Translations of
MATHEMATICAL MONOGRAPHS

Volume 191

Methods of Information Geometry

Shun-ichi Amari
Hiroshi Nagaoka

AMS
American Mathematical Society

OXFORD
UNIVERSITY PRESS

Translations of
MATHEMATICAL
MONOGRAPHS

Volume 191

Methods of Information Geometry

Shun-ichi Amari
Hiroshi Nagaoka

Translated by
Daishi Harada

AMERICAN MATHEMATICAL SOCIETY

OXFORD
UNIVERSITY PRESS

情報幾何の方法

JOHO KIKA NO HOHO
(Methods of Information Geometry)
by Shun-ichi Amari and Hiroshi Nagaoka

Originally published in Japanese by Iwanami Shoten, Publishers, Tokyo, 1993

The authors, Oxford University Press, and the American Mathematical Society gratefully acknowledge the financial support provided by the Daido Life Foundation for the editing of this work.

Translated from the Japanese by Daishi Harada

2000 *Mathematics Subject Classification.* Primary 00A69, 53-02, 53B05, 53A15, 62-02, 62F05, 62F12, 93C05, 81Q70, 94A15.

Library of Congress Cataloging-in-Publication Data

Amari, Shun'ichi.
[Joho kika no hoho. English]
Methods of information geometry / Shun-ichi Amari, Hiroshi Nagaoka ; [translated from the Japanese by Daishi Harada].
p. cm. — (Translations of mathematical monographs, ISSN 0065-9282 ; v. 191)
Includes bibliographical references and index.
ISBN 0-8218-0531-2 (alk. paper)
1. Mathematical statistics. 2. Geometry, Differential. I. Nagaoka, Hiroshi, 1955– II. Title. III. Series.

QA276.A56313 2000
519.5—dc21 00-059362

∞ The paper used in this book is acid-free and falls within the guidelines
established to ensure permanence and durability.
Visit the AMS home page at URL: http://www.ams.org/

10 9 8 7 6 5 4 3 2 1 05 04 03 02 01 00

Selected Titles in This Series

191 **Shun-ichi Amari and Hiroshi Nagaoka,** Methods of information geometry, 2000
190 **Alexander N. Starkov,** Dynamical systems on homogeneous spaces, 2000
189 **Mitsuru Ikawa,** Hyperbolic partial differential equations and wave phenomena, 2000
188 **V. V. Buldygin and Yu. V. Kozachenko,** Metric characterization of random variables and random processes, 2000
187 **A. V. Fursikov,** Optimal control of distributed systems. Theory and applications, 2000
186 **Kazuya Kato, Nobushige Kurokawa, and Takeshi Saito,** Number theory 1: Fermat's dream, 2000
185 **Kenji Ueno,** Algebraic Geometry 1: From algebraic varieties to schemes, 1999
184 **A. V. Mel′nikov,** Financial markets, 1999
183 **Hajime Sato,** Algebraic topology: an intuitive approach, 1999
182 **I. S. Krasil′shchik and A. M. Vinogradov, Editors,** Symmetries and conservation laws for differential equations of mathematical physics, 1999
181 **Ya. G. Berkovich and E. M. Zhmud′,** Characters of finite groups. Part 2, 1999
180 **A. A. Milyutin and N. P. Osmolovskii,** Calculus of variations and optimal control, 1998
179 **V. E. Voskresenskiĭ,** Algebraic groups and their birational invariants, 1998
178 **Mitsuo Morimoto,** Analytic functionals on the sphere, 1998
177 **Satoru Igari,** Real analysis—with an introduction to wavelet theory, 1998
176 **L. M. Lerman and Ya. L. Umanskiy,** Four-dimensional integrable Hamiltonian systems with simple singular points (topological aspects), 1998
175 **S. K. Godunov,** Modern aspects of linear algebra, 1998
174 **Ya-Zhe Chen and Lan-Cheng Wu,** Second order elliptic equations and elliptic systems, 1998
173 **Yu. A. Davydov, M. A. Lifshits, and N. V. Smorodina,** Local properties of distributions of stochastic functionals, 1998
172 **Ya. G. Berkovich and E. M. Zhmud′,** Characters of finite groups. Part 1, 1998
171 **E. M. Landis,** Second order equations of elliptic and parabolic type, 1998
170 **Viktor Prasolov and Yuri Solovyev,** Elliptic functions and elliptic integrals, 1997
169 **S. K. Godunov,** Ordinary differential equations with constant coefficient, 1997
168 **Junjiro Noguchi,** Introduction to complex analysis, 1998
167 **Masaya Yamaguti, Masayoshi Hata, and Jun Kigami,** Mathematics of fractals, 1997
166 **Kenji Ueno,** An introduction to algebraic geometry, 1997
165 **V. V. Ishkhanov, B. B. Lur′e, and D. K. Faddeev,** The embedding problem in Galois theory, 1997
164 **E. I. Gordon,** Nonstandard methods in commutative harmonic analysis, 1997
163 **A. Ya. Dorogovtsev, D. S. Silvestrov, A. V. Skorokhod, and M. I. Yadrenko,** Probability theory: Collection of problems, 1997
162 **M. V. Boldin, G. I. Simonova, and Yu. N. Tyurin,** Sign-based methods in linear statistical models, 1997
161 **Michael Blank,** Discreteness and continuity in problems of chaotic dynamics, 1997
160 **V. G. Osmolovskiĭ,** Linear and nonlinear perturbations of the operator div, 1997
159 **S. Ya. Khavinson,** Best approximation by linear superpositions (approximate nomography), 1997
158 **Hideki Omori,** Infinite-dimensional Lie groups, 1997
157 **V. B. Kolmanovskiĭ and L. E. Shaĭkhet,** Control of systems with aftereffect, 1996
156 **V. N. Shevchenko,** Qualitative topics in integer linear programming, 1997

Preface

Information geometry provides the mathematical sciences with a new framework for analysis. This framework is relevant to a wide variety of domains, and it has already been usefully applied to several of these, providing them with a new perspective from which to view the structure of the systems which they investigate. Nevertheless, the development of the field of information geometry can only be said to have just begun.

Information geometry began as an investigation of the natural differential geometric structure possessed by families of probability distributions. As a rather simple example, consider the set S of normal distributions with mean μ and variance σ^2:

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}.$$

By specifying (μ, σ) we determine a particular normal distribution, and hence S may be viewed as a 2-dimensional space (manifold) which has (μ, σ) as a coordinate system. However, this is not a Euclidean space, but rather a Riemannian space with a metric which naturally follows from the underlying properties of probability distributions. In particular, when S is a family of normal distributions, it is a space of constant negative curvature. The underlying characteristics of probability distributions lead not only to this Riemannian structure; an investigation of the structure of probability distributions leads to a new concept within differential geometry: that of mutually dual affine connections. In addition, the structure of dual affine connections naturally arises in the framework of affine differential geometry, and has begun to attract the attention of mathematicians researching differential geometry.

Probability distributions are the fundamental element over which fields such as statistics, stochastic processes, and information theory are developed. Hence not only is the natural dualistic differential geometric structure of the space of probability distributions beautiful, but it must also play a fundamental role in these information sciences. In fact, considering statistical estimation from a differential geometric viewpoint has provided statistics with a new analytic tool which has allowed several previously open problems to be solved; information geometry has already established itself within the field of statistics. In the fields of information theory, stochastic processes, and systems, information geometry is being currently applied to allow the investigation of hitherto unexplored possibilities.

The utility of information geometry, however, is not limited to these fields. It has, for example, been productively applied to areas such as statistical physics and the mathematical theory underlying neural networks. Further, dualistic differential geometric structure is a general concept not inherently tied to probability distributions. For example, the interior method for linear programming may be analyzed from this point of view, and this suggests its relation to completely integrable dynamical systems. Finally, the investigation of the information geometry of quantum systems may lead to even further developments.

This book presents for the first time the entirety of the emerging field of information geometry. To do this requires an understanding of at least the fundamental concepts in differential geometry. Hence the first three chapters contain an introduction to differential geometry and the recently developed theory of dual connections. An attempt has been made to develop the fundamental framework of differential geometry as concisely and intuitively as possible. It is hoped that this book may serve generally as an introduction to differential geometry. Although differential geometry is said to be a difficult field to understand, this is true only of those texts written by mathematicians for other mathematicians, and it is not the case that the principal ideas in differential geometry are hard. Nevertheless, this book introduces only the amount of differential geometry necessary for the remaining chapters, and endeavors to do so in a manner which, while consistent with the conventional definitions in mathematical texts, allows the intuition underlying the concepts to be comprehended most immediately.

On the other hand, a comprehensive treatment of statistics, system theory, and information theory, among others, from the point of view of information geometry is for each distinct, relying on properties unique to that particular theory. It was beyond the scope of this book to include a thorough description of these fields, and inevitably, many of the relevant topics from these areas are rather hastily introduced in the latter half of the book. It is hoped that within these sections the reader will simply gather the flavor of the research being done, and for a more complete analysis refer to the corresponding papers. To complement this approach, many topics which are still incomplete and perhaps consist only of vague ideas have been included.

Nothing would make us happier than if this book could serve as an invitation for other researches to join in the development of information geometry.

Preface to the English Edition

Information geometry provides a new method applicable to various areas including information sciences and physical sciences. It has emerged from investigating the geometrical structures of the manifold of probability distributions, and has applied successfully to statistical inference problems. However, it has been proved that information geometry opens a new paradigm useful for elucidation of information systems, intelligent systems, control systems, physical systems, mathematical systems, and so on.

There have been remarkable progresses recently in information geometry. For example, in the field of neurocomputing, a set of neural networks forms a neuro-manifold. Information geometry has become one of fundamental methods for analyzing neurocomputing and related areas. Its usefulness has also been recognized in multiterminal information theory and portfolio, in nonlinear systems and nonlinear prediction, in mathematical programming, in statistical inference and information theory of quantum mechanical systems, and so on. Its mathematical foundations have also shown a remarkable progress.

In spite of these developments, there were no expository textbooks covering the methods and applications of information geometry except for statistical ones. Although we published a booklet to show the wide scope of information geometry in 1993, it was unfortunately written in Japanese. It is our great pleasure to see its English translation. Mr. Daishi Harada has achieved an excellent work of translation.

In addition to correction of many misprints and errors found in the Japanese edition, we have made revision and rearrangement throughout the manuscript to make it as readable as possible. Also we have added several new topics, and even new sections and a new chapter such as §2.5, §3.2, §3.5, §3.8 and Chapter 7. The bibliography and the guide to it have largely been extended as well. These works were done by the authors after receiving the original translation, and it is the authors, not the translator, who should be responsible for the English writing of these parts.

This is a small booklet, however. We have presented a concise but comprehensive introduction to the mathematical foundation of information geometry in the first three chapters, while the other chapters are devoted to an overview of wide areas of applications. Even though we could not show detailed and comprehensive explanations for many topics, we expect that the readers feel its flavor and prosperity from the description. It is our pleasure if the book would play a key role for further developments of information geometry.

Year 2000

Shun-ichi Amari
Hiroshi Nagaoka

Contents

Chapter 1

Elementary differential geometry

Differential geometry is a mature field of mathematics and has many introductory texts; still, it is not an easy field to master. However, in this book we shall require only the fundamental ideas and methodologies of differential geometry. The main theme of modern differential geometry has been to characterize the global properties of manifolds, and much theory has been developed towards this end. At this time, the field of information geometry (mostly) requires only the theory of the locally characterizable properties of manifolds.

For information geometry the most important aspects of differential geometry are those which allow us to take problems from a variety of fields: statistics, information theory, and control theory; visualize them geometrically; and from this develop novel tools with which to extend and advance these fields. In this chapter we present an introduction to differential geometry from this point of view.

1.1 Differentiable manifolds

A **differentiable manifold** is a mathematical concept denoting a generalization/abstraction of geometric objects such as smooth curves and surfaces in an n-dimensional space. Intuitively, a manifold S is a "set with a coordinate system." Since S is a set, it has elements. It does not matter what these elements are (these elements are also called the **points** of S.) For example, in this book, we shall introduce manifolds whose points are probability distributions and also those whose points are linear systems. S must also have a **coordinate system**. By this we mean a one-to-one mapping from S (or its subset) to $\mathbb{R}^n$, which allows us to specify each point in S using a vector of n real numbers (this vector is called the **coordinates** of the corresponding point). We call the natural number n the **dimension** of S, and write $n = \dim S$.

We call a coordinate system that has S as its domain a global coordinate

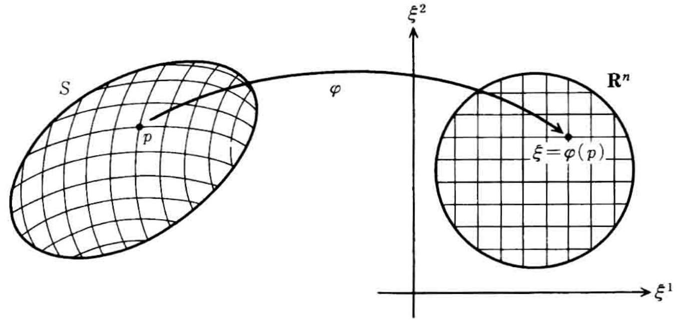

Figure 1.1: A coordinate system for S.

system. In our analysis below, we shall consider only the case where there exists a global coordinate system. However, in general there are many manifolds which do not have global coordinate systems. Examples of such a manifold include the surface of a sphere and the torus (the surface of a donut). These manifolds have only local coordinate systems. This may be viewed informally in the following way. Consider an open subset U of S, and suppose that U has a coordinate system. This provides a local coordinate system for those points contained in U. For a point not contained in U, consider another open subset V containing that point which also has a coordinate system. Repeat this process until the original set S is covered, so that each point in S is contained in an open subset which has a coordinate system. Then this collection of open subsets of S and their corresponding coordinate systems would allow us to express any point in S using coordinates. However, as mentioned above, in this chapter we shall consider only the case when there exists a global coordinate system. This will suffice to prepare us for the later chapters. Indeed, since in this chapter we principally develop the local theory of manifolds, this assumption does not typically affect the generality of the analysis.

Let S be a manifold and $\varphi : S \to \mathbb{R}^n$ be a coordinate system for S. Then φ maps each point p in S to n real numbers: $\varphi(p) = [\xi^1(p), \cdots, \xi^n(p)] = [\xi^1, \cdots, \xi^n]$. These are the coordinates of the point p. Each ξ^i may be viewed as a function $p \to \xi^i(p)$ which maps a point p to its i^{th} coordinate; we call these n functions $\xi^i : S \to \mathbb{R}$ $(i = 1, \cdots, n)$ the **coordinate functions**.[1] We shall write the coordinate system φ in ways such as $\varphi = [\xi^1, \cdots, \xi^n] = [\xi^i]$ (Figure 1.1).

Let $\psi = [\rho^i]$ be another coordinate system for S. Then the same point $p \in S$ has both the coordinates $[\xi^i(p)] = [\xi^i] \in \mathbb{R}^n$ with respect to the coordinate system φ, and the coordinates $[\rho^i(p)] = [\rho^i] \in \mathbb{R}^n$ with respect to the coordinate system ψ. The coordinates $[\rho^i]$ may be obtained from $[\xi^i]$ in the following way. First apply the inverse mapping φ^{-1} to $[\xi^i]$; this gives us a point p in S. Then apply ψ to this point; this result is $[\rho^i]$. In other words, we apply the

[1] We shall use ξ^i, ρ^i to denote both (the variable representing) the i^{th} coordinate of a point and a coordinate function. This is similar to writing "the function $y = y(x)$."

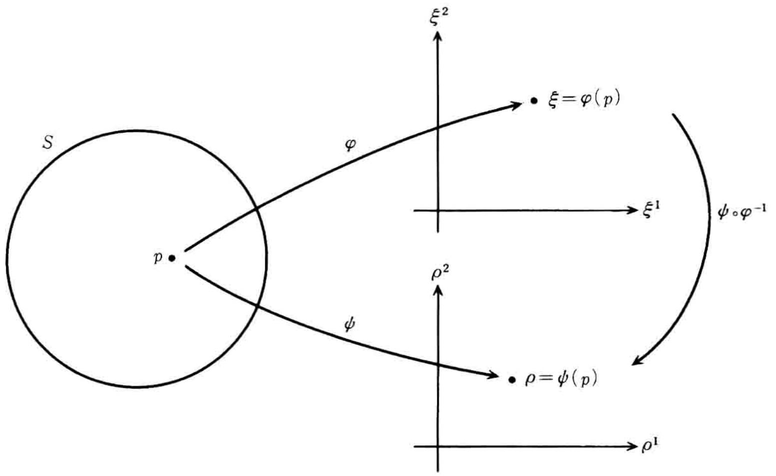

Figure 1.2: Coordinate transformation.

transformation on $\mathbb{R}^n$ given by

$$\psi \circ \varphi^{-1} : [\xi^1, \cdots, \xi^n] \mapsto [\rho^1, \cdots, \rho^n]. \tag{1.1}$$

This is called the coordinate transformation from $\varphi = [\xi^i]$ to $\psi = [\rho^i]$ (Figure 1.2).

To consider S as a manifold means that one is interested in investigating those properties of S which are invariant under coordinate transformations. In particular, differential geometry analyzes the geometry of objects using differential operators with respect to a variety of functions on S, and it would be problematic if these operators depended fundamentally on the choice of coordinates. Hence it is necessary to restrict the coordinate systems to those which allow smooth transformations between each other.

In order to properly formalize the concepts described above, let us now formally define manifolds for which there exists a global coordinate system.

Let S be a set. If there exists a set of coordinate systems $\mathcal{A}$ for S which satisfies the conditions (i) and (ii) below, we call S (more properly, $(S, \mathcal{A})$) an n-dimensional C^∞ **differentiable manifold**, or more simply, a **manifold**.

(i) Each element φ of $\mathcal{A}$ is a one-to-one mapping from S to some open subset of $\mathbb{R}^n$.

(ii) For all $\varphi \in \mathcal{A}$, given any one-to-one mapping ψ from S to $\mathbb{R}^n$, the following holds:

$$\psi \in \mathcal{A} \iff \psi \circ \varphi^{-1} \text{ is a } C^\infty \text{ diffeomorphism.}$$

Here, by a C^∞ **diffeomorphism** we mean that $\psi \circ \varphi^{-1}$ and its inverse $\varphi \circ \psi^{-1}$ are both C^∞ (infinitely many times differentiable). From these conditions, and given the coordinate transformation described in Equation (1.1), it follows that we may take the partial derivative of the function $\rho^i = \rho^i(\xi^1, \cdots, \xi^n)$ with respect to its variable arguments as many times as needed, and that the same holds for $\xi^i = \xi^i(\rho^1, \cdots, \rho^n)$. In this book, the condition C^∞ is used a number of times, but in fact it is usually not necessary; it would suffice for the relevant functions to be differentiable some appropriate number of times. Intuitively, then, we may consider C^∞ to simply mean "sufficiently smooth".

Let S be a manifold and φ be a coordinate system for S. Let U be a subset of S. If the image $\varphi(U)$ is an open subset of $\mathbb{R}^n$, then we say that U is an open subset of S. From condition (ii) above, we see that this property is invariant over the choice of coordinate system φ. This allows us to consider S as a topological space. For any non-empty open subset U of S, we may restrict φ, the coordinate system of S, to obtain $\varphi|_U$ (the mapping $U \to \mathbb{R}^n$ obtained by restricting the domain of φ to U), which may be taken as a coordinate system for U. Hence we see that U is a manifold whose dimension is the same as that of S.

Let $f : S \to \mathbb{R}$ be a function on a manifold S. Then if we select a coordinate system $\varphi = [\xi^i]$ for S, this function may be rewritten as a function of the coordinates; i.e., letting $[\xi^i]$ denote the coordinates of the point p, we have $f(p) = \bar{f}(\xi^1, \cdots, \xi^n)$, where $\bar{f} = f \circ \varphi^{-1}$. Note that $\bar{f}$ is a real-valued function whose domain is $\varphi(S)$, an open subset of $\mathbb{R}^n$. Now suppose that $\bar{f}(\xi^1, \cdots, \xi^n)$ is partially differentiable at each point in $\varphi(S)$. Then the partial derivative $\frac{\partial}{\partial \xi^i} \bar{f}(\xi^1, \cdots, \xi^n)$ is also a function on $\varphi(S)$. By transforming the domain back to S, we may define the partial derivatives of f to be $\frac{\partial f}{\partial \xi^i} \stackrel{\text{def}}{=} \frac{\partial \bar{f}}{\partial \xi^i} \circ \varphi : S \to \mathbb{R}$. We write $\left(\frac{\partial f}{\partial \xi^i}\right)_p$ to denote the value of this function at point p (the partial derivative at point p).

When $\bar{f} = f \circ \varphi^{-1}$ is C^∞, in other words when $\bar{f}(\xi^1, \cdots, \xi^n)$ can be partially differentiated with respect to its variables an unbounded number of times, we call f a $\boldsymbol{C^\infty}$ **function** on S. This definition does not depend on the choice of coordinate system φ. The partial derivatives $\frac{\partial f}{\partial \xi^i}$ of a C^∞ function f are also C^∞ functions. We may similarly define the higher-order partial derivatives, e.g. $\frac{\partial^2 f}{\partial \xi^j \partial \xi^i} = \frac{\partial}{\partial \xi^j} \frac{\partial f}{\partial \xi^i}$. These will also be C^∞. As with the case of C^∞ functions on $\mathbb{R}^n$, $\frac{\partial^2 f}{\partial \xi^j \partial \xi^i} = \frac{\partial}{\partial \xi^j} \frac{\partial f}{\partial \xi^i}$ holds.

Let us denote the class of C^∞ functions on S by $\mathcal{F}(S)$, or simply $\mathcal{F}$. For all f and g in $\mathcal{F}$ and a real number c, we define the sum $f + g$ as $(f + g)(p) = f(p) + g(p)$, the scaling cf as $(cf)(p) = cf(p)$, and the product $f \cdot g$ as $(f \cdot g)(p) = f(p) \cdot g(p)$; these functions are also members of $\mathcal{F}$.

Let $[\xi^i]$ and $[\rho^j]$ be two coordinate systems. Since the coordinate functions ξ^i and ρ^j are clearly C^∞, the partial derivatives $\frac{\partial \xi^i}{\partial \rho^j}$ and $\frac{\partial \rho^j}{\partial \xi^i}$ are well defined, and they satisfy

$$\sum_{j=1}^{n} \frac{\partial \xi^i}{\partial \rho^j} \frac{\partial \rho^j}{\partial \xi^k} = \sum_{j=1}^{n} \frac{\partial \rho^i}{\partial \xi^j} \frac{\partial \xi^j}{\partial \rho^k} = \delta^i_k, \tag{1.2}$$

where δ^i_k is 1 if $k = i$, and 0 otherwise (the Kronecker delta). In addition, for any C^∞ function f, we have

$$\frac{\partial f}{\partial \rho^j} = \sum_{i=1}^{n} \frac{\partial \xi^i}{\partial \rho^j}\frac{\partial f}{\partial \xi^i} \quad \text{and} \quad \frac{\partial f}{\partial \xi^i} = \sum_{j=1}^{n} \frac{\partial \rho^j}{\partial \xi^i}\frac{\partial f}{\partial \rho^j}. \tag{1.3}$$

Note: In this book there often appear equations which contain indices such as $i, j, \cdots$, and are to be summed over those indices that are both super and subscripted. For these equations we shall abbreviate by omitting the summation sign $\sum$ corresponding to these indices. For example, Equations (1.2) and (1.3) above would be written as

$$\frac{\partial \xi^i}{\partial \rho^j}\frac{\partial \rho^j}{\partial \xi^k} = \frac{\partial \rho^i}{\partial \xi^j}\frac{\partial \xi^j}{\partial \rho^k} = \delta^i_k$$

$$\frac{\partial f}{\partial \rho^j} = \frac{\partial \xi^i}{\partial \rho^j}\frac{\partial f}{\partial \xi^i}, \qquad \frac{\partial f}{\partial \xi^i} = \frac{\partial \rho^j}{\partial \xi^i}\frac{\partial f}{\partial \rho^j}.$$

We shall also abbreviate $\sum_{i=1}^n \sum_{j=1}^n A^{ij}_{jk} B^h_i$ as $A^{ij}_{jk} B^h_i$. Hence (unless there is ambiguity), whenever there appears such an equation we shall assume that there is an implicit $\sum$ (i.e., there is a summation over the relevant indices). Note therefore that $A^i_j X^j = A^i_k X^k$, for instance, is always true. This notation is known as Einstein's convention.

Let S and Q be manifolds with coordinate systems $\varphi : S \to \mathbb{R}^n$ and $\psi : Q \to \mathbb{R}^m$. A mapping $\lambda : S \to Q$ is said to be C^∞ or smooth if $\psi \circ \lambda \circ \varphi^{-1}$ is a C^∞ mapping from an open subset of $\mathbb{R}^n$ to $\mathbb{R}^m$. A necessary and sufficient condition for λ to be C^∞ is that $f \circ \lambda \in \mathcal{F}(S)$ for all $f \in \mathcal{F}(Q)$. If a C^∞ mapping λ is a bijection (i.e., one-to-one and $\lambda(S) = Q$) and the inverse λ^{-1} is also C^∞, then λ is called a C^∞ diffeomorphism from S onto Q.

1.2 Tangent vectors and tangent spaces

The tangent space T_p at a point $p \in S$ of a manifold S is intuitively the vector space obtained by "locally linearizing" S around p. Let $[\xi^i]$ be some coordinate system for S, and let $\boldsymbol{e}_i$ denote the "tangent vector" which goes through point p and is parallel to the i^{th} coordinate curve (coordinate axis). By the i^{th} coordinate curve we mean the curve which is obtained by fixing the values of all ξ^j for $j \neq i$ and varying only the value of ξ^i. The n-dimensional space spanned by the n tangent vectors $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n$ is the tangent space T_p at point p (Figure 1.3). Let p' be a point "very close" to p, and let $[\xi^i]$ and $[\xi^i + \mathrm{d}\xi^i]$ (where $\mathrm{d}\xi^i$ is an infinitesimal) be the coordinates of p and p', respectively. Then the segment joining these two points may be described by $\overrightarrow{pp'} = \mathrm{d}\xi^i \boldsymbol{e}_i$, an infinitesimal vector in T_p.

Let us make the above concepts more precise. To do so, we must first formally define what we mean by curves and the tangent vector of curves on a

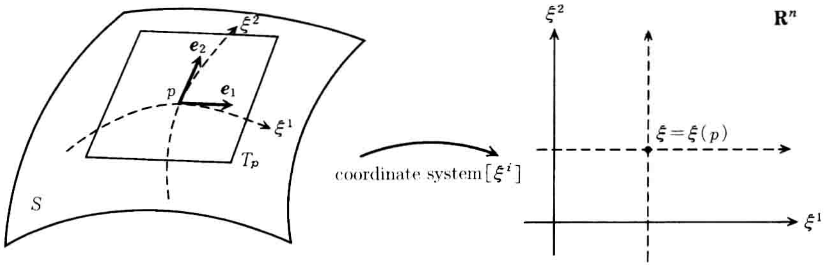

Figure 1.3: Tangent Space

manifold. Consider a one-to-one function $\gamma : I \to S$ from some interval I $(\subset \mathbb{R})$ to S. By defining $\gamma^i(t) \stackrel{\text{def}}{=} \xi^i(\gamma(t))$ we may express the point $\gamma(t)$ $(t \in I)$ using coordinates as $\bar{\gamma}(t) = [\gamma^1(t), \cdots, \gamma^n(t)]$. If $\bar{\gamma}(t)$ is C^∞ for $t \in I$, we call γ a $\boldsymbol{C^\infty}$ **curve** on S. This definition is independent of coordinate system choice.

Now, given a curve γ and a point $\gamma(a) = p$, let us consider what is meant by the "derivative" of γ at p, or alternatively the "tangent vector" $\left(\frac{\mathrm{d}\gamma}{\mathrm{d}t}\right)_p = \dot{\gamma}(a)$. When S is simply an open subset of $\mathbb{R}^n$, or can be embedded smoothly into $\mathbb{R}^\ell$ $(\ell \geq n)$, the range of γ is contained within a single linear space, and hence it suffices to consider the standard derivative

$$\dot{\gamma}(a) = \lim_{h\to 0} \frac{\gamma(a+h) - \gamma(a)}{h}. \tag{1.4}$$

In general, however, the equation above is not meaningful. On the other hand, if we take a C^∞ function $f \in \mathcal{F}$ on S and consider the value of $f(\gamma(t))$ on the curve, since this is a real-valued function, we may define the derivative $\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))$ in the usual way. Using coordinates, we have $f(\gamma(t)) = \bar{f}(\bar{\gamma}(t)) = \bar{f}(\gamma^1(t), \cdots, \gamma^n(t))$, and the derivatives may be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t)) = \left(\frac{\partial \bar{f}}{\partial \xi^i}\right)_{\bar{\gamma}(t)} \frac{\mathrm{d}\gamma^i(t)}{\mathrm{d}t} = \left(\frac{\partial f}{\partial \xi^i}\right)_{\gamma(t)} \frac{\mathrm{d}\gamma^i(t)}{\mathrm{d}t}. \tag{1.5}$$

We call this the directional derivative of f along the curve γ. Let us consider this directional derivative as an expression of the tangent vector of γ. In other words, we take the operator : $\mathcal{F} \to \mathbb{R}$ which maps $f \in \mathcal{F}$ to $\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))|_{t=a}$, and simply define the tangent vector $\left(\frac{\mathrm{d}\gamma}{\mathrm{d}t}\right)_p = \dot{\gamma}(a)$ to be this operator. Then we may rewrite Equation (1.5) as

$$\dot{\gamma}(a) = \left(\frac{\mathrm{d}\gamma}{\mathrm{d}t}\right)_p = \dot{\gamma}^i(a) \left(\frac{\partial}{\partial \xi^i}\right)_p \tag{1.6}$$

$(\dot{\gamma}^i(a) = \frac{\mathrm{d}}{\mathrm{d}t}\gamma^i(t)|_{t=a})$. Here $\left(\frac{\partial}{\partial \xi^i}\right)_p$ is an operator which maps $f \mapsto \left(\frac{\partial f}{\partial \xi^i}\right)_p$. It is possible to show that when the tangent vectors can be defined using Equation (1.4), there is a natural one-to-one correspondence between Equations (1.4) and (1.6). Hence the definition of tangent vectors as operators may be viewed as a generalization of Equation (1.4).

Since a partial derivative is simply a directional derivative along a coordinate axis, the operator $\left(\frac{\partial}{\partial \xi^i}\right)_p$ is the tangent vector at point p of the i^{th} coordinate curve. The $\boldsymbol{e}_i$ mentioned previously corresponds to this $\left(\frac{\partial}{\partial \xi^i}\right)_p$. From Equation (1.3), we see that

$$\left(\frac{\partial}{\partial \rho^j}\right)_p = \left(\frac{\partial \xi^i}{\partial \rho^j}\right)_p \left(\frac{\partial}{\partial \xi^i}\right)_p \quad \text{and} \quad \left(\frac{\partial}{\partial \xi^i}\right)_p = \left(\frac{\partial \rho^j}{\partial \xi^i}\right)_p \left(\frac{\partial}{\partial \rho^j}\right)_p. \tag{1.7}$$

Consider all curves which pass through the point p. We denote the set of all tangent vectors corresponding to these curves by T_p, or $T_p(S)$. From Equation (1.6), we see that

$$T_p(S) = \left\{ c^i \left(\frac{\partial}{\partial \xi^i}\right)_p \middle| [c^1, \cdots, c^n] \in \mathbb{R}^n \right\}. \tag{1.8}$$

This forms a linear space, and since the operators $\left\{\left(\frac{\partial}{\partial \xi^i}\right)_p ; i = 1, \cdots, n\right\}$ are clearly linearly independent, the dimension of this space is n ($= \dim S$). We call $T_p(S)$ and its elements the **tangent space** and **tangent vectors**, of S at the point p , respectively. In addition, we call $\left(\frac{\partial}{\partial \xi^i}\right)_p$ the **natural basis** of the coordinate system $[\xi^i]$.

Let $D \in T_p$ be some tangent vector. Then for all $f, g \in \mathcal{F}$ and all $a, b \in \mathbb{R}$, D satisfies the following:

$$\text{[Linearity]} \qquad D(af + bg) = aD(f) + bD(g). \tag{1.9}$$

$$\text{[Leibniz's rule]} \qquad D(f \cdot g) = f(p)D(g) + g(p)D(f). \tag{1.10}$$

Conversely, it can be shown that any operator $D : \mathcal{F} \to \mathbb{R}$ satisfying these properties is an element of T_p. Hence, it is possible to define tangent vectors in terms of these properties.

Let $\lambda : S \to Q$ be a smooth mapping from a manifold S to another manifold Q. Given a tangent vector $D \in T_p(S)$ of S, the mapping $D' : \mathcal{F}(Q) \to \mathbb{R}$ defined by $D'(f) = D(f \circ \lambda)$ satisfies Equations (1.9) (1.10) with p replaced with $\lambda(p)$, and hence D' belongs to $T_{\lambda(p)}(Q)$. Representing this correspondence as $D' = (\mathrm{d}\lambda)_p(D)$, we may define a linear mapping $(\mathrm{d}\lambda)_p : T_p(S) \to T_{\lambda(p)}(Q)$, which is called the **differential** of λ at p. When S and Q are provided with coordinate systems $[\xi^i]$ and $[\rho^j]$ respectively, we have

$$(\mathrm{d}\lambda)_p(\left(\frac{\partial}{\partial \xi^i}\right)_p) = \left(\frac{\partial(\rho^j \circ \lambda)}{\partial \xi^i}\right)_p \left(\frac{\partial}{\partial \rho^j}\right)_{\lambda(p)}. \tag{1.11}$$

Moreover, for any curve $\gamma(t)$ on S passing through the point p it follows that

$$(\mathrm{d}\lambda)_p(\left(\frac{\mathrm{d}\gamma}{\mathrm{d}t}\right)_p) = \left(\frac{\mathrm{d}(\lambda \circ \gamma)}{\mathrm{d}t}\right)_{\lambda(p)}. \tag{1.12}$$