Exercise 14.7.2: In the section, we have explained why given any nonzero $p \in \mathbb{R}^{n*}$, there exists a unique $v \in \mathrm{S}$ such that $p(v) = F^*(p)$.

(a) Along the ray λv, $\lambda \geqslant 0$ generated by this v, explain why one has

$$p(y) = F(y)\; F^*(p) \,.$$

(b) Prove that corresponding to the given nonzero p, there is a unique nonzero $y \in \mathbb{R}^n$ such that

$$p(y) = F(y)\; F^*(p) \quad \text{and} \quad F(y) = F^*(p) \,.$$

(c) Equivalently, prove that there is a unique nonzero $y \in \mathbb{R}^n$ at which one has

$$p(y) = F^2(y) \quad \text{and} \quad F^*(p) = F(y) \,.$$

Exercise 14.7.3:

(a) Show that $F^*(p)$ is positively homogeneous of degree 1 in p.

(b) Prove that $F^*(p)$ is absolutely homogeneous in p if and only if $F(y)$ is absolutely homogeneous in y.

Exercise 14.7.4:

(a) Show that if $z \in \overline{\mathrm{B}}^o$, then the linear functional $z \bullet (\,\cdot\,)$ is in $\overline{\mathrm{B}}^*$.

(b) Conversely, suppose $p \in \overline{\mathrm{B}}^*$ is given. Explain why it must be of the form $z \bullet (\,\cdot\,)$ for some $z \in \mathbb{R}^n$. Then show that such z must be in $\overline{\mathrm{B}}^o$.

(c) Check that the map $z \mapsto z \bullet (\,\cdot\,)$ has derivative matrix given by the Kronecker delta; hence its Jacobian determinant is equal to 1.

14.8 The Legendre Transformation

Given $y \in \mathbb{R}^n \setminus 0$, define the element $y^\flat \in \mathbb{R}^{n*} \setminus 0$ by

$$\boxed{y^\flat(v) \ := \ g_{ij}(y)\, y^i\, v^j \ =: \ y_j\, v^j} \ . \tag{14.8.1}$$

So, with each nonzero y, we have associated a nonzero linear functional $y^\flat$. Note, however, that the map $y \mapsto y^\flat$ is typically nonlinear because of the presence of y in g_{ij}.

Let us motivate the above definition. See also Miron [Mir2], [Mir3].

* Start with the given Minkowski norm F. As in classical mechanics, we form the **Lagrangian** function

$$\mathcal{L}(y) \ := \ \frac{1}{2}\, F^2(y)$$

on the velocity phase space $\mathbb{R}^n$. The coordinates y^i are our **velocity** variables.

* Next, we switch from the velocities-Lagrangian description to the momenta-Hamiltonian description. To this end, define the **conjugate** (or **canonical**) **momenta** p_j by

$$p_j \ := \ \frac{\partial \mathcal{L}}{\partial y^j} \ = \ y_j \ .$$

The last equality comes from $(\frac{1}{2}F^2)_{y^j} = F\, F_{y^j} = F\, \ell_j = y_j$. So the conjugate momenta of any nonzero y^j is simply the y_j in (14.8.1).

* The corresponding **Hamiltonian** $\mathcal{H}$ is a function on the momentum phase space $\mathbb{R}^{n*}$. Its value at $p = y^\flat$ is given (via Euler's theorem) by

$$\mathcal{H}(p) \ := \ p_j\, y^j - \mathcal{L}(y) \ = \ 2\,\mathcal{L} - \mathcal{L} \ = \ \frac{1}{2}\, F^2(y) \ ,$$

which is numerically the same as $\mathcal{L}(y)$!

What we have just described is known as the **Legendre transformation.**

Although it is true that the process $y \mapsto y^\flat$ maps $\mathbb{R}^n \setminus 0$ onto $\mathbb{R}^{n*} \setminus 0$, we have not established the said surjectivity yet. How then would one define the Hamiltonian $\mathcal{H}$ on all of $\mathbb{R}^{n*}$? For that purpose, let us recall the dual norm F^* defined in §14.7. By definition, we have

$$F^*(y^\flat) = \sup_{v \in \mathrm{S}} y^\flat(v) .$$

For each v in the indicatrix S, the fundamental inequality (1.2.16) says that

$$y^\flat(v) = g_{ij}(y)\, y^i v^j \leqslant F(y)\, F(v) = F(y) .$$

Thus $F(y)$ is an upper bound of $\{ y^\flat(v) : v \in \mathrm{S} \}$. This upper bound is actually attained at $v := \frac{y}{F(y)}$. Therefore,

$$\boxed{F^*(y^\flat) = F(y)} . \tag{14.8.2}$$

Consequently, we can define the Hamiltonian on all of $\mathbb{R}^{n*}$ as

$$\mathcal{H}(p) := \frac{1}{2} F^{*2}(p) ,$$

and that would be consistent with what we obtained through the Legendre transformation.

Let us also set

$$g^{ij}(p) := \left[\frac{1}{2} F^{*2}(p) \right]_{p_i p_j} . \tag{14.8.3}$$

Note that since $[\frac{1}{2}F^{*2}(p)]_{p_i}$ is homogeneous of degree 1 in p, Euler's theorem (Theorem 1.2.1) implies that

$$g^{ij}(p)\, p_j = \left[\frac{1}{2} F^{*2}(p) \right]_{p_i} . \tag{14.8.4}$$

We are now ready to state the following useful proposition. Some of our arguments are based on insights from the treatment in Rund [R].

Proposition 14.8.1. *Let F be a Minkowski norm on $\mathbb{R}^n$ that is possibly only positively homogeneous. Let F^* denote the dual norm, in the functional analysis sense, on $\mathbb{R}^{n*}$. Then:*

- *The Legendre transformation $y \mapsto y^\flat$, namely,*

$$y^j \mapsto y_j := g_{ij}(y)\, y^i ,$$

is a smooth diffeomorphism from $\mathbb{R}^n \setminus 0$ onto $\mathbb{R}^{n*} \setminus 0$.

- *It is norm preserving. That is,*

$$F^*(y^\flat) = F(y) .$$

- *The inverse of the Legendre transformation is given by*

$$p \mapsto p^\sharp \ , \quad \text{namely,} \quad p_i \mapsto p^i := g^{ij}(p)\, p_j \ .$$

- *At $p = y^\flat$, we have*

$$g^{ij}(p) = g^{ij}(y) \ ,$$

where $g^{ij}(y)$ denotes the matrix inverse of $g_{ij}(y)$.

- *The dual norm F^* in the functional analysis sense is actually a bonafide Minkowski norm on $\mathbb{R}^{n*}$.*

Proof.

- The Minkowski norm is by hypothesis smooth at all nonzero y. Hence the same holds for the fundamental tensor $g_{ij}(y)$. This now gives the smoothness of the map $y^j \mapsto y_j := g_{ij}(y)\, y^i$.
- According to (14.8.2), the said map is also norm preserving. That is, $F^*(\, y^\flat \,) = F(y)$.

We now turn our attention to the remaining assertions of the proposition.

Establishing injectivity:

Suppose $y_1^\flat = p = y_2^\flat$ for nonzero y_1, y_2 in $\mathbb{R}^n$. In particular,

$$p(y_1) = F^2(y_1) \quad \text{and} \quad p(y_2) = F^2(y_2) \ .$$

The elements

$$v_1 := \frac{y_1}{F(y_1)} \ , \qquad v_2 := \frac{y_2}{F(y_2)}$$

belong to the indicatrix S; they re-express the above equalities as

$$(*) \qquad p(v_1) = F(y_1) \quad \text{and} \quad p(v_2) = F(y_2) \ .$$

As remarked at the beginning of the proof, the map $y \mapsto y^\flat$ is norm preserving. So $F(y_1) = F^*(p) = F(y_2)$, thereby converting (*) to read:

$$(**) \qquad p(v_1) = F^*(p) = p(v_2) \ .$$

The element p is nonzero. This is made manifestly so through (*). Statement (14.7.5) then guarantees a unique v in the indicatrix S such that $p(v) = F^*(p)$. This uniqueness forces v_1 and v_2 to be equal, which implies that $y_2 = \lambda\, y_1$ for some positive λ. Substituting such conclusion into $y_1^\flat = y_2^\flat$, we get

$$g_{ij}(y_1)\; y_1^i = g_{ij}(\lambda\, y_1)\; \lambda\; y_1^i \ .$$

But $g_{ij}(y)$ is invariant under positive rescaling in y, thus the above becomes

$$g_{ij}(y_1)\; y_1^i = g_{ij}(y_1)\; \lambda\; y_1^i \ .$$

In other words,

$$(1 - \lambda)\, y_1^\flat \;=\; (1 - \lambda)\, p \;=\; 0\,.$$

Since $p \neq 0$, we must have $\lambda = 1$. That is, $y_2 = y_1$.

Establishing surjectivity:

Let us be given any nonzero $p \in \mathbb{R}^{n*}$. By (14.7.5), there exists a unique v in the indicatrix S such that $p(v) = F^*(p)$. Define

$$y \;:=\; F^*(p)\, v\,.$$

Then

$$(***) \qquad F(y) \;=\; F^*(p)\,.$$

We want to show that $y^\flat = p$. To this end, fix any $w \in \mathbb{R}^n$. In view of (14.7.4) and (14.7.5), we have

$$p\left(\frac{y + tw}{F(y + tw)}\right) \;\leqslant\; p(v) \;=\; F^*(p)\,.$$

Therefore the function

$$f(t) \;:=\; p(y + tw) \;-\; F(y + tw)\, F^*(p)$$

is nonpositive for all t, and achieves its absolute maximum value of zero at $t = 0$. In particular, $f'(0) = 0$. Using the fact that p is linear, together with (***) above and formula (1.4.3) for F_{y^i}, we have

$$f'(0) \;=\; p(w) \;-\; \frac{y_i}{F(y)}\, w^i\, F^*(p) \;=\; p(w) \;-\; y^\flat(w)\,.$$

This holds for any fixed w. Therefore $p = y^\flat$ as claimed.

We have shown that:

The Legendre transformation $y \mapsto y^\flat$ is a smooth, norm preserving diffeomorphism from $(\mathbb{R}^n \smallsetminus 0\,,\, F)$ onto $(\mathbb{R}^{n*} \smallsetminus 0\,,\, F^*)$.

The inverse of the Legendre transformation:

We wish to demonstrate that

$$g^{ij}(p)\, p_j \;=\; y^i \quad \text{at} \quad p = y^\flat\,.$$

The computation consists of the following steps, all to be carried out at $p = y^\flat$.

* With the help of (14.8.4) and then (14.8.2), we have

$$g^{ij}(p)\, p_j \;=\; \left[\frac{1}{2}\, F^{*2}(p)\right]_{p_i} \;=\; \left[\frac{1}{2}\, F^2(y)\right]_{p_i}\,.$$

* Since the Legendre transformation is a smooth diffeomorphism, the relation $p = y^\flat$ can be inverted, albeit only implicitly at the moment,

to give y as a smooth function of p. In particular, the chain rule is applicable to the last term in the above. The result reads

$$\begin{aligned} g^{ij}(p)\, p_j &= \left[\frac{1}{2}\, F^2(y)\right]_{y^j} \frac{\partial y^j}{\partial p_i} \\ &= \left[\frac{1}{2}\, F^2(y)\right]_{y^j y^k} y^k\, \frac{\partial y^j}{\partial p_i} \\ &= g_{jk}(y)\, y^k\, \frac{\partial y^j}{\partial p_i} \,. \end{aligned}$$

Here, the second last equality follows from Euler's theorem (Theorem 1.2.1) and the fact that $[\frac{1}{2}F^2(y)]_{y^j}$ is homogeneous of degree 1 in y.

* Differentiate the statement $p_i = g_{ij}(y)\, y^j$ with respect to p_k:

$$\begin{aligned} \delta_i^{\ k} &= \frac{\partial g_{ij}(y)}{\partial y^s}\, \frac{\partial y^s}{\partial p_k}\, y^j + g_{ij}(y)\, \frac{\partial y^j}{\partial p_k} \\ &= \left[\frac{1}{2}\, F^2(y)\right]_{y^i y^s y^j} y^j\, \frac{\partial y^s}{\partial p_k} + g_{ij}(y)\, \frac{\partial y^j}{\partial p_k} \\ &= g_{ij}(y)\, \frac{\partial y^j}{\partial p_k} \,. \end{aligned}$$

The last equality comes from a direct application of Euler's theorem (Theorem 1.2.1) to the quantity $[\frac{1}{2}F^2(y)]_{y^i y^s}$, which is homogeneous of degree zero in y. Hence

$$\frac{\partial y^j}{\partial p_k} = g^{jk}(y) := \text{the inverse of } g_{ij}(y) \,.$$

* Putting all these ingredients together, we have:

$$g^{ij}(p)\, p_j = g_{jk}(y)\, y^k\, g^{ji}(y) = \delta_k^{\ i}\, y^k = y^i \,,$$

as desired.

This proves that the map $p \mapsto p^\sharp$, namely, $p_i \mapsto g^{ij}(p)\, p_j$, is the inverse of the Legendre transformation.

Showing that at $p = y^\flat$, we have $g^{ij}(p) = g^{ij}(y)$:

* We have just shown that

$$y^i = g^{ij}(p)\, p_j = \left[\frac{1}{2}\, F^{*2}(p)\right]_{p_i} \,.$$

Differentiating with respect to y^k and using the chain rule, we get:

$$\delta^i_{\ k} = \left[\frac{1}{2}\, F^{*2}(p)\right]_{p_i p_j} \frac{\partial p_j}{\partial y^k} = g^{ij}(p)\, \frac{\partial p_j}{\partial y^k} \,.$$

* On the other hand, differentiating the relation $p_j = g_{js}(y)\, y^s$ with respect to y^k gives

$$\frac{\partial p_j}{\partial y^k} = g_{jk}(y) \ .$$

The computation is similar to (but simpler than) that for $\frac{\partial y^j}{\partial p_k}$, which we carried out in detail a moment ago.

* Therefore,

$$\delta^i{}_k = g^{ij}(p)\ g_{jk}(y) \ .$$

We have just shown that at $p = y^\flat$, the matrix $(\, g^{ij}(p)\,)$ is the inverse of the matrix $(\, g_{jk}(y)\,)$. In other words, $g^{ij}(p) = g^{ij}(y)$.

That $F^*(p)$ is a Minkowski norm on $\mathbb{R}^{n*}$:
At any nonzero p, we have $p = y^\flat$.

* The inverse $p \mapsto y$ of the Legendre transformation is a smooth diffeomorphism from $\mathbb{R}^{n*} \setminus 0$ onto $\mathbb{R}^n \setminus 0$. Also, $F^*(p) = F(y)$ and F is smooth at any nonzero y. Therefore, the chain rule tells us that F^* is smooth at every nonzero p.
* It can be checked from its definition (14.7.4) that $F^*(p)$ is positively homogeneous of degree 1 in p. See Exercise 14.7.3.
* We have seen that

$$\left[\frac{1}{2}\ F^{*2}(p) \right]_{p_i p_j} = g^{ij}(p) = g^{ij}(y) \ .$$

Since F is strongly convex, $g_{ij}(y)$ is positive-definite. Thus so is its inverse $g^{ij}(y)$. It follows that $[\frac{1}{2}F^{*2}(p)]_{p_i p_j}$ is positive-definite.

We now know that F^* is regular, positively homogeneous of degree 1, and strongly convex. Therefore it is a Minkowski norm on $\mathbb{R}^{n*}$. Note that given these three properties, its positivity is automatic, by Theorem 1.2.2. Such is also apparent from the statement $F^*(p) = F(y)$. □

Exercises

Exercise 14.8.1: Show that the Jacobian determinant of the Legendre transformation is simply $\det(\, g_{ij}(y)\,)$.

Exercise 14.8.2: By differentiating the relation $p_j = g_{js}(y)\, y^s$ with respect to y^k, show that

$$\frac{\partial p_j}{\partial y^k} = g_{jk}(y) \ .$$

14.9 A Mixed-Volume Inequality, and Brickell's Theorem

This section has two goals:

* We first illustrate how the Blaschke–Santaló inequality and the Legendre transformation can join forces to produce a beautiful mixed-volume inequality.
* This mixed-volume inequality is then used, in conjunction with Hopf's classification of Riemannian space forms, to give an elegant proof of Brickell's theorem.

The plan we have just outlined is due to Schneider [Sch1].

We begin with some preliminaries. Let F be any Minkowski norm on $\mathbb{R}^n$. Recall from §14.2 the Riemannian metric $\hat{g} := g_{ij}(y)\, dy^i \otimes dy^j$ that it induces on the punctured space $\mathbb{R}^n \setminus 0$. Worth emphasizing is the fact that these y^i are global linear coordinates, discussed in detail in §14.2. This $\hat{g}$, in turn, induces a Riemannian metric $\dot{g}$ on the indicatrix S.

Those global linear coordinates, like Cartesian coordinates in Euclidean space, are sometimes not the most convenient to use. Such is the case in these preliminaries. So, as in §14.3, let us introduce the radial variable $r := F(y)$. The coordinates θ^α $(\alpha = 1, \ldots, n-1)$ on S are then our angular variables. Formula (14.3.5) tells us that, in terms of the spherical coordinates just described, the metric $\hat{g}$ takes the form

$$dr \otimes dr \ + \ r^2 \ \dot{g} \ .$$

In matrix language:

$$\begin{pmatrix} \hat{g}_{rr} & \hat{g}_{r\beta} \\ \hat{g}_{\alpha r} & \hat{g}_{\alpha\beta} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & r^2\, \dot{g}_{\alpha\beta} \end{pmatrix} . \tag{14.9.1}$$

Formula (14.9.1) immediately gives

$$\det \hat{g} \ = \ (r^2)^{n-1} \ \det \dot{g} \ ,$$

which can equivalently be re-expressed as

$$\sqrt{\hat{g}} \ = \ r^{n-1} \ \sqrt{\dot{g}} \ .$$

Consequently, we have the following relationship between volume forms:

$$\sqrt{\hat{g}}\ dr \wedge d\theta \ = \ r^{n-1}\ dr \wedge \left(\sqrt{\dot{g}}\ \ d\theta \right) .$$

Here, $d\theta := d\theta^1 \wedge \cdots \wedge d\theta^{n-1}$. In other words,

$$dV_{\hat{g}} \ = \ r^{n-1}\ dr \ \wedge \ dV_{\dot{g}} \ . \tag{14.9.2}$$

We now use (14.9.2) to calculate the $\hat{g}$ Riemannian volume of

$$\overline{\mathrm{B}}(r) \ := \ \{ y \in \mathbb{R}^n : \ F(y) \leqslant r \} \ .$$

This is to be distinguished from its Euclidean volume. See Exercise 14.9.1. Anyway, we have

$$\mathrm{vol}_{\hat{g}}[\,\overline{\mathrm{B}}(r)\,] \;=\; \int_{0<F(y)\leqslant r} dV_{\hat{g}} \;=\; \lim_{\epsilon\to 0} \int_S \left(\int_\epsilon^r \rho^{n-1}\, d\rho \right) dV_{\dot{g}} \;.$$

It is necessary to have improper integrals here because $\hat{g}$ is typically singular at the origin of $\mathbb{R}^n$. But the singularity is mild, and simple calculus gives

(14.9.3) $$\boxed{\mathrm{vol}_{\hat{g}}[\,\overline{\mathrm{B}}(r)\,] \;=\; \frac{r^n}{n}\; \mathrm{vol}_{\dot{g}}(\,\mathrm{S}\,)} \;,$$

where S is the indicatrix, *not* $\mathrm{S}(r)$. In particular, for $r = 1$, we have

(14.9.4) $$\boxed{\mathrm{vol}_{\hat{g}}(\,\overline{\mathrm{B}}\,) \;=\; \frac{1}{n}\; \mathrm{vol}_{\dot{g}}(\,\mathrm{S}\,)} \;.$$

There is a way to derive (14.9.4) without using spherical coordinates. See Exercise 14.9.3. Also, combining (14.9.3) with Exercise 14.5.4 gives something worth noting: $\mathrm{vol}[\,\overline{\mathrm{B}}(r)\,] = \frac{r}{n}\, \mathrm{vol}[\,\mathrm{S}(r)\,]$.

We have finished the requisite preliminaries. Now we turn to the first goal of this section, namely, the **mixed-volume inequality**.

Proposition 14.9.1. *Let $(\,\mathbb{R}^n\,,\, F\,)$ be a Minkowski space. Suppose the Minkowski norm F is* **absolutely homogeneous** *of degree one. Then*

$$\boxed{\mathit{vol}_{\hat{g}}(\,\overline{B}\,) \;\leqslant\; \mathit{vol}(\,\overline{\mathbb{B}}\,)} \;.$$

Equivalently,

$$\boxed{\mathit{vol}_{\dot{g}}(\,S\,) \;\leqslant\; \mathit{vol}(\,\mathbb{S}\,)} \;.$$

In either case, equality holds if and only if F is the norm induced by an inner product on $\mathbb{R}^n$.

Remarks:

* For emphasis, we reiterate that

$$\begin{aligned} \overline{\mathrm{B}} &:= \{\, y \in \mathbb{R}^n : \; F(y) \;\leqslant\; 1 \,\} \\ \overline{\mathbb{B}} &:= \{\, y \in \mathbb{R}^n : \; y \bullet y \;\leqslant\; 1 \,\} \\ \mathrm{S} &:= \{\, y \in \mathbb{R}^n : \; F(y) \;=\; 1 \,\} \\ \mathbb{S} &:= \{\, y \in \mathbb{R}^n : \; y \bullet y \;=\; 1 \,\}\,, \end{aligned}$$

  where $\bullet$ denotes the dot product.

* The closed balls are n-dimensional, but the spheres are $(n-1)$-dimensional.

* The upper bounds $\mathrm{vol}(\,\overline{\mathbb{B}}\,)$ and $\mathrm{vol}(\,\mathbb{S}\,)$ are Euclidean volumes. See Exercise 14.5.5 for explicit formulas of $\mathrm{vol}(\,\mathbb{S}\,)$.

* By contrast, the quantities they are bounding, namely, $\mathrm{vol}_{\hat{g}}(\overline{\mathrm{B}})$ and $\mathrm{vol}_{\dot{g}}(\mathrm{S})$, are (typically non-Euclidean) Riemannian volumes. Hence the name "mixed-volume" inequality.

Proof. Formula (14.9.4) says that $\mathrm{vol}_{\hat{g}}(\overline{\mathrm{B}}) = \frac{1}{n}\mathrm{vol}_{\dot{g}}(\mathrm{S})$. Specializing it to Euclidean space gives $\mathrm{vol}(\overline{\mathbb{B}}) = \frac{1}{n}\mathrm{vol}(\mathbb{S})$. The two asserted inequalities are thus equivalent.

We prove the inequality for the closed balls. In order to reduce clutter, let us introduce the abbreviation

$$dy \ := \ dy^1 \wedge \cdots \wedge dy^n \ .$$

We first write out the definition of $\mathrm{vol}_{\hat{g}}(\overline{\mathrm{B}})$ and apply the standard Cauchy–Schwarz inequality to it:

$$\begin{aligned} \mathrm{vol}_{\hat{g}}(\overline{\mathrm{B}}) &= \lim_{\epsilon\to 0} \int_{\epsilon\leqslant F\leqslant 1} \sqrt{\det[\,g_{ij}(y)\,]}\ dy \\ (*) \qquad &\leqslant \lim_{\epsilon\to 0} \left[\int_{\epsilon\leqslant F\leqslant 1} \det[\,g_{ij}(y)\,]\ dy\right]^{1/2} \left[\int_{\epsilon\leqslant F\leqslant 1} dy\right]^{1/2} \\ &= \left[\lim_{\epsilon\to 0} \int_{\epsilon\leqslant F\leqslant 1} \det[\,g_{ij}(y)\,]\ dy\right]^{1/2} [\,\mathrm{vol}(\overline{\mathrm{B}})\,]^{1/2} \ . \end{aligned}$$

Note that $\mathrm{vol}(\overline{\mathrm{B}})$ is the Euclidean volume of the subset $\overline{\mathrm{B}}$ of $\mathbb{R}^n$.

Next, recall from Proposition 14.8.1 that the Legendre transformation $y \mapsto y^\flat$ is a norm preserving diffeomorphism from $\mathbb{R}^n \setminus 0$ onto $\mathbb{R}^{n*} \setminus 0$. Hence it takes the subset $\epsilon \leqslant F(y) \leqslant 1$ diffeomorphically onto the subset $\epsilon \leqslant F^*(p) \leqslant 1$. Also, Exercise 14.8.1 tells us that this map has Jacobian determinant equal to $\det[\,g_{ij}(y)\,]$. Therefore, by the change-of-variables theorem, we can re-express the *Euclidean* volume of the dual closed unit ball $\overline{\mathrm{B}}^*$ [see (14.7.6)] as follows:

$$\begin{aligned} \mathrm{vol}(\,\overline{\mathrm{B}}^*\,) &= \lim_{\epsilon\to 0} \int_{\epsilon\leqslant F^*\leqslant 1} dp_1 \wedge \cdots \wedge dp_n \\ &= \lim_{\epsilon\to 0} \int_{\epsilon\leqslant F\leqslant 1} \det[\,g_{ij}(y)\,]\ dy \ . \end{aligned}$$

Consequently, we can rewrite (*) simply as

$$(**) \qquad \mathrm{vol}_{\hat{g}}(\overline{\mathrm{B}}) \ \leqslant\ [\,\mathrm{vol}(\,\overline{\mathrm{B}}^*\,)\,]^{\frac{1}{2}}\ [\,\mathrm{vol}(\overline{\mathrm{B}})\,]^{\frac{1}{2}} \ .$$

By hypothesis, the Minkowski norm F is absolutely homogeneous of degree 1. Thus the closed convex subset $\overline{\mathrm{B}}$ is centrally symmetric. As a result, the Blaschke–Santaló inequality (14.7.8) is applicable to the right-hand side of (**). It refines (**) to read:

$$(***) \qquad \mathrm{vol}_{\hat{g}}(\overline{\mathrm{B}}) \ \leqslant\ [\,\mathrm{vol}(\,\overline{\mathrm{B}}^*\,)\,]^{\frac{1}{2}}\ [\,\mathrm{vol}(\overline{\mathrm{B}})\,]^{\frac{1}{2}} \ \leqslant\ \mathrm{vol}(\overline{\mathbb{B}}) \ .$$

This is the inequality we wanted to deduce.

Finally:

* Suppose $\mathrm{vol}_{\hat{g}}(\overline{\mathrm{B}})$ is equal to $\mathrm{vol}(\overline{\mathbb{B}})$. This implies that in (***), the Blaschke–Santaló inequality becomes an equality. As stated at the end of §14.7, such forces $\overline{\mathrm{B}}$ to be an ellipsoid, and hence F must be the norm induced by an inner product on $\mathbb{R}^n$.
* Conversely, suppose our Minkowski norm arises from an inner product on $\mathbb{R}^n$. Say, $F = \sqrt{c_{ij}\, y^i y^j}$, where the c_{ij} are *constants.* The fundamental tensor is then c_{ij}, and

$$\mathrm{vol}_{\hat{g}}(\overline{\mathrm{B}}) \;=\; \int_{0\leqslant F\leqslant 1} \sqrt{\det(c_{ij})}\; dy^1 \wedge \cdots \wedge dy^n \;.$$

  Choose an orthonormal basis for this inner product and use the resulting global linear coordinates z^a. Then $F = \sqrt{\delta_{ab}\, z^a z^b} = \sqrt{z \bullet z}$. Express the y coordinates in terms of the z coordinates:

$$y^i \;=\; P^i{}_a\, z^a \;.$$

  This means that the y basis vectors are equal to the inverse matrix P^{-1} acting on the z basis vectors. Linear algebra then gives $c = (P^{-1})^t\, I\, P^{-1}$ because the fundamental tensor in the z coordinates is simply the identity matrix. In particular, $\sqrt{\det\, c} = (\det\, P)^{-1}$. On the other hand, $dy^1 \wedge \cdots \wedge dy^n = (\det\, P)\, dz^1 \wedge \cdots \wedge dz^n$. These maneuvers transform the above integral to

$$\int_{0\leqslant\sqrt{z\bullet z}\leqslant 1} dz^1 \wedge \cdots \wedge dz^n \;,$$

  which is $\mathrm{vol}(\overline{\mathbb{B}})$. We have thus shown that when the Minkowski norm comes from an inner product, (***) reduces to an equality.

The proof of the proposition is therefore complete. □

We are now fully prepared to state and prove **Brickell's theorem** [B2]. The proof we give is due to Schneider [Schl].

Theorem 14.9.2 (Brickell). *Suppose:*

- $(\mathbb{R}^n, F)$ *is a Minkowski space, with F smooth and strongly convex on all of* $\mathbb{R}^n \setminus 0$.
- $n \geqslant 3$.
- F *is* **absolutely** *homogeneous of degree one.*
- *The curvature tensor of* $\hat{g} := g_{ij}(y)\, dy^i \otimes dy^j$ *is zero on* $\mathbb{R}^n \setminus 0$.

Then F must be the norm of some inner product on $\mathbb{R}^n$.

Remarks:

(a) The hypothesis $n \geqslant 3$ is essential, because the conclusion blatantly fails for Minkowski surfaces. As we can see from the 2-D examples given in §14.1, none of the Minkowski norms there is induced by any inner product. Yet, given any Minkowski norm F on $\mathbb{R}^2 \setminus 0$, the curvature of $\hat{g}$ is *always* zero. This can be verified as in Exercise 14.2.3. Or one can use part (b) of Exercise 4.1.1, followed by the same of Exercise 13.2.2.

(b) There has been some speculation by a number of researchers (for example, Matsumoto) that perhaps the absolute homogeneity hypothesis can be weakened to positive homogeneity. This issue has yet to be settled. We feel that Brickell's original proof is more amenable to this purpose. However, a certain reference to Cartan in that proof needs to be closely scrutinized.

Proof. The argument we describe involves three major ingredients:

* The Gauss equation (14.6.6).
* Hopf's classification of Riemannian space forms (Theorem 13.4.1).
* The mixed-volume inequality (Proposition 14.9.1). This in turn comes from the Blaschke–Santaló inequality (14.7.8) and a detailed understanding of the Legendre transformation (Proposition 14.8.1).

The Riemannian metric $\hat{g}$ is defined by the fundamental tensor $g_{ij}(y)$ of F. It induces a Riemannian metric $\dot{g}$ on the indicatrix S. In view of Proposition 14.6.1, the hypothesized flatness of $\hat{g}$ is equivalent to the statement that $(\mathrm{S}, \dot{g})$ has constant sectional curvature 1. This is a direct consequence of the Gauss equation (14.6.6).

Since $n \geqslant 3$, the indicatrix S has dimension at least two. It is therefore simply connected. It is also compact, hence complete. By Hopf's classification theorem of Riemannian space forms (Theorem 13.4.1), $(\mathrm{S}, \dot{g})$ must be isometric to the standard sphere of radius 1 in Euclidean $\mathbb{R}^n$.

Because of the said isometry, the Riemannian volume of $(\mathrm{S}, \dot{g})$ is equal to the Euclidean value $\mathrm{vol}(\mathbb{S})$. This means that our mixed-volume inequality becomes an equality for the case in question. According to Proposition 14.9.1, this can only happen when the Minkowski norm F is induced by an inner product on $\mathbb{R}^n$. $\square$

Exercises

Exercise 14.9.1: The difference between the Riemannian and the Euclidean volume of $\overline{\mathrm{B}}(r)$ is most easily seen in terms of global linear coordinates y^i $(i = 1, \ldots, n)$.

(a) Let $\sqrt{g}$ denote $\sqrt{\det(g_{ij})}$, where g_{ij} is the fundamental tensor of the Minkowski norm F, in the said global linear coordinates. Check that the $\hat{g}$ Riemannian volume of $\overline{\mathrm{B}}(r)$ is

$$\int_{F(y)\leqslant r} \sqrt{g}\; dy^1 \wedge \cdots \wedge dy^n \ .$$

Hint: see §14.2.

(b) On the other hand, explain why the Euclidean volume of $\overline{\mathrm{B}}(r)$ is simply

$$\int_{F(y)\leqslant r} dy^1 \wedge \cdots \wedge dy^n \ .$$

Hint: see §14.7.

Exercise 14.9.2: Prove that

$$\boxed{(\sqrt{g}\,)_{y^k} \;=\; \frac{1}{2}\,\sqrt{g}\; g^{ij}\; (g_{ij}\,)_{y^k} \;=\; \frac{1}{F}\,\sqrt{g}\; A_k} \ .$$

Exercise 14.9.3: Define

$$\mathrm{S}(r) \;:=\; \{\, y \in \mathbb{R}^n :\ F(y) = r \,\} \ .$$

Recall from (1.4.8) or Exercise 1.4.3 that the volume form of the Riemannian manifold $\mathrm{S}(r)$ is

$$\sqrt{g}\, \sum_{j=1}^{n} (-1)^{j-1}\, \frac{y^j}{F}\; dy^1 \wedge \cdots \wedge dy^{j-1} \wedge dy^{j+1} \wedge \cdots \wedge dy^n \ ,$$

where that F is actually equal to the constant value r. Introduce, for the sake of this exercise, the following $(n-1)$-form:

$$\Phi \;:=\; \sqrt{g}\, \sum_{j=1}^{n} (-1)^{j-1}\; y^j\; dy^1 \wedge \cdots \wedge dy^{j-1} \wedge dy^{j+1} \wedge \cdots \wedge dy^n \ .$$

(a) Show that the integral of Φ over $\mathrm{S}(\epsilon)$ is ϵ^n times its integral over $\mathrm{S} = \mathrm{S}(1)$.

(b) Prove that

$$d\Phi \;=\; n\, \sqrt{g}\; dy^1 \wedge \cdots \wedge dy^n \;=\; n\; dV_{\hat{g}} \ .$$

Hint: use Exercise 14.9.2.

(c) Identify the origin of the statement

$$\int_{\mathrm{S}(1)} \Phi \;-\; \int_{\mathrm{S}(\epsilon)} \Phi \;=\; n \int_{\epsilon\leqslant F\leqslant 1} dV_{\hat{g}} \ .$$

(d) Let $\epsilon \to 0$ in part (c). Check that (14.9.4) results.

How would you modify the above strategy to deduce (14.9.3)?

References

[B1] F. Brickell, *A new proof of Deicke's theorem on homogeneous functions*, Proc. AMS **16** (1965), 190–191.

[B2] F. Brickell, *A theorem on homogeneous functions*, J. London Math. Soc. **42** (1967), 325–329.

[BL1] D. Bao and B. Lackey, *Randers surfaces whose Laplacians have completely positive symbol*, Nonlinear Analysis **38** (1999), 27–40.

[Bla] W. Blaschke, *Vorlesungen über Differentialgeometrie*, vol. II, Springer, 1923.

[D] A. Deicke, *Über die Finsler-Raume mit $A_i = 0$*, Arch. Math. **4** (1953), 45–51.

[GT] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Grundlehren der Mathematischen Wissenschaften, vol. 224, Springer-Verlag, 1983.

[Kawa] A. Kawaguchi, *On the theory of non-linear connections II. Theory of Minkowski spaces and of non-linear connections in a Finsler spàce*, Tensor, N.S. **6** (1956), 165–199.

[Kik] S. Kikuchi, *Theory of Minkowski space and of non-linear connections in Finsler space*, Tensor, N.S. **12** (1962), 47–60.

[M8] M. Matsumoto, *The main scalar of a two-dimensional Finsler space with special metric*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **32(4)** (1992), 889–898.

[MP] M. Meyer and A. Pajor, *On Santaló's inequality*, Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics, vol. 1376, J. Lindenstrauss and V. D. Milman, eds., Springer-Verlag, 1989.

[Mir2] R. Miron, *Cartan spaces in a new point of view by considering them as duals of Finsler spaces*, Tensor, N.S. **46** (1987), 330–334.

[Mir3] R. Miron, *Hamilton geometry*, Ann. şt. Univ. Al.I.Cuza, Iaşi **35** (1989), 33–67.

[R] H. Rund, *The Differential Geometry of Finsler Spaces*, Springer-Verlag, 1959.

[Roy] H. L. Royden, *Real Analysis*, Macmillan, 1963.

[San] L. A. Santaló, *Un invariante afín para los cuerpos convexos del espacio de n dimensiones*, Portugal Math. **8** (1949), 155–161.

[Sch1] R. Schneider, *Über die Finslerräume mit $S_{ijkl} = 0$*, Arch. Math. **19** (1968), 656–658.

[Sch2] R. Schneider, *Convex Bodies: The Brunn–Minkowski Theory*, Encyclopedia of Mathematics and its Applications, vol. 44, Cambridge University Press, 1993.

[Smo] J. Smoller, *Shock Waves and Reaction-Diffusion Equations*, Grundlehren der Mathematischen Wissenschaften, vol. 258, 2nd ed., Springer-Verlag, 1994.

Bibliography

Here, we consolidate all the cited references into one place. This is intended to give a sense of the book's scope. We ask that our readers not feel offended by any omission.

References

[A] M. Anastasiei, *A historical remark on the connections of Chern and Rund*, Cont. Math. **196** (1996), 171–176.

[AB] P. L. Antonelli and R. H. Bradbury, *Volterra–Hamilton Models in the Ecology and Evolution of Colonial Organisms*, World Scientific, 1996.

[AIM] P. L. Antonelli, R. S. Ingarden, and M. Matsumoto, *The Theory of Sprays and Finsler Spaces with Applications in Physics and Biology*, FTPH 58, Kluwer Academic Publishers, 1993.

[AL] P. L. Antonelli and B. Lackey (eds.), *The Theory of Finslerian Laplacians and Applications*, MAIA 459, Kluwer Academic Publishers, 1998.

[AP] M. Abate and G. Patrizio, *Finsler Metrics — A Global Approach, with applications to geometric function theory*, Lecture Notes in Mathematics, vol. 1591, Springer-Verlag, 1994.

[AZ] H. Akbar-Zadeh, *Sur les espaces de Finsler à courbures sectionnelles constantes*, Acad. Roy. Belg. Bull. Cl. Sci. (5) **74** (1988), 281–322.

[AZas] P. L. Antonelli and T. J. Zastawniak, *Fundamentals of Finslerian Diffusion with Applications*, FTPH 101, Kluwer Academic Publishers, 1999.

[Ant] P. L. Antonelli, *Preface for Applications of Finsler Differential Geometry to Biology, Engineering, and Physics*, Cont. Math. **196** (1996), 199–202.

[Asan] G. S. Asanov, *Finsler Geometry, Relativity and Gauge Theories*, FTPH 12, D. Reidel and Kluwer Academic Publishers, 1985.

[Au] L. Auslander, *On curvature in Finsler geometry*, Trans. AMS **79** (1955), 378–388.

[B1] F. Brickell, *A new proof of Deicke's theorem on homogeneous functions*, Proc. AMS **16** (1965), 190–191.

[B2] F. Brickell, *A theorem on homogeneous functions*, J. London Math. Soc. **42** (1967), 325–329.

[BC1] D. Bao and S. S. Chern, *On a notable connection in Finsler geometry*, Houston J. Math. **19** (1993), 135–180.

[BC2] D. Bao and S. S. Chern, *A note on the Gauss–Bonnet theorem for Finsler spaces*, Ann. Math. **143** (1996), 233–252.

[BCS1] D. Bao, S. S. Chern, and Z. Shen, *On the Gauss–Bonnet integrand for 4-dimensional Landsberg spaces*, Cont. Math. **196** (1996), 15–25.

[BCS2] D. Bao, S. S. Chern, and Z. Shen (eds.), *Finsler Geometry*, Joint Summer Research Conference on Finsler Geometry, July 16–20, 1995, Seattle, Washington, AMS, Cont. Math. **196**, 1996.

[BCS3] D. Bao, S. S. Chern, and Z. Shen, *Rigidity issues on Finsler surfaces*, Rev. Roumaine Math. Pures Appl. **42** (1997), 707–735.

[BG] M. Berger and B. Gostiaux, *Differential Geometry: Manifolds, Curves, and Surfaces*, Graduate Texts in Mathematics **115**, Springer-Verlag, 1988.

[BL1] D. Bao and B. Lackey, *Randers surfaces whose Laplacians have completely positive symbol*, Nonlinear Analysis **38** (1999), 27–40.

[BL2] D. Bao and B. Lackey, *A Hodge decomposition theorem for Finsler spaces*, C.R. Acad. Sci. Paris **323** (1996), 51–56.

[BL3] D. Bao and B. Lackey, *A geometric inequality and a Weitzenböck formula for Finsler surfaces*, The Theory of Finslerian Laplacians and Applications, P. L. Antonelli and B. Lackey (eds.), MAIA 459, Kluwer Academic Publishers, 1998, pp. 245–275.

[BM] S. Bácsó and M. Matsumoto, *On Finsler spaces of Douglas type, a generalization of the notion of Berwald space*, Publ. Math. Debrecen **51** (1997), 385–406.

[BS] D. Bao and Z. Shen, *On the volume of unit tangent spheres in a Finsler manifold*, Results in Math. **26** (1994), 1–17.

[Bej] A. Bejancu, *Finsler Geometry and Applications*, Ellis Harwood, 1990.

[Ber1] L. Berwald, *Atti Congresso Internal dei Mate., Bologna 3–10, Sept.* (1928).

[Ber2] L. Berwald, *Two-dimensional Finsler spaces with rectilinear extremals*, Ann. Math. **42** (1941), 84–112.

[Bl] R. G. Beil, *Preface for Applications of Finsler Geometry to Relativistic Field Theory*, Cont. Math. **196** (1996), 261–263.

[Bla] W. Blaschke, *Vorlesungen über Differentialgeometrie*, vol.II, Springer, 1923.

[Br1] R. Bryant, *Finsler structures on the 2-sphere satisfying* $K = 1$, Cont. Math. **196** (1996), 27–42.

[Br2] R. Bryant, *Projectively flat Finsler 2-spheres of constant curvature*, Selecta Mathematica, New Series **3** (1997), 161–203.

[Br3] R. Bryant, *Finsler surfaces with prescribed curvature conditions*, Aisenstadt Lectures, in preparation.

[BuMa] H. Busemann and W. Mayer, *On the foundations of the calculus of variations*, Trans. AMS **49** (1941), 173–198.

[CE] J. Cheeger and D. Ebin, *Comparison Theorems in Riemannian Geometry*, North Holland/American Elsevier, 1975.

[Ch1] S. S. Chern, *Local equivalence and Euclidean connections in Finsler spaces*, Sci. Rep. Nat. Tsing Hua Univ. Ser. A **5** (1948), 95–121; or Selected Papers, vol. II, 194–212, Springer 1989.

[Ch2] S. S. Chern, *Historical remarks on Gauss–Bonnet*, Analysis et Cetera, volume dedicated to Jürgen Moser, Academic Press, 1990, pp. 209–217.

[Ch3] S. S. Chern, *A simple intrinsic proof of the Gauss–Bonnet formula for closed Riemannian manifolds*, Ann. Math. **45(4)** (1944), 747–752.

[Ch4] S. S. Chern, *On the curvatura integra in a Riemannian manifold*, Ann. Math. **46(4)** (1945), 674–684.

[D] A. Deicke, *Über die Finsler-Raume mit* $A_i = 0$, Arch. Math. **4** (1953), 45–51.

[Daz] P. Dazord, *Propriétés globales des géodésiques des Espaces de Finsler*, Theses, Université de Lyon, 1969.

[delR] L. del Riego, *Tenseurs de Weyl d'un spray de directions*, Theses, Université Scientifique et Medicale de Grenoble, 1973.

[doC1] M. P. do Carmo, *Differential Forms and Applications*, Springer-Verlag, 1994.

[doC2] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, 1976.

[doC3] M. P. do Carmo, *Riemannian Geometry*, Mathematics: Theory and Applications, Birkhäuser, 1992.

[E] D. Egloff, *Uniform Finsler Hadamard manifolds*, Ann. Inst. Henri Poincaré **66** (1997), 323–357.

[F1] P. Funk, *Über zweidimensionale Finslersche Räume, insbesondere über solche mit geradlinigen Extremalen und positiver konstanter Krümmung*, Math. Zeitschr. **40** (1936), 86–93.

[F2] P. Funk, *Eine Kennzeichnung der zweidimensionalen elliptischen Geometrie*, Österreichische Akad. der Wiss. Math., Sitzungsberichte Abteilung II **172** (1963), 251–269.

[Fou] P. Foulon, *Géométrie des équations différentielles du second ordre*, Ann. Inst. Henri Poincaré **45(1)** (1986), 1–28.

[Fult] C. M. Fulton, *Parallel vector fields*, Proc. Amer. Math. Soc. **16** (1965), 136–137.

[GHL] S. Gallot, D. Hulin, and J. Lafontaine, *Riemannian Geometry*, Universitext, 2nd ed., Springer-Verlag, 1990.

[GT] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Grundlehren der Mathematischen Wissenschaften, vol. 224, Springer-Verlag, 1983.

[GW] R. Gardner and G. Wilkens, *Preface for Applications of Finsler Geometry to Control Theory*, Cont. Math. **196** (1996), 227–229.

[Gr] A. Gray, *Modern Differential Geometry of Curves and Surfaces with Mathematica*, 2nd ed., CRC Press, 1998.

[HI] M. Hashiguchi and Y. Ichijyō, *On some special (α, β)-metrics*, Rep. Fac. Sci. Kagoshima Univ. **8** (1975), 39–46.

[HS1] D. Hrimiuc and H. Shimada, *On the $\mathcal{L}$-duality between Lagrange and Hamilton manifolds*, Nonlinear World **3** (1996), 613–641.

[HS2] D. Hrimiuc and H. Shimada, *On some special problems concerning the $\mathcal{L}$-duality between Finsler and Cartan spaces*, Tensor, N.S. **58** (1996), 48–61.

[I] Y. Ichijyō, *Finsler manifolds modeled on a Minkowski space*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **16-3** (1976), 639–652.

[Ing] R. S. Ingarden, *On physical applications of Finsler geometry*, Cont. Math. **196** (1996), 213–223.

[J] J. Jost, *Riemannian Geometry and Geometric Analysis*, Universitext, Springer-Verlag, 1995.

[KN1] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry*, vol. I, Wiley-Interscience, 1963 (1996).

[Kawa] A. Kawaguchi, *On the theory of non-linear connections II. Theory of Minkowski spaces and of non-linear connections in a Finsler space*, Tensor, N.S. **6** (1956), 165–199.

[Ki] S. Kikuchi, *On the condition that a space with (α, β)-metric be locally Minkowskian*, Tensor, N.S. **33** (1979), 242–246.

[Kik] S. Kikuchi, *Theory of Minkowski space and of non-linear connections in Finsler space*, Tensor, N.S. **12** (1962), 47–60.

[Ko] S. Kobayashi, *On conjugate and cut loci*, Global Differential Geometry, S. S. Chern, ed., Math. Assoc. America, 1989, pp. 140–169.

[M1] M. Matsumoto, *A slope of a mountain is a Finsler surface with respect to a time measure*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **29-1** (1989), 17–25.

[M2] M. Matsumoto, *Foundations of Finsler Geometry and Special Finsler Spaces*, Kaiseisha Press, Japan, 1986.

[M3] M. Matsumoto, *Theory of curves in tangent planes of two-dimensional Finsler spaces*, Tensor, N.S. **37** (1982), 35–42.

[M4] M. Matsumoto, *On Finsler spaces with Randers' metric and special forms of important tensors*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **14** (1974), 477–498.

[M5] M. Matsumoto, *Randers spaces of constant curvature*, Rep. on Math. Phys. **28** (1989), 249–261.

[M6] M. Matsumoto, *Remarks on Berwald and Landsberg spaces*, Cont. Math. **196** (1996), 79–82.

[M7H] M. Matsumoto and Hōjō, *A conclusive theorem on C-reducible Finsler spaces*, Tensor, N.S. **32** (1978), 225–230.

[M8] M. Matsumoto, *The main scalar of a two-dimensional Finsler space with special metric*, J. Math. Kyoto Univ. (Kyoto Daigaku J. Math.) **32(4)** (1992), 889–898.

[M9] M. Matsumoto, *Theory of Finsler spaces with m-th root metric II*, Publ. Math. Debr. **49** (1996), 135–155.

[M10] M. Matsumoto, *Berwald connection of a Finsler space with an* (α, β)*-metric*, Tensor, N.S. **50** (1991), 18–21.

[M11] M. Matsumoto, *A History of Finsler Geometry*, Proc. of the 33rd Symp. on Finsler Geometry (ed. Okubo), Lake Yamanaka, 1998, pp. 71–97.

[MA] R. Miron and M. Anastasiei, *The Geometry of Lagrange Spaces: Theory and Applications*, FTPH 59, Kluwer Academic Publishers, 1994.

[MP] M. Meyer and A. Pajor, *On Santaló's inequality*, Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics, vol. 1376, J. Lindenstrauss and V. D. Milman, eds., Springer-Verlag, 1989.

[Mir1] R. Miron, *General Randers spaces*, Lagrange and Finsler Geometry, P. L. Antonelli and R. Miron (eds.), Kluwer Academic Publishers, 1996, pp. 123–140.

[Mir2] R. Miron, *Cartan spaces in a new point of view by considering them as duals of Finsler spaces*, Tensor, N.S. **46** (1987), 330–334.

[Mir3] R. Miron, *Hamilton geometry*, Ann. şt. Univ. Al.I.Cuza, Iaşi **35** (1989), 33–67.

[Nu] S. Numata, *On the torsion tensors* R_{jhk} *and* P_{hjk} *of Finsler spaces with a metric* $ds = (g_{ij}(dx)\, dx^i dx^j)^{1/2} + b_i(x)\, dx^i$, Tensor, N.S. **32** (1978), 27–31.

[Num] S. Numata, *On Landsberg spaces of scalar curvature*, J. Korean Math. Soc. **12** (1975), 97–100.

[ON] B. O'Neill, *Semi-Riemannian Geometry, with Applications to Relativity*, Academic Press, 1983.

[Ok] T. Okada, *On models of projectively flat Finsler spaces of constant negative curvature*, Tensor, N.S. **40** (1983), 117–124.

[Oku] K. Okubo, Lecture at the Symposium on Finsler Geometry, 1977, unpublished (communicated to us by M. Matsumoto).

[On] B. O'Neill, *Elementary Differential Geometry*, 2nd ed., Academic Press, 1997.

[Op] J. Oprea, *Differential Geometry and its Applications*, Prentice-Hall, 1997.

[P] M. Pinl, *In memory of Ludwig Berwald*, translated by P. Bergmann and M. Grumet, Scripta Math. **27** (1965), 193–203.

[R] H. Rund, *The Differential Geometry of Finsler Spaces*, Springer-Verlag, 1959.

[Ra] G. Randers, *On an asymmetric metric in the four-space of general relativity*, Phys. Rev. **59** (1941), 195–199.

[Roy] H. L. Royden, *Real Analysis*, Macmillan, 1963.

[SK] C. Shibata and M. Kitayama, *On Finsler spaces of constant positive curvature*, Proceedings of the Romanian–Japanese Colloquium on Finsler Geometry, Braşov, 1984, pp. 139–156.

[SS] J. Schouten and D. Struik, *On some properties of general manifolds relating to Einstein's theory of gravitation*, Amer. J. Math. **43** (1921), 213–216.

[SSAY] C. Shibata, H. Shimada, M. Azuma, and H. Yasuda, *On Finsler spaces with Randers' metric*, Tensor, N.S. **31** (1977), 219–226.

[ST] I. M. Singer and J. A. Thorpe, *Lecture Notes on Elementary Topology and Geometry*, Undergraduate Texts in Mathematics, Springer-Verlag, 1976.

[SY] R. Schoen and S. T. Yau, *Lectures on Differential Geometry*, Conference Proceedings and Lecture Notes in Geometry and Topology, vol. I, International Press, 1994.

[SaS] S. Sabau and H. Shimada, *Classes of Finsler spaces with (α, β)-metrics*, 1999 preprint.

[San] L. A. Santaló, *Un invariante afín para los cuerpos convexos del espacio de n dimensiones*, Portugal Math. **8** (1949), 155–161.

[Sch1] R. Schneider, *Über die Finslerräume mit $S_{ijkl} = 0$*, Arch. Math. **19** (1968), 656–658.

[Sch2] R. Schneider, *Convex Bodies: The Brunn–Minkowski Theory*, Encyclopedia of Mathematics and its Applications, vol. 44, Cambridge University Press, 1993.

[Sh] Z. Shen, *Curvature, distance and volume in Finsler geometry*, unpublished.

[Smo] J. Smoller, *Shock Waves and Reaction–Diffusion Equations*, Grundlehren der Mathematischen Wissenschaften, vol. 258, 2nd ed., Springer-Verlag, 1994.

[Sp1] M. Spivak, *Differential Geometry*, vol. I, Publish or Perish, 1975.

[Sp2] M. Spivak, *Differential Geometry*, vol. II, Publish or Perish, 1970.

[Sp4] M. Spivak, *Differential Geometry*, vol. IV, Publish or Perish, 1975.

[Sz] Z. Szabó, *Positive definite Berwald spaces (structure theorems on Berwald spaces)*, Tensor, N.S. **35** (1981), 25–39.

[U1] C. Udriste, *Appendix 4. Completeness and convexity on Finsler manifolds*, Convex Functions and Optimization Methods on Riemannian Manifolds, MAIA 297, Kluwer Academic Publishers, 1994, pp. 318–330.

[U2] C. Udriste, *Completeness of Finsler manifolds*, Publ. Math. Debrecen **42** (1993), 45–50.

[W] J. H. C. Whitehead, *Convex regions in the geometry of paths*, Quarterly J. Math. Oxford, Ser. 3 (1932), 33–42.

[Wa] F. Warner, *Foundations of Differentiable Manifolds and Lie Groups*, Scott-Foresman, 1971.

[Y] H. Yasuda, *On the indicatrices of a Finsler space*, Tensor, N.S. **33** (1979), 213–221.

[YS] H. Yasuda and H. Shimada, *On Randers spaces of scalar curvature*, Rep. on Math. Phys. **11** (1977), 347–360.

Index

Graduate Texts in Mathematics

(continued from page ii)

62 KARGAPOLOV/MERLZJAKOV. Fundamentals of the Theory of Groups.
63 BOLLOBAS. Graph Theory.
64 EDWARDS. Fourier Series. Vol. I 2nd ed.
65 WELLS. Differential Analysis on Complex Manifolds. 2nd ed.
66 WATERHOUSE. Introduction to Affine Group Schemes.
67 SERRE. Local Fields.
68 WEIDMANN. Linear Operators in Hilbert Spaces.
69 LANG. Cyclotomic Fields II.
70 MASSEY. Singular Homology Theory.
71 FARKAS/KRA. Riemann Surfaces. 2nd ed.
72 STILLWELL. Classical Topology and Combinatorial Group Theory. 2nd ed.
73 HUNGERFORD. Algebra.
74 DAVENPORT. Multiplicative Number Theory. 2nd ed.
75 HOCHSCHILD. Basic Theory of Algebraic Groups and Lie Algebras.
76 IITAKA. Algebraic Geometry.
77 HECKE. Lectures on the Theory of Algebraic Numbers.
78 BURRIS/SANKAPPANAVAR. A Course in Universal Algebra.
79 WALTERS. An Introduction to Ergodic Theory.
80 ROBINSON. A Course in the Theory of Groups. 2nd ed.
81 FORSTER. Lectures on Riemann Surfaces.
82 BOTT/TU. Differential Forms in Algebraic Topology.
83 WASHINGTON. Introduction to Cyclotomic Fields. 2nd ed.
84 IRELAND/ROSEN. A Classical Introduction to Modern Number Theory. 2nd ed.
85 EDWARDS. Fourier Series. Vol. II. 2nd ed.
86 VAN LINT. Introduction to Coding Theory. 2nd ed.
87 BROWN. Cohomology of Groups.
88 PIERCE. Associative Algebras.
89 LANG. Introduction to Algebraic and Abelian Functions. 2nd ed.
90 BRØNDSTED. An Introduction to Convex Polytopes.
91 BEARDON. On the Geometry of Discrete Groups.
92 DIESTEL. Sequences and Series in Banach Spaces.
93 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part I. 2nd ed.
94 WARNER. Foundations of Differentiable Manifolds and Lie Groups.
95 SHIRYAEV. Probability. 2nd ed.
96 CONWAY. A Course in Functional Analysis. 2nd ed.
97 KOBLITZ. Introduction to Elliptic Curves and Modular Forms. 2nd ed.
98 BRÖCKER/TOM DIECK. Representations of Compact Lie Groups.
99 GROVE/BENSON. Finite Reflection Groups. 2nd ed.
100 BERG/CHRISTENSEN/RESSEL. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.
101 EDWARDS. Galois Theory.
102 VARADARAJAN. Lie Groups, Lie Algebras and Their Representations.
103 LANG. Complex Analysis. 3rd ed.
104 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part II.
105 LANG. $SL_2(\mathbf{R})$.
106 SILVERMAN. The Arithmetic of Elliptic Curves.
107 OLVER. Applications of Lie Groups to Differential Equations. 2nd ed.
108 RANGE. Holomorphic Functions and Integral Representations in Several Complex Variables.
109 LEHTO. Univalent Functions and Teichmüller Spaces.
110 LANG. Algebraic Number Theory.
111 HUSEMÖLLER. Elliptic Curves.
112 LANG. Elliptic Functions.
113 KARATZAS/SHREVE. Brownian Motion and Stochastic Calculus. 2nd ed.
114 KOBLITZ. A Course in Number Theory and Cryptography. 2nd ed.
115 BERGER/GOSTIAUX. Differential Geometry: Manifolds, Curves, and Surfaces.
116 KELLEY/SRINIVASAN. Measure and Integral. Vol. I.
117 SERRE. Algebraic Groups and Class Fields.
118 PEDERSEN. Analysis Now.
119 ROTMAN. An Introduction to Algebraic Topology.

120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic *K*-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory. 2nd ed.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.
179 DOUGLAS. Banach Algebra Techniques in Operator Theory. 2nd ed.

180 SRIVASTAVA. A Course on Borel Sets.
181 KRESS. Numerical Analysis.
182 WALTER. Ordinary Differential Equations.
183 MEGGINSON. An Introduction to Banach Space Theory.
184 BOLLOBAS. Modern Graph Theory.
185 COX/LITTLE/O'SHEA. Using Algebraic Geometry.
186 RAMAKRISHNAN/VALENZA. Fourier Analysis on Number Fields.
187 HARRIS/MORRISON. Moduli of Curves.
188 GOLDBLATT. Lectures on the Hyperreals: An Introduction to Nonstandard Analysis.
189 LAM. Lectures on Modules and Rings.
190 ESMONDE/MURTY. Problems in Algebraic Number Theory.
191 LANG. Fundamentals of Differential Geometry.
192 HIRSCH/LACOMBE. Elements of Functional Analysis.
193 COHEN. Advanced Topics in Computational Number Theory.
194 ENGEL/NAGEL. One-Parameter Semigroups for Linear Evolution Equations.
195 NATHANSON. Elementary Methods in Number Theory.
196 OSBORNE. Basic Homological Algebra.
197 EISENBUD/ HARRIS. The Geometry of Schemes.
198 ROBERT. A Course in p-adic Analysis.
199 HEDENMALM/KORENBLUM/ZHU. Theory of Bergman Spaces.
200 BAO/CHERN/SHEN. An Introduction to Riemann–Finsler Geometry.
201 HINDRY/SILVERMAN. Diophantine Geometry.