信息几何导引

An Elementary Introduction to Information Geometry

孙华飞　张真宁　彭林玉　段晓敏　著

科学出版社

信息几何导引

An Elementary Introduction to Information Geometry

孙华飞　张真宁　彭林玉　段晓敏　著

科学出版社
北京

内 容 简 介

本书简要介绍经典信息几何与矩阵信息几何的基本内容及其应用. 全书共八章：第 1 章概述信息的发展历史；第 2 章简要介绍作为信息几何理论基础的微分几何的基本内容，没有涉及太多复杂的定义；第 3 章介绍经典信息的基本内容；第 4 章介绍矩阵信息几何，着重介绍相关的李群、李代数以及一般线性群的重要子群和子流形的性质，而且介绍各种流形上的自然梯度算法；第 5~7 章介绍经典信息几何的应用；第 8 章介绍矩阵信息几何的应用.

本书可供从事数学、信息科学研究的研究生以及教师参考使用.

图书在版编目(CIP)数据

信息几何导引/孙华飞等著. —北京：科学出版社, 2016. 3

ISBN 978-7-03-047435-3

Ⅰ. ①信…　Ⅱ. ①孙…　Ⅲ. ①微分几何　Ⅳ. ①O186.1

中国版本图书馆 CIP 数据核字(2016) 第 046244 号

责任编辑：李　欣 / 责任校对：钟　洋
责任印制：张　倩 / 封面设计：陈　敬

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

三河市骏杰印刷有限公司 印刷
科学出版社发行　各地新华书店经销
*
2016 年 3 月第　一　版　开本：720 × 1000 1/16
2016 年 3 月第一次印刷　印张：10
字数：194 000

定价：58.00 元

(如有印装质量问题，我社负责调换)

前　言

近年来, 信息几何在统计推断、神经网络、随机分布控制、信息理论、密码学、物理学和医学成像等领域得到广泛应用, 引起越来越多人的关注. 经典信息几何在处理随机的非线性问题时已经获得了很大的成功, 成为解决信息领域中各种问题的重要工具之一. 最近, 矩阵信息几何的诞生极大地丰富了信息几何的内容, 其中李群理论发挥了重要的作用. 信息几何的理论基础是现代的微分几何理论, 涉及诸多深刻的数学分支. 我们撰写本书的宗旨是希望有更多人来关注信息几何, 希望利用它来有效地解决信息领域的问题, 诸如信号处理、图像处理、系统的稳定性与最优控制、流形上的优化等问题. 同时, 我们也期待人们能够将纤维丛、代数拓扑等深刻的数学理论引入到信息几何研究中来, 给信息几何的发展带来新的方法, 为信息几何的应用提供强有力的工具. 本书分以下几个部分: 首先概述信息几何的内容, 然后介绍微分几何的基本内容, 接下来介绍经典信息几何和矩阵信息几何的内容, 最后介绍信息几何理论的一些应用.

本书并不追求严格的数学定义, 也没有概括信息几何的全貌, 有兴趣的读者可以阅读各章所附的参考文献. 建议侧重于应用研究的读者可以不必拘泥于一些抽象的数学概念, 可以直接使用已有的理论来解决实际问题, 而侧重于理论研究的读者可以仔细阅读参考文献的内容, 力求在理论研究方面有所创新. 本书的主要内容来源于作者在北京理工大学授课的讲义, 包含北京理工大学信息几何研究组的部分研究成果.

本书的出版获得北京理工大学数学与统计学院的资助. 作者感谢国家自然科学基金委员会的大力资助 (资助号: 61179031, 10871218, 61401058, 10932002).

由于作者水平所限, 书中不当之处在所难免, 恳请读者批评指正.

作　者

2015 年 10 月

目　　录

第 1 章　信息几何概述

信息几何是利用微分几何的方法来研究信息领域中的问题的学问. 几何方法在处理非线性问题时往往发挥重要的作用, 例如, 利用微分几何建立广义相对论, 以及利用微分几何把仿射非线性系统精确线性化等都是很好的例子. 对于非线性问题, 如果一味地线性化有时达不到所需的精度, 所以可以考虑利用几何的方法来解决.

人们要研究的目标本身可能是非线性的, 但是作为微分流形, 其每一点的切空间都是线性的, 可以像在欧氏空间那样在切空间上定义内积, 从而获得需要的度量. 在经典信息几何理论中, Rao 把概率密度函数全体看成统计流形, 并用 Fisher 信息阵来定义流形上的黎曼度量, 从而构建了黎曼流形. Amari 计算了正态分布流形在黎曼联络下的黎曼曲率, 惊奇地发现它是带有负常曲率的双曲空间. 既然概率分布全体是弯曲的流形, 人们就设法研究各种概率分布的几何性质, 并希望利用这些几何性质来研究各种随机问题.

流形的几何性质取决于所选取的几何度量及其联络 (求导数的方式). 保持无挠性和相容性的黎曼联络在微分几何理论中是最理想的联络, 但是在经典信息几何中却“好得过火”, 不太容易派上用场. 于是人们设法定义新的联络来代替黎曼联络, 获得“没那么好, 但很有用”的联络. Chentsov 引入了一族仿射联络, Efron 给出了统计流形上的曲率. Amari 引入了对偶联络的概念, 这个概念是经典微分几何中所不具有的新内容. 利用这个对偶联络, 学者们已经获得了很多新成果. 这种对偶联络本身既没有无挠性, 与黎曼度量之间也没有相容性, 对信息几何的研究没有直接的贡献, 而 Amari 由此提出的 α-联络却是神来之笔, 因为 α-联络与 $-\alpha$-联络是对偶联络, 它们保证了 α-联络的无挠性, 这对问题的研究带来了极大的方便. 众所周知, 爱因斯坦的广义相对论的能量-动量方程就基于无挠性的假设. 那么, 引入了无挠的 α-联络究竟有什么好处呢? 指数分布族和混合分布族是两大重要的分布族, 它们包含了许多已知的重要的概率分布. 利用 α-联络可以计算指数分布族

的几何量, 计算结果表明: 指数分布族的几何量由实数 α 和势函数完全确定, 特别地, 当 $\alpha = \pm 1$ 时, 指数分布族流形的曲率和挠率都为零, 也就是说该流形为既不扭曲, 也不弯曲的平坦的流形, 但却不是欧氏空间, 因为此时联络与度量之间没有相容性.

在 ± 1-平坦的流形上, 可以找到对偶势函数, 于是就可以引入散度作为距离函数, 来测量流形上两点之间的差异. 在信息几何中, Kullback-Leibler 散度经常被用来测量统计流形上两点的差异. 虽然该散度只满足距离函数的非负性, 不满足对称性和三角不等式, 但该散度在研究具有统计特性的信息问题时却非常好用.

在实际问题中经常需要求目标函数的极值. 在欧氏空间中求目标函数的最小值通常采用最速下降法, 而在弯曲的黎曼流形中最速下降方向是由自然梯度给出的. 要得到自然梯度需借助于所研究的黎曼流形的几何结构, 并可将自然梯度视为欧氏空间中的最速下降方向在黎曼流形中的推广.

上述的这些理论被成功地应用于统计推断、神经网络、信号处理、纠错码、量子理论、控制理论等领域. 因此, 利用这些理论解决信息领域中的问题时, 首先需要把所研究的问题构建成微分流形, 定义适当的黎曼度量, 选择适当的距离函数, 给出合适的算法. 迄今为止, 除了 $\alpha = 0$ 对应着黎曼联络外, 仅仅当 $\alpha = \pm 1$ 时获得了广泛的应用. 另外, 既然完备、单连通及带有常曲率的流形等距于欧氏空间、球面或者双曲空间其中之一, 而且 Amari 已经得到一元正态分布流形等距于双曲空间, 我们希望分类统计流形, 看看有多少统计流形是带有常曲率的. 但这个问题至今没有解决, 即便是对于比较特殊的指数分布族也没有得到完整的分类.

信息几何发展的时间较短, 它的理论和应用研究在许多方面仍然处于起步阶段. 毫无疑问, 几何上的许多重要成果在信息几何理论中都应该有对应的结果, 其应用的范围还远远不够广泛. 迄今为止, 人们在研究信息几何的过程中主要使用测地线、测地距离等概念, 而几何的其他核心概念, 比如曲率, 却仅仅用于研究 Jacobi 场的稳定性等, 在许多实际问题的研究中曲率并没有派上用场. 例如, 当人们利用几何研究系统的稳定性时, 最希望得到的结论是: 当系统所构成的流形的曲率满足某种条件时系统就是稳定的. 但遗憾的是, 至今人们仍得不到这种理想的结论, 期望有人能够做到这一点. 另外, 利用建立在纤维丛上的和乐群理论, 我们已经获得

了对指数分布族的一种分类, 特别是给多元正态分布族一个完全的分类, 对于其他的统计流形还没有结果. 我们希望微分几何中深刻而优美的理论都能够在信息几何的研究中发挥作用.

我们称上述内容为经典信息几何的范畴. 最近 Barbaresco, Nielsen, Pennec 等提出了矩阵信息几何的概念, 主要用来研究雷达信号处理、流形学习、系统的稳定性与最优化、图像处理等问题. 特别是一般线性群的子群, 如正交群、酉群、特殊辛群、特殊欧几里得群, 以及一般线性群的子流形, 如正定矩阵流形、Stiefel 流形以及 Grassman 流形等在信息领域中都有重要的应用. 其中, 左不变 (或右不变) 度量作为黎曼度量被采用, 测地距离被用于定义距离函数, 更重要的是测地线能够通过流形上的指数映射和对数映射有一个显式的表达, 这在应用上很方便.

对于紧李群, 比如正交群 $SO(n)$、酉群 $U(n)$ 以及特殊辛群 $Sp(n,\mathbb{R})$, 它们上面任意两点都可以用测地线来连接, 利用测地线的测地距离可以测量其上两点间的距离. 紧李群上存在双不变的黎曼度量, 这使得紧李群的截面曲率是非负的.

由 n 阶正定矩阵全体构成的正定矩阵流形 $SPD(n)$, 其切空间为 n 阶对称矩阵全体 $Sym(n)$. 首先在其上可以定义 Frobenius 内积, 但是此内积在实际应用中有时会有一定的局限性. 同时, 定义仿射不变度量, 可以使 $SPD(n)$ 构成完备的 Hadamard 空间, 其上面的指数映射和对数映射是可逆的映射. $SPD(n)$ 上任意两点可以用测地线连接, 由该测地线可以获得由特征值表示的测地距离. 在研究线性系统稳定性和最优控制等问题时, 经常需要在 $SPD(n)$ 上求解 Lyapunov 方程或代数 Riccati 方程, 利用测地距离和自然梯度我们给出了求方程数值解的自然梯度算法. 另一方面, 为了使计算变得简单, 可以在 $SPD(n)$ 上定义对数欧氏度量, 通过引入矩阵乘法运算使得 $SPD(n)$ 具有群的结构, 在 $SPD(n)$ 和它的切空间之间可以建立等距关系, 使得 $SPD(n)$ 是平坦的空间, 此时测地线也变得简单. 对于几何平均问题, 在一般的流形上较难获得唯一解. 但是由于 $SPD(n)$ 在引入仿射黎曼度量后是 Hadamard 空间, 在其上求几何平均可以获得唯一解.

利用矩阵信息几何, 有时可将所研究的问题转化为矩阵群或矩阵流形上的优化问题. 具体地, 可以用自然梯度给出解决方案, 其好处就是自然梯度中已不再需要求度量矩阵的逆矩阵了. 特别值得一提的是, Fiori 提出了广义 Hamilton 算法, 在自然梯度算法的基础上加了动量项, 使得在求黎曼流形特别是在 $SPD(n)$, Stiefel 流

形以及 Grassman 流形上的优化问题时避免了局部最小的现象, 并且利用技巧在计算关键的测地线时免去了计算黎曼度量的烦琐过程, 使得计算变得简单.

最近, Ollivier 等开展了对离散空间的曲率的研究, 类似于在微分流形上的定义, 给出了网络的曲率的定义, 为信息几何的研究开创了全新的领域.

许多深刻的数学理论可以在信息几何理论研究中充分发挥作用. 纤维丛是微分几何的重要内容, 其在数学、物理以及其他学科中发挥着重要的作用, 几何学大师陈省身正是利用纤维丛给出了 Gauss-Bonnet 定理的内蕴证明方法, 这一创新引发了诺贝尔物理学奖得主杨振宁和物理学家 Mills 的杨–米尔斯规范场理论的微分几何版的诞生. 纤维丛的概念比较抽象和复杂, 涉及诸多数学领域的内容, 掌握起来比较困难, 但是它具有一般微分流形所不具有的“居高临下”的优势, 即带有附加的李群结构等, 正是这些附加的群与拓扑兼备的李群结构使得纤维丛的功能变得异常强大. 纤维丛的引入会给信息几何的发展带来新的活力, 可以进一步拓宽其应用的范围. 例如, 可以利用建立在切丛上的和乐群对统计流形进行分类; 可以在丛空间上计算统计流形的曲率等几何量, 从本质上克服了在底空间上计算的复杂性. 这些都受益于作用在丛空间上的李群结构的优良性质.

另外, 我们的研究也可以在无穷维空间上进行. 利用算子的分解建立主纤维丛结构, 其中的结构群是酉算子或正交算子等. 可以给出相应的纤维丛的切空间的分解, 其中的水平子空间与底空间的切空间同构, 竖直子空间与结构群的李代数同构. 同时利用复 Hilbert 空间上的内积可以给出类似于有限维空间上的 Finsler 度量和测地距离等表达式.

随着信息领域各种实际应用的需求, 与之相适应的信息几何理论也会不断产生, 从而数学各学科的内容将更有用武之地. 信息几何是一个综合数学各分支以及信息学科知识的学问, 研究者如果掌握了数学中的几何、代数、方程、拓扑、分析、计算、概率统计等内容, 又通晓信息学科的内容诸如信息论、信号处理、神经网络、图像处理、控制理论等知识, 对从事信息几何的研究将很有益处.

参 考 文 献

[1] Amari S. Differential-Geometrical Methods in Statistics, Lecture Notes in Statistics. Berlin: Springer-Verlag, 1985.

[2] Amari S, Nagaoka H. Methods of Information Geometry, TMM 191. Oxford: Oxford University Press, 2000.

[3] Amari S. Information geometry of the EM and em algorithm for neural networks. Neural Networks, 1995, 8: 1379–1408.

[4] Amari S. Differential geometry of a parametric family of invertible linear systems–Riemannian metric, dual affine connections, and divergence. Math. Systems Theory, 1987, 20: 53–82.

[5] Amari S. Differential geometry of curved exponential families-curvatures and information loss. Ann. Statist., 1982, 10: 357–385.

[6] Amari S. Fisher information under restriction of Shannon information. Ann. Inst. Statist. Math., 1989, 41: 623–648.

[7] Amari S. Information geometry on hierarchy of probability distributions. IEEE Trans. Inf. Theory, 2001, 47: 1701–1711.

[8] Amari S. Natural gradient works efficiently in learning. Neural Comput., 1998, 10: 251–276.

[9] Amari S. Superefficiency in blind source separation. IEEE Trans. Signal Process, 1999, 47: 936–944.

[10] Amari S, Cardoso J. Blind source separation-semiparametric statistical approach. IEEE Trans. Signal Process., 1997, 45: 2692–2700.

[11] Amari S, Cichocki A. Information geometry of divergence functions. Bull. Pol. Acad. Sci., Tech. Sci., 2010, 58: 183–195.

[12] Amari S, Douglas S C. Why natural gradient. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, 1998, 2: 1213–1216.

[13] Amari S, Kawanabe M. Information geometry of estimating functions in semiparametric statistical models. Bernoulli, 1997, 3: 29–54.

[14] Amari S, Kurata K, Nagaoka H. Information geometry of Boltzmann machines. IEEE Trans. Neural Netw., 1992, 3: 260–271.

[15] Amari S, Park H, Fukumizu K. Adaptive method of realizing natural gradient learning for multilayer perceptrons. Neural Comput., 2000, 12: 1399–1409.

[16] Amari S. Natural gradient works effeciently in learning. Neural Comput., 1998, 10: 251–276.

[17] Arsigny V, Fillard P, Pennec X, et al. Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl., 2007, 29: 328–347.

[18] Arwini K, Dodson C T J. Neighbourhoods of randomness and geometry of McKay bivariate gamma 3-manifold. Sankhya: The Indian Journal of Statistics, 2004, 66: 213–233.

[19] Arwini K, Dodson C T J. Information Geometry: Near Randomness and Near Independence. Berlin Heidelberg: Springer-Verlag, 2008.

[20] Barbaresco F. Interactions between symmetric cones and information geometrics: Bruhat-Tits and Siegel spaces models for high resolution autoregressive Doppler imagery. ETCV08 Conference, Ecole Polytechnique, Nov. 2008, published by Springer in Lecture Notes in Computer Science, 2009, 5416: 124–163.

[21] Barbaresco F, Roussigny H. Innovative tools for Radar signal processing based on Cartan's geometry of SPD matrices and information geometry. IEEE International Radar Conference, 2008.

[22] Bregman L M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. USSR Comp. Math. & Math. Phys., 1967, 7: 200–217.

[23] Cafaro C, Ali S A. Jacobi fields on statistical manifolds of negative curvature. Physica D, 2007, 234: 70–80.

[24] Censor Y, Iusem A N, Zenios S A. An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. Math. Program., 1998, 81: 373–400.

[25] Cheng Y, Wang X, Caelli T, et al. Tracking and localizing moving targets in the presence of phase measurement ambiguities. IEEE Trans. Signal Process., 2011, 59: 3514–3525.

[26] Cheng Y, Wang X, Morelande M, et al. Information geometry of target tracking sensor networks. Inform. Fusion, 2013, 14: 311–326.

[27] Efron B. Defining the curvature of a statistical problem. Ann. Stat., 1975, 3: 1189–1242.

[28] Efron B. The geometry of exponential families. Ann. Stat., 1978, 6: 362–376.

[29] Fiori S. A theory for learning by weight flow on Stiefel-Grassman manifold. Neural

Comput., 2001, 13: 1625–1647.

[30] Fiori S, Tanaka T. An algorithm to compute averages on matrix Lie groups. IEEE Trans. Signal Process., 2009, 57: 4734–4743.

[31] Fiori S. Extended Hamiltonian learning on Riemannian manifolds: theoretical aspects. IEEE Trans. Neural Netw., 2011, 22: 687–700.

[32] Fiori S. Extended Hamiltonian learning on Riemannian manifolds: numerical aspects. IEEE Trans. Neural Netw. Learn. Syst., 2012, 23: 7–21.

[33] Fletcher P, Joshi S. Riemannian geometry for the statistical analysis of diffusion tensor data. Signal Process., 2007, 87: 250–262.

[34] Jeffreys H. Theory of Probability. 3rd ed. Oxford: Clarendon Press, 1961.

[35] Karcher H. Riemannian center of mass and mollifier smoothing. Comm. Pure Appl. Math., 1977, 30: 509–541.

[36] Moakher M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl., 2005, 26: 735–747.

[37] Moakher M. On the averaging of symmetric positive-definite tensors. J. Elasticity, 2006, 82: 273–296.

[38] Nielsen F, Bhatia R. Matrix Information Geometry. Berlin: Springer, 2013.

[39] Ohara A, Amari S. Differential geometric structures of stable state feedback systems with dual connections. Kybernetika, 1994, 30: 369–386.

[40] Ollivier Y. Ricci curvature of Markov chains on metric spaces. J. Funct. Anal., 2009, 256: 810–864.

[41] Qi M, Li B, Sun H. Image watermarking using polar harmonic transform with parameters in $SL(2, R)$. Signal Process-Image, 2015, 31: 161–173.

[42] Qi M, Li B, Sun H. Image watermarking via fractional polar harmonic transforms. J. Electron. Imaging, 2015, 24: 013004.

[43] Rao C R. Information and accuracy attainable in the estimation of statistical parameters. Bull. Calcutta. Math. Soc., 1945, 37: 81–91.

[44] Sandhu R, Georgiou T, Reznik E, et al. Graph curvature for differentiating cancer networks. Sci. Rep., 2015, 5: 12323.

[45] Skovgaard L T. A Riemannian geometry of the multivariate normal model. Scand. J. Stat., 1984, 11: 211–223.

[46] Yang C N, Mills R. Conservation of isotopic spin and isotopic gauge invariance. Phys. Rev., 1954, 96: 191–195.

[47] 甘利俊一, 長岡浩司. 情報幾何の方法. 东京: 岩波書店, 1993.

[48] 黎湘, 程永强, 王宏强, 秦玉亮. 雷达信号处理的信息几何方法. 北京: 科学出版社, 2014.

[49] 罗四维. 大规模人工神经网络理论基础. 北京: 清华大学出版社, 2005.

[50] Amari S. 信息几何. 孙华飞, 译. 数学译林, 2003, 2: 112–120.

[51] 孙华飞, 彭林玉, 张真宁. 信息几何及其应用. 数学进展, 2011, 40: 257–269.

[52] 陶然, 邓兵, 王越. 分数阶傅里叶变换及其应用. 北京: 清华大学出版社, 2009.

[53] 韦博成. 统计推断与微分几何. 北京: 清华大学出版社, 1988.

[54] 许天周, 李炳照. 线性正则变换及其应用. 北京: 科学出版社, 2013.

[55] 张贤达. 矩阵分析与应用. 北京: 清华大学出版社, 2004.

第 2 章　微分几何基础

本章将简要地介绍信息几何中所涉及的微分几何的基础内容, 细节可以参照本章后面所附的参考文献.

对于三维欧氏空间中的曲线和曲面, 情况相对简单, 曲线的曲率和挠率分别描述曲线在一点附近的弯曲程度和扭曲程度, 例如, 直线的曲率和挠率都为零, 单位圆是均匀弯曲的曲线, 其曲率是 1. 曲率和挠率在刚体运动下是不变的. 高斯曲率是描述曲面一点附近弯曲度的几何量, 它仅仅依赖于曲面自身的度量, 与曲面位于的欧氏空间没有关系. 例如, 平面的高斯曲率为零, 二维单位球面的高斯曲率是 1, 而单叶双曲面的高斯曲率是负的. 作为曲线和曲面的推广, 人们引入微分流形的概念. 微分流形对一些读者来说也许是比较陌生的, 而且其严格定义更是要花费一番气力才能深刻地理解. 粗略地讲, 所谓微分流形就是局部上可以安装一个欧氏空间的这样一个空间, 其上面具有拓扑结构和微分结构, 可以进行求导和积分运算等, 有兴趣的读者可参照本章后面的参考文献系统地学习. 例如, 欧氏空间 $\mathbb{R}^n$ 本身就是最简单的微分流形, 半径为 1 的球面

$$S^n(1) = \left\{x = (x_1, x_2, \cdots, x_{n+1}) \mid x_1^2 + x_2^2 + \cdots + x_{n+1}^2 = 1\right\},$$

半径为 1 的开球

$$B^n(1) = \left\{x = (x_1, x_2, \cdots, x_n) \mid x_1^2 + x_2^2 + \cdots + x_n^2 < 1\right\},$$

以及一般线性群

$$GL(n, \mathbb{R}) = \left\{A \in \mathbb{R}^{n \times n} \mid \det(A) \neq 0\right\}$$

都是微分流形. 特别地, 正态分布的全体所构成的集合

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

在一定条件下也是一个微分流形, 该流形在信息领域中发挥着重要的作用. 以后我们要研究的流形都假设是光滑的.

设 M 为 n 维微分流形, T_pM 表示流形 M 在 p 处的切空间. 切空间的无交并为切丛 $TM=\bigcup_{p\in M}T_pM$. M 上的光滑向量场为切丛的光滑截面, 记为 $X\in\mathfrak{X}(M)$. 局部地, 对于流形 M 上的任一点 p, $X(p)\in T_pM$. 类似于在欧氏空间上求函数的方向导数, 在流形 M 上有相应的运算 ∇ 称为联络或协变导数:

$$\begin{aligned}\nabla:\mathfrak{X}(M)\times\mathfrak{X}(M)&\to\mathfrak{X}(M),\\ (X,Y)&\mapsto\nabla_XY\in\mathfrak{X}(M),\end{aligned}$$

满足以下条件:

$$\begin{aligned}&\nabla_X(Y+Z)=\nabla_XY+\nabla_XZ,\\ &\nabla_{X+Y}Z=\nabla_XZ+\nabla_YZ,\\ &\nabla_{fX}Y=f\nabla_XY,\\ &\nabla_X(fY)=X(f)Y+f\nabla_XY,\end{aligned}$$

其中 $X,Y,Z\in\mathfrak{X}(M)$, $f\in C^\infty(M)$ 为光滑函数. 那么, 用局部坐标如何来表示 ∇_XY 呢? 对于 n 维流形 M 上的一点 p 的一个邻域, 存在坐标系 $\{x^i\},i=1,2,\cdots,n$, 则切空间可以表示为

$$T_pM=\operatorname{span}\left\{\frac{\partial}{\partial x^1},\frac{\partial}{\partial x^2},\cdots,\frac{\partial}{\partial x^n}\right\},$$

其中 $\left\{\dfrac{\partial}{\partial x^i}\right\}_{i=1}^n$ 称为 T_pM 的自然基底. 设 $X=X^i\dfrac{\partial}{\partial x^i},Y=Y^j\dfrac{\partial}{\partial x^j}$, 利用联络的性质, 则有

$$\begin{aligned}\nabla_XY&=\nabla_{\left(X^i\frac{\partial}{\partial x^i}\right)}\left(Y^j\frac{\partial}{\partial x^j}\right)\\ &=X^i\left(\nabla_{\frac{\partial}{\partial x^i}}\left(Y^j\frac{\partial}{\partial x^j}\right)\right)\\ &=X^i\left(\frac{\partial Y^j}{\partial x^i}\frac{\partial}{\partial x^j}+Y^j\nabla_{\frac{\partial}{\partial x^i}}\frac{\partial}{\partial x^j}\right)\\ &=X^i\left(\frac{\partial Y^k}{\partial x^i}+Y^j\Gamma_{ij}^k\right)\frac{\partial}{\partial x^k},\end{aligned}$$

其中 $\nabla_{\frac{\partial}{\partial x^i}}\dfrac{\partial}{\partial x^j}=\Gamma_{ij}^k\dfrac{\partial}{\partial x^k}$, 称 Γ_{ij}^k 为联络系数.

注2.1　在本书中爱因斯坦求和约定广泛使用.

定义2.1　设 $f:M\to N$ 是两个光滑流形间的光滑映射, 则称

$$f_{*p}:T_pM\to T_{f(p)}N$$

为切空间 T_pM 到切空间 $T_{f(p)}N$ 的切映射. 如果 f_* 是单射, 称 f 为浸入. 如果 f_* 是满射, 称 f 为淹没. 如果 f 是同胚的浸入, 则称 f 为嵌入.

定义2.2　定义黎曼度量

$$g:\mathfrak{X}(M)\times\mathfrak{X}(M)\to\mathbb{R}$$

为满足对称性和正定性的双线性函数, 即对于任意的 $X,Y\in\mathfrak{X}(M)$, 有

$$g(X,Y)=g(Y,X),$$
$$g(X,X)\geqslant 0,$$

其中第二式中等号成立的充要条件是 $X=0$.

因此, 局部地, 我们可以定义流形上任意一点 p 处的切空间上的内积. 用分量可以表示为 $g(X,Y)=g_{ij}X^iX^j$, 其中矩阵 (g_{ij}) 的元素为 $g\left(\dfrac{\partial}{\partial x^i},\dfrac{\partial}{\partial x^j}\right)$. 特别地, 当流形为欧氏空间时, $g_{ij}=\delta_{ij}$, 从而 $g(X,Y)=\sum\limits_i X^iY^i$, 这就是通常的欧氏空间的内积.

注2.2　任意的一个光滑流形上总存在黎曼度量. 带有黎曼度量 g 的微分流形 M 称为黎曼流形 (M,g).

定义2.3　如果 $f:(M,g)\to(N,h)$ 是两个黎曼流形之间的光滑同胚, 并且对于任意的 $p\in M$, 以及 $u,v\in T_pM$, 都有

$$h(f_*(u),f_*(v))=g(u,v),$$

则称 (M,g) 与 (N,h) 等距.

现在用曲率张量来描述流形 M 的弯曲程度.