

引力

一本广义相对论教程

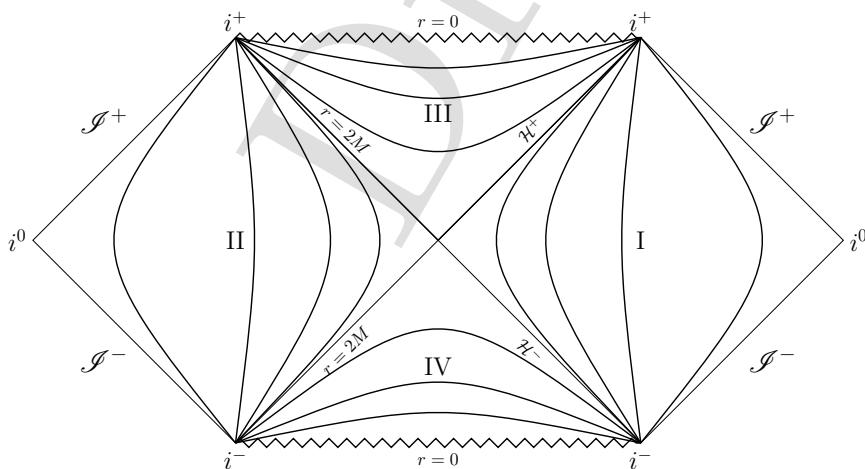
夏草

Draft

引力

一本广义相对论教程

夏草



CC0 license
NOT PUBLISHED

静待花开时.

Draft

Copyright © 2023 夏草 NOT PUBLISHED.

License: 知识共享署名国际许可协议 (CC0)

<http://creativecommons.org/licenses/by/4.0/>

仅限教学科研使用，使用与传播应严格遵守相关协议及法律规定。

Version: 2023.4.22, First Edition(Draft).

序言

人们普遍接受了广义相对论（下称“广相”）在引力现象上所取得的巨大成功。Albert Einstein 为了推广相对性原理，使之也能够囊括引力，选择了另辟蹊径，从引力同惯性、时空几何学的关系上入手，终于在 1915 年 11 月完成了他 8 年奋斗的最后一幕：

The image shows two handwritten mathematical equations on aged paper. The top equation is $G_{im} = -\kappa \left(T_{im} - \frac{1}{2} g_{im} T \right)$. The bottom equation is $\sum_{\sigma} g^{\sigma\tau} T_{\sigma\tau} = \sum_{\sigma} T_{\sigma} = T$.

图：取自《引力的场方程》(*Die Feldgleichungen der Gravitation*)

后人称此为 *Einstein* 场方程，并以“体系的精髓”冠之。在先后经历三大实验验证后，各大报纸相继对此刊登头条：“现代物理学赢来了一次伟大胜利。”

这些“神秘”数学符号固然令人费解，但我们暂时不需要完全知晓这些东西的含义¹，只需知道，方程的左边关乎着引力，而右边关乎着物质。这与此前 Newton 的描述方式是类似的，他选择让引力同场源的质量成正比，在距离上满足平方反比律。其经后人改写为所谓的 *Poisson* 方程

$$\nabla^2 \phi = 4\pi\rho,$$

其中 ϕ 是引力势， ρ 是质量密度。可见确实亦为“引力”正比于“物质”的形式。

比较 *Einstein* 和 *Poisson* 的方程，很明显后者的记号要通俗些，前者那可怕的符号诚然多少“败坏”了广相的名声。不过大可不必妄自菲薄，因为它们都代表了人类智慧漫漫长征的里程碑，且都以必要的概念框架为前提：*Newton* 理论需要微积分，而 *Einstein* 理论需要额外的几何学罢了。为了陈述 *Newton* 理论所必须的知识，已被纳入到了本科一年级教学中，譬如 3 维 Euclid 空间、坐标系、函数、导数、质

¹需要指出，图中的 G_{im} 在现代写法中实际是 Ricci 曲率张量 R_{im} ，现在人们更多把 G_{im} 称作 *Einstein* 张量。

量和力这样一些概念，相信诸位具有一般的数学和物理背景的读者并不陌生。至于广相的结构，或许人们就远不是这样熟悉了，但其所必须的知识并不比微积分难到哪儿去，一旦接受了，这两个方程都将变得更自然、更简单。

想要真正了解一个理论，仅仅知道怎样写出控制它的框架还远不够丰富。广相是与 20 世纪物理学所作的一些最蔚为奇观的预言联系在一起的，也许读者或多或少在 Hawking 的科普贡献中听说过一些概念，比如引力坍缩、黑洞、奇点、宇宙膨胀……这些东西在 1915 年时还未为人知，所以它们与场方程的提出没有关系。只有当后来人们懂得围绕场方程的解，去研究整体的动力学应用时，才会注意到这些现象。这件事花的时间长得相当惊人，可能人们对其中的艰辛事迹的熟悉程度，甚至还远不及对场方程背后孤勇奋斗的了解程度。因而本书将会着重讲解广相的经典应用和前沿话题，而对某些历史作简短一瞥。

本书致力对引力及时空的话题做一个全面、深入的讲述，由其教程之出发点，仅为了帮助相关专业的学生及从业人士建立必要的数学、物理基础，为理解前沿领域中相关论文作准备。本书主要内容分为三个部分：必要的基础（主要为经典内容）、对许多话题的概述（上世纪末以及千禧年的前沿、基础概念的探讨）以及随时可供查阅的附录。书的末尾附上了参考资料及索引以飨读者。

谨以此序阐明 Part I 内容之深度、广度，有如下考虑：

以理工类专业大一学习的微积分和线性代数为基础。具体地，笔者将假定读者已完整地、熟练地掌握多元微积分的场论符号，了解一点分析学和线性空间，故本书所穿插的数学知识只会是上述内容中没有完整涉及的部分。在介绍新内容时，笔者将在教程中尽可能多地提示读者，新内容与其所学过的哪些内容相关（甚至精确到某些书籍的章、节、页码）。希望读者在比较后能够同意，笔者在此教程中介绍的新知识不是完全陌生的。

穿插剖析后续物理理论所需要的基础数学概念。例如，在引入时空的概念（第 X 章）时，需要用到集合的划分与等价类的知识，因此笔者会在期间穿插集合关系的知识。亦如，作为基础概念的各类张量，在本教程中会事先简化成“向量复合出来的数组”，并引入微分几何（第 X 章）后将其定义为线性变换。按照上一条假定考虑读者的数学基础，仍需要比较详尽地介绍有限维向量空间及其上的线性变换的知识。

采用坐标依赖（详细讲解并使用 Einstein 求和约定）与无坐标依赖（数学系符号为主，抽象指标仅作简介）交互的表达风格。笔者希望这种做法本身能彰显广相的建立者希望实现的物理客观性，既物理规律（数学表达式）不依赖包括坐标系的选择在内的任何主观选择；但同时在面对具体应用问题时，仍然保留建立曲线坐标系的权利。此时向量和线性变换的坐标关系需要曲线坐标系的知识。

本书将不会设置习题，仅讲解足够读者学习进程的例题，为了使过程详尽将会使用彩色标注的方式。仅为了证明某定理所需的数学概念和引理，要么给出简介，要

么放在附录中讲解.

定理的证明过程, 仅供有兴趣的读者参考, 故未必都提供. 不提供时, 笔者尽可能提出其他提供了证明过程的参考资料 (具体到章、节、页码). 例如: Reynolds 传输定理的证明 (附录 XX)、等距变换的表示定理的证明 (附录 XX).

本教程用不到的数学知识, 不作介绍, 部分放入附录. 例如, 在拓扑学 (第 X 节) 中提到了部分公理集合论的公理, 但又假定关于自然数、算术运算、偏序、全序、数学归纳法等知识为暂时无需, 包括 Peano 公设在内的相关的内容. 故关于微分几何章节的最终状态, 深度上似乎要介绍完微分拓扑, 但广度上又不完整. 这是笔者故意为之的. 因为本教程的目标不是要向读者大肆介绍数学知识, 而是补充看懂广相前沿进展的最少必要的、作了严格定义的数学知识.

一般情况下, 概念的定义仅通过字体的改变来暗示. 仅在需要时, 定义才以带编号的方式引入. 而定理、引理和例子则均带编号. 定义是极其重要的. 它在文中只出现一次, 因此难免要经常反复回顾. 不采用带编号式的引入, 只是因为需要定义的概念很多, 如果每个定义都带编号定义将会严重打断行文的流畅性, 让本就不易的内容雪上加霜, 而代价则是使定义失去了引用链接的便利, 故在此敬请读者在学习时要不厌其烦地翻阅回顾定义.

Part II 将在其完成时再做序.

夏草

2023 年 4 月于北京

Draft

目录

序言

v

I 基础	1
1 概览	3
1.1 何为引力?	3
1.1.1 很矮的引力	4
1.1.2 很高的引力	5
1.1.3 很远的引力	6
1.1.4 很近的引力	7
1.1.5 很大的引力	10
1.1.6 很小的引力	18
1.2 为何需要相对论?	18
1.2.1 经典电磁理论与经典力学的矛盾	18
1.2.2 Einstein 对狭相的理解	21
1.2.3 boost 变换	22
1.2.4 质能方程	26
1.2.5 Minkowski 对狭相的理解	26
1.3 怎么描述弯曲几何?	29
1.3.1 度规	29
1.3.2 闵氏几何	32
1.3.3 Einstein 对其之理解	32
1.3.4 Lorentz 几何	39
1.3.5 测地偏离	54
1.4 广相应该是怎样的一套理论?	58
1.4.1 理想流体	58

1.4.2 广义协变性	61
1.4.3 场方程	62
1.4.4 作用量	64
1.4.5 时空的拓扑结构	65
1.4.6 这套理论有哪些应用?	65
2 平直时空	69
2.1 代数学拾遗	69
2.1.1 集合与映射	69
2.1.2 矢量空间	71
2.1.3 对偶空间	74
2.1.4 张量	77
2.1.5 坐标系	84
2.1.6 微分形式	88
2.1.7 积分	96
2.2 时空结构	113
2.2.1 时空的几何学	113
2.2.2 时空的物理学	115
2.3 惯性运动	118
2.3.1 对钟	118
2.3.2 运动学	122
2.3.3 力学	125
2.3.4 Döppler 效应	128
2.3.5 接触作用	129
2.4 非惯性运动	130
2.4.1 瞬时观者	130
2.4.2 无自转观者	134
2.5 经典电磁理论	135
2.5.1 回顾	135
2.5.2 几何语言	149
2.6 流体动力学	155
2.6.1 尘埃	155
2.6.2 连续介质的能动张量	155

3 弯曲时空	157
3.1 几何学拾遗	157
3.1.1 协变微分	157
3.1.2 克氏符	159
3.1.3 Levi-Civita 联络	163
3.1.4 曲率	163
3.2 基本原理探讨	163
3.3 参考系	163
3.4 作用量理论	163
3.5 线性引力论	163
3.6 引力辐射	163
3.7 引力波简介	163
3.7.1 洞穴论据	163
4 现代语言	165
4.1 拓扑流形	165
4.1.1 什么是拓扑?	165
4.1.2 流形	173
参考文献	175
索引	177

Draft

Part I

基础

Draft

Chapter 1

概览



图 1.1: 2022 年 7 月 12 日 (美国时间 11 日), NASA 公布了 Webb 空间望远镜第一张全彩深空照片——这是迄今为止最深的宇宙星系的红外图像, 展示了 130 多亿年前的宇宙景象, 以及可观测宇宙边缘的素未谋面的星系.

1.1 何为引力?

引力, 是人尽皆知的吸引作用. 每个人自蹒跚学步起便与之“抗衡”.

1.1.1 很矮的引力

我们身边随时随地都存在着这样一种力量，没有支撑的物体会因为它而下坠，有支撑的物体会因为它而稳定。即便如此，生活中还是有许多驾驭引力的高手，比如那些轻松掷中篮板的职业球员；或者挥拍自如的羽毛球手，当然，他们大抵是不需要计算什么物理学的；再如鸟儿，它们和羽毛球手都得借助气流，但鸟儿更为厉害的是，它自身就生长有一对契合空气动力学的机翼，从而高耸入云、自由翱翔。但我们知道，气流在大范围上的无扰压强分布仍需借助引力而稳定，因此结论或许稍显意外：引力虽可能使鸟儿丧命，但也正是引力使得鸟儿脱身。

是的，这皆拜引力所赐，它是如此地普遍，人们的文明从很早开始就在思考它的缘由了。自古以来，世界各族人民都对其作了不同色彩的描述。

面对落体运动，中国上古时期的人们将其解释为“物”趋向于回归到它“原本的地方”，这或许是对的，现在我们知道，引力大概也只有吸引作用，但不知若他们瞧见后来的孔明灯¹或者西洋人更大的热气球会作何解释。现在我们知道，正是引力造就了压强差。

受哲学及逻辑学的启蒙，西方世界对引力的描述要更为精准一些。面对自由落体，古希腊时期的 Aristotle 尝云，“置两个物体于同一高度释放，重的那个将率先落地”。公元 5~6 世纪，一位希腊哲学家 Ioannes Philoponus 记载了这样一个发现，即“在同一点的各种物体都以完全相同的方式运动着，无论大小，无论重量”。虽说同一“高度”不一定代表着“同一点”，但同一点很显然属于同一高度，这即对 Aristotle 学派的物理思想做出了挑战。

时间来到 16 世纪，彼时的意大利弥漫着复兴的气息。一位研究自然的学者 Galileo 得知 Aristotle 学派所秉持的思想有难以弥补的矛盾²后，利用斜坡和单摆实验发现了“距离正比于时间平方”，且“与其自身性质无关”的规律，随后合理外推³至自由落体得出这样的结论：物体从一定高度 h 下落到地面所需的时间 t 与其质量 m 无关，且 $h \propto t^2$ ，或者准确地说

$$\frac{d^2h}{dt^2} = -\underbrace{g}_{\text{引力加速度}}, \quad (1.1.1)$$

因此“在同一点的各种物体都以完全相同的方式”被引力加速。他在西欧致力于传播这一事实，得到了广泛回应与讨论。当然，后来人们知道，在高度变化过大时，其对加速度的影响亦即愈发明显了起来，故 (1.1.1) 式只能是个局部的规律。

¹ 不一定真为诸葛亮的“发明”，但文献显示孔明灯的使用年代确实很早。

² 所谓关于两个相互牵制的物体，下落加速应该归因于谁的问题。

³ 物理学的连续性思想以及忽略不必要阻力的思想。

1.1.2 很高的引力

经过数十年的思索与竞争, Newton 找到了统一地上学问和天上规律的方法, 于 17 世纪末在他的《自然哲学的数学原理》[1] 中全面地阐述了一个更为精确的规律:

$$\text{万有引力常数 } \frac{\text{场源质量}}{\text{质点与场源位矢之差}} \rightarrow \mathbf{F} = -\frac{G M m}{|\mathbf{r}|^3} \mathbf{r},$$

结合其第二定律

$$\mathbf{F} = \frac{d}{dt} \left(m \frac{d\mathbf{r}}{dt} \right) = m \frac{d^2\mathbf{r}}{dt^2},$$

两个式子中所带相同的 m 将可以约去, 从而与之无关:

$$\ddot{\mathbf{r}} = -\frac{GM}{|\mathbf{r}|^3} \mathbf{r}. \quad (1.1.2)$$

对于一个密度分布不均的物体而言, 虽不能确定 \mathbf{r} 的选取, 但可将其视作多个小质点的组合. 首先, 当引力源有多个时, 应该写为

$$\ddot{\mathbf{r}}(\mathbf{r}, t) = -G \sum_i \frac{\Delta M_i (\mathbf{r} - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}_i|^3},$$

因此, 如果我们知道其密度的分布情况⁴ $\rho(\mathbf{r}, t)$, 则整体的总和是积分

$$\ddot{\mathbf{r}}(\mathbf{r}, t) = -G \int_M \frac{\rho(\mathbf{r}', t)(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} d^3 r'. \quad (1.1.3)$$

这其实是一个更普遍的写法, 因为它不仅可以描述按密度分布的连续质量, 也包括了原来的质点模型.

一个质点的密度就是⁵ $\rho_i(\mathbf{r}, t) = M_i \delta(\mathbf{r} - \mathbf{r}_i(t))$, 将 n 个质点密度加在一起就是总密度 ρ , 再代入到 (1.1.3) 中就有

$$\begin{aligned} \ddot{\mathbf{r}}(\mathbf{r}, t) &= -G \int_M \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \sum_{i=1}^n M_i \delta(\mathbf{r}' - \mathbf{r}_i(t)) d^3 r' \\ &= -G \sum_{i=1}^n M_i \int_M \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \delta(\mathbf{r}' - \mathbf{r}_i(t)) d^3 r' \\ &= -G \sum_{i=1}^n M_i \frac{\mathbf{r} - \mathbf{r}_i(t)}{|\mathbf{r} - \mathbf{r}_i(t)|^3}, \end{aligned}$$

↑ 挑选性, δ 函数定义推论

因此这就又回到了质点的万有引力定律.

⁴ 密度是测度微元之比 dM/dV , 关于其严格的数学定义, 请查阅 Radon-Nykodim 导数.

⁵ 这里 δ 为 Dirac δ 函数, (非严格地) 定义为 $\mathbf{r} \neq \mathbf{0} \Leftrightarrow \delta(\mathbf{r}) = 0$ 且 $\int_{\mathbb{R}^3} \delta(\mathbf{r}) d^3 r = 1$.

1.1.3 很远的引力

实际上，我们还可以继续扩大“高度”——或者说把视线放到“更远的距离”去，那么一个很明显的问题就出来了：引力将会是一种能够传播无限远的超距作用。Newton 本人拒绝这种看法，但也不能提供合理的物理实质，而始终认为其使用平方反比律只是在“数学地描述”引力吸引现象。那么，在很远的地方，引力还将是平方反比吗？

想要解决超距作用，一种思想便是想象整个空间弥漫着一种媒介性质的物质，Faraday 称其为场，它可以将作用传播到很远很远的地方。18~19 世纪，经由 Gauss、Maxwell 等人对场论数学的发展，场逐渐作为一种物理实质而在物理学里占据一席。狭义上用第二定律定义的“力”只是一种“须要接触”的相互作用，而在宏观上不“接触”也具有相互作用的力适合用势能 ϕ 描述⁶

$$\mathbf{F} = -m \nabla \phi.$$

因此，假设试验质点放置在恒稳的引力势场 ϕ 中，这时经典的“引力”由上式给出，根据 Newton 第二定律，此即是说加速度是引力势的负梯度。类比到电磁学的电场强度 $\mathbf{E} = \mathbf{F}/q$ 上，加速度就亦可称作引力场强。密度的分布是一种标量场，我们将势能与 (1.1.3) 式联立有

$$\nabla \phi(\mathbf{r}, t) = G \int_M \frac{\rho(\mathbf{r}', t)(\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3} d^3 r'. \quad (1.1.4)$$

如果对其按 \mathbf{r} 取散度，则

$$\begin{aligned} \nabla^2 \phi(\mathbf{r}, t) &= G \nabla \cdot \int_M \rho(\mathbf{r}', t) \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} d^3 r' \\ &= G \int_M \rho(\mathbf{r}', t) d^3 r' \nabla \cdot \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \\ &= G \int_M \rho(\mathbf{r}', t) d^3 r' \cdot 4\pi \delta(\mathbf{r} - \mathbf{r}') \\ &= 4\pi G \int_M \rho(\mathbf{r}', t) \delta(\mathbf{r} - \mathbf{r}') d^3 r' \\ &= 4\pi G \rho(\mathbf{r}, t), \end{aligned}$$

↑ 挑选性

关于 \mathbf{r}' 的积分和密度与散度无关
由 Gauss 定理得

即

$$\nabla^2 \phi = 4\pi G \rho, \quad (1.1.5)$$

⁶ 负号纯粹是习惯，但确实是个好习惯。

此即引力场的 Poisson 方程⁷；如果对其求解 ϕ ，则注意

$$\nabla \phi(\mathbf{r}, t) = G \int_M \rho(\mathbf{r}', t) \nabla \frac{-1}{|\mathbf{r} - \mathbf{r}'|} d^3 r',$$

由于梯度是按 \mathbf{r} 取的，提出积分后可以有⁸

$$\phi(\mathbf{r}, t) = -G \int_M \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d^3 r',$$

这再一次说明了引力场传播无限快，因此即使将引力视作力场，由于我们默认任何一点都满足 Newton 平方反比，场的存在仍然不能根除其超距作用的矛盾。即使如此，想要修改 Newton 的引力理论，借助场这个模型是依旧有希望的——只要我们先不承认平方反比的正确性，而是企图直接研究引力场的物理性质，这样就有望研究出一些比较整体的东西出来，但别忘了在较小距离内还原到平方反比去。

按理来讲，场中相互作用（作为一种波动）的传播速度总该是一个有限数（但可能是一个没有上限的有限数）。按照基本的类比思想，人们自然容易想到另外一种场：19 世纪，人们已经发现了其中波以有限速度传播的场——电磁场。可见，要想找到引力在距离非常之大的地方有何规律，我们似乎得考虑从电磁场上抽象出一点性质出来。在 1.2.2 节将会知道，所选择出来的抽象性质是相对性原理 (*principle of relativity*)。

1.1.4 很近的引力

Newton 引力理论不仅在“很远”的地方出了矛盾，在“很近”的地方虽然“这种矛盾不大”⁹，但是出现了另一个与事实“误差很大”的矛盾：19 世纪末，水星近日点进动现象被首次观测。毫无干扰的情况下，Newton 理论当然不能解释进动，因

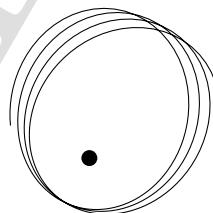


图 1.2: 轨道进动（离心率和进动角均夸大）

此必须要考虑其它因素，如同当年从天王星轨道扰动发现的“笔尖下的行星”海王星

⁷序言中的 Poisson 方程是 $G = 1$ 的版本，这涉及到单位制 (*system of units*) 选取的问题，我将在后文中详细解释。

⁸只是其中一个解，可以加减一个常数（相当于零点选取，类比电磁学的 Lorenz 规范）。

⁹较近距离引力的传播速度确实可以视作瞬间。

那样. 由于地球自转轴的进动, 从地球上观测水星的进动大部分都是所谓的“岁差”, 但这纯粹是参考系的视觉效果, 因此我们就先扣除掉它. 那么在经典引力下, “其它因素” 不外乎就是其余行星造成的摄动及太阳本身的实际形状了.

我们先处理摄动¹⁰. 完全看懂需要读者接触一些数学技巧, 但不懂也无伤大雅, 读者大可跳过这段并接受结论. 取太阳为惯性系, 水星所处的外势场写作

$$V(r, t) = -\frac{GM_{\odot}m}{r} - Gm \sum_i \frac{M_i}{\sqrt{R_i(t)^2 + r^2 - 2R_i(t)r \cos \theta_i(t)}},$$

行星位置 二者在该系的夹径夹角

前一项是太阳的引力势, 后面是各行星的作用总和. 忽略其余行星轨道的离心率之影响 (假定做周期圆周运动), 设单个行星角速度为 ω_i , 将其势场在时域上做 Fourier 展开 $\sum_{n \in \mathbb{Z}} a_n(r) e^{in\omega_i t}$, 其中系数为

$$a_n = -Gm \frac{\omega_i}{2\pi} \int_0^{2\pi/\omega_i} \frac{M_i e^{-in\omega_i t} dt}{\sqrt{R_i^2 + r^2 - 2R_i r \cos \theta_i}}.$$

我们只保留最低阶的 $n = 0$, 并用 $d\theta_i$ 替换 $\omega_i dt$ 从而使得引力势同时间无关, 就可得到外势场的近似表达

$$V(r) = -\frac{GM_{\odot}m}{r} - Gm \sum_i \int_0^{2\pi} \frac{M_i}{\sqrt{R_i^2 + r^2 - 2R_i r \cos \theta}} \frac{d\theta}{2\pi},$$

此即是说, 在长期作用下, 其余行星对水星的作用相当于均匀分布在其轨道上的圆环¹¹. 问题现在相当于, 在太阳的 $-1/r$ 势基础之上, 计算附加这样一个微扰后行星的进动. 根号不方便处理, 则将其按 r/R_i 展开为

$$V(r) = -\frac{GM_{\odot}m}{r} - Gm \sum_{i,j} \frac{M_i}{R_i} \int_0^{2\pi} \left(\frac{r}{R_i} \right)^j P_j(\cos \theta) \frac{d\theta}{2\pi}, \quad (1.1.6)$$

↑ P 为 Legendre 多项式

计算该积分有¹²

$$\sum_j \int_0^{2\pi} \left(\frac{r}{R_i} \right)^j P_j(\cos \theta) \frac{d\theta}{2\pi} = 1 + \frac{1}{4} \left(\frac{r}{R_i} \right)^2 + \frac{9}{64} \left(\frac{r}{R_i} \right)^4 + \frac{25}{256} \left(\frac{r}{R_i} \right)^6 + \dots,$$

接下来处理有心力问题. 通常从能量、角动量守恒出发:

$$\begin{aligned} \frac{1}{2} m(\dot{r}^2 + r^2 \dot{\theta}^2) + V(r) &= E, \\ mr^2 \dot{\theta} &= L. \end{aligned}$$

¹⁰ 关于更精细的考虑, 参见 Goldstein 《经典力学》第 5~8 章, 以及相关的天体力学书籍.

¹¹ 另外也可注意, 高阶项仍然保留着有心力的形式, 故可继续使用有心力的理论处理.

¹² 另一方面, 岁差具体来源于地球本身非球体而受太阳及月亮引力而导致的进动力矩, 这样下来处理的是刚体转动, 所得引力势的结果与此处用 Legendre 多项式作为生成函数的展开结果一致.

消去 t 得到 θ, r 的关系

$$d\theta = \frac{L dr}{r^2 \sqrt{2m(E - V(r)) - L^2/r^2}}.$$

将上式从近日点积分至远日点后乘以两倍，减去 2π ，便是行星一个周期内的进动角。假定其他行星的作用相对与太阳是个小量，便能得到一个更为直接的表达式

$$\delta\theta = 2m \frac{\partial}{\partial L} \int_{r_{\min}}^{r_{\max}} \frac{\delta V(r) dr}{\sqrt{2m(E + GM_{\odot}m/r) - L^2/r^2}},$$

为方便计算该积分，引入参数（所谓的偏近点角） ψ 使得 $r = a(1 + \epsilon \cos \psi)$ ，有

$$\delta\theta = \sqrt{\frac{2m}{-E}} \frac{\partial}{\partial L} \int_0^\pi a(1 + \epsilon \cos \psi) \delta V(r) d\psi.$$

将势能 (1.1.6) 代入，计算得到水星一个周期内的进动角为

$$\begin{aligned} \delta\theta = \pi \sqrt{1 - \epsilon^2} \sum_i \frac{M_i}{M_{\odot}} & \left(\frac{3}{2} \left(\frac{r}{R_i} \right)^3 + \frac{45}{16} \left(1 + \frac{3}{4} \epsilon^2 \right) \left(\frac{r}{R_i} \right)^5 \right. \\ & \left. + \frac{525}{64} \left(1 + \frac{5}{2} \epsilon^2 + \frac{5}{8} \epsilon^4 \right) \left(\frac{r}{R_i} \right)^7 + \dots \right). \end{aligned}$$

仅考察第一项，结果显示：进动速度与“外行星、太阳的质量比”成正比，与“水星、外行星轨道半径比”的三次方成正比。我们所考虑的是在 100 个地球周期内的进动角，则

$$\Delta\theta = \frac{180 \times 3600}{\pi} \times 100 \left(\frac{a_{\text{地}}}{r} \right)^{3/2} \delta\theta.$$

根据该结果可算得水星一个世纪内的进动约为 $600''$ 。但是，将外行星近似为均匀分布的一个环，这对于金星而言是及其不靠谱的。金星和水星轨道周期比在高精度下为 $5:2$ ，意味着相对水星而言金星的运动是有明显周期的，这将导致完全不同的摄动结果。这也是上面写下的结果只写了一位有效数字的原因。但没有关系，继续保留一些更高项，将按类似思路处理有心问题。最终所得精确值约为 $532''$ ，不能与事实相符。

关于太阳的实际形状，Shapiro 发现太阳的引力势展开

$$\phi_{\odot} = -\frac{GM_{\odot}}{r} \left(1 - \sum_{l=2}^{\infty} J_l \left(\frac{R_{\odot}}{r} \right)^l P_l(\cos \theta) \right)$$

中起主要作用的是四极矩¹³： $J_2 = (-0.8 \pm 2.5) \times 10^{-5}$ 。而 Dicke 和 Goldenberg 发现的太阳扁率则对应于 $J_2 = (2.7 \pm 2.5) \times 10^{-5}$ 。如果限定其为零，则 Shapiro 的分析给出水星及火星近日点额外进动值分别为实际值的 0.99 ± 0.01 和 1.07 ± 0.1 倍。

¹³偶极矩可以通过平移坐标原点消除，因此太阳的非球形可以看作四极矩。

可见，在经过一系列类似尝试后，很难从 Newton 理论精确地计算出进动角的实际观测结果，总是无法用 Newton 引力算出那相差的“ $43''/\text{世纪}$ ”。这说明，水星由于离太阳“很近”，以至于其所处的引力场强烈到不符合平方反比律。修改 Newton 引力理论仿佛迫在眉睫。

1.1.5 很大的引力

引力在较大尺度上有什么效果？前文提到，要想从局部的现象揣测出宏观的性质，似乎得从电磁场本身抽象出一种性质，并将其移到引力场上。当时的情况是，人们已经普遍接近于抛弃经典力学中的相对性原理：物理规律在所有惯性系中相同（详见 1.2.1 节），企图寻找一种可以与其它惯性系相区分的特殊惯性系——“以太”，作为包括光在内各种电磁波的传播介质，进而说明 Maxwell 的方程只适用于这种参考系。随着光速测量实验的进展，人们逐渐意识到以太理论难以克服的困难。Einstein 本人在 1946 年的《自述》(*Autobiographisches*) 中说，他不太记得光速测量实验是否影响了他的思索，但他一心想要“调和经典力学和电动力学”，选择保留相对性原理。如果是这样，那么问题就只能出在经典力学的另一根基——Newton 的绝对时空观¹⁴ 上。

为了理解这件事，我们必须再次回到现代物理学的起源。在西方文化中，有两种理解“空间”的传统方式：作为一个实体 (*entity*) 或作为一种关系 (*relation*)。

“空间是一个实体”意味着当除了空间之外没有其他东西时，空间仍然存在。它本身就存在，物体在其中运动。这是 Newton 描述空间的方式，称为绝对空间。后面我们将强调，这也是狭义相对论中理解时空（不是空间）的方式。尽管自古以来（在民主传统中）就被认为是这样，但这种理解空间的方式并不是西方文化中的传统主导观点。从 Aristotle 到 Descartes，主流观点是把空间理解为一种关系。

“空间是一种关系”意味着世界是由物理物体，或物理实体组成的。这些物体具有这样的属性：它们可以彼此接触，亦可不接触。空间就是物体之间的这种“接触”、“毗连”或“紧邻”的关系。例如，Aristotle 将一个物体的空间位置定义为围绕它的物体集合的（内部）边界。这就是关系空间。

与这两种理解空间的方式严格相联系，有两种理解运动的方式。

“绝对运动 (*absolute motion*)”：如果空间是一个实体，运动可以被定义为从空间的一个部分到空间的另一个部分。这就是 Newton 对运动的定义。

“相对运动 (*relative motion*)”：如果空间是一种关系，运动只能被定义为从一个

¹⁴Newton 本人主张绝对空间观而其它哲学家则不然，但在 Einstein 之前，几乎所有人都支持绝对时间观，其中 Poincaré 和 Lorentz 在观念上的革新仍不够彻底。

物体的紧邻性到另一个物体的紧邻性. 这就是 Descartes¹⁵ 和 Aristotle¹⁶ 对运动的定义.

物理学家的问题在于, 这两种关于空间和运动的思维方式中, 哪一种可以更有效地描述世界. 对 Newton 来说, 空间是绝对的, 运动也是绝对的¹⁷. 这是对 Descartes 主义的第二次违反. 当然, Newton 也并不是怀揣轻松的心态迈出这一步的: 他在《原理》中用了很长的开头部分来解释他的选择的原因. 对 Newton 有利的最有力论据完全是马后炮: 他的理论结构当然运行得特别好, 而 Descartes 的物理学从未如此有效. 但这不是 Newton 的论点.

Newton 采用了经验证据, 讨论了一个著名的水桶思想实验. 将一盛有清水的桶静置, 并用一根长绳悬吊 (图中略去) 使桶可以靠绳索的扭力而转动; 随后, 尽可能使桶很快地加速旋转至某一角速度 ω 并保持. 起初, 桶相对于我们旋转, 但 (大部分) 水由于惯性仍将保持静止, 并且液面是平坦的; 待桶的运动通过摩擦力传导至水, 使水开始渐渐地随桶旋转, 液面逐渐凹陷. 在某个时刻, 水和桶都将以 ω 旋转, 即相对静止, 液面凹陷呈抛物面状: 用等势面可算出液面方程在柱坐标系下为 $z = \frac{\omega^2}{2g}r^2$; 尽可能很快地停止旋转, 这个时候水由于惯性仍然要旋转一会儿才会静止, 液面仍然凹陷. 虽说是思想实验, 但其可行性还是挺大的: Newton 本人就声称

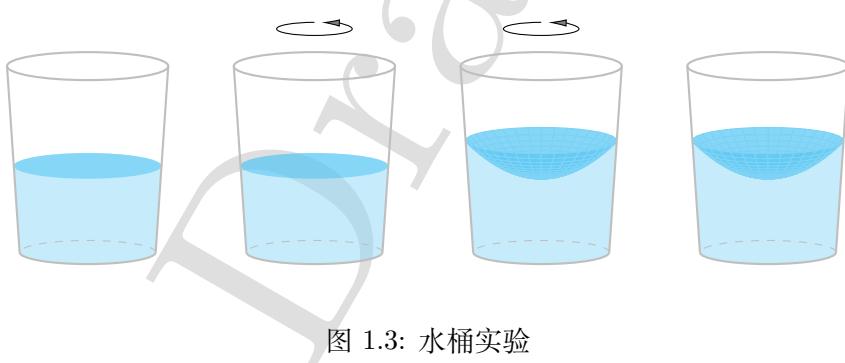


图 1.3: 水桶实验

“已经做过实验”, 并对该结果表示肯定.

¹⁵ “我们可以说, 运动是物质的一部分或物体的一部分, 从与它紧邻的、被认为是静止的那些物体附近转移到其他一些物体附近.” —— Descartes, 《哲学原理》[2], 第 II-25 节, 第 51 页

¹⁶ Aristotle 坚持认为, 运动是相对的. 他用一个人在船上行走的例子来说明这一点. 这个人是相对于船运动的, 而船又是相对于河水运动的, 而河水又是相对于地面运动的……Aristotle 的关系主义因以下事实而有所缓和: 有一些可以作为首选参考的物体: 位于“宇宙中心”的地球, 以及天球 (即星空, 共同组合为天球音乐), 特别是其中一颗固定的星星. 因此, 我们可以说, 如果某物相对于地球运动, 它就是在“绝对地”运动. 然而, 在古代宇宙学中, 有两个首选框架: 地球和固定恒星, 而且这两个框架是相对于对方旋转的. 中世纪的思想家们并没有错过这一点, 他们详细讨论了恒星是围绕地球旋转还是地球围绕恒星旋转. 值得注意的是, 在 14 世纪, Buridan 基于理性的理由得出结论, 两种观点都不比另一种更真实, 而 Oresme 研究了地球的旋转, 比 Copernicus 的日心说早了一个多世纪.

¹⁷ “所以, 只要空间被视为真正有别于运动的物体, 那么就有必要把位置的定义和局部的运动, 认定为是关于一些不运动的东西而言的, 如单独的延伸或空间.” —— [3], 这与脚注 15 中 Descartes 的定义形成了鲜明对比.

问题在于，究竟是什么原因使得液面凹陷？我们从经验中知道，可能是水在相对于某个参考系旋转，进而产生了离心作用，但这个参考系是谁呢？Newton 的水桶实验显示了该问题的微妙之处：如果像 Descartes 所要求的那样，运动是相对于周围物体的位置变化，那么我们必须说，起初水相对于环绕它的桶在旋转，但随后水与桶一同旋转，那么水相对于桶就是静止的。如果水的凹陷是因为旋转，那为什么当水与桶相对静止时，液面却是凹陷的？这样从 Descartes 的思想就似乎有一个矛盾，表面的凹陷反而出现于“静止”而非“旋转”。因此，产生凹陷的旋转不是相对于水桶的旋转。那它还能相对于什么呢？是我们观测者还是……Newton 回答道，它是相对于空间本身的旋转。水面的凹陷是水的绝对运动的效果，即相对于绝对空间的运动，而非相对于周围物体的运动。Newton 企图利用这个实验说明绝对空间的存在性，即如果水在其中旋转会凹陷时，这个空间就是绝对背景，而“不受人为因素所决定”。

为了正确理解它，我们应该阐明一个常见误解——不应该将相对主义，即运动只能在与其他物体的关系中被定义的想法与 Galileo 相对性相混淆。Galileo 相对性是指“匀速直线运动”先验地与静止无异的说法。也就是说，速度是相对于其他物体而言的；而另一方面，关系主义却认为，任何运动（无论如何曲折）都是先验的，与静止无法区分。Galileo 相对性的表述本身就假定了运动的非关系性定义。“直线和匀速”是相对于什么而言的呢？这应当视作惯性系的定义。

Newton 声称相对于绝对空间的运动是真实且物理的，在某种意义上，他做得过头了，他坚持认为即使是直线均匀运动也是绝对的。这引起了一场痛苦的争论，因为惯性运动没有物理效应，因此，对这一类特殊的运动来说，水桶实验是失败的¹⁸。Newton 的动力学基础所需要的——亦正是这里讨论的——不是惯性运动的相对性：而是加速运动，如水桶里的水的旋转，是相对的还是绝对的。问题不是“是否存在一个其中速度可以定义的绝对空间”，而是“是否存在一个其中加速度可以定义的绝对空间”。Newton 的答案在水桶实验的支持下是肯定的。如果没有这个回答，Newton 第二定律将不再有意义。

三个世纪以来，没有人能够彻底驳斥 Newton 的论点。期间先后有诸多科学家和哲学家试图批判绝对空间，甚至强烈于对其超距引力的反对。Leibniz 和他的学派关于 Newton 的绝对运动及其所用的绝对加速度进行了激烈的争论¹⁹。随后，人们的疑虑从未真正消失过，一种挥之不去的感觉仍然存在，总觉着 Newton 的论证有问题²⁰。19 世纪末，Ernst Mach 回到了这个问题上，做出了显赫的一步，他认为，水并非相对于某种绝对介质旋转，而是相对于宇宙的全部物质组分（星空背景）。我们

¹⁸Newton 很清楚这一点，这在《原理》的推论 V 中有明确的说明，但他在导言中选择忽略这一点。或许他这样做只是为了简化他的论证，这对他同时代的人来说已经够难了。

¹⁹当然 Leibniz 还有其他抱怨的理由，比如两人在争夺微积分发明的优先权。或许历朝历代的科学家的弱点都殊途同归吧。

²⁰对于远距作用，Newton 主义在经验上的巨大胜利是难以磨灭的。或者还是可以“磨灭”的？毕竟前面讲过，Newton 的理论似乎无法解释水星轨道那 43'' 弧度的进动。

知道，水凹陷并趋于溢出应当源于水的惯性，在旋转的参考系中出现了惯性力。它是所谓“运动学的部分”，而反之，假使真的有能力让“所有物质”都绕着水桶旋转，从相对性的推广上讲，确实应当产生同样的效果，尽管这个实验确实很难做。

我们甚至可以更为激进地假设：一个毫无其它物质的空间中有一颗弹性球，由于没有任何参照，它的旋转便无法定义，那么随之产生的惯性效果显然不会出现。随着我在周围加入的物质越多，它的旋转也容易通过参照而观测，进而才体现出了惯性（如逐渐扁平），直到我加入了“所有物质”，其惯性便“完全”体现了。Mach 如是说，“惯性需要在物质的相对运动中体现”，以驳斥绝对空间。与物质质量联系的“动力学的部分”此刻产生了惯性力的效果。随“物质的加入”而“愈发明显的”作用，我们知道，已经有一个被称为引力的相互作用 (*interaction*) 具有极其相似的性质。从简而言，惯性（力）和引力确实都与“质量”的大小成正比。但是它们是正比于相同的“质”量吗²¹？

让我们回到 (1.1.2)，这里我们约去了“相同的 m ”。这本应只是 Newton 的搪塞，默认了相同而回避了进一步的哲学思索，但实际上我们完全可以据此定义出两种不同的质量出来：18~19 世纪，已经有许多物理学家认识到了这一点，即第二定律中定义的应叫惯性质量 (*inertial mass*)：

$$\mathbf{F} = m_I \ddot{\mathbf{r}};$$

而万有引力定律中定义的应叫引力质量 (*gravitational mass*)²²：

$$\mathbf{F} = -m_G \nabla \phi,$$

无心插柳柳成荫，这一点被 Einstein 敏锐地察觉到了，随后于 1907 年决定在今后寻找相对论囊括引力的方法时，贯穿这样一条准则：引力质量与惯性质量精确地成正比，且比例与物质本身无关。因此，选取适当的单位制，可使引力质量等于惯性质量，那此二质量便可不再作区分。这便萌芽了后人所称的等效原理 (*equivalence principle*)。我将在 1.2.2 以及 3.2 节更加深入地讨论这些观点及其对 Einstein 的影响。

总之，Einstein 对水桶实验的新解释很简单，也很有说服力：

水相对于一个局域的物理实体旋转：引力场。

是引力场“告诉”物体是否在加速或旋转，而非绝对空间。没有像绝对空间这样的呆滞的背景实体，只有动态的物理实体。这些实体中包括场。在这些场中存在引力场部分。Newton 的水桶中液面的平坦或凹陷并非由水相对于绝对空间的运动决定的，它

²¹或者说，总是可以假定是同“种”质量，但它们的比例是否对任意物体都是一致的？

²²Newton 力学中还将其分为主动 (active) 与被动 (passive) 两类。前者指作为场源时的质量，后者指作为试探质点的质量。本文此处指后者。

是由水和引力场之间的物理互动决定的。Einstein 关于引力的两条思路（为 Newton 引力找到场论、摆脱绝对加速度）在此相遇。关键在于，Newton 把重力场误认为是一个绝对空间。

那是什么使得 Einstein 有了这一想法？为什么 Newton 的加速度要相对于引力场来定义？Einstein 所关注的引力的特殊属性——等效原理给出了这个答案²³。这一深邃的思想说明了，任何物质在引力场中应该具有相同的加速度，就像当年 Galileo 所说的那样。恰恰是 Galileo 所研究的那“矮”的局部现象，才能从引力的整体现象中剥离出重要的特征出来。

请把这纷繁的世界想象成一个“舞台”，物质在其中“舞蹈”着。现在让我们引入某种力场，对舞台中同一“位置”，任何“舞者”都将感受到相同的由力场带来的加速度。这很可能说明这个力场本身就是这“舞台”的内在性质，要研究这个力，自然就要研究这个“舞台”的几何学 (*geometry*)。

我们知道，惯性力的引入是为了使 Newton 定律在非惯性系中保持形式。也就是说，在选择了“不正确的参考系”的情况下引入一个“起源不明的力”解释运动学。这就是所谓的赝力 (*pseudo force*)。这些力的出现，只是因为观者没有站在惯性系的角度去观测。既然赝力和引力都与“质量”成正比，那也就有可能将引力本身解释为一种赝力。换种说法就是，引力是由于没有选对参考系而引起的。

Einstein 首先把这件事解释为对于从惯性系到非惯性系的坐标变换的一种协变性 (*covariance*)。我们通过其电梯思想实验来揭示它，它直接导致了支撑广相的物理直觉。

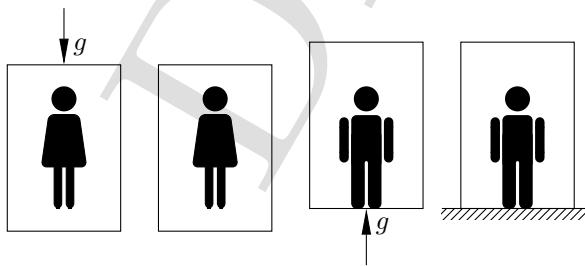


图 1.4: 电梯实验

设想一部封闭的电梯²⁴和一位勇敢的女乘客 。起初，这部电梯是由缆绳牵制而静置于空中的，乘客发现始终有一种作用让他待在电梯的地板上，这个力与它的体

²³引力是“特殊”的，因为 Newton 的绝对空间是引力场的一个构型。一旦我们摆脱了绝对空间的概念，引力相互作用就不再是特别的了。它是形成世界的场之一。但这与 Newton 和 Maxwell 的世界不太一样。

²⁴Einstein 本身是瑞士专利局职员，其很多思想实验都与电车机械有关。为了描述一种自由落体的参考框架，他自然便会想到类似方型箱匣的事物：一部断线的有些骇人的电梯，并且当时电梯恰恰刚诞生不久。

重成正比。根据 (1.1.1)，我们可以说，这是乘客测量到了地球所产生的恒稳的引力势场

$$\phi(h, t) = gh,$$

但若缆绳不幸断裂，这部电梯将顷刻间具有相对大地朝下的加速度 g ，这样，原来的电梯就从 h 变到加速下坠的电梯参考系

$$\tilde{h} = h + \frac{1}{2}gt^2. \quad (1.1.7)$$

在新的参考系下，Newton 第二定律就可以写成

$$\ddot{\mathbf{r}} = \mathbf{0}. \quad (1.1.8)$$

这说明，乘客就似乎失重而漂浮了起来。也可以把这个过程颠倒过来看。假如根本没有地球，飞船自由地漂浮，那么男乘客 也会同样漂浮其中，自然也就没有出现什么引力现象。此时若施加一个加速度 g 向上拖动电梯，那么乘客应当会发现一个赝力，如同引力一样把人拉向地板。这样我们就成功在一个加速的参考系中用 (1.1.7) 式来“模拟”引力场。关键在于，在这些实验中，乘客并不知道外界是否有地球，亦或电梯（飞船）是否加速，那么赝力和引力对其而言是难以区分的，说不清楚总的作用之中“有哪些是惯性力，有哪些是引力”。

下面将用 Newton 宇宙论再介绍另一个更现代、更现实的“电梯实验”。这个模型很简单。在 Newton 物理学的背景下，考虑一个由非常大的球形星系云组成的宇宙。假设这些星系均匀地分布在空间中，并保持着与时间有关的密度 $\rho(t)$ ，星系间在引力作用下相互吸引。设 C 是云的中心。考虑一个星系 A （例如，咱们的银河系），与中心 C 的距离为 $r(t)$ 。众所周知， C 周围半径为 r 的球体外的星系对 A 的引力互相抵消，而该球体内的星系的引力与集中在 C 内的相同质量的质点的引力效果相同。因此 A 所受引力为

$$\mathbf{F} = -G \frac{m_A \frac{4}{3}\pi |\mathbf{r}(t)|^3 \rho(t)}{|\mathbf{r}(t)|^3} \mathbf{r}(t),$$

结合第二定律，此即

$$\ddot{\mathbf{r}} = -\frac{4}{3}G\pi\rho(t)\mathbf{r}(t).$$

若质量在空间上保持常数，则其密度标度将一致依赖于 $|\mathbf{r}|^{-3}$ ，即 $\rho(t) = \rho_0|\mathbf{r}(t)|^{-3}$ 。因此

$$\ddot{\mathbf{r}} = -\frac{4}{3}\pi G\rho_0 \frac{\mathbf{r}(t)}{|\mathbf{r}(t)|^3} = -\frac{c}{|\mathbf{r}(t)|^3} \mathbf{r}(t), \quad (1.1.9)$$

这里定义的 $c = 4\pi G\rho_0/3$ 当然是个常数。在后续内容中读者将知道，方程 (1.1.9) 实际上就是支配宇宙膨胀的 Friedmann 方程，因为这与在平直近似下从广相中得到的方程相同。

所考虑的 Newton 模型中，星系 C 位于宇宙的中心且定义了一个惯性系，而星系 A 不在中心且不是惯性的。假设星云如此之大，以至于从 C 或 A 无法观察到它的边界，那么如果你在这两个星系中的一个，你怎么能知道你在哪个星系中？即，你怎么能知道你是在惯性系 C 中还是在加速系 A ？答案很明显：你不能。由于整个星云均匀地膨胀或收缩，所以从所有星系来看，本地天空的画面都是以相同方式均匀地膨胀或收缩的。而且，你也不能通过局域实验，检测出你是在惯性星系 C 还是在加速星系 A ！事实上，要检测出你是否在一个加速系中，你必须观察惯性力。 A 系的加速度是 (1.1.9) 式，则引入一个作用在所有移动质量上的惯性力

$$\mathbf{F}_{\text{惯性}} = \frac{c}{|\mathbf{r}(t)|^3} \mathbf{r}(t),$$

这就是允许我们检测出非惯性系的那种力。除此之外，所有质量还受到局域的其它力 $\mathbf{F}_{\text{局域}}$ ，也应当受到一个指向 C 的引力

$$\mathbf{F}_{\text{中心}} = -\frac{c}{|\mathbf{r}(t)|^3} \mathbf{r}(t),$$

进而它们在加速系 A 中的运动由

$$m\ddot{\mathbf{r}} = \mathbf{F}_{\text{局域}} + \mathbf{F}_{\text{惯性}} + \mathbf{F}_{\text{中心}} = \mathbf{F}_{\text{局域}}$$

决定。这样， A 中局域的动力学看起来就像是在惯性系中一样。从外界看来，在 A 中坠落的石头将沿着抛物线运动，但在加速系 A 看来就是一条直线。因此，你不可能判断是否处在宇宙的中心，也不可能判断所在参考系是惯性的还是非惯性的。

我们该怎样理解这一“不可能”？根据 Newton 理论， C, A 的动力学应当是截然不同的，但这种不同却不是可物理观测的 (*physically observable*)！在 Newton 的概念结构中， A 是非惯性系，因为同时出现了引力和惯性力，但二者背后藏有某种隐蔽的联系。实际上，情况是完全普适的：在一个足够小的邻域内，惯性力和引力在一个自由落体参考系中，是精确抵消的。我们再一次说明了等效原理的深邃性²⁵。

很明显，可以不借助于所有这些不可物理观测的力：理解该物理情形有个更好的办法，那就是扔掉 Newton 所偏好的整体 (*global*) 参考系，并认识到每个星系都有其自己的局部 (*local*) 惯性系。局部惯性系可以定义为其中没有可观测惯性效应的参考系，正如 Newton 所做的那样。这样，每个星系就都有其局部惯性系。这些参考系由引力所确定，也就是说，正是引力在每个点决定了什么东西该是惯性的。惯性运动应当像这样相对于局部引力场而言，而非绝对空间。

²⁵ 顺便提一句，Newton 这个天才是知道这件事的：“无论诸物体彼此间以何种方式运动，若它们被沿着平行线的相等加速力所推动，则它们都将继续彼此间的运动，遵循的方式就如同没有那些力作用一样。”——《原理》，推论 VI “运动定律” [1]。Newton 用这个推论来计算太阳系中月球的复杂运动。在地球系下，惯性力和太阳引力在良好近似下相互抵消，那么月球就遵循一个 Kepler 轨道。

进而地，引力决定了不同星系参考系相对彼此坠落的方式。引力场可以阐明了不同惯性系间的关系，因为正是引力场决定了惯性运动。Newton 的“绝对运动”不是相对于绝对空间的运动，而是相对于引力场所决定的参考系的运动。或者说，运动应当相对于引力场。那么思想很明显了，第二定律应该支配的正是物体相对于引力场的运动。

关于为何不能找到整体惯性系，读者不妨作如下设想：在地球某一局部找到 Einstein 电梯使得引力变成惯性力，固然是很轻松的。况且颠倒地来说，待在地表上的我们都相当于一个“向上加速的参考系”，而在这个参考系中我们观察到了“惯性力”。然而，如果我们扩大视野，观察两个相距较远的电梯，将会立即发现问题：很显然，在球面的另一些地方，甚至是另一侧，这一部电梯的加速参考系当然就解释不了另一部电梯的情况了。Einstein 如是说，“只能在某一点及其附近能局部地找到这样等效的参考系”。想要找到普适的“参考系”，或许说明，这个世界的几何学要比欧氏几何复杂的多。在 1.3 节将会知道，我们可能需要弯曲的 *Lorentz* 几何。一旦引入了弯曲几何，那么就总可以找到参考系使得引力变得“不再受力”，如同 (1.1.8) 式那样。

综上我们发现，其实引力虽然在局部上与惯性力无法区分，但是只要稍微扩大一点范围，由于不同地方所受的引力方向不同，因而在普通的几何里找不到一个共同的等效参考系，进而就区分出了引力和惯性力。

此外，即使方向是相同的，也有可能区分。如果我们再扩大这部电梯，使得不同高度的点由于其与地心距离不同，而具有明显不同的加速度。即势场不再取 $\phi(h, t) = gh$ ，而是

$$\phi(r, t) = -\frac{GM}{r}.$$

这部电梯的不同部分所受引力是不均匀的，因而也就会产生形变，具体来讲是一种竖直方向的拉伸。由于这有些类似于月球引力对地表海洋影响所产生的潮汐，那么这种形变的力也叫作潮汐力 (*tidal force*)。很明显，这个潮汐力是不能用惯性力解释的，因此这个方法同样分辨出了引力与惯性力。

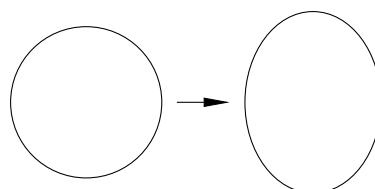


图 1.5: 潮汐效应

1.1.6 很小的引力

在量子尺度下，引力的强度是很弱的，这在一定程度上暗示了经典引力理论在量子领域可能的失败。现在我们知道，自 Einstein 开始，至今仍未完全地调和广相同量子理论的各种矛盾。由于深入讲解这个话题必将需要读者了解相关的陌生概念，故我将不在此赘述，而只简单地指出：引力同经典电磁作用的统一理论实际上已经完成，其中 Kaluza-Klein 理论和 Einstein 的（经典）统一场论²⁶是完全经典的，而其它的诸如弦论、非交换几何 (CLCB 模型) 只部分地回答了量子引力的疑难。现在所谓的“大一统理论”通常指的是沿着标准模型的思路，将强相互作用同电弱作用统一的尝试。狭义上的量子引力理论只涉足引力的量子化问题，但这在数学上有巨大的困难；广义的量子引力理论或许是人们所希望的“万物理论”。此外，受黑洞热力学的启发，现在也有企图从统计物理角度寻找引力同热学共通之处的理论，诸如熵力理论。这些话题将部分地放入本书 Part II 中去。

1.2 为何需要相对论？

现在我们知道，的确可以否认绝对空间极其绝对运动了，那绝对时间呢？如前文所述，除 Einstein 外几乎无人质疑过绝对时间，但要想在电磁场上找灵感，我们得先来看看当时关于对电磁场的认识发生了怎样一回事。

1.2.1 经典电磁理论与经典力学的矛盾

19 世纪，光作为“光介质”或发光以太的扰动的波动理论被广泛接受，该理论在 James Maxwell 的著作中达到了最发达的形式。根据 Thomas Young(1804) 和 Augustin-Jean Fresnel(1816) 的工作，人们认为光在称为发光以太的弹性介质中以横波的形式传播。然而，光学现象和电动力学现象是有区别的，因此有必要为所有现象创建特定的以太模型。试图统一这些模型或创建对它们的完整机械描述的尝试没有成功，但经过诸如 Faraday、Kelvin 男爵等许多科学家的大量工作，Maxwell 于 1864 年提出了一个准确的电磁理论，推导出一组电、磁和电感方程，其经由 Oliver Heaviside 用矢量微积分的语言重新书写为我们当今熟悉的形式：

$$\begin{cases} \nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}, & \nabla \times \mathbf{E} = -\partial_t \mathbf{B}, \\ \nabla \cdot \mathbf{B} = 0, & \nabla \times \mathbf{B} = \mu_0 \epsilon_0 \partial_t \mathbf{E} + \mu_0 \mathbf{J}, \end{cases} \quad (1.2.1)$$

²⁶受到 Kaluza 理论的启发的 Teleparallel 引力，其数学语言采用 Weitzenbock 几何。

这里我还原了国际单位制 (SI): μ_0 称为真空磁导率 (*vacuum permeability*), ϵ_0 称为真空介电常数 (*vacuum permittivity*). 在真空无电荷的假设下有:

$$\begin{cases} \nabla \cdot \mathbf{E} = 0, & \nabla \times \mathbf{E} = -\partial_t \mathbf{B}, \\ \nabla \cdot \mathbf{B} = 0, & \nabla \times \mathbf{B} = \mu_0 \epsilon_0 \partial_t \mathbf{E}, \end{cases} \quad (1.2.2)$$

则

$$\mu_0 \epsilon_0 \partial_{tt} \mathbf{E} = \nabla \times (-\nabla \times \mathbf{E}) = \nabla^2 \mathbf{E} - \nabla \nabla \cdot \mathbf{E} = \nabla^2 \mathbf{E}; \quad (1.2.3)$$

同理

$$\mu_0 \epsilon_0 \partial_{tt} \mathbf{B} = \nabla^2 \mathbf{B}, \quad (1.2.4)$$

皆为真空电磁场的波动方程. 于是我们发现, 真空中的电磁波以某种常速率 c 传播, 数值上是

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \approx 299\,792\,458 \text{ m/s}, \quad (1.2.5)$$

这与光速的实验测定高度吻合, 这预言了“光是一种电磁波”的结论, 并在 1887 年由 Hertz 实验证实. 随后, 以 Maxwell 方程组为核心的经典电磁理论的正确性被大量实验所证实.

但是这里的光速是相对于谁的? 注意到我们并没有事先假设任何特定参考系, 就已经得到了一个“绝对”的数值. 这不符合当时人们的直觉. 我们知道, Newton 的经典物理学建立在绝对时间和空间的基础上, 坚持 Galileo 的相对性原理, 将其精确地重述为机械系²⁷. 这可表述为: 就力学定律而言, 所有处于惯性运动中的参考系都享有同等的特权, 不能将任何特定的惯性参考系归因于偏好的运动状态. 不可通过观察粒子及其他物体的动力学行为识别一个参考系是静止的还是在做匀速运动. 因此在经典力学的世界里力学定律在所有的惯性系中都成立:

相对性原理 不存在静止的绝对标准; 只有相对运动是可观察的.

Galileo 在其《关于两大世界体系的对话》中通过辩论说明这一原理, “如同在甲板下的船舱中观察到的那样, 在平静海面上船的匀速行驶不影响鱼、蝴蝶和其他运动物体的行为”. 相对性理论将这一原理作为基础, 正如关于空间和时间本性的表述等同于关于 Newton 运动方程特性的表述. 但若给予其这种普遍意义, 则必须适用于全部物理学, 而不只是适用于 Newton 力学. 最初这似乎不成问题, 很难想像它在这种基本水平上成立, 但不适用于更复杂的物理相互作用. 尽管如此, 当尝试将其扩展到电磁学时, 由于 Galileo 协变性与 Maxwell 方程组相抵触, 深度的问题还是显露出来.

对含有低速运动电荷及低速变化的电场、磁场的系统, 一切显得直截了当. 它们看来像是被不同匀速运动参考系之间变换的不变性定律所支配. 可以想像, Galileo 船的现代方案也可以携带若干磁体、电池和半导体以及其他电性元件. 主人公“Salviati”对相对性的论证似乎是强有力的.

²⁷ 《原理》运动定律之推论 V.

但当考虑迅变场，尤其是考虑光的传播时，问题就出现了。后来的 Einstein 在他的原始论文《论动体的电动力学》²⁸中指出，“Maxwell 电动力学……，当应用到运动物体时导致一些不对称性，而该不对称性似乎并非现象所固有的”。如前文所述，主要的困难是 Maxwell 方程组给光和其他电磁波一个确定的速度，在真空中它以相同的速度沿各方向传播，与源的运动无关，这一事实与 Galileo 变换不相容，光在一个参考系中以速率 c 传播，在向着光源以速率 u 运动的参考系中必定具有速率 $c+u$ 。因此，对于光以任意速率传播必定是可能的，光在一个光源静止的参考系中以速率 c 传播，在运动参考系中应有某种其他速率，所以 Galileo 变换意味着光速依赖于源的运动。

为了解决这一抵触，Maxwell 放弃相对性原理，首先提出光实际上是在同一个以太介质中的波动（电磁辐射），认为方程组只对一个绝对参考系（以太）成立，它是电磁现象的起因。然而，Maxwell 的理论在运动物体的光学方面并不令人满意，虽然他能够提出一个完整的数学模型，但他无法提供对以太的连贯机械描述。据这一假说，由方程组计算得到的真空光速是相对于以太的速度，在相对于“以太”运动的参考系中，光速具有不同的数值。根据 Maxwell 的理论，所有光学和电学现象都通过该介质传播，这表明应该可以通过实验确定相对于以太的运动。然而，任何已知的探测以太运动的实验都失败了。Fizeau 实验和 Michelson-Morley 实验表明光速与参考系的运动无关。该实验结果否定了以太假说，表明相对性原理的正确性。

这导致 Hendrik Lorentz 从 1892 年开始发展了一种基于不动的发光以太（虽然没有推测其物质构成）、物理长度收缩和“当地时间”的理论，其中 Maxwell 方程在所有惯性系中保持其形式。Lorentz 的假说解决了上述矛盾，但他不能对 Lorentz 变换的物理本质做出合理的解释。在 Lorentz 的以太理论之后，Henri Poincaré 猜测 Lorentz 变换和时空性质有关，他早先提出了“相对性原理”作为物理学的“一般”规律（包括电动力学和万有引力），在 1905 年利用这一原理修正了 Lorentz 的初步变换公式，也即后来的 boost 变换。然而，只有 Einstein 的深刻洞察力，才把这种不变性提升到基本原理的高度。

Einstein 从 16 岁就开始思考的光与以太的问题，在此之前他已经明白电磁理论与经典力学的 Galileo 相对性不相容，他对其印象深刻：单个物体的速度没有意义、只有物体之间的相对速度才有意义。请注意，从某种意义上说，这是 Newton 证明“真正的运动”的企图一个微小但重要的失败。对 Einstein 来说，这是一个暗示，表明 Newton（以及用经典力学看待电动力学）的概念有问题。

尽管 Newton 理论在经验上取得了成功，但其关于绝对空间的想法仍令人深感

²⁸ [9]，这是狭义相对论（下称狭相）确立的重要标志。第一部分主要解决运动学上的矛盾，他没完全使用初等数学而独立地得到了 Lorentz 的结果；而在第二部分中，他用半边篇幅系统地描述了 Maxwell 方程组在不同惯性参考系之间的变换，但由于并非使用矢量微积分语言，略显繁琐。论文发表于 1905 年 6 月 30 日。这一年他共发表 6 篇论文，在 5 个不同领域中取得 4 个重大历史性成就，故亦史称“奇迹年”。

不安. 正如 Leibniz、Mach 和其他许多人所强调的, 空间是一种超感官的实体, 它“作用”于物体, 但物体不能“作用”于它. Einstein 确信, 这种绝对空间的想法是错误的. 不可能有绝对的空间, 不可能有“真正的运动”. 只有相对运动, 因此相对加速度必须具有物理意义而绝对加速度不应加入物理方程. 通过狭义相对论, Einstein 成功地将伽利略有关于速度的相对论从 Maxwell 理论的挑战中解放出来. 然后, 他确信他可以为整个 Aristote-Descartes 的相对论进行辩护. 用 Einstein 的话说, “不仅是在惯性系中, 运动定律在所有参考系中都应该是相同的”. 事物的运动是相对于彼此而言的, 而不是相对于一个绝对的空间而言的; 不可能有任何绝对运动的物理效应²⁹.

1.2.2 Einstein 对狭相的理解

Einstein 意识到 Galileo 变换实际上是 Newton 经典时空观的体现, 如果承认“真空光速独立于参考系”这一实验事实为基本原理, 可以建立起一种新的相对时空观, 由相对性原理即可导出 boost 变换 (*transformation*)³⁰:

$$\begin{cases} \hat{y} = y, & \hat{x} = \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \\ \hat{z} = z, & \hat{t} = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}. \end{cases} \quad (1.2.6)$$

进而地就可以明确规定: 描述物理定律的基本方程都应当在 Lorentz 变换中保持形式不变, 其结果是时空平移、空间旋转、boost 变换以及三者的任意复合. 这样, Galileo 变换

$$\begin{cases} \hat{y} = y, & \hat{x} = x - vt, \\ \hat{z} = z, & \hat{t} = t \end{cases} \quad (1.2.7)$$

其实就可以看作 boost 在 $c \rightarrow \infty$ 时的极限. 如此便调和了经典力学和电动力学. 这一理论完满解决如前所述的矛盾.

承认光速的不变将导致一些有趣的现象. Einstein 首先提出了一个关键性的概念: “同时”具有相对性. “那列火车 7 点钟到达这里”, 这实际上表述的是“火车到达”和“表的短针指向 7”是同时的事件. 但我们在这一参考系中观察到的同时事件, 在另一个参考系中不一定也“同时”. 在另外一个参考系中, 火车有可能不是在 7 点钟到达的, 而是“在 6 点或者 8 点到达的”.

由于同时的相对性, 当我们无法与某物体保持相对静止时, 比如你想要测量一辆高速行驶的列车某一车厢的“长度”, 你必然会使用到计时器、标尺等工具, 我们

²⁹根据许多当代物理学家的说法, 这是对“哲学”思想的过度重视, 它不应该在物理学中发挥作用. 但 Einstein 在物理学上的成就远比这些物理学家获得的成就更为重要.

³⁰其含义取时间轴上的推动之意. 一个较好的翻译是“伪转动” (pseudo rotation), 我们将在本节末解释这一点.

就会注意到以下现象：物体运动时，它的一切（物理、化学变化等）从参照系的时间来看都会变慢；重新审视长度的真正含义，并会发现，如上述的列车在某一参考系的运动测量值会短于静止测量值；速度将不再是满足线性叠加的矢量，因为我们所熟知的“位移矢量对时间求导”的“时间”将不再绝对；光速不变避免了在电动力学中对发光以太的任何参考的需要，并提出所谓的相对论限制 (*restriction*)：任意惯性参考系下，均不容许有质粒子以快于或等于光速的速度运动……关于这些有趣现象的理解，请在学习必要内容后查阅第 2 章，尤其是那些更为直观通俗的图片！

1.2.3 boost 变换

下面我将补充 boost 变换的推导，为后续的一个更深刻的“几何”视角做预备。

我们期望找到这样的关系，它连接着原惯性系 S 所测量的 x, y, z, t 和新惯性系 \hat{S} 所测量的 $\hat{x}, \hat{y}, \hat{z}, \hat{t}$ 。不考虑原点的移动，那么结合“相对性原理”可以推知两个惯性系之间的时空变换是线性变换³¹。假设这个线性变换是四阶方阵 Λ ，那么应该有

$$\begin{bmatrix} ct \\ \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = \Lambda \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix}, \quad (1.2.8)$$

其中给时间乘上这么一个常数 c 以使量纲一致，这样 Λ 的分量就是无量纲的。为了学习以后的张量知识，我们现在可以开始介绍一点用以运算的符号了。我们给如上的列矢量的分量进行编号：

$$x^0 = ct, x^1 = x, x^2 = y, x^3 = z, \quad (1.2.9)$$

当然，这里的编号不是乘幂，而只是指标 (*index*)，并且是上指标 (*upper/raised indices*)，因为在给行矢量、矩阵分量编号的时候还可以用下指标 (*lower indices*)，为此而作上下之区分。不必担心，在很多情况下，我们是不会与乘幂相混淆的；实在冲突时，只需文字标注即可。现在用 Λ^{μ}_{ν} 来表示 Λ 的矩阵元，其中位于左边的 μ 称为行指标、右边的 ν 称为列指标，那么根据矩阵乘法，我们很容易发现每个 μ 分量都满足等式

$$\hat{x}^{\mu} = \sum_{\nu=0}^3 \Lambda^{\mu}_{\nu} x^{\nu},$$

可见，对一个列矢量左乘一个矩阵 Λ ，相当于与其矩阵元 Λ^{μ}_{ν} 下方的列指标 ν “相靠”后，在上方留一个行指标 μ 。Einstein 认为，既然线性代数的求和运算都已事先

³¹由于说清这件事需要更多数学，我们只能暂时先默认这一点，而将细节放在 2.3 节。

蕴含在矩阵乘法的定义之中，那何必不直接“甩掉求和号”，而只关注分量本身呢？毕竟这样的话

$$\hat{x}^\mu = \Lambda^\mu{}_\nu x^\nu \quad (1.2.10)$$

似乎更像是原来的矩阵方程。而矩阵的乘法体现在哪儿？体现在这两个实数的相乘，以及指标 ν 的“相靠”和“一上一下”之中。像这样，如果在一个式子中，同一个指标既作为上标出现，又作为下标出现，就意味着要对该指标求和。Einstein 这个懒惰得非常巧妙但又十分深邃，所以后人给其取名为“求和约定 (*summation convention*)”。关于“一上一下”这件事，或许有些书会选择统一放置在下方成为 $\Lambda_{\mu\nu}$ 而只强调指标“总共重复两次”的条件，更有甚者会选择将矩阵元直接写为 Λ^μ_ν 而不区分其左右位置。这样在讨论更深刻的问题时都会出现缺陷：前者容易混淆矩阵的转置和逆，后者说不清楚“谁与谁相靠”，并且在“升降指标”这个操作时出现顺序混淆的问题³²。因此强烈建议只在特殊情况不会混淆时，再采取这种更为简化的做法。

回到问题上，光速不变不外乎说明这样一件事：所有惯性系所测光速一致。设光速在 S 系测得

$$\frac{\sqrt{dx^2 + dy^2 + dz^2}}{dt} = c, \quad (1.2.11)$$

此即是说

$$dx^2 + dy^2 + dz^2 - c^2 dt^2 = 0, \quad (1.2.12)$$

给左边取个名字叫线元平方 (*square of line element*) ds^2 ，这个等式实际上就是如下写法：

$$ds^2 = dx^\mu \eta_{\mu\nu} dx^\nu = 0,$$

这里 $\eta_{\mu\nu}$ 是对角矩阵 $\text{diag}(-1, 1, 1, 1)$ 的矩阵元。可见这里一个因式就出现了两个指标 μ, ν 。不过，当多对指标重复时，由乘法的交换、结合及分配律知，多组求和顺序是无所谓的，即 $\sum_{\mu,\nu} = \sum_\mu \sum_\nu = \sum_\nu \sum_\mu$ 。这再一次体现了求和约定对其分量关注的精神，而一般而言是难以混淆的。

我们要求在 \hat{S} 下同样有

$$d\hat{s}^2 = d\hat{x}^\mu \eta_{\mu\nu} d\hat{x}^\nu = 0,$$

这两个量由线性变换所联系，固然也是同阶的无穷小，只能说明二者彼此是成比例的：

$$d\hat{s}^2 = k ds^2, \quad (1.2.13)$$

³² 我们以后会知道，指标的上下代表着矢量、对偶矢量或者逆变、协变的区分，指标的左右代表着缩并作用、基底张量积的顺序。

其中系数 $k \neq 0$ 只与两惯性系的相对速率 $|v|$ 有关, 但不可能与坐标或者时间有关, 否则不同空间点或者不同时刻就不再等价, 这与时空的均匀性相悖; 也不可能与相对速度之方向有关, 否则将与空间的各向同性相悖³³.

用 $\hat{x}^\nu = \Lambda^\nu_\lambda x^\lambda$ 替换掉:

$$d\hat{s}^2 = \Lambda^\mu_\sigma dx^\sigma \eta_{\mu\nu} \Lambda^\nu_\lambda dx^\lambda = 0,$$

关于矩阵的转置, 比如对于 Λ^μ_σ , 我们像这样左右交换, 即对调行列指标为 $(\Lambda^\top)_\sigma^\mu$, 有时简写为 Λ_σ^μ . 则有

$$d\hat{s}^2 = dx^\sigma \Lambda_\sigma^\mu \eta_{\mu\nu} \Lambda^\nu_\lambda dx^\lambda = 0,$$

可见中间的 $\Lambda_\sigma^\mu \eta_{\mu\nu} \Lambda^\nu_\lambda$ 是指标相靠的, 相当于按这个顺序乘出了一个新矩阵, 并且根据比例关系为

$$\Lambda_\sigma^\mu \eta_{\mu\nu} \Lambda^\nu_\lambda = k \eta_{\sigma\lambda}.$$

为了方便, 不妨还原为矩阵写法

$$\Lambda^\top(v) \eta \Lambda(v) = k(|v|) \eta,$$

当然希望物理意义上 $\Lambda(-v)\Lambda(v) = I$, 因此对两遍同时左乘 $\Lambda^\top(-v)$ 并右乘 $\Lambda(-v)$ 有

$$\eta = k(|v|) \Lambda^\top(-v) \eta \Lambda(-v) = k(|v|) k(|-v|) \eta = k^2(|v|) \eta.$$

因此 $k = \pm 1$, 显然应该取 $k(0) = 1$. 此即是说

$$d\hat{s}^2 = ds^2. \quad (1.2.14)$$

这个性质就称作间隔不变性 (*interval invariance*). 如

$$\Lambda^\top \eta \Lambda = \eta \quad (1.2.15)$$

这样的 Λ (再包括原点的平移) 就称作 Lorentz 变换. 为此, 假设这样一个简单情形: x^2, x^3 不变而只观察 x^0, x^1 , 这样 Λ 在实质上就成为了一个二阶矩阵 (对另两个空间坐标而言是恒等变换 I). 于是方程 (1.2.15) 成为了

$$\begin{bmatrix} (\Lambda^0_1)^2 - (\Lambda^0_0)^2 & \Lambda^0_1 \Lambda^1_1 - \Lambda^0_0 \Lambda^1_0 \\ \Lambda^0_1 \Lambda^1_1 - \Lambda^0_0 \Lambda^1_0 & (\Lambda^1_1)^2 - (\Lambda^1_0)^2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix},$$

³³必须注意, 这实际上是除“相对性原理”和“光速不变”以外的“公理”, 因此我们常说从二者出发实际上是不能严格导出 boost 变换的, 更别提整个狭相论, 比如所默认的“线性变换”就并不能严格推出, 我们将在 2.3 节仔细探讨.

这里括号外的“2”当然就表示平方了。显然解有无穷多组，但这无穷多组都可以看作一类关于某实数变量的函数。根据物理意义（或者说那些额外的“公理”），一个 v 就决定了一个变换，因此这个变量一定唯一，而且无论怎么替换，总应该是含有 v 。当然这不是说只含有这个“字母”，实际上也是含有 c 的，但那些都是常数。我们可以先代入试解：由于对角元形似双曲线方程，因此实际上，解总是形如

$$\begin{bmatrix} \cosh \theta & -\sinh \theta \\ -\sinh \theta & \cosh \theta \end{bmatrix}, \quad (1.2.16)$$

但为何不再考虑负的系数呢？我们需要做如下限制：排除时空反演。排除时间反演意味着 $\Lambda^0_0 > 0$ ，排除空间反演意味着 $\Lambda^1_1 > 0$ 。如此对角元为相等正数，再由于方程中 η 的零元，则 Λ 成为对称矩阵，就能保证行列式 $|\Lambda| = 1 > 0$ 。由于 $\sinh \theta$ 前添正号相当于给整个矩阵代入 $-\theta$ ，因此实质上也属于如上形式。这样，解就只有这一种形式。

随后期望寻找 θ, v 的关系。不失一般性，设 \hat{S}, S 初态 $t = \hat{t} = 0$ 时，空间轴亦

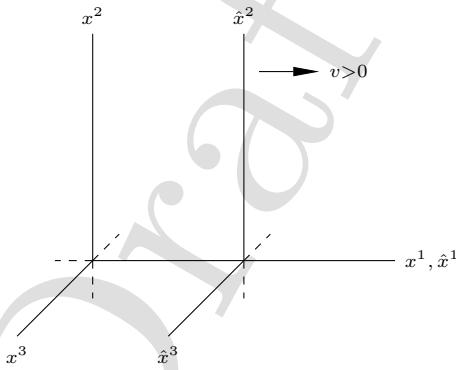


图 1.6: boost 变换示意图

重合，但 \hat{x} 相对于 x 以沿共同 x^1, \hat{x}^1 轴正方向的 v 运动。则

$$\begin{bmatrix} ct \\ 0 \end{bmatrix} = \begin{bmatrix} \cosh \theta & -\sinh \theta \\ -\sinh \theta & \cosh \theta \end{bmatrix} \begin{bmatrix} ct \\ vt \end{bmatrix},$$

此即

$$\begin{bmatrix} ct \\ vt \end{bmatrix} = \begin{bmatrix} \cosh \theta & \sinh \theta \\ \sinh \theta & \cosh \theta \end{bmatrix} \begin{bmatrix} ct \\ 0 \end{bmatrix},$$

上下两行一比即得

$$v = c \tanh \theta, \quad (1.2.17)$$

满足这种关系的 θ 称为快度。进而 $\cosh \theta = \frac{1}{\sqrt{1-v^2/c^2}}$, $\sinh \theta = \frac{v/c}{\sqrt{1-v^2/c^2}}$ 。这样也就得到了 boost 变换。

1.2.4 质能方程

相对性原理除了引出新的时空观念外，还导致了质量和能量的深刻重组。同年9月27日，Einstein专门发表了另一篇论文《物体的惯性同它所含的能量有关吗？》[10]，以讨论力学中的能量问题：他企图讨论Newton第二定律推广至这个新时空观时发现了如下说法：

一个物体的能量在自然意义上可以定义为

$$E = \frac{mc^2}{\sqrt{1 - v^2/c^2}}. \quad (1.2.18)$$

其中 m 为物体的质量。Einstein 为使 Newton 定律推广而为此捏造了一个“动质量”概念 $m_k = m/\sqrt{1 - v^2/c^2}$ ，此外又称“静能”为 $E_0 = mc^2$ 。现在人们一般不再使用这种“余孽”定义，而选择将其扔进历史的纸篓桶里。注意，在相对论中，这代表着质量和能量的直接等同，而非所谓的“转化”——而且在后面我们将会解释这样一种单位值，它使得 $c = 1$ 是无量纲量，那么将直接得到 $E_0 = m$ ，此即说任何静止物体当然都含有“能量”，毕竟其有内蕴的“质量”。这意味着能量、质量等物理量或许是某种更深刻事物的不同侧面。

1.2.5 Minkowski 对狭相的理解

正是这种“深刻事物”的“不同侧面”引导狭相成为了一门几何学：Einstein 对狭相的理解是“代数”的，而 Minkowski 于 1908 年给 Einstein 的代数理论提供了深藏其下的“几何”学解释。

如前文所述的一切“奇特”现象，皆是“光速”与相关运动定律在任意惯性系不变所带来的，究其直觉，就是“同时的相对性”。这使我们构建了一个颠覆性的先进观念：那就是要区分“绝对”和“相对”！虽然确实可以说在物体相对静止系下所测数据为原有的“实际值”，而在运动系下所测为“观测值”

然而，讨论物理量必经由实验测量，而从测量角度而言，没有任何种的测量结果能鹤立鸡群，独特地作为“原有的”、“实际的”，尽管确实在不同参考系中测量“同一对象的物理量”，却得到“不同结果”。这意味着，我们有必要认为或许测量值本身是可以相对变化的，而不变的应该就是那个“深刻事物”。这些测量只是该事物的不同侧面。

Minkowski 发现，“间隔不变性”实际不仅对光速是成立的。若对 boost 变换求微分（就单个变换而言， v, c 当然是常数），那么我们将发现 $d\hat{s}^2 = ds^2$ 是一个恒等式。当然，在此前求得 $k = 1$ 时已经暗示了这一发现。如果定义这样一种物理量，它是物体在其静止系所测的本征时间，称固有时 (*proper time*) τ ，则根据定义取空间

分量为零，则

$$ds^2 = -c^2 d\tau^2, \quad (1.2.19)$$

那么就有可能通过给空间再“增添”一个时间维度，使我们所讨论的物理背景为一个4维的时空连续统 (*space-time continuum*)，简称时空 (*spacetime*)³⁴，而找到那些“深刻事物”。

或许负数开方总是会让读者有些不适，但不必担心，我们不需要讨论 ds 的虚实。为了消除这一点，部分书籍会选择令 $\eta = \text{diag}(1, -1, -1, -1)$ ，从而使得 $ds^2 = c^2 d\tau^2 > 0$ 。如果定义 η 的 Lorentz 号差 (*signature*) 为其矩阵的迹，则这种选择就是 $\eta^\mu_\mu = -2$ ，而本书一般选择 +2。更有甚者选择将时间分量排在最后作为第4分量，这样就有 $\text{diag}(1, 1, 1, -1)$ ，但这并不影响号差。无论怎样，这些选择纯粹只是偏好，不会带来任何实质区别³⁵。

我们知道，Newton 时空观认为空间平直线性，其上运动可由三维速度描述，因此时间是割裂于空间的自变量。但正如前文讨论，狭相时空观已经如此地暗示了时间和空间的深刻联系，那么 Minkowski 对 Einstein 理论的理解是非常自然的。

可见，光速不变不仅保证了间隔不变性，甚至还保证了固有时的存在性。我们可以说，正是有光速不变，我们才有定义均匀流逝“标准钟”的可能，使得这个固有时是绝对事物，就像质量一样，而真正成为可以表征物体自身固有属性的内禀 (*intrinsic*) 物理量。

对初学者来说，关键总是在于究竟如何想象时空。还记得 x - t 图吗？为便于理解，我们就以这样一张2维白纸 \mathbf{R}^2 举例，其上只有两个任意选取的自然坐标。当选定某种原点及其坐标系时，坐标数字才具有位置的意义。图上的点都被赋予了两个坐标，因此也称作事件 (*event*)，因为为了叙述一个生活意义上的事件，其实和写记叙文没什么区别：那便是地点和时间³⁶。经典力学中，两个事件 $(t, x), (t, x')$ 相当于自然语言的“在 x 处及另一不同 x' 处于相同时刻 t 发生的事”——Newton 习惯像这样把“处”和“时刻”分割开来看，但在狭相中很容易找到一个 Lorentz 变换使得这两个事件不再具有相同的 t 。

我故意没有提及“人物”：不同的人当然可以在同一个时空点干不同的“事”，但为了区分这种“事”，只需要说清楚对应的是谁即可。实际上，如果我们定义世界线 (*worldline*) 就是物体所经历的一系列事件构成的曲线，那在一部小说中，当然也就

³⁴Minkowski 于 1908 年在德国科隆的自然哲学家大会 (*Naturforscher Versammlung*) 上以“空间与时间”为题发表了一篇演讲。它的开篇是：“我打算提给你们的空间和时间的概念，是从实验物理学的土壤上生长出来的，它的力量正在于此。这种观念是革命性的。由于有了这个观念，空间本身和时间本身，注定了会逐渐消逝，成为仅仅是影子，只有二者的联合才会保持为独立的现实性。”该演说又发表为了论文 [11]。

³⁵可能会带来哲学区别的是这样的“Wick 技巧”：如果设 $x^0 = ict$ ，则 $\eta = I$ ，我们后续会知道，这意味着区别于闵氏几何的欧氏勾股定理。那么 boost 变换就变成了 x - ict 图的“旋转”。一般除讨论量子力学复结构的哲学外，不建议使用这种做法。

³⁶至于所谓的起因、经过和结果，对应着的是动力学以及因果结构的内容，要在很后面的章节中才会讲到了。

可以区分谁干了什么事，因为不同的事在不同的“故事线”上。因此，我们就不再专门设置这样一个鸡肋的概念，而把时空的点就称作事件，而把物体等同于一条世界线。

考虑到我们所处的是三维空间，暂且不严谨地说，像这样的事件 (ct, x, y, z) 所构成的集合就叫时空。这里乘上 c 的理由如前文所述。等等？那我们在推导 boost 时，矢量 $(ct, x, y, z)^\top$ 不正对应着时空坐标吗？的确，我们那时已经隐约窥见了时空概念。并且还可以告诉读者一个好消息：矢量正好属于我们所寻觅的“深刻事物”，其不同侧面正是分量。如前文所述，对 boost 的微分将会得到间隔不变性，我们这里就尝试一下把 x^μ 对 τ 求导会怎么样：

$$\left(\frac{dx^0}{d\tau}, \frac{dx^1}{d\tau}, \frac{dx^2}{d\tau}, \frac{dx^3}{d\tau} \right)^\top,$$

似乎得到某种速度矢量。但另一方面，间隔不变性意味着

$$-\left(\frac{dx^0}{d\tau}\right)^2 + \left(\frac{dx^1}{d\tau}\right)^2 + \left(\frac{dx^2}{d\tau}\right)^2 + \left(\frac{dx^3}{d\tau}\right)^2 = -c^2 \quad (1.2.20)$$

是恒等的。也就是说，我们最好认为至少这个矢量是个“几何绝对”的东西³⁷，因为其“模长”恒为光速。而相对体现在其分量的投影上。我们知道三维空间里速度是

$$\vec{v} = \frac{\sqrt{dx^2 + dy^2 + dz^2}}{dt}, \quad (1.2.21)$$

上下同除 $d\tau$ 得

$$\vec{v} = \frac{\sqrt{\left(\frac{dx}{d\tau}\right)^2 + \left(\frac{dy}{d\tau}\right)^2 + \left(\frac{dz}{d\tau}\right)^2}}{dt/d\tau} = \frac{c\sqrt{(dt/d\tau)^2 - 1}}{dt/d\tau},$$

因此实际上

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - v^2/c^2}} = \cosh \theta. \quad (1.2.22)$$

我们可以单独给它取个名字叫 *Lorentz 因子 (factor) γ* 。可见，相对速率 $|v|$ 的不同所带来的，不外乎就是 t, τ 间相对流逝“快慢”的变化。

不过还没细究一件事：究竟什么是模长呢？虽然坐标可以定位一个事件，但它并没给出几何信息，比如长度、夹角什么的。Minkowski 类比了一下几何学知识，指出狭相实际上给出了几何信息：正是这个“模长”！并且他还强调，间隔不变性已包含了狭相的所有物理学！是的，我们回忆一下，间隔不变性就是指线元

$$ds^2 = dx^\mu \eta_{\mu\nu} dx^\nu$$

³⁷对此，Einstein 曾表示相比于“相对论”这个名字，他更喜欢“绝对论”。

在不同惯性系中是绝对的，这里我们略去了转置记号。而我们之前的所有物理结论确实可从该“公理”推出。上式可看作二次型，类比一下空间的勾股定理，它是

$$d\ell^2 = dx^2 + dy^2 + dz^2 = dx^i \delta_{ij} dx^j,$$

其中 δ_{ij} 是恒等变换 I 的矩阵元。这里有必要解释一下：一般像 i, j 这种拉丁字母默认取遍 1, 2, 3，而希腊字母 α, μ 等默认取遍 0, 1, 2, 3。

可见，要想定义距离，我们有必要重新审视一下几何学。

1.3 怎么描述弯曲几何？

1.3.1 度规

为了测量距离，当然需要一个“度量衡”标准，或者说一把“规尺”。在几何学中，定义距离的是其相应的度规 (*metric*) 概念，这个命名十分自然。

假设球面上有两点 P, Q ，它们之间的“距离”可以怎样确定？如果我们定义这个球面是所有使得 $x^2 + y^2 + z^2 = 1$ 的点 (x, y, z) 之集合，那么 P, Q 就是 \mathbf{R}^3 中的点。所以我们可以用勾股定理确定它们的距离，例如 $(1, 0, 0)$ 和 $(0, 1, 0)$ 的距离就是 $\sqrt{2}$ 。然而我们真的想去计算直线段 PQ 的长度吗？这个直线段并不完全躺在球面上，所以用直线段来定义出来的长度似乎“什么也不是”。在球面几何学里，我们知道，可以定义 P, Q 之间的距离为连接这两点，而完全贴于球面上的“最短路径”之长。但无论如何，在球面的局部上总是有勾股定理的。因此我们可以尝试用空间的勾股定理去“诱导 (*induce*)”出在球面上的“勾股定理”。

现在用上微积分的语言。考虑一个 \mathbf{R}^3 上的路径长度。路径可以看成是一个动点 $(x(t), y(t), z(t))$ 之轨迹，假设 t 从 0 开始而于 1 停止。如果 Δt 很小，则 $x(t + \Delta t)$ 可以近似为 $x(t) + \dot{x}(t)\Delta t = x + \Delta x$ ；对于 y, z 同理。所以我们知道附近两点的距离就相当于“最短的直线长” $\Delta t \sqrt{\dot{x}^2(t) + \dot{y}^2(t) + \dot{z}^2(t)}$ ，或者说 $\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$ 。要计算光滑路径的长度，就在此路径上取很多很多点，而相邻二者非常接近。将这些线元加起来，可以给出一个很好的近似。取得越多，逼近越好。当 $\Delta t \rightarrow 0$ 时，所有沿路径的距离之和会得到一个积分，即

$$\ell = \int_0^1 \sqrt{\dot{x}^2(t) + \dot{y}^2(t) + \dot{z}^2(t)} dt.$$

如果我们换成微分记号 $\frac{d}{dt}$ ，则可以改写成

$$\int \sqrt{dx^2 + dy^2 + dz^2} = \int d\ell.$$

这样我们发现，想要诱导球面上的线长 ℓ ，只需要讨论空间线元平方 $d\ell^2$ 在球面上的表达式即可，有时在数学中叫 *Riemann* 度量，但是就叫它度规也未尝不可。为了找

到球面度规，我们只需要如下的球坐标变换即可：

$$\begin{cases} x = r \sin \theta \cos \phi, \\ y = r \sin \theta \sin \phi, \\ z = r \cos \theta. \end{cases} \quad (1.3.1)$$

代入到 $dx^2 + dy^2 + dz^2$ 之中，可以计算得到

$$d\ell^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2, \quad (1.3.2)$$

设球面在球坐标下是 $r = R$ ，则

$$d\ell^2 = R^2(d\theta^2 + \sin^2 \theta d\phi^2),$$

有时候会把括号写作 $d\Omega^2$ ，这里 Ω 就是所谓的立体角 (*solid angle*) 了³⁸。至此，给

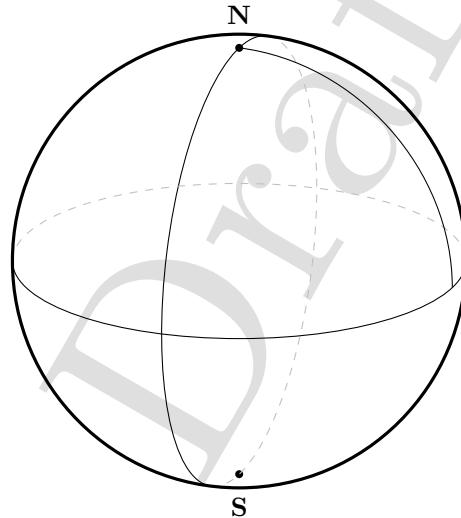


图 1.7: 球面

定球面上的曲线，即经纬度参数式或关系式，那么就可利用上述线元计算球面距离。比如，假定曲线是整个赤道 ε ，此即 $\phi = 2\pi t$ 而 $\theta = \pi/2$ ，则线元化简为 $d\ell = R d\phi$ ，积分得到 $\int_{\varepsilon} d\ell = 2\pi R$.

然而，刚才只是应用这个公式作为开始，这个例子还没有显示出度规的真正好处。在几何学中，度规的选择直接决定了该体系内的几何知识，因此上述所讨论的都

³⁸在单位球中，以圆心为顶点延伸出一个锥体（一般选圆锥），则所截出的球面面积即为立体角，范围显然是 $[0, 4\pi]$ 。

称作欧氏几何 (*Euclidean geometry*). 历史上为论证平面几何“第五公理”的独立性，曾为此构造了另一种度规：

$$d\ell^2 = \frac{dx^2 + dy^2}{1 - x^2 - y^2},$$

为使距离为实数，定义域要求 $x^2 + y^2 < 1$ ，而这个区域是一块不含边界的单位圆盘. 类似地，路径在该度规下的长度为

$$\int_0^1 \sqrt{\frac{\dot{x}(t)^2 + \dot{y}(t)^2}{1 - x(t)^2 - y(t)^2}} dt.$$

当接近圆盘边缘时，这个“距离”比欧氏距离会越来越大. 这个度规所对应的体系称作双曲几何 (*hyperbolic geometry*) 的圆盘模型.

无论怎样，上述例子具有这样一种共性：由于在局部上线元都是“直”的，因此距离的表达式必须是二次型：

$$d\ell^2 = E(x, y)dx^2 + 2F(x, y)dx dy + G(x, y)dy^2,$$

使线元均为正只需使得 $EG > F^2$ 即可. 交叉项只是因为两个坐标的线元不一定“垂直”，但总归还是个“勾股定理”. 此即说，无论这个几何多么扭曲，至少在局部上要还原至欧氏几何. 当然，最好这三个函数也具有一些光滑性.

这个定义可以直截了当地推广至更高维的几何，因为勾股定理的高维推广仍然是二次型. 对于 n 维而言，它就是

$$d\ell^2 = g_{\mu\nu} dx^\mu dx^\nu, \quad (1.3.3)$$

其中 $g_{\mu\nu}$ 是 $n \times n$ 的对称矩阵 g 的矩阵元，而且还可能是各坐标 x^μ 的函数. 这个 g 就叫度规，其分量一般而言可以直接指代矩阵本身，故也可以叫做度规. 后面我们会知道，把这个线元平方叫做度规，也是可以的. 在这种情况下，为了迎合欧氏几何的直观，我们自然要求度规是正定的 (*positive definite*)，即 $g_{\mu\nu} > 0$ ，可作强调地称其 *Riemann* 度规；否则就叫伪 *Riemann* 度规 (*pseudo Riemannian*)，包括负定的 (*negative definite*)，以及不定的 (*indefinite*) 即正负都有的.

度规不仅仅是定义了距离概念，实际上也定义了“夹角”。如果我们选择带有直线轴的坐标系，那么 x^μ 的均匀变化将给出真正的直线距离 Δx^μ ，因此

$$\Delta\ell^2 = g_{\mu\nu} \Delta x^\mu \Delta x^\nu,$$

如果我们把坐标差值理解为矢量模长，则实际上这就给出了某种实数

$$g_{\mu\nu} v^\mu u^\nu.$$

直接使用直角坐标系，则 $g_{\mu\nu}$ 就成为了欧氏度规 δ_{ij} ，那么

$$v^i \delta_{ij} u^j = v_j u^j.$$

可见这正是列矢量再左乘行矢量，也即平面几何中内积 (*inner product*) 的定义，当然也就给出了夹角信息。因此我们说，实际上定义度规就是在定义在该几何体系中，两个矢量的内积是什么。既然度规在一般意义上决定的是两个矢量的“乘法”，那我们不妨就将行列指标均写在下方，以遵循求和约定。

1.3.2 阁氏几何

对比一下间隔不变性和度规的定义，可以迅速发现这样一件事实： η 可以看作度规，而且是个伪 Riemann 度规！受到如上启发，我们将线元写作

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \quad (1.3.4)$$

而所谓的间隔就是在直角坐标系下时空直线“距离”，其平方即 Δs^2 ，线元不变固然也就意味着“间隔”不变。既然度规决定了该体系的所有几何知识，而一方面间隔不变决定了狭相的所有物理知识，我们就可以说，狭相可以等价于一门几何学，这门几何就按“主人”取名为 Minkowski 几何，简称阁氏几何，与欧氏几何相对照。 η 就叫做阁氏度规。

可见，狭相中“合法”的参考系变换就相当于几何里的保度规变换 (*isometry*)，这样当然有间隔不变性：

$$\eta_{\hat{\mu}\hat{\nu}} d\hat{x}^{\hat{\mu}} d\hat{x}^{\hat{\nu}} = \eta_{\mu\nu} dx^\mu dx^\nu,$$

由于坐标变换满足全微分 $dx^\mu = \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} d\hat{x}^{\hat{\mu}}$ （指标可任选），因此对任意 $\hat{x}^{\hat{\mu}}$ ，必有

$$\eta_{\hat{\mu}\hat{\nu}} = \eta_{\mu\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}}, \quad (1.3.5)$$

此即 η 的坐标变换式 (*formula of coordinate transformation*)，分母下方是上指标，而这相当于分子下指标，因此这样写是指标平衡的；而一方面， $\Lambda^\mu_{\hat{\mu}} = \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}}$ ，因此该式就是

$$\eta_{\hat{\mu}\hat{\nu}} = \Lambda^\mu_{\hat{\mu}} \eta_{\mu\nu} \Lambda^\nu_{\hat{\nu}}.$$

我们说，Lorentz 变换就是狭相中保持度规 η 不变的所有变换，称 η 具有 *Lorentz* 协变性。

1.3.3 Einstein 对其之理解

物理过程的时间和一支杆所度量的长度，是时空自然的几何结构的两个相互联系的侧面。这个观念极具革命性。甚至在这趟通往广相旅程的最初阶段，Einstein 本人是拒绝时空概念的，宁可保持一个“空间”作为确定的实体，而只需考虑到这个空间里的同时性是相对的。然而，随后在 1907~1909 年间寻求等效原理的数学表述时，才发现引力这一最深刻的内涵是很难在狭义相对论的框架下，尤其还带着割裂时空

观，旅途是步步艰辛，极为困难，不得不重新考虑用“时空连续统”研究非惯性系的物理，再借助闵氏几何的思想，逐渐接受了时空整体观，寻求将引力纳入到时空几何学之中的解释。

“不受力”却仍然改变了相对轨迹，这作何解释？原来，的确可以考虑“舞台”具有复杂几何的可能：在平直的普通“舞台”中，“舞者”的惯性运动当然是直线或静止。我们可以说这是其“固有的自然方向”。一旦出现了运动状态之改变，Newton 的解释是受力³⁹，而 Einstein 选择用“舞台”几何的扭曲来解释，即弯曲几何当然会改变平直几何的性质，这包括改变“固有的自然方向”。

此处有个小细节——可不可以是空间弯曲呢？我们不妨这样想：“舞者”的惯性运动“要么是曲线要么是个点”，这只是默认了“舞台”是空间。但果真如此吗？我的意思是，如果我们想要研究的是“某条曲线”，那是否应该想办法把“点”的情况也转化成一条线？这样我们才能统一地描述“惯性运动”。

其实，只单独地考虑空间弯曲性对引力的解释，Poincaré 尝试过但已经失败。或许确实如 Minkowski 所说的那样，要想统一描述，可能得在空间里再加一条时间维度：毕竟一个静止物体的 x - t 图象当然会是一条直线，匀速运动时也还是。平直的时空中，不受力的自然世界线当然就是直线了。这种自然的世界线有个几何学里的名字——测地线 (*geodesic*)，因此弯曲时空就是在改变测地线的形状罢了。

时空这个新概念带来了很多有趣的思考。譬如，由于静止物体的世界线仍然是一条直线，我们可以说，哪怕物体在空间上静止，其也在时间上“运动”，即不断地往未来迁移，这种情况下 4 维速度的分量全部给予了时间⁴⁰。我们 also 可以说，物体的静能总是向着未来“传递”，但在相对运动的参考系中，这些能量会部分地传递向空间，于是我们测量到了动能，因为它在相对运动。

虽然等效原理一般的含义多少有些模糊，自 Einstein 引入以来一直处于辩论之中，但由此前讨论可见，甚至在没有引力的平直时空中，知道各对象在任意弯曲的坐标系下是什么样子，是至关重要的。这就是说，要从闵氏时空的直角坐标系变到最一般的甚至是弯曲时空中的坐标系。这样做就是要把 η 换成一个一般的 g ，取名为 *Lorentz* 度规，并且要求在时空的每一点处， g 都可以找到坐标基使得其分量等于 $\eta_{\mu\nu}$ 。像这样的一点处的坐标基就称作局部惯性坐标系 (*local inertial coordinate system*)，这样便可支撑起等效原理。形象地说，*Lorentz* 度规局部看来和闵氏度规一样，正如 Riemann 度规局部看来和欧氏度规一样。

对于在引力作用下的粒子，总存在一个与之自由降落 (*free fall*) 的局部惯性系

³⁹ 此处语境只考虑引力场。

⁴⁰ 那种“断掉”的世界线在广相中称为不可延世界线，暂时还遇不到。

ξ^α , 那么它在该坐标系里的世界线满足“不受力”的直线方程⁴¹, 也就是:

$$\frac{d^2\xi^\alpha}{d\tau^2} = 0, \quad (1.3.6)$$

这里我们把参数 τ 看成固有时, 但它一定要满足 $ds^2 = -c^2 d\tau^2$ 吗? 设曲线是 α 在任意坐标系的表达式是 $x^\mu(\tau)$, 则其速度矢量就是 $U = \frac{dx^\mu}{d\tau} e_\mu$, 这里 e_μ 是坐标系的基矢。 $ds^2 = -c^2 d\tau^2$ 要求速度模长平方就是不变的

$$\eta_{\alpha\beta} U^\alpha U^\beta = -c^2,$$

因而我们才说速度仿佛是“深刻事物”, 而我们借助参考系观测其不同情况的分量. 但使得速度矢量模长不变的所谓仿射参数 (*affine parameter*) 当然有很多任意选择, 而且它们实际上都是固有时的线性函数, 对应的矢量模长也只是 $-c^2$ 的倍数. 我们当然也可以有一个更简单的做法: 选择参数 $\tau \mapsto c\tau$, 就像 $x^0 = ct$ 那样! 这样 τ 的量纲就彻底成了空间长度及线长的单位, 即 $ds^2 = -d\tau^2$, 模长也就“归一化”了:

$$\eta_{\alpha\beta} U^\alpha U^\beta = -1.$$

这样我们的物理方程中就将再也不会出现 c , 因为它全部被收进了时间里 (就像万有引力常数 G 收进场源质量里), 不影响其它任何物理量的计算. 但有一个更为深刻的想法: 这实际上是在选取不同的单位制! 令 $c = 1$ (甚至 $G = 1$) 这个做法是在利用额外的自由度, 因为根本不会有问 $1s$ 等于多少 m . 于是我们说, 这些数据并不需要收进任何一个物理量的定义中, 而是直接收进单位 (的相对关系) 里! 这样的好处是, 可以甚至彻底地统一时空整体. 关于单位制的问题, 我们在 2.3 中还会继续讲解.

回到问题上, 现在看看把它从 ξ^α 换到别的坐标系会怎样:

$$\frac{d}{d\tau} \left(\frac{\partial \xi^\alpha}{\partial x^\mu} \frac{dx^\mu}{d\tau} \right) = \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{d^2x^\mu}{d\tau^2} + \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0,$$

这里只要保证单个因式内至多出现两个上下重复指标, 那多个因式之间存在重复的指标对是没问题的. (因为它们是“加”的关系!) 乘上 $\frac{\partial x^\lambda}{\partial \xi^\alpha}$ 以消除系数, 则有测地线方程 (*geodesic equations*)

$$\frac{d^2x^\lambda}{d\tau^2} + \frac{\partial x^\lambda}{\partial \xi^\alpha} \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0. \quad (1.3.7)$$

可见, 在局部坐标系中的直线就是一般坐标系的测地线, 这里定义 *Christoffel 符号* (*symbol*) (简称克氏符) 为

$$\Gamma^\lambda{}_{\mu\nu} = \frac{\partial x^\lambda}{\partial \xi^\alpha} \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu}. \quad (1.3.8)$$

⁴¹由于等效原理只对一点而言, 因此测地线上的局部坐标系只能恰好在这一条线上, 而不包含附近.

只要我们知道坐标系每一处的克氏符，我们就得到了关于 $x^\mu(\tau)$ 的常微分方程。那这是否意味着 $\Gamma_{\mu\nu}^\lambda$ 就是我们想要找的“引力”呢？毕竟

$$\frac{d^2x^\lambda}{d\tau^2} = -\Gamma_{\mu\nu}^\lambda \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}$$

长得像第二定律似的。答案是否！切记 $\Gamma_{\mu\nu}^\lambda$ 只是坐标系自身的信息（每一处都有这 $4 \times 4 \times 4$ 个实数！），很明显平面上也可以选取坐标系使得克氏符一般不为零，但由于是平面，测地线却依然是直线，当然也就没有“引力”。只是在该坐标系下，直线的表达式变复杂了一些。看来，想要真正地找到引力，我们可能得想想哪些几何对象是不依赖于坐标系的，以排除坐标系选取不同带来的干扰。这很符合我们对相对性原理的追求：要是作为物理对象的引力是依赖于坐标系的，那我们现在身处的是怎样一个独特的“以太”呢？

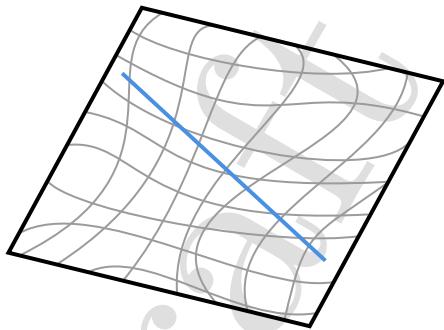


图 1.8: 一个完整的坐标系由处处的基矢决定，而非仅仅只是原点处的坐标轴。图中所示为一张平面上的曲线坐标网格，蓝线为一条测地线。

下面可以看到，式 (1.3.8) 的右方与度规有联系，因此又称作与度规相适配的 (*be associated with metric*) 克氏符。首先不要忘记这样一个细节：上式 $\eta_{\alpha\beta}$ 只是度规在该局部坐标系的分量，而不是一个整体的度规（我们现在处于任意弯曲的白纸上）。设任意坐标系下度规分量是 $g_{\mu\nu}$ ，则

$$d\tau^2 = -\eta_{\alpha\beta} d\xi^\alpha d\xi^\beta = -g_{\mu\nu} dx^\mu dx^\nu,$$

因此

$$g_{\mu\nu} = \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{\partial \xi^\beta}{\partial x^\nu} \eta_{\alpha\beta}. \quad (1.3.9)$$

可见在这种情况下，度规应该与克氏符有关联：对其求导即有

$$\begin{aligned} \frac{\partial g_{\mu\nu}}{\partial x^\lambda} &= \frac{\partial^2 \xi^\alpha}{\partial x^\lambda \partial x^\mu} \frac{\partial \xi^\beta}{\partial x^\nu} \eta_{\alpha\beta} + \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{\partial^2 \xi^\beta}{\partial x^\lambda \partial x^\nu} \eta_{\alpha\beta} \\ &= \Gamma^\rho_{\lambda\mu} \frac{\partial \xi^\alpha}{\partial x^\rho} \frac{\partial \xi^\beta}{\partial x^\nu} \eta_{\alpha\beta} + \Gamma^\rho_{\lambda\nu} \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{\partial \xi^\beta}{\partial x^\rho} \eta_{\alpha\beta} \\ &= \Gamma^\rho_{\lambda\mu} g_{\rho\nu} + \Gamma^\rho_{\lambda\nu} g_{\rho\mu}, \end{aligned}$$

即

$$\frac{\partial g_{\mu\nu}}{\partial x^\sigma} = g_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} + g_{\nu\lambda}\Gamma^\lambda_{\sigma\mu}. \quad (1.3.10)$$

我们当然也可以从中“解”出 $\Gamma^\lambda_{\mu\nu}$: 运用一个以后可能很常用的技巧“指标轮换”，利用对称性加减消元。根据目标，我们做 $\mu \rightarrow \nu \rightarrow \sigma \rightarrow \mu$ 的轮换，然后像这样加减：

$$\frac{\partial g_{\sigma\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma} = 2g_{\sigma\lambda}\Gamma^\lambda_{\mu\nu},$$

即可得到

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2}g^{\sigma\lambda} \left(\frac{\partial g_{\sigma\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \right). \quad (1.3.11)$$

因而称这个 $\Gamma^\lambda_{\mu\nu}$ 为与度规相适配的⁴²，这里 $g^{\sigma\lambda}$ 是其逆矩阵的分量，此即规定 $g_{\mu\sigma}g^{\sigma\lambda} = \delta_\mu^\lambda$.

例 1.1. 验证过球心截圆为测地线，并借其劣弧参数式估算从北京到纽约的最短路径长。已知北京、纽约经纬度分别约 116°E, 40°N, 74°W, 40°N。地球平均半径约 6371 km。

解。笔者“醉翁之意不在酒”，只为展示如何计算出测地线方程，而不通过 3 维空间来分析（上帝的外禀视角），一切只在球面的 ϕ, θ 两个坐标上做文章。先找到 ϕ, θ, τ 关系。思路很清晰： $g_{\mu\nu}, g^{\mu\nu} \rightarrow \Gamma^\lambda_{\mu\nu} \rightarrow$ 研究方程。剩下就开始略微繁琐的硬算。3 维欧式空间中球面度规为 $g_{\theta\theta} = R^2, g_{\phi\phi} = R^2 \sin^2 \theta$ ，因此 $g^{\theta\theta} = R^{-2}, g^{\phi\phi} = R^{-2} \sin^{-2} \theta$ ，求导得

$$g_{\theta\theta,\theta} = g_{\theta\theta,\phi} = 0; \quad g_{\phi\phi,\theta} = R^2 \sin 2\theta, \quad g_{\phi\phi,\phi} = 0.$$

加逗号表示求对应坐标的偏导，这种写法很便于草稿计算。代入公式得到

$$\begin{aligned} \Gamma_{\theta\theta}^\sigma &= \frac{1}{2}g^{\sigma\rho}(g_{\rho\theta,\theta} + g_{\rho\theta,\theta} - g_{\theta\theta,\rho}) = 0, \\ \Gamma_{\theta\phi}^\sigma &= \frac{1}{2}g^{\sigma\rho}(g_{\rho\theta,\phi} + g_{\rho\phi,\theta} - g_{\theta\phi,\rho}) = \frac{1}{2}g^{\sigma\phi}g_{\phi\phi,\theta} = \frac{1}{2}(\delta^{\sigma\phi}R^{-2}\sin^{-2}\theta)(R^2\sin 2\theta) \\ &= \delta^{\sigma\phi}\sin^{-1}\theta\cos\theta = \delta^{\sigma\phi}\cot\theta; \\ \Gamma_{\phi\phi}^\sigma &= \frac{1}{2}g^{\sigma\rho}(g_{\rho\phi,\phi} + g_{\rho\phi,\phi} - g_{\phi\phi,\rho}) = -\frac{1}{2}g^{\sigma\theta}g_{\phi\phi,\theta} = -\frac{1}{2}(\delta^{\sigma\theta}R^{-2})(R^2\sin 2\theta) \\ &= -\delta^{\sigma\theta}\sin\theta\cos\theta. \end{aligned}$$

求得非零分量如下：

$$\Gamma_{\theta\phi}^\phi = \Gamma_{\phi\theta}^\phi = \cot\theta = \frac{\cos\theta}{\sin\theta}, \quad \Gamma_{\phi\phi}^\theta = -\sin\theta\cos\theta.$$

⁴²以上所有情况甚至包括整个经典的广义相对论都默认了这一点，那就是 $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$ 。在讨论曲率几何含义时，我们会知道这意味着挠率 (torsion) 为零，而它当然在更一般的情况下不为零。借此我们可以说，实际上广义相对论只考虑无挠时空 (torsion-free)，它是一个无挠引力理论。



图 1.9: 从北京到纽约的最短航线

因此测地线方程为:

$$\begin{aligned} 0 &= \frac{d^2\theta}{d\tau^2} + \Gamma_{\phi\phi}^\theta \frac{d\phi}{d\tau} \frac{d\phi}{d\tau} = \frac{d^2\theta}{d\tau^2} - \sin\theta \cos\theta \left(\frac{d\phi}{d\tau} \right)^2, \\ 0 &= \frac{d^2\phi}{d\tau^2} + \Gamma_{\phi\theta}^\phi \frac{d\phi}{d\tau} \frac{d\theta}{d\tau} + \Gamma^{\phi\theta} \frac{d\theta}{d\tau} \frac{d\phi}{d\tau} = \frac{d^2\phi}{d\tau^2} + \frac{2\cos\theta}{\sin\theta} \frac{d\theta}{d\tau} \frac{d\phi}{d\tau}. \end{aligned}$$

作旋转变换以降低运算量. 先绕 O 系 x 轴旋转 α 得到 O' 系:

$$\begin{cases} x' = x, \\ z' = y \cos \alpha + z \sin \alpha, \\ z' = -y \sin \alpha + z \cos \alpha. \end{cases} \quad \begin{cases} x = x', \\ z = y' \cos \alpha - z' \sin \alpha, \\ z = z' \sin \alpha + z' \cos \alpha. \end{cases}$$

再绕 O' 系的 z' 轴旋转 β 角度得 O'' 系, 这两系之间的坐标关系为

$$\begin{cases} x'' = x' \cos \beta + y' \sin \beta, \\ y'' = -x' \sin \beta + y' \cos \beta, \\ z'' = z'. \end{cases} \quad \begin{cases} x' = x'' \cos \beta - y'' \sin \beta, \\ y' = x'' \sin \beta + y'' \cos \beta, \\ z' = z''. \end{cases}$$

由此可得 O 系与 O'' 系的坐标关系:

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} \cos \beta & \cos \alpha \sin \beta & \sin \alpha \sin \beta \\ -\sin \beta & \cos \alpha \cos \beta & \sin \alpha \cos \beta \\ 0 & -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

正交矩阵的逆为转置, 则

$$\begin{aligned} \begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \begin{bmatrix} \cos \beta & -\sin \beta & 0 \\ \cos \alpha \sin \beta & \cos \alpha \cos \beta & -\sin \alpha \\ \sin \alpha \sin \beta & \sin \alpha \cos \beta & \cos \alpha \end{bmatrix} \begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} \\ &= \begin{bmatrix} \cos \beta & -\sin \beta & 0 \\ \cos \alpha \sin \beta & \cos \alpha \cos \beta & -\sin \alpha \\ \sin \alpha \sin \beta & \sin \alpha \cos \beta & \cos \alpha \end{bmatrix} \begin{bmatrix} R \sin \theta'' \cos \phi'' \\ R \sin \theta'' \sin \phi'' \\ R \cos \theta'' \end{bmatrix}, \end{aligned}$$

此即

$$R \begin{bmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{bmatrix} = R \begin{bmatrix} \sin \theta'' \cos(\phi'' + \beta) \\ \sin \theta'' \sin(\phi'' + \beta) \cos \alpha - \cos \theta'' \sin \alpha \\ \sin \theta'' \sin(\phi'' + \beta) \sin \alpha + \cos \theta'' \cos \alpha \end{bmatrix},$$

因此

$$\cos \theta = \sin \theta'' \sin(\phi'' + \beta) \sin \alpha + \cos \theta'' \cos \alpha,$$

$$\tan \phi = \tan(\phi'' + \beta) \cos \alpha - \frac{\cot \theta'' \sin \alpha}{\cos(\phi'' + \beta)}.$$

O 系任意过球心圆弧都可用 O'' 系的赤道或经线描述. 不妨选择赤道, 则 $\theta'' = \frac{\pi}{2}, \phi'' = a\tau + b$:

$$\cos \theta = \sin(a\tau + b + \beta) \sin \alpha = \sin \alpha \sin \varphi',$$

$$\tan \phi = \tan(a\tau + b + \beta) \cos \alpha = \cos \alpha \tan \varphi',$$

其中 $\varphi' = a\tau + b + \beta$, 即

$$\theta(\tau) = \arccos(\sin \alpha \sin \varphi'),$$

$$\phi(\tau) = \arctan(\cos \alpha \tan \varphi').$$

可验证其满足测地线方程. 从球坐标纬度到地理纬度应做 $\theta \mapsto \pi/2 - \theta$ 的替换:

$$\theta(\tau) = \frac{\pi}{2} - \arccos(\sin \alpha \sin \varphi'),$$

设北京 $(\phi(0), \theta(0))$ 、纽约 $(\phi(1), \theta(1))$. 设航线为曲线 $\alpha = (\phi(\tau), \theta(\tau))$, 代入数据以确定常数 (参数 τ 可看意愿选取). 沿航线的距离为

$$\ell(\alpha) = R \int_{\alpha} \sqrt{d\theta^2 + \sin^2 \theta d\phi^2} = R \int_0^1 \sqrt{\dot{\theta}^2 + \sin^2 \theta \dot{\phi}^2} d\tau.$$

最终计算积分得距离 $\ell \approx 11062 \text{ km}$.

□

1.3.4 Lorentz 几何

把固定的 η 换成弯曲时空中动态的 g , 这个深刻的洞察把 Einstein 从闵氏几何带到了 Lorentz 几何. 在狭相中, Einstein 把相对性原理表述为: 所有惯性观者对物理定律陈述等价, 而 Minkowski 把相对性原理用几何转述为: 物理学的方程只有包含闵氏几何量, 才与时空有关联. 很自然就会想到要寻求这个原理的推广, 而一个适当的说法就立即显现出来了: 物理学只有通过与度规自然相关的几何量去涉及时空. 更准确地说:

相对性原理 物理学方程只有包含仅由 g 定义的几何量才能与时空发生联系. [4]

1912 年左右 Einstein 开始从狭相转向广相, 遇到了数学上的重大困难. 例如 Newton 理论中引力位势是一个函数, 而在 Einstein 的考虑中却应该是 10 个函数 (实际上就是度规的 10 个分量), 这是古典数学无法解决的. 为了推广闵氏几何, 使之成为 Lorentz 几何, Einstein 求助于一位出生于匈牙利的瑞士数学家 Marcel Grossmann (1878-1936)——他是 Einstein 在苏黎世高工 (ETH) 的同班同学和朋友. 他们选择了仿造 Riemann 几何的蓝图, 并意外发现所需数学工具非常齐备. Einstein 说“这几乎就像是命中注定的工具”. 为此, Grossmann 帮助他研读 Riemann 的著作, 并熟悉其它大数学家在微分几何上的工作: 19 世纪末和 20 世纪初, 由 Riemann、Bianchi(1856-1928, 意大利)、Christoffel(1829-1900, 德国)、Ricci⁴³ (1853-1925, 意大利) 和 Levi-Civita(1873-1941, 意大利) 这样一批数学家, 就 Riemann 几何发展了许多概念, 它们都可以直接移到 Lorentz 几何里来. 这就是我们这一章在讲的内容. 而后他们便成功抓住了广相数学基础的要害. 这样的合作成为一件美谈⁴⁴.

⁴³在他和他的学生 Levi-Civita 合写的一本出版于 1927 年的名著《绝对微分学》(以后会解释, 所谓绝对微分学就是张量计算和微分几何) 中, 他把自己的名字 Gregorio Ricci-Curbastro 简写为 Ricci, 所以后人就这样称呼他, 一些有关的名词, 如 Ricci 流、Ricci 曲率等等也这样命名, 这多少造成一些误会.

⁴⁴为纪念这个合作, 从 1975 年开始, 大约每 3 年就举行一次 Marcel Grossmann 国际会议, 讨论与广相有关的物理特别是数学物理的进展. 一部关于 Einstein 的学术历程的传记 [12] 对 Einstein 创立广相的思想历程 (包括他与 Grossmann 的合作) 作了极为精彩的解说, 对了解这段历史大有助益.

协变性

受 η 协变性的启发，可以推广这样一件事： g 也应该在更加随意的坐标系中这样变换：

$$\hat{g}_{\hat{\mu}\hat{\nu}} = g_{\mu\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}}, \quad (1.3.12)$$

就可以说 g 是 *Lorentz* 协变的、无坐标依赖的 (*coordinate independent*) 或者内禀的，就像 η 那样。而度规的逆所满足的关系可用升降指标得

$$\hat{g}^{\hat{\sigma}\hat{\lambda}} = g^{\sigma\lambda} \frac{\partial \hat{x}^{\hat{\sigma}}}{\partial x^\sigma} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda}. \quad (1.3.13)$$

标量函数 f 仅为时空事件本身附带的实数，因而与坐标系无关，即使从 η 推广到 g 后，它也没有什么坐标变换的必要。但矢量 U 会怎样呢？不妨构造这样一个标量，那就是内积 $g_{\mu\nu} U^\mu U^\nu$ 。标量没什么分量的说法，在任何坐标系都是一样的，因此

$$g_{\mu\nu} U^\mu U^\nu = \hat{g}_{\hat{\mu}\hat{\nu}} \hat{U}^{\hat{\mu}} \hat{U}^{\hat{\nu}},$$

根据 g 的协变性，说明矢量分量应满足

$$\hat{U}^{\hat{\mu}} = \frac{\partial \hat{x}^{\hat{\mu}}}{\partial x^\mu} U^\mu, \quad (1.3.14)$$

显然这些新的分量 $U^{\hat{\mu}}$ 不仅依赖于老分量 U^μ ，还依赖于其起点。推广到弯曲时空时，我们就不能再像平直时空那样，直接把一点的矢量空间 (*vector space*) 等同于时空本身了⁴⁵，因为这就相当于随便平移矢量，而后将看到弯曲时空中平移是很微妙的概念。矢量又可写作 $U = U^\mu e_\mu$ ，因此坐标基的变换应当满足：

$$\hat{e}_{\hat{\mu}} = \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} e_\mu, \quad (1.3.15)$$

可见这在狭相中也是说得通的。

再如更多指标的物理量，比如克氏符，也可按照类似思路构造标量，比如说

$$f = \Gamma^\lambda_{\mu\nu} U^\mu U^\nu U^\rho g_{\rho\lambda},$$

因此

$$\hat{\Gamma}^{\hat{\lambda}}_{\hat{\mu}\hat{\nu}} = \Gamma^\lambda_{\mu\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda},$$

当然，这只能视作相应于它协变性的判据，但究竟成立与否还需从定义证明或证否。下面证明克氏符实际上不协变！利用定义

$$\Gamma^\lambda_{\mu\nu} = \frac{\partial x^\lambda}{\partial \xi^\alpha} \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu},$$

⁴⁵一点的矢量空间就是指该点所有矢量所构成的集合，由于其矢量的线性性质，这个集合实际上就是个 \mathbf{R}^4 ；而另一方面，狭相的平直时空也是个 \mathbf{R}^4 。

容易得到其在 \hat{S} 系下 $\hat{\Gamma}$ 同原来 Γ 的关系

$$\hat{\Gamma}^{\hat{\lambda}}_{\hat{\mu}\hat{\nu}} = \Gamma^\lambda_{\mu\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} + \frac{\partial^2 x^\lambda}{\partial \hat{x}^{\hat{\mu}} \partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda}. \quad (1.3.16)$$

可见由于多出来一个 $\frac{\partial^2 x^\lambda}{\partial \hat{x}^{\hat{\mu}} \partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda}$, $\Gamma^\lambda_{\mu\nu}$ 不是 Lorentz 协变的. 这是可预料的, 因为此前已定性说明过克氏符是依赖于坐标系的。

对于其它任意情况, 确实可继续用度规、标量的协变性去推演, 但其实在结果上, 坐标变换就是满足指标平衡的偏微分乘积. 这样在其它情况下就可以快速地写出判据了. 只要将该性质代入定义之中, 即可检验相应量的协变性.

与度规相适配的协变导数 ∇

现在我们要说明这样一件事: 标量 f 直接对坐标 x^ν 的求偏导当然是协变的:

$$\frac{\partial f}{\partial x^\nu} = \frac{\partial f}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\nu}}}{\partial x^\nu}.$$

矢量有分量和基矢之分. 若直接对其分量求偏导, 一般情况下会不协变:

$$\frac{\partial U^\mu}{\partial x^\nu} = \frac{\partial}{\partial x^\nu} \left(\hat{U}^{\hat{\mu}} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \right) = \frac{\partial \hat{U}^{\hat{\mu}}}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} + \hat{U}^{\hat{\mu}} \frac{\partial^2 x^\mu}{\partial \hat{x}^{\hat{\mu}} \partial x^\nu}.$$

欲从其构造出一个协变的导数来。借助于克氏符 $\Gamma^\mu_{\sigma\nu}$, 考察

$$\begin{aligned} \Gamma^\mu_{\sigma\nu} U^\sigma &= \left(\hat{\Gamma}^{\hat{\mu}}_{\hat{\sigma}\hat{\nu}} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial \hat{x}^{\hat{\sigma}}}{\partial x^\sigma} \frac{\partial \hat{x}^{\hat{\nu}}}{\partial x^\nu} + \frac{\partial^2 \hat{x}^{\hat{\mu}}}{\partial x^\sigma \partial x^\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \right) \frac{\partial x^\sigma}{\partial \hat{x}^{\hat{\alpha}}} \hat{U}^{\hat{\alpha}} \\ &= \hat{\Gamma}^{\hat{\mu}}_{\hat{\sigma}\hat{\nu}} \hat{U}^{\hat{\sigma}} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial \hat{x}^{\hat{\nu}}}{\partial x^\nu} + \frac{\partial^2 \hat{x}^{\hat{\mu}}}{\partial \hat{x}^{\hat{\sigma}} \partial x^\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \hat{U}^{\hat{\sigma}} \end{aligned}$$

将其加在原本并不协变的导数 $\frac{\partial U^\mu}{\partial x^\nu}$ 上做修正, 使之得到一个具有协变性的导数:

$$\begin{aligned} \frac{\partial U^\mu}{\partial x^\nu} + \Gamma^\mu_{\sigma\nu} U^\sigma &= \left(\frac{\partial \hat{U}^{\hat{\mu}}}{\partial \hat{x}^{\hat{\nu}}} + \hat{\Gamma}^{\hat{\mu}}_{\hat{\sigma}\hat{\nu}} \hat{U}^{\hat{\sigma}} \right) \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial \hat{x}^{\hat{\nu}}}{\partial x^\nu} + \left(\frac{\partial^2 x^\mu}{\partial \hat{x}^{\hat{\sigma}} \partial x^\nu} + \frac{\partial^2 \hat{x}^{\hat{\mu}}}{\partial \hat{x}^{\hat{\sigma}} \partial x^\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \right) \hat{U}^{\hat{\sigma}} \\ &= \left(\frac{\partial \hat{U}^{\hat{\mu}}}{\partial \hat{x}^{\hat{\nu}}} + \hat{\Gamma}^{\hat{\mu}}_{\hat{\sigma}\hat{\nu}} \hat{U}^{\hat{\sigma}} \right) \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial \hat{x}^{\hat{\nu}}}{\partial x^\nu} + \frac{\partial}{\partial \hat{x}^{\hat{\sigma}}} \left(\frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial \hat{x}^{\hat{\mu}}}{\partial x^\nu} \right) \hat{U}^{\hat{\sigma}} \\ &= \left(\frac{\partial \hat{U}^{\hat{\mu}}}{\partial \hat{x}^{\hat{\nu}}} + \hat{\Gamma}^{\hat{\mu}}_{\hat{\sigma}\hat{\nu}} \hat{U}^{\hat{\sigma}} \right) \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial \hat{x}^{\hat{\nu}}}{\partial x^\nu} + \frac{\partial \delta_\nu^\mu}{\partial \hat{x}^{\hat{\sigma}}} \hat{U}^{\hat{\sigma}} \\ &= \left(\frac{\partial \hat{U}^{\hat{\mu}}}{\partial \hat{x}^{\hat{\nu}}} + \hat{\Gamma}^{\hat{\mu}}_{\hat{\sigma}\hat{\nu}} \hat{U}^{\hat{\sigma}} \right) \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial \hat{x}^{\hat{\nu}}}{\partial x^\nu}. \end{aligned}$$

称

$$\nabla_\nu U^\mu = \frac{\partial U^\mu}{\partial x^\nu} + \Gamma^\mu_{\sigma\nu} U^\sigma \quad (1.3.17)$$

为 U 的协变导数的第 μ, ν 分量, 这里将导数算符 (*derivative operator*) 记作 ∇ . 由于标量 f 的偏导数已经是协变的, 因此对 f 而言,

$$\nabla_\nu f = \frac{\partial f}{\partial x^\nu}. \quad (1.3.18)$$

我们的协变性是从度规同坐标变换的联系上自然推广的, 那我们不妨来研究它应该具有怎样的协变导数. 首先其偏导 $\frac{\partial g_{\mu\nu}}{\partial x^\sigma}$ 是同 $\Gamma^{\lambda}_{\mu\nu}$ 相联系的:

$$\frac{\partial g_{\mu\nu}}{\partial x^\sigma} = g_{\mu\lambda}\Gamma^{\lambda}_{\nu\sigma} + g_{\nu\lambda}\Gamma^{\lambda}_{\sigma\mu}.$$

可以预料到, 式右相加不能消除 $\Gamma^{\lambda}_{\mu\nu}$ 的不协变部分. 因此度规的偏导也是不协变的. 那我们只能仿造 f, U^μ 那样去构建一个合理的定义了. 为此, 我们构造 $f = g_{\mu\nu}U^\mu U^\nu$, 则肯定有对任意 U^μ 有

$$\nabla_\sigma(g_{\mu\nu}U^\mu U^\nu) = \frac{\partial}{\partial x^\sigma}(g_{\mu\nu}U^\mu U^\nu).$$

∇ 是个导数, 特点就是满足 *Leibniz* 乘法法则 (*rule*)! 因此两边都可以拆开:

$$U^\mu U^\nu \nabla_\sigma g_{\mu\nu} + 2g_{\mu\nu}U^\mu \nabla_\sigma U^\nu = U^\mu U^\nu \frac{\partial g_{\mu\nu}}{\partial x^\sigma} + 2g_{\mu\nu}U^\mu \frac{\partial U^\nu}{\partial x^\sigma},$$

代入 $\nabla_\sigma U^\nu$ 可消去一部分

$$U^\mu U^\nu \nabla_\sigma g_{\mu\nu} + 2g_{\mu\nu}U^\mu \Gamma^{\nu}_{\sigma\lambda} U^\lambda = U^\mu U^\nu \frac{\partial g_{\mu\nu}}{\partial x^\sigma},$$

为了“解”出 $\nabla_\sigma g_{\mu\nu}$, 需使中间这个二倍项“提出” $U^\mu U^\nu$. 那我们把二倍项拆开, 由指标任意性知

$$2g_{\mu\nu}\Gamma^{\nu}_{\sigma\lambda} U^\mu U^\lambda = (g_{\mu\lambda}\Gamma^{\lambda}_{\sigma\nu} + g_{\nu\lambda}\Gamma^{\lambda}_{\sigma\mu})U^\mu U^\nu,$$

则

$$\left(\nabla_\sigma g_{\mu\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma} + g_{\mu\lambda}\Gamma^{\lambda}_{\sigma\nu} + g_{\nu\lambda}\Gamma^{\lambda}_{\sigma\mu} \right) U^\mu U^\nu = 0$$

对任意 U^μ 成立, 因此

$$\nabla_\sigma g_{\mu\nu} = \frac{\partial g_{\mu\nu}}{\partial x^\sigma} - g_{\mu\lambda}\Gamma^{\lambda}_{\sigma\nu} - g_{\nu\lambda}\Gamma^{\lambda}_{\sigma\mu}. \quad (1.3.19)$$

比较一下, 由 $\Gamma^{\lambda}_{\sigma\nu} = \Gamma^{\lambda}_{\nu\sigma}$, 我们发现这样一个很简洁的事实:

$$\nabla_\sigma g_{\mu\nu} = 0. \quad (1.3.20)$$

原来与度规相适配的 $\Gamma^{\lambda}_{\mu\nu}$ 就意味着 $g_{\mu\nu}$ 的协变导数是零! 由于协变导数带着 $\Gamma^{\lambda}_{\mu\nu}$ 走, 所以我们也可以称满足 (1.3.20) 式的是与度规相适配的协变导数算符.

定义 $U_\mu = g_{\mu\nu}U^\nu$, 其亦可看作矢量, 但分量上不完全等同于原矢量, 鉴于其对应关系又称作对偶矢量 (*dual vector*). 任意对偶矢量 ω_μ 可按内积 $v^\mu \omega_\mu = g_{\mu\nu}v^\mu \omega^\nu$ 求得其协变导数为

$$\nabla_\nu \omega_\mu = \frac{\partial \omega_\mu}{\partial x^\nu} - \Gamma^{\sigma}_{\mu\nu}\omega_\sigma. \quad (1.3.21)$$

曲率

如前文所述，把度规合理地定义为二次型，从而满足某些直观性质，但其中最重要的其实是说：弯曲几何在小距离尺度上必须非常接近欧式几何。举个例子，设平面 M 上中心为 x 、半径为 r 的圆盘 $B(x, r) = \{y \in M | d(x, y) < r\}$ ，这里 $d(x, y)$ 表示按度规 g 衡量的两点距离。因为 g 带来了几何学，因而也自然有面积的概念。我们知道，欧氏平面上 B 的面积是 πr^2 ，但是在一般曲面上则不一定。例如由于整个地球表面积是有限的，则地表任意圆盘面积当然也就有限，所以在整体上看，至少表达式不能是 πr^2 ，因为 $r \rightarrow \infty$ 时其发散。不过，我们其实只要求在 r 变小时，面积越来越接近 πr^2 ——换句话说，要求对 $\forall x$ 有

$$\lim_{r \rightarrow 0} \frac{|B(x, r)|}{\pi r^2} = 1.$$

这就把我们导向了曲率概念，具体而言是标量曲率 (*scalar curvature*) R 。有些情况下，例如球面，小圆盘面积总是稍小于 πr^2 ，这种情况称 $R > 0$ ；其他情况如马鞍面上则相反；更有甚者，比如柱面，二者总是相等的，或者说在更高阶下近似，此即说“弯曲”空间也可以有零曲率。标量曲率定义为

$$R_{(x)} = \lim_{r \rightarrow 0} \frac{\pi r^2 - |B(x, r)|}{\pi r^4 / 24}. \quad (1.3.22)$$

接下来我们进一步把这个概念精确地化为 Ricci 曲率 Ric 。在 B 内考虑一扇形，即中心为 x 、以单位矢量（即某条半径方向） v 为中线、开口为 θ 的扇形 $A(x, r, \theta, v) = \{y \in B(x, r) | \frac{g(y-x, v)}{d(x, y)} < \theta/2\}$ 。在平面上 $|A| = \frac{1}{2}r^2\theta$ ，而一般曲面则可以稍小于或稍大于，我们就说曲面在 x 点和 v 方向上，有正的或负的 Ricci 曲率，其具体定义为

$$\text{Ric}_{(x)}(v, v) = \lim_{r \rightarrow 0} \lim_{\theta \rightarrow 0} \frac{\frac{1}{2}r^2\theta - |A|}{r^4\theta / 24}. \quad (1.3.23)$$

现在可以证明，对于曲面来说，这个比较复杂的 Ricci 曲率实际上是标量曲率的一半。这说明 v 对 2 维 Ricci 曲率不起作用。然而，把上面所说的推广到高维，比如 3 维中就相应用到球体和立体角，这带来的效果是，一个点可能在某一方向有正 Ricci 曲率，但在另一方向是负的——这意味着，在前者上很窄的扇形“向内弯曲”，而在后者中的窄扇形则“向外弯曲”。

为了像 Newton 定律那样描述时空几何学的引力场方程，我们务必要弄清楚曲率是否能“代表”引力。不妨想象这样一种 2 维时空，它主要是一张平面，但是有一个凹陷部分，而这个凹陷可以看作一个下半球面，边缘平滑地连接平面部分。我们来看看，在赤道处相隔一定经度（比如 30° ），垂直于赤道出发的两粒子运动有何现象。它们当然都沿着“不受力”的测地线走，但后面我们会证明：球面上的测地线就是过球心的大圆弧，因此它们实际上都沿着经线走，但经线们很显然都将汇聚在南极点。

可见，“曲率”确实影响了测地线的轨迹，使得二者仿佛受到“吸引”般相互靠近。这个例子虽然不严谨⁴⁶，但的确可以给曲率和引力的关系一个直观感受。

其实，咱们新认识的这俩伙伴： R, Ric ，再配合上老旧识 g ，就已经足以“理解”场方程的几何部分了。但由于缺乏对这些曲率的准确定义，我们还暂不能“推理”出场方程。要为了找到这些曲率的定义，我们不妨再换个思路来表达曲率，使之至少在思想上是简单的（具体表达式当然不一定简单）。借助先前思想我们发现， R 相比于 Ric 由于缺少引入一个方向矢量，而失去了描述各方向不同曲率的能力。看来我们可以模仿这个思路，来描述不同方向所有的曲率情况！

由于曲率直观地改变了测地线，此即说“产生”了 $\Gamma^\lambda_{\mu\nu}$ （而无法通过改变坐标系消除⁴⁷），我们不妨来回看测地线方程中的这个 $\Gamma^\lambda_{\mu\nu}$ ，它可以描述一个对偶矢量 ω_μ 沿着各 x^ν 方向的变化。但似乎这还不够融洽，因为各个方向似乎是独立的，而且我们知道这些变化在平面上都可以通过坐标系变换消除。可见，为使之融洽且不随坐标系变化，我们不能让这些东西太“独立”了，得给我“耦合”一些。或许“两两”或“三三”坐标相互牵制，共同造就的东西就不容易被坐标系“分心”了。

那么究竟是选几个坐标“牵连”且要“牵连”到何种程度呢？考虑这样一种简单的耦合：我先将 ω_σ 沿 x^ν 走一小段以比较出 $\nabla_\nu \omega_\sigma$ ，随后再将其沿 x^μ 走一小段以比较出 $\nabla_\mu \nabla_\nu \omega_\sigma$ 。如果是个平面的普通坐标系，那么在微积分中我们学过 Fubini 定理，偏微分顺序将不影响结果。但现在由于平面可以弯曲，基矢不能通过改变坐标系而保持不变 ($\Gamma^\lambda_{\mu\nu}$ 不能消除)，再次求导必将带着导数走，则颠倒顺序会得到不同结果。我们可以比较此二顺序的结果之差：

$$R^\lambda_{\mu\nu\sigma} \omega_\lambda = [\nabla_\mu, \nabla_\nu] \omega_\sigma, \quad (1.3.24)$$

其中方括号称作对易子 (commutator)，也就是交叉相减 $[\nabla_\mu, \nabla_\nu] = \nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu$ 。先计算 $\nabla_\mu \nabla_\nu \omega_\sigma$ ，考虑到 $\nabla_\nu \omega_\sigma$ 是协变的双指标量，可仿照此前做法证明其协变导数类似于度规那样：

$$\nabla_\mu \nabla_\nu \omega_\sigma = \frac{\partial}{\partial x^\mu} (\nabla_\nu \omega_\sigma) - \Gamma^\lambda_{\mu\nu} \nabla_\lambda \omega_\sigma - \Gamma^\lambda_{\mu\sigma} \nabla_\nu \omega_\lambda,$$

接下来继续拆成

$$\begin{aligned} \nabla_\mu \nabla_\nu \omega_\sigma &= \frac{\partial}{\partial x^\mu} \left(\frac{\partial \omega_\sigma}{\partial x^\nu} - \Gamma^\lambda_{\nu\sigma} \omega_\lambda \right) - \Gamma^\lambda_{\mu\nu} \nabla_\lambda \omega_\sigma - \Gamma^\lambda_{\mu\sigma} \left(\frac{\partial \omega_\lambda}{\partial x^\nu} - \Gamma^\gamma_{\nu\lambda} \omega_\gamma \right) \\ &= \frac{\partial^2 \omega_\sigma}{\partial x^\mu \partial x^\nu} - \frac{\partial \Gamma^\lambda_{\nu\sigma}}{\partial x^\mu} \omega_\lambda - \Gamma^\lambda_{\nu\sigma} \frac{\partial \omega_\lambda}{\partial x^\mu} - \Gamma^\lambda_{\mu\nu} \nabla_\lambda \omega_\sigma - \Gamma^\lambda_{\mu\sigma} \frac{\partial \omega_\lambda}{\partial x^\nu} + \Gamma^\lambda_{\mu\sigma} \Gamma^\gamma_{\nu\lambda} \omega_\gamma \end{aligned}$$

⁴⁶ 因为不是先天存在这么一个时空和曲率去影响物质运动，而是物质和曲率相互影响着，这意味着这个曲率是物质自己产生的。这便是我们找场方程的目标。

⁴⁷ 由于弯曲时空，尤其是无挠时空，在局部上总有平直的坐标系，那么总可至少在一点上变换坐标消除 $\Gamma^\lambda_{\mu\nu}$ 。因此这里指的是不能整体地处处消除。后面将看到这是等效原理的数学表述。

由 $\frac{\partial^2 \omega_\sigma}{\partial x^\mu \partial x^\nu} = \frac{\partial^2 \omega_\sigma}{\partial x^\nu \partial x^\mu}$, $\Gamma^\lambda_{\mu\nu} = \Gamma^\lambda_{\nu\mu}$, 则

$$\begin{aligned} R^\lambda_{\mu\nu\sigma}\omega_\lambda &= -\frac{\partial \Gamma^\lambda_{\sigma\nu}}{\partial x^\mu}\omega_\lambda + \Gamma^\lambda_{\sigma\mu}\Gamma^\gamma_{\lambda\nu}\omega_\gamma + \frac{\partial \Gamma^\lambda_{\sigma\mu}}{\partial x^\nu}\omega_\lambda - \Gamma^\lambda_{\sigma\nu}\Gamma^\gamma_{\lambda\mu}\omega_\gamma \\ &= \frac{\partial \Gamma^\lambda_{\sigma\mu}}{\partial x^\nu}\omega_\lambda - \frac{\partial \Gamma^\lambda_{\sigma\nu}}{\partial x^\mu}\omega_\lambda + \Gamma^\gamma_{\sigma\mu}\Gamma^\lambda_{\gamma\nu}\omega_\lambda - \Gamma^\gamma_{\sigma\nu}\Gamma^\lambda_{\gamma\mu}\omega_\lambda, \end{aligned}$$

可见

$$R^\lambda_{\mu\nu\sigma} = \frac{\partial \Gamma^\lambda_{\sigma\mu}}{\partial x^\nu} - \frac{\partial \Gamma^\lambda_{\sigma\nu}}{\partial x^\mu} + \Gamma^\gamma_{\sigma\mu}\Gamma^\lambda_{\gamma\nu} - \Gamma^\gamma_{\sigma\nu}\Gamma^\lambda_{\gamma\mu}, \quad (1.3.25)$$

这称作 *Riemann* 曲率 (的分量). 其中容易发现 $R^\lambda_{\mu\nu\sigma} = -R^\lambda_{\nu\mu\sigma}$. 当然, 这个分量实在是太多了, 而且某一些分量往往会重复或恒为零. 为方便遍这样定义了一个新东西:

$$R_{\mu\nu} = R^\lambda_{\mu\lambda\nu}. \quad (1.3.26)$$

这个操作将四阶表格“压缩到”了一个二阶表格, 因此又常被单独叫做缩并 (*contraction*). 只缩这两个是因为其他情况要么为零“没啥用”, 要么反号或相等“没区别”. 可以证明, 这就是那个 *Ricci* 曲率 (的分量). 但假如我们没有事先知晓任何数学知识或者几何意义⁴⁸, 完全可以将此看作一个定义, 而不去关心它是否真的在 2 维情况是 (1.3.23) 式. 此外, 我们所说的标量曲率其实就是按照类似思路定义的

$$R = R^\nu_{\nu}. \quad (1.3.27)$$

可我们只知道 $R_{\mu\nu}$ 呀. 原来, 之前所谓的“升降指标”操作都默认了这样一件事: 在数值上它就是 R^μ_{ν} , 此即是说 $R^\mu_{\nu} = \delta^{\mu\sigma}R_{\sigma\nu}$. 但我们现在不一定在欧氏空间里办事了, 所以应该用 g 而且还是 $g^{\mu\nu}$ 来升指标. 以后做升降操作一般都用当下所用的 g 来办就妥. 因此 (1.3.27) 又是

$$R = g^{\mu\nu}R_{\mu\nu}. \quad (1.3.28)$$

一般其刻画沿所有方向的“平均”曲率. 可见对于球面这种对称性高的空间, $R_{\mu\nu}$ 沿所有方向都一样, 自然和失去自由度后的 R 没有区别了.

至此足够了吗? 是否还需在 *Riemann* 曲率的基础上继续“套娃”下去? 其实不必了: 第一, 受到 *Newton* 方程的启发, 我们同样只期望场方程是个二阶微分方程⁴⁹; 第二, 我们可以证明, 这已经是我们想要的不依赖坐标系的东西了——区别于 $\Gamma^\lambda_{\mu\nu}$. 下面推 $R^\lambda_{\mu\nu\sigma}$ 的坐标变换. 我们写下 \hat{S} 系的

$$\hat{R}^\lambda_{\hat{\mu}\hat{\nu}\hat{\sigma}} = \frac{\partial \hat{\Gamma}^\lambda_{\hat{\sigma}\hat{\mu}}}{\partial \hat{x}^{\hat{\nu}}} - \frac{\partial \hat{\Gamma}^\lambda_{\hat{\sigma}\hat{\nu}}}{\partial \hat{x}^{\hat{\mu}}} + \hat{\Gamma}^{\hat{\alpha}}_{\hat{\sigma}\hat{\mu}}\hat{\Gamma}^\lambda_{\hat{\alpha}\hat{\nu}} - \hat{\Gamma}^{\hat{\alpha}}_{\hat{\sigma}\hat{\nu}}\hat{\Gamma}^\lambda_{\hat{\alpha}\hat{\mu}},$$

⁴⁸ 不过我们可以剧透其物理意义: “时空体积 (spacetime volume)” 沿不同方向的变化方式.

⁴⁹ $\Gamma^\lambda_{\mu\nu}$ 是 g 的一阶方程: 克氏符就像力, 而度规就像其势.

观察一下，前两项相减是能够把 $\Gamma^\lambda_{\mu\nu}$ 不协变的地方消掉的。其次还能再消掉一些：

$$\begin{aligned} \frac{\partial \hat{\Gamma}^{\hat{\lambda}}_{\hat{\sigma}\hat{\mu}}}{\partial \hat{x}^{\hat{\nu}}} - \frac{\partial \hat{\Gamma}^{\hat{\lambda}}_{\hat{\sigma}\hat{\nu}}}{\partial \hat{x}^{\hat{\mu}}} &= \frac{\partial}{\partial \hat{x}^{\hat{\nu}}} \left(\Gamma^\lambda{}_{\sigma\mu} \frac{\partial x^\sigma}{\partial \hat{x}^{\hat{\sigma}}} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} \right) - \frac{\partial}{\partial \hat{x}^{\hat{\mu}}} \left(\Gamma^\lambda{}_{\sigma\nu} \frac{\partial x^\sigma}{\partial \hat{x}^{\hat{\sigma}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} \right) \\ &= \left(\frac{\partial \Gamma^\lambda{}_{\sigma\mu}}{\partial x^\nu} - \frac{\partial \Gamma^\lambda{}_{\sigma\nu}}{\partial x^\mu} \right) \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial x^\sigma}{\partial \hat{x}^{\hat{\sigma}}} \\ &\quad + \frac{(\Gamma^\lambda{}_{\sigma\mu} \partial^2 x^\mu - \Gamma^\lambda{}_{\sigma\nu} \partial^2 x^\nu) \partial x^\sigma \partial \hat{x}^{\hat{\lambda}}}{\partial \hat{x}^{\hat{\mu}} \partial \hat{x}^{\hat{\nu}} \partial \hat{x}^{\hat{\sigma}} \partial x^\lambda} \\ &\quad + \frac{(\Gamma^\lambda{}_{\sigma\mu} \partial x^\mu - \Gamma^\lambda{}_{\sigma\nu} \partial x^\nu) \partial (\partial x^\sigma \partial \hat{x}^{\hat{\lambda}})}{\partial \hat{x}^{\hat{\mu}} \partial \hat{x}^{\hat{\nu}} \partial \hat{x}^{\hat{\sigma}} \partial x^\lambda} \\ &= \left(\frac{\partial \Gamma^\lambda{}_{\sigma\mu}}{\partial x^\nu} - \frac{\partial \Gamma^\lambda{}_{\sigma\nu}}{\partial x^\mu} \right) \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial x^\sigma}{\partial \hat{x}^{\hat{\sigma}}}, \end{aligned}$$

从而只剩下协变部分。同理

$$\begin{aligned} \hat{\Gamma}^{\hat{\alpha}}_{\hat{\sigma}\hat{\mu}} \hat{\Gamma}^{\hat{\lambda}}_{\hat{\alpha}\hat{\nu}} - \hat{\Gamma}^{\hat{\alpha}}_{\hat{\sigma}\hat{\nu}} \hat{\Gamma}^{\hat{\lambda}}_{\hat{\alpha}\hat{\mu}} &= (\Gamma^\alpha{}_{\sigma\mu} \Gamma^\lambda{}_{\alpha\nu} - \Gamma^\alpha{}_{\sigma\nu} \Gamma^\lambda{}_{\alpha\mu}) \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial x^\sigma}{\partial \hat{x}^{\hat{\sigma}}} \\ &\quad + (\Gamma^\lambda{}_{\alpha\nu} \partial x^\nu - \Gamma^\lambda{}_{\alpha\mu} \partial x^\mu) \frac{\partial x^\alpha \partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda \partial \hat{x}^{\hat{\mu}} \partial \hat{x}^{\hat{\nu}} \partial \hat{x}^{\hat{\sigma}}} \frac{\partial^2 x^\gamma}{\partial x^\gamma} \\ &\quad + (\Gamma^\alpha{}_{\sigma\mu} \partial x^\mu - \Gamma^\alpha{}_{\sigma\nu} \partial x^\nu) \frac{\partial x^\sigma \partial \hat{x}^{\hat{\lambda}}}{\partial x^\alpha \partial \hat{x}^{\hat{\mu}} \partial \hat{x}^{\hat{\nu}} \partial \hat{x}^{\hat{\sigma}}} \frac{\partial^2 x^\lambda}{\partial x^\lambda} \\ &= (\Gamma^\alpha{}_{\sigma\mu} \Gamma^\lambda{}_{\alpha\nu} - \Gamma^\alpha{}_{\sigma\nu} \Gamma^\lambda{}_{\alpha\mu}) \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial x^\sigma}{\partial \hat{x}^{\hat{\sigma}}}. \end{aligned}$$

故

$$\hat{R}^{\hat{\lambda}}_{\hat{\mu}\hat{\nu}\hat{\sigma}} = R^\lambda{}_{\mu\nu\sigma} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial x^\sigma}{\partial \hat{x}^{\hat{\sigma}}}. \quad (1.3.29)$$

既然 $R^\lambda{}_{\mu\nu\sigma}$ 是内禀的，那很显然

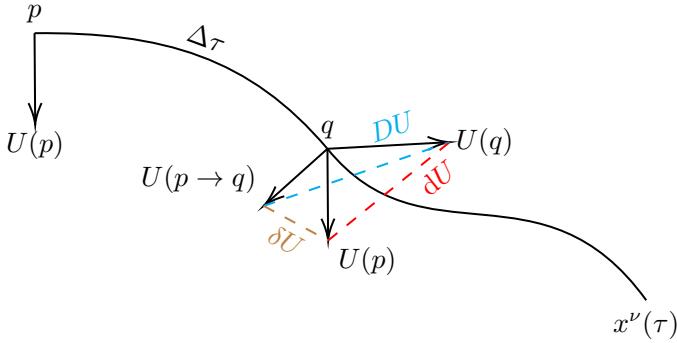
$$\hat{R}_{\hat{\mu}\hat{\nu}} = \hat{R}^{\hat{\lambda}}_{\hat{\mu}\hat{\lambda}\hat{\nu}} = R_{\mu\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \quad (1.3.30)$$

也是内禀的。而 R 是标量，当然内禀。因此我们说曲率是内禀的。

下面解释曲率的几何含义。在构造曲率时，我们选择的方式是将基矢按不同顺序求导后比较差值。由定义，曲率其实在比较基矢沿不同顺序求导的结果，而我们知道，在曲率为零的平直白纸上，先沿某轴再沿某轴地“平移”矢量，所得结果与颠倒顺序后的结果一致。但如果曲率不为零，矢量的“平移”会因为路径的不同而不同，进而存在二者之差。

不过沿不同方向“平移”矢量究竟是什么意思呢？为什么笔者在这里要加个引号，而之前也有这个词但没加？原来之前我们提到的平移 (translation) 就是指原封不动地照搬分量，而接下来要讲的“平移”是一个总与路径有关的移动。

不妨来研究这件事： $\nabla_\nu U^\mu$ 比 $\frac{\partial U^\mu}{\partial x^\nu}$ 多出来的那个 $\Gamma^\mu{}_{\sigma\nu} U^\sigma$ 究竟是什么？如图

图 1.10: 图中 $\Delta\tau$ 等差值均夸大

1.10, 设弯曲空间中某曲线 $x^\nu(\tau)$ 上临近两点 p, q 各有一个 $U(p), U(q)$. 它们是任意选取的, 可看作同一个矢量场 $U(x^\nu(\tau))$ 在 p, q 的取值. 先来看看 $\frac{\partial U^\mu}{\partial x^\nu}$ 的偏微分定义:

$$\frac{\partial U^\mu}{\partial x^\nu} = \lim_{q \rightarrow p} \frac{U^\mu(q) - U^\mu(p)}{\Delta x^\nu}, \quad (1.3.31)$$

原来它应该是直接拿 $U^\mu(p), U^\mu(q)$ 比较了, 但是这减出来的东西既不满足 p 点的坐标协变, 也不满足 q 的. 可见这便是该导数不协变原因的几何含义. 我们在图中把 $U(p)$ 照搬到 q 点, 于是它们的差值可以标注如图中红线:

$$dU^\mu = \frac{\partial U^\mu}{\partial x^\nu} dx^\nu. \quad (1.3.32)$$

但 $\nabla_\nu U^\mu$ 求的究竟是什么呢? 我们知道它是 $\frac{\partial U}{\partial x^\nu}$ 的分量, 即

$$\nabla_\nu U^\mu = \lim_{q \rightarrow p} \frac{U^\mu(q) - U^\mu(p)}{\Delta x^\nu} + \Gamma^\mu_{\sigma\nu} U^\sigma,$$

而它准确而言就是对 p 点求的, 其协变性就是按照 p 点的, 因此我们说后面的 $\Gamma^\mu_{\sigma\nu}, U^\sigma$ 实际上均取 p 点值. 为了找到它实际上用的是谁的差值, 我们尝试着通分后并入极限中:

$$\nabla_\nu U^\mu = \lim_{q \rightarrow p} \frac{U^\mu(q) - (U^\mu(p) - \Gamma^\mu_{\sigma\nu} U^\sigma(p) \Delta x^\nu)}{\Delta x^\nu}, \quad (1.3.33)$$

可见 $U^\mu(q)$ 应该减去的是一个暂不明意义的新矢量, 但因为在 q “还没趋近” p 时, 它们减出来的东西就已在 q 处协变, 因此这个新矢量可以看作是 q 点的, 但又是从 p 衍生的, 我们记其分量为

$$U^\mu(p \rightarrow q) = U^\mu(p) - \Gamma^\mu_{\sigma\nu} U^\sigma \Delta x^\nu. \quad (1.3.34)$$

这样, 其差值就可以直接表示为二者箭端连线了, 如图中蓝线:

$$DU^\mu = dx^\nu \nabla_\nu U^\mu. \quad (1.3.35)$$

而这个矢量 $U(p \rightarrow q)$ 从 $U(p)$ 衍生或者移动，我们也标注其比 $U(p)$ 多的差值，如图中棕线：

$$\delta U^\mu = U^\mu(p \rightarrow q) - U^\mu(p). \quad (1.3.36)$$

由 (1.3.34) 或者图示知

$$\delta U^\mu = dU^\mu - DU^\mu = -\Gamma^\mu_{\sigma\nu} U^\sigma dx^\nu. \quad (1.3.37)$$

可见，原来 $\nabla_\nu U^\mu$ 比 $\frac{\partial U^\mu}{\partial x^\nu}$ 多出来的那个修正项就是

$$\Gamma^\mu_{\sigma\nu} U^\sigma = -\frac{\delta U^\mu}{\partial x^\nu}, \quad (1.3.38)$$

多亏了它，我们才有了协变的 $\nabla_\nu U^\mu$. 那这个 $U^\mu(p \rightarrow q)$ 究竟是怎样一种移动结果呢？我们不妨令 U 自己就处处按这种方式移动，即使二者之差 DU 始终为零：

$$DU^\mu = dx^\nu \nabla_\nu U^\mu = 0. \quad (1.3.39)$$

欲研究 U^μ 在这种移动下有何性质，不外乎就从其特征上入手. 最典型地，比如其模长平方：

$$\begin{aligned} D(g_{\mu\nu} U^\mu U^\nu) &= U^\mu U^\nu Dg_{\mu\nu} + 2g_{\mu\nu} U^\mu D U^\nu \\ &= U^\mu U^\nu dx^\sigma \nabla_\sigma g_{\mu\nu} + 2g_{\mu\nu} U^\mu D U^\nu, \end{aligned}$$

可见， $\nabla_\sigma g_{\mu\nu} = 0$ 和 $D U^\nu = 0$ 保证了

$$D(g_{\mu\nu} U^\mu U^\nu) = 0, \quad (1.3.40)$$

因此 U 的模长不变，的确可以看作一种打引号的“平移”，但以后为避免与照搬分量的平移相混淆，我们称这种移动为平行移动 (*parallel transport*)⁵⁰. 直观上，它不仅保持其自身模长；而且如果我们能够将弯曲背景上含有路径的部分“撕下”并“展平”，便会发现这些矢量“精确地”按照平直空间的方式平移. 由于这与“撕下”背景的哪一块儿有关，因此当然与路径有关. 我们以后处理的，都是更一般的“一点儿也不平”的弯曲背景，其中平行移动处处与路径及附近相关，因而可以良好定义，但对平移而言，则没有一个整体标准了，并且照搬分量显然要依靠坐标系. 综上所述，在弯曲背景中，平行移动是一个比平移更自然的“平移”.

说了这么多，可总算是把平移和平行移动的区别弄清楚了. 那这所有一切与曲率有什么关系呢？不妨回到最初定义曲率的地方，我们利用的是协变导数的对易子，

⁵⁰亦称作 *Levi-Civita* 移动，因为首先由他于 1917 年提出. 但客观讲这实在是“太晚了”，因为现代微分几何的大部分基本概念、计算工具、张量分析已经由他和 Ricci 等人在 1901~1915 完善殆尽，而且 Einstein 完成他理论构建的时间也惊奇地早于此. 可见 Einstein 以一种不需要平移任何东西的方式理解了协变导数，但其方式也许对初学者来说难以想象. 或者说，他很可能从未且无需真正理解曲率？

而另一方面，一个矢量沿着两条不同路径的平行移动结果确实可能不同。在一张平直的白纸上，矢量的平行移动与平移无异，因而与路径无关，或者说，平直白纸上矢量可沿任意回路平行移动，而回到原处时与原矢量无区别；但在有曲率的地方，比如一个球面，矢量从北极沿任意经线平行移动至南极，再沿任意不同的经线平行移动回北极，所得结果就不同于原矢量了。当然，我们一般不需要那么大的回路，因为我们想找的是一点局部曲率的几何意义。

下面，我们来看能否构造这样的微小回路。设两条曲线 $x^\mu(\tau), x^\mu(s)$ 交于 o ，但

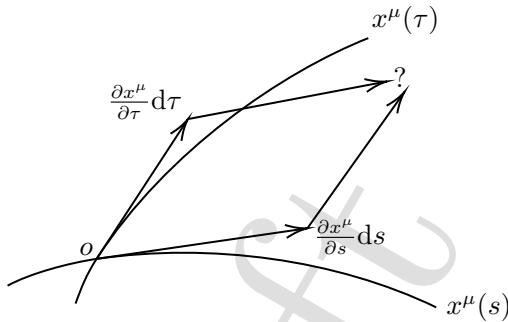


图 1.11：无穷小测地回路

由于我们研究的是局部情况，因此两条线就成了微小直线 $\frac{\partial x^\mu}{\partial \tau} d\tau, \frac{\partial x^\mu}{\partial s} ds$ ，放大场景如图 1.11 所示。因此，要形成微小回路，不外乎想用两条微小切矢构造“平行”四边形。我们用平行移动来阐明“平行”的含义，即要求各自对边为原矢量的平行移动。根据定义， $\frac{\partial x^\mu}{\partial \tau} d\tau$ 的对边矢量就是其沿着 $\frac{\partial x^\nu}{\partial s} ds$ 平行移动，则与原来相比差值是：

$$\frac{\delta}{\partial s} ds = -\Gamma^{\mu}_{\lambda\nu} \frac{\partial x^\lambda}{\partial \tau} \frac{\partial x^\nu}{\partial s} d\tau ds,$$

同理 $\frac{\partial x^\mu}{\partial s} ds$ 的对边矢量与原来相比的差值是

$$\frac{\delta}{\partial \tau} d\tau = -\Gamma^{\mu}_{\lambda\nu} \frac{\partial x^\lambda}{\partial s} \frac{\partial x^\nu}{\partial \tau} d\tau ds,$$

则二者之差是

$$\begin{aligned} \Delta &= \Gamma^{\mu}_{\lambda\nu} \frac{\partial x^\lambda}{\partial s} \frac{\partial x^\nu}{\partial \tau} d\tau ds - \Gamma^{\mu}_{\lambda\nu} \frac{\partial x^\lambda}{\partial \tau} \frac{\partial x^\nu}{\partial s} d\tau ds \\ &= (\Gamma^{\mu}_{\lambda\nu} - \Gamma^{\mu}_{\nu\lambda}) \frac{\partial x^\lambda}{\partial s} \frac{\partial x^\nu}{\partial \tau} d\tau ds. \end{aligned}$$

因而我们终于明白 $\Gamma^{\mu}_{\lambda\nu} = \Gamma^{\mu}_{\nu\lambda}$ 的意义了：这保证我们有一个无穷小的四边形回路！数学家认为，如果在一般情况下没有这个条件，那么定义这样一个量加以区分就很有帮助了：

$$\mathcal{T}^{\mu}_{\lambda\nu} = \Gamma^{\mu}_{\lambda\nu} - \Gamma^{\mu}_{\nu\lambda} \quad (1.3.41)$$

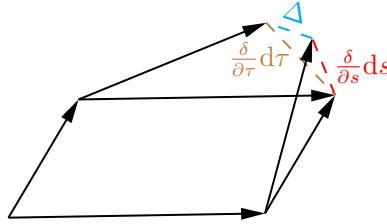


图 1.12: 挠率导致的差值

称为挠率. 从名字上很好理解, 正是因为平行移动路径像一股麻绳似的扭转 (“挠”), 才不能保证平行移动结果与原矢量在平直空间的意义下平行; 而另一方面, 一个矢量绕一圈很可能怎么也回不到原来的位置, 越 “挠” 越远.

出于对等效原理的需求, 广相选择了研究无挠时空, 这样我们处处都能构造出一个无穷小的四边形回路⁵¹. 现在研究矢量 U^μ 沿着回路的平行移动. 根据定义, 先

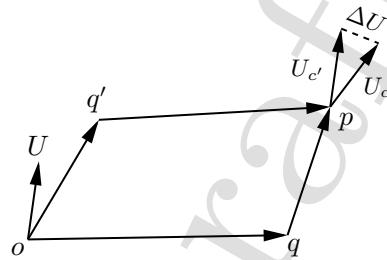


图 1.13: 曲率导致的差值

移动至 q 处有

$$U^\mu(q) = U^\mu - \Gamma^\mu{}_{\sigma\nu} U^\sigma \frac{\partial x^\nu}{\partial s} ds,$$

再移动至 p 点有

$$\begin{aligned} U_c^\mu &= U^\mu(q) - \Gamma^\mu{}_{\nu\kappa}(q) U^\kappa(q) \left(\frac{\partial x^\nu}{\partial \tau} d\tau - \Gamma^\nu{}_{\alpha\beta} \frac{\partial x^\alpha}{\partial \tau} \frac{\partial x^\beta}{\partial s} d\tau ds \right) \\ &= U^\mu - \Gamma^\mu{}_{\nu\kappa} U^\kappa \frac{\partial x^\nu}{\partial s} ds - \left(\Gamma^\mu{}_{\nu\kappa} + \frac{\partial \Gamma^\mu{}_{\nu\kappa}}{\partial x^\lambda} \frac{\partial x^\lambda}{\partial s} ds \right) \left(U^\kappa - \Gamma^\kappa{}_{\sigma\rho} U^\rho \frac{\partial x^\sigma}{\partial s} ds \right) \frac{\partial x^\nu}{\partial \tau} d\tau \\ &\quad + \Gamma^\nu{}_{\alpha\beta} \frac{\partial x^\alpha}{\partial \tau} \frac{\partial x^\beta}{\partial s} U^\kappa(q) \Gamma^\mu{}_{\nu\kappa}(q) d\tau ds \\ &= U^\mu - \Gamma^\mu{}_{\nu\kappa} U^\kappa \left(\frac{\partial x^\nu}{\partial s} ds + \frac{\partial x^\nu}{\partial \tau} d\tau \right) - U^\kappa \left(\frac{\partial \Gamma^\mu{}_{\nu\kappa}}{\partial x^\lambda} - \Gamma^\rho{}_{\lambda\kappa} \Gamma^\mu{}_{\nu\rho} \right) \frac{\partial x^\lambda}{\partial s} \frac{\partial x^\nu}{\partial \tau} ds d\tau \\ &\quad + \frac{\partial \Gamma^\mu{}_{\nu\kappa}}{\partial x^\lambda} \frac{\partial x^\lambda}{\partial s} U^\rho \Gamma^\kappa{}_{\sigma\rho} \frac{\partial x^\sigma}{\partial \tau} \frac{\partial x^\nu}{\partial s} ds^2 d\tau + \Gamma^\nu{}_{\alpha\beta} \frac{\partial x^\alpha}{\partial \tau} \frac{\partial x^\beta}{\partial s} U^\kappa(q) \Gamma^\mu{}_{\nu\kappa}(q) ds d\tau, \end{aligned}$$

⁵¹如不闭合仍可构造五边形回路, 在第二章对此详解.

同理沿 o, q', p 有

$$\begin{aligned} U_{c'}^\mu &= U^\mu - \Gamma^\mu{}_{\nu\kappa} U^\kappa \left(\frac{\partial x^\nu}{\partial s} ds + \frac{\partial x^\nu}{\partial \tau} d\tau \right) - U^\kappa \left(\frac{\partial \Gamma^\mu{}_{\lambda\kappa}}{\partial x^\nu} - \Gamma^\rho{}_{\nu\kappa} \Gamma^\mu{}_{\lambda\rho} \right) \frac{\partial x^\lambda}{\partial s} \frac{\partial x^\nu}{\partial \tau} ds d\tau \\ &\quad + \frac{\partial \Gamma^\mu{}_{\lambda\kappa}}{\partial x^\nu} \frac{\partial x^\nu}{\partial \tau} U^\rho \Gamma^\kappa{}_{\sigma\rho} \frac{\partial x^\sigma}{\partial \tau} \frac{\partial x^\lambda}{\partial s} d\tau^2 ds + \Gamma^\nu{}_{\alpha\beta} \frac{\partial x^\alpha}{\partial s} \frac{\partial x^\beta}{\partial \tau} U^\kappa(q') \Gamma^\mu{}_{\nu\kappa}(q') d\tau ds. \end{aligned}$$

相减得到

$$\Delta U^\mu = U_{c'}^\mu - U_c^\mu \approx R^\mu{}_{\nu\lambda\kappa} U^\kappa \frac{\partial x^\lambda}{\partial s} \frac{\partial x^\nu}{\partial \tau} ds d\tau,$$

其中略去了更高的无穷小，并注意挠率为零使得邻边平行移动贡献抵消，即可知曲率之几何意义。正是因为有了曲率，矢量的平行移动才与路径有关，哪怕是在这样的无穷小路径上。

测地线

现在我们要借测地线说明度规与协变导数的适配性。在前面我们默认了速度矢量模长的不变性，但这应当只是这个性质的等价表述，请看：

$$\begin{aligned} \frac{d}{d\tau} \left(g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \right) &= \frac{\partial g_{\mu\nu}}{\partial x^\lambda} \frac{dx^\lambda}{d\tau} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} + g_{\mu\nu} \frac{d^2 x^\mu}{d\tau^2} \frac{dx^\nu}{d\tau} + g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{d^2 x^\nu}{d\tau^2} \\ &= \left(\frac{\partial g_{\kappa\sigma}}{\partial x^\lambda} - g_{\mu\sigma} \Gamma^\mu{}_{\kappa\lambda} - g_{\nu\kappa} \Gamma^\kappa{}_{\sigma\lambda} \right) \frac{dx^\kappa}{d\tau} \frac{dx^\sigma}{d\tau} \frac{dx^\lambda}{d\tau} \\ &= \nabla_\lambda g_{\kappa\sigma} \frac{dx^\kappa}{d\tau} \frac{dx^\sigma}{d\tau} \frac{dx^\lambda}{d\tau} = 0, \end{aligned}$$

也就是说，当我一开始选定其模长归一，只有当度规与协变导数适配时，沿着测地线才会保持均匀的固有时流逝。

不仅如此，这样做还可以使得运动轨迹成为事实上几何上的测地线。回顾一下，几何学将测地线定义为直线的推广，而定义一条直线方程的方式有很多，比如“矢量沿自身平行移动所得轨迹”或者“两点间长度最短的连线”。前者说明克氏符可绕开度规，直接从协变导数定义；后者说明克氏符进而将同度规适配，因为度规决定着距离。

先讨论前者。无非是说沿着曲线，速度矢量满足⁵²

$$\frac{DU^\mu}{d\tau} = 0, \tag{1.3.42}$$

则

$$\frac{dx^\nu}{d\tau} \nabla_\nu \frac{dx^\mu}{d\tau} = 0,$$

⁵²如果定义直线使得矢量始终平行，那我们仅要求速度的导数是速度的倍数，但这只是对参数 τ 的重新选取罢了。实际上总可选取参数使得速度的导数压根儿为零，这种参数正是仿射参数。

也即

$$\frac{dx^\nu}{d\tau} \left(\frac{\partial}{\partial x^\nu} \frac{dx^\mu}{d\tau} + \Gamma^\mu_{\sigma\nu} \frac{dx^\sigma}{d\tau} \right) = 0.$$

因此

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\sigma\nu} \frac{dx^\sigma}{d\tau} \frac{dx^\nu}{d\tau} = 0,$$

可见一种思路便是先定义协变导数和克氏符，再利用平行移动给出测地线方程。

再讨论后者。测地线其实并不一定是最短的。对于球面而言，两点之间的测地线方程不仅给出了劣弧，它也包括了优弧，而优弧既不是最短的也不是最长的，似乎“什么也不是”……果真如此吗？我们知道微积分学中，某种函数在某点取最值的必要条件是函数在该点的导数为零。优弧应该也满足“导数为零”，但只是不保证最值罢了。（甚至都不是极值，而是个“鞍点 (*saddle point*)”。）总之，虽说只是个必要条件，但我们不妨先来研究研究怎么“求导”。

我们的研究目标是两点之间“线长”，或者准确说是所经历的固有时 (*proper time elapsed along a path*):

$$T(\alpha) = \int d\tau = \int_\alpha \sqrt{-g_{\mu\nu} dx^\mu dx^\nu} = \int_{\tau_0}^{\tau_1} \sqrt{-g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu} d\tau, \quad (1.3.43)$$

这当然是个函数，因为值域在实数 \mathbf{R} 内，但自变量是啥？自变量是待求测地线 α 的参数表达 $x^\mu(\tau)$ ！数学家们给这种自变量并非某些分立的变量，而是似乎具有“无穷自由度”（因为 $\tau \in \mathbf{R}$ ）的函数 $x^\mu(\tau)$ 的函数取名为泛函 (*functional*)，对其微分叫做变分 (*variation*) δ 。唯一的限制是要求这些可能的解都固定了两端，即点 $x^\mu(\tau_0), x^\mu(\tau_1)$ 是给定而不随函数改变而改变的，这样变分的代数运算将同微分一致，且由于端点不变，变分可以挪动积分之内。假设欲求路径是 x^μ ，则对其变分一点到 $x^\mu + \delta x^\mu$ （但固定端点），所引起的泛函改变是

$$\delta T = 0, \quad (1.3.44)$$

为零是因为应当不变，或者说相对于路径变化是高阶无穷小。处理里面的函数，不外乎是

$$\begin{aligned} \delta T &= \frac{1}{2} \int \frac{\delta(-g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu)}{\sqrt{-g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu}} d\tau \\ &= -\frac{1}{2} \int (\dot{x}^\mu \dot{x}^\nu \delta g_{\mu\nu} + g_{\mu\nu} \dot{x}^\nu \delta \dot{x}^\mu + g_{\mu\nu} \dot{x}^\mu \delta \dot{x}^\nu) d\tau \\ &= -\frac{1}{2} \left(\int \dot{x}^\mu \dot{x}^\nu \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \delta x^\sigma d\tau + \int g_{\mu\nu} \dot{x}^\nu \delta \dot{x}^\mu d\tau + \int g_{\mu\nu} \dot{x}^\mu \delta \dot{x}^\nu d\tau \right), \end{aligned}$$

由于这里面含有咱不太想要的 $\delta \dot{x}^\nu$ ，利用微积分学中的“分部积分法”即可转化它。

且由于端点固定, 有 $\delta x^\nu(\tau_0) = \delta x^\nu(\tau_1) = 0$:

$$\begin{aligned}\int g_{\mu\nu} \dot{x}^\mu d\delta x^\nu &= g_{\mu\nu} \dot{x}^\mu \delta x^\nu \Big|_{\tau_0}^{\tau_1} - \int \delta x^\nu d(g_{\mu\nu} \dot{x}^\mu) \\ &= - \int \delta x^\nu \left(\frac{dg_{\mu\nu}}{d\tau} \dot{x}^\mu + g_{\mu\nu} \ddot{x}^\mu \right) d\tau \\ &= - \int \delta x^\nu \left(\frac{\partial g_{\mu\nu}}{\partial x^\sigma} \dot{x}^\sigma \dot{x}^\mu + g_{\mu\nu} \ddot{x}^\mu \right) d\tau,\end{aligned}$$

出现了 $\dot{x}^\sigma \dot{x}^\mu$ 和 \ddot{x}^μ , 可见已有测地线方程的雏形了. 把这些全部代回去有

$$\begin{aligned}\delta T = -\frac{1}{2} \int &\left(\delta x^\sigma \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \dot{x}^\mu \dot{x}^\nu - \delta x^\mu \left(\frac{\partial g_{\mu\nu}}{\partial x^\sigma} \dot{x}^\sigma \dot{x}^\nu + g_{\mu\nu} \ddot{x}^\nu \right) \right. \\ &\left. - \delta x^\nu \left(\frac{\partial g_{\mu\nu}}{\partial x^\sigma} \dot{x}^\sigma \dot{x}^\mu + g_{\mu\nu} \ddot{x}^\mu \right) \right) d\tau.\end{aligned}$$

我们把后两项都换成 δx^σ , 这样就可以提出来:

$$\begin{aligned}\delta T = -\frac{1}{2} \int \delta x^\sigma &\left(\frac{\partial g_{\mu\nu}}{\partial x^\sigma} \dot{x}^\mu \dot{x}^\nu - \left(\frac{\partial g_{\sigma\nu}}{\partial x^\mu} \dot{x}^\mu \dot{x}^\nu + g_{\sigma\nu} \ddot{x}^\nu \right) \right. \\ &\left. - \left(\frac{\partial g_{\mu\sigma}}{\partial x^\nu} \dot{x}^\nu \dot{x}^\mu + g_{\mu\sigma} \ddot{x}^\mu \right) \right) d\tau,\end{aligned}$$

再将里面的 $\dot{x}^\mu \dot{x}^\nu$ 提出来, 并合并剩下两项的 μ, ν 都换成 λ , 则

$$\delta T = -\frac{1}{2} \int \delta x^\sigma \left(\left(\frac{\partial g_{\mu\nu}}{\partial x^\sigma} - \frac{\partial g_{\sigma\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\sigma}}{\partial x^\nu} \right) \dot{x}^\mu \dot{x}^\nu - 2(g_{\sigma\lambda} \ddot{x}^\lambda) \right) d\tau.$$

该式对任意变分 δx^σ 成立, 当且仅当括号恒为零, 因此

$$\frac{1}{2} \left(\frac{\partial g_{\mu\nu}}{\partial x^\sigma} - \frac{\partial g_{\sigma\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\sigma}}{\partial x^\nu} \right) \dot{x}^\mu \dot{x}^\nu = g_{\sigma\lambda} \ddot{x}^\lambda,$$

把 $g_{\sigma\lambda}$ 移过去即得测地线方程:

$$\frac{d^2 x^\lambda}{d\tau^2} + \frac{1}{2} g^{\sigma\lambda} \left(\frac{\partial g_{\sigma\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \right) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0. \quad (1.3.45)$$

可见, 遵从测地线的质点总使得其所经历的固有时为极值, 并且这正是与度规适配的克氏符.

当然, 接触过变分法的读者可能发现了, 以上过程就相当于证明 *Euler-Lagrange* 方程

$$\frac{d}{d\tau} \frac{\partial L}{\partial \dot{x}^\mu} - \frac{\partial L}{\partial x^\mu} = 0, \quad (1.3.46)$$

它是为“变分为零”程序化操作专设的充要条件. 将 $L = \sqrt{-g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}$ 代入该方程也是可以的, 于是

$$\frac{\partial L}{\partial \dot{x}^\mu} = -g_{\mu\nu}\dot{x}^\nu, \quad \frac{\partial L}{\partial x^\mu} = -\frac{1}{2}\frac{\partial g_{\nu\sigma}}{\partial x^\mu}\dot{x}^\nu\dot{x}^\sigma,$$

代入有

$$-g_{\mu\nu}\ddot{x}^\nu - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\dot{x}^\sigma\dot{x}^\nu + \frac{1}{2}\frac{\partial g_{\nu\sigma}}{\partial x^\mu}\dot{x}^\nu\dot{x}^\sigma = 0,$$

此即

$$g_{\mu\nu}\ddot{x}^\nu + \frac{1}{2}\left(\frac{\partial g_{\mu\nu}}{\partial x^\sigma} + \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\nu\sigma}}{\partial x^\mu}\right)\dot{x}^\nu\dot{x}^\sigma = 0.$$

1.3.5 测地偏离

现在我们终于要研究曲率怎么“代表”引力了! 前面说过, 引力可以将两条“不受力”的局部惯性系轨迹相互吸引, 就好像球面上的曲率, 使两条最初平行的测地线不能永远保持不相交. 下面就来具体研究曲率如何改变测地线间的距离.

设想一块儿饼干或者一部 Einstein 电梯, 回忆一下, 它们当然不再是一个局部的点, 因此等效原理不能整体成立, 引力和惯性力可以在整体上区分出来, 因为它们会形变, 这就是潮汐效应 (*tidal effect*). 我们用测地线的观点来看待这件事. 我们说两条测地线间距会随时间变化, 而一块形变的饼干上任意两质点的世界线间的距离确实是在变化的.

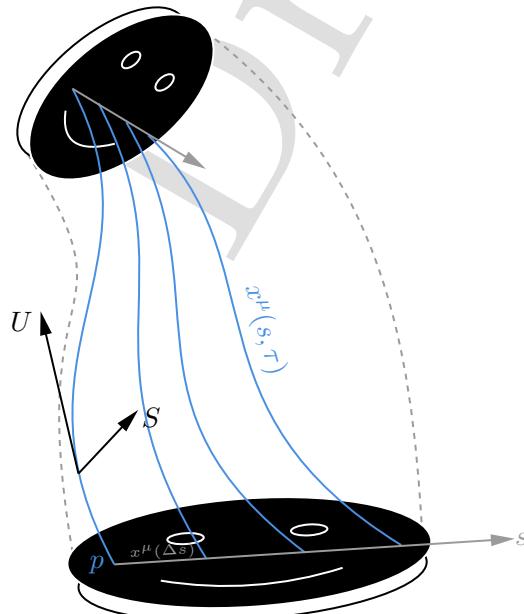


图 1.14: 测地偏离 (s 轴不一定是直的)

如图，让我们研究其中一个质点 p 及其局部，从该点延伸出的不是这块饼干的霉菌毛，那是该点的世界线，咱给它编个号 $s = 0$ ，这样的话线条就记作 $x^\mu(0, \tau)$ ， p 点坐标就规定是 $x^\mu(0, 0)$ 。为了与之比较，我们就得在它局部再任意找条测地线，如图中点 $p' = x^\mu(\Delta s, 0)$ 发出的 $x^\mu(\Delta s, \tau)$ 。此两条测地线就好似两条河岸一般，而我们就是要打算研究这个河宽的变化。

这条河构成了一张窄窄的曲面 $x^\mu(s, \tau)$ ，窄是因为目前取 $s \in [0, \Delta s]$ 。但我们不妨继续按照这个思路走，将其沿着 s 方向延伸，于是我们就相当于在饼干某点 $p = x^\mu(0, 0)$ 的任意方向上取了一张世界面 (*worldsheet*)。它一方面是一张 2 维曲面 (*surface*)，另一方面它是这个方向上的质点所发出的世界线，这样河的流线整体又叫作单参测地线汇 (*one parameter congruence*)，因为区分每条线只要一个参数 s 就行。可以想象，因为只随便取了一个方向，而如果这饼干是 2 维的，其上任意点都有 s, k 两个参数，我们就有了这块饼干作为质点系 (*system of point mass*) 的世界线汇 (*congruence of worldlines*)，同时也可看作一个立体的世界管 (*worldtube*)。3 维电梯也有其世界管，只不过这个管当然就是个 4 维体积了⁵³。

世界面上任意一点当然都对应着一条测地线，因此每一个点都有一个关于测地线的切失或者说速度

$$U^\mu = \frac{\partial x^\mu}{\partial \tau}, \quad (1.3.47)$$

取偏导是考虑到有两个参数，因而 U^μ 是关于 s, τ 的。按之前的思想，考虑 p, p' 河宽变化，其实就是在研究 $x^\mu(\Delta s, \tau) - x^\mu(0, \tau)$ 怎么随 τ 变，而它正是

$$S^\mu = \frac{\partial x^\mu}{\partial s} \quad (1.3.48)$$

的分子。直观上，分母如果都看作相等的 Δs ，那这些流线就有点像是坐标系的等距常数 s 线。比较每个等距点 s 的流线间距，当然就相当于比较这个矢量的大小，因为分母是一样的。所以这个矢量可以代表测地线的相离程度，此亦正是物体间的相对距离，因此 S 称作偏移矢量 (*separation vector*)。

现在我们期望直接研究 S 沿着测地线的“加速度”，也正是物体间的相对加速度。那到时候引力（的差距）或者说潮汐力就容易表述了。对 S 沿着测地线求协变的导数，这就是说

$$\frac{D^2 S^\mu}{d\tau^2} = U^\rho \nabla_\rho U^\sigma \nabla_\sigma S^\mu, \quad (1.3.49)$$

答案是

$$\frac{D^2 S^\mu}{d\tau^2} = R^\mu_{\rho\nu\sigma} U^\rho U^\nu S^\sigma, \quad (1.3.50)$$

⁵³以后我们将会提及，在弯曲时空中最好把坐标系和参考系区分开来：参考系就是指物体或者一个场的世界线汇（当然，还要配上处处的世界线切矢，观者用之以测量），而坐标系则可很任意地选择了。

这称作测地偏离方程 (*geodesic deviation equation*). 可见曲率直接描述的就是潮汐力大小. 证明大可直接拆开协变导数做硬算, 然后将 $\Gamma^\lambda_{\mu\nu}$ 们整理成曲率. 但为了避免太多冗杂的符号, 我们先证明这样一个结论:

$$U^\rho \nabla_\rho S^\mu = S^\rho \nabla_\rho U^\mu. \quad (1.3.51)$$

这个拆开是很好证的:

$$\begin{aligned} \frac{\partial x^\rho}{\partial \tau} \nabla_\rho \frac{\partial x^\mu}{\partial s} &= \frac{\partial x^\rho}{\partial \tau} \left(\frac{\partial^2 x^\mu}{\partial x^\rho \partial s} + \Gamma^\mu_{\sigma\rho} \frac{\partial x^\sigma}{\partial s} \right) \\ &= \frac{\partial^2 x^\mu}{\partial \tau \partial s} + \Gamma^\mu_{\sigma\rho} \frac{\partial x^\sigma}{\partial s} \frac{\partial x^\rho}{\partial \tau} \\ &= \frac{\partial x^\rho}{\partial s} \left(\frac{\partial^2 x^\mu}{\partial x^\rho \partial \tau} + \Gamma^\mu_{\sigma\rho} \frac{\partial x^\sigma}{\partial \tau} \right) \\ &= \frac{\partial x^\rho}{\partial s} \nabla_\rho \frac{\partial x^\mu}{\partial \tau}, \end{aligned}$$

这样的话

$$\begin{aligned} \frac{D^2 S^\mu}{d\tau^2} &= U^\rho \nabla_\rho (U^\sigma \nabla_\sigma S^\mu) \\ &= U^\rho \nabla_\rho (S^\sigma \nabla_\sigma U^\mu) \\ &= (U^\rho \nabla_\rho S^\sigma) (\nabla_\sigma U^\mu) + U^\rho S^\sigma \nabla_\rho \nabla_\sigma U^\mu, \end{aligned}$$

就出现了二阶协变导数, 我们就可以将其往曲率上靠. 一方面根据微积分学,

$$[\nabla_\mu, \nabla_\nu] (U^\lambda \omega_\lambda) = \left[\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu} \right] (U^\lambda \omega_\lambda) = 0,$$

而另一方面根据定义,

$$\begin{aligned} [\nabla_\mu, \nabla_\nu] (U^\lambda \omega_\lambda) &= \nabla_\mu (U^\lambda \nabla_\nu \omega_\lambda + \omega_\lambda \nabla_\nu U^\lambda) - \nabla_\nu (U^\lambda \nabla_\mu \omega_\lambda + \omega_\lambda \nabla_\mu U^\lambda) \\ &= U^\lambda \nabla_\mu \nabla_\nu \omega_\lambda + \omega_\lambda \nabla_\mu \nabla_\nu U^\lambda - U^\lambda \nabla_\nu \nabla_\mu \omega_\lambda - \omega_\lambda \nabla_\nu \nabla_\mu U^\lambda, \end{aligned}$$

从而

$$\begin{aligned} \omega_\lambda [\nabla_\mu, \nabla_\nu] U^\lambda &= -U^\lambda [\nabla_\mu, \nabla_\nu] \omega_\lambda \\ &= -U^\lambda R^\sigma_{\mu\nu\lambda} \omega_\sigma \\ &= -\omega_\lambda R^\lambda_{\mu\nu\sigma} U^\sigma, \end{aligned}$$

此即

$$R^\lambda_{\mu\nu\sigma} U^\sigma = -[\nabla_\mu, \nabla_\nu] U^\lambda, \quad (1.3.52)$$

则继续有

$$\begin{aligned}\frac{D^2 S^\mu}{d\tau^2} &= (S^\rho \nabla_\rho U^\sigma) (\nabla_\sigma U^\mu) + U^\rho S^\sigma (\nabla_\sigma \nabla_\rho U^\mu - R^\mu_{\rho\sigma\nu} U^\nu) \\ &= (S^\rho \nabla_\rho U^\sigma) (\nabla_\sigma U^\mu) + S^\sigma \nabla_\sigma (U^\rho \nabla_\rho U^\mu) - (S^\sigma \nabla_\sigma U^\rho) \nabla_\rho U^\mu - R^\mu_{\rho\sigma\nu} U^\nu U^\rho S^\sigma \\ &= -R^\mu_{\rho\sigma\nu} U^\nu U^\rho S^\sigma = R^\mu_{\rho\nu\sigma} U^\rho U^\nu S^\sigma,\end{aligned}$$

至此就完成了证明.

测地偏离方程又称为 *Jacobi* 方程, 因此测地线上的偏离矢量场又称作 *Jacobi* 场. 该方程的解是很多的, 不过我们对其中一种解很感兴趣: 一般的 S 自然是不必同 U 正交的, 但总可构造出一个与其正交的偏离矢量

$$\eta^\mu = S^\mu + g_{\sigma\nu} U^\sigma S^\nu U^\mu, \quad (1.3.53)$$

可以验证:

$$\begin{aligned}g_{\mu\nu} U^\nu \eta^\mu &= g_{\mu\nu} U^\nu S^\mu + g_{\sigma\nu} U^\sigma S^\nu g_{\mu\nu} U^\nu U^\mu \\ &= g_{\mu\nu} U^\nu S^\mu - g_{\sigma\nu} U^\sigma S^\nu = 0;\end{aligned}$$

而且也满足测地偏离方程. 以后我们将看到, 如果测地线切矢就是时间轴的方向, 那么与之正交就代表着空间方向 (进而彻底成为 3 维空间矢量!), 因此这是个性质很良好的矢量.

说了这么多几何学的事儿, 只为了弄清楚一件东西: 那个“深刻事物”. 现在我们知道引力很可能可以用绝对的曲率表述, 因此是一种“深刻事物”. 而其“另一半”——物质, 是其产生的原因. 为了弄清物质如何影响曲率, 我们可以再次回到潮汐力上. 一方面, 潮汐力用曲率的语言写就是

$$\frac{D^2 \eta^\mu}{d\tau^2} = R^\mu_{\rho\nu\sigma} U^\rho U^\nu \eta^\sigma, \quad (1.3.54)$$

其中我们选择与切矢正交的 *Jacobi* 场; 而另一方面, Newton 力学可以这样描述两个相邻物体的潮汐力: 在引力势场 ϕ 中一个物体的 3 维加速度是

$$\frac{d^2 x^i}{dt^2} = -\eta^{ij} \frac{\partial \phi}{\partial x^j}, \quad (1.3.55)$$

其中注意到梯度 ∇ 的分量就是偏导, 而利用 $\eta^{\mu\nu}$ 升指标在 3 维空间分量下和 δ^{ij} 无异. 则与其临近物体的相对加速度是

$$\begin{aligned}\frac{d^2 \eta^i}{dt^2} &= \frac{d^2(x^i + \eta^i)}{dt^2} - \frac{d^2 x^i}{dt^2} \\ &\approx \frac{\partial}{\partial x^k} \left(\frac{d^2 x^i}{dt^2} \right) \eta^k,\end{aligned}$$

即

$$\frac{d^2\eta^i}{dt^2} = -\eta^{ij}\frac{\partial^2\phi}{\partial x^j\partial x^k}\eta^k, \quad (1.3.56)$$

这个近似是考虑到其临近性. 要使得曲率的语言在低速情景上 (此即 $t \sim \tau$) 还原到 Newton 理论, 即是想要类比一下二者:

$$\eta^{ij}\frac{\partial^2\phi}{\partial x^j\partial x^k} \sim R^\mu_{\rho\sigma\nu}U^\rho U^\nu,$$

而左边若取 $k = i$, 则立即成为了 Poisson 方程:

$$\eta^{ij}\frac{\partial^2\phi}{\partial x^i\partial x^j} = 4\pi G\rho, \quad (1.3.57)$$

而这个操作对应着右边缩并为

$$R^\mu_{\rho\mu\nu}U^\rho U^\nu = R_{\rho\nu}U^\rho U^\nu, \quad (1.3.58)$$

可见, 这个所谓的密度 ρ 若理解为一组线汇或者说物质流线 (*streamline*) 的 3 维密度, 为了描述其与 $R_{\mu\nu}$ 相同多自由度的质量属性, 我们必须为之也同样构造一个量使得:

$$R_{\mu\nu}U^\mu U^\nu \sim 4\pi GT_{\mu\nu}U^\mu U^\nu, \quad (1.3.59)$$

或者说

$$T_{\mu\nu}U^\mu U^\nu = \rho. \quad (1.3.60)$$

也就是说, 我们希望这个 $T_{\mu\nu}$ 能全面地描述线汇的物质属性, 包括 ρ , 但当然不止 ρ ! 因为我们想要用场的语言描述物质团体, 而无论是物体 (*body*) 还是流体 (*fluid*), 都没有质点模型那样简单. 尤其拿流体来说, 除了 ρ 外, 当然还有个属性叫压强, 这是不能单靠质量囊括的, 而只能指望 $T_{\mu\nu}$. 读者或许会误会, 比较按一般的认知压强应该看作动力效应而非物质的内禀属性, 但正由于流体的无穷自由度 (*infinite freedom*) 导致了我们只能将其看作平均的集合体, 而所有你所熟知的诸如质点轨迹、质量等物理量就没了意义, 因为我们说不出哪个质点能代表整个东西的运动. 这个时候, 诸如质量、压强、速度场 (*velocity field*) 等概念就更为有用了⁵⁴.

1.4 广相应该是怎样的一套理论?

1.4.1 理想流体

下面的目的便是找到一张能重组统一密度、压强等所有概念的矩阵数表 $T_{\mu\nu}$. 为此, 可以回到狭相中, 看看它对于物质及物理量做了怎样的诠释, 这里便蕴含着统一

⁵⁴ 不仅如此, 压强应当视作彻底的内禀属性有很多支撑性证据, 比如, 以后会知道, 压强与能量密度之间的关系 $p = w\rho$ 直接决定了辐射、普通物质、宇宙学常数等不同成分对宇宙膨胀速度的影响.

思想. 前面提到, Einstein 提出了著名的质能关系, 根据现在所用的单位制是

$$E = \frac{m}{\sqrt{1 - v^2}} = \gamma m, \quad (1.4.1)$$

并且看到在静止系中质量和能量等同: $E_0 = m$. 这也是最为人所知的一个关系, 并且注意 $T_{\mu\nu}U^\mu U^\nu = \rho$ 就是质能关系的推广! 其似乎揭示了这些量或许是背后某个“深刻事物”的不同侧面, 正如时间和空间之于时空、各三维测量值之于四维内禀量那样.

这些观念是我们在 1.1.5 节讨论“很大尺度”的引力时思索出来的. 在这个尺度下的引力对象就无所谓刚体了, 所以一旦引入物质场的概念, 就必须引入密度、压强等概念以推广为流体的概念. Einstein 注意这样一件事, 之前我们只将 3 维速度 v 推广到了 4 维中的 U , 关系是

$$U^\mu = \frac{dx^\mu}{d\tau}, \quad (1.4.2)$$

其与 3 维速度分量 v^i 的关系是

$$U^i = \gamma \frac{dx^i}{dx^0} = \gamma v^i. \quad (1.4.3)$$

但这只是运动学的部分, 我们还并没有将整个动力学搬到时空上来. 比如动量, 类比 $p^i = mv^i$, 它在时空中可以定义为

$$P^\mu = mU^\mu, \quad (1.4.4)$$

这样甚至能统一能量(质量)和动量, 因为 $P^0 = \gamma m = E$ 而 $P^i = \gamma p^i$. 可见, 如果我们寻求密度和压强的统一, 可以试着观察一下动量流密度 (*momentum flux density*), 简称动流密度:

$$T_{\mu\nu} \propto \rho U_\mu U_\nu = \frac{dm U_\mu U_\nu}{dx^1 dx^2 dx^3}, \quad (1.4.5)$$

这里 $U_\mu = g_{\mu\nu}U^\nu$ 是其对偶. 这样立即有 $T_{00} \propto \frac{dm}{dV} = \rho$ 且 $T_{ii} \propto f^i/dS_i$, 这里 dS_i 表示 dx^i 所对面积, 比如 dx^1 对应 $dx^2 dx^3$ 的法向量, 那么这即是正压强的 i 分量. 可见, 一方面动流密度就是压强, 而在时间分量上它就是质量或者能量密度. 这启发我们至少在低速情景下(这样压强很小) 定义

$$T_{\mu\nu} \approx \rho U_\mu U_\nu, \quad (1.4.6)$$

那么当然有 $T_{\mu\nu}U^\mu U^\nu = \rho$.

Einstein 首先考虑理想流体的引力效应. 针对流体某一点处, 若取与之相对静止的观者, 所见应当是静态而各向同性的流体⁵⁵, 这称作理想流体 (*perfect fluid*). 这

⁵⁵这样压强在所有方向相同, 将是球对称的标量. 这样认为各部分之间的作用力都是在界面的法线方向上的, 所以没有剪切力, 单位面积上的压力称为正压强 (*pressure*).

样根据此前讨论，在某点的相对静止系 \tilde{x}^β 中 T 就成为了对角矩阵：

$$\tilde{T}_{00} = \rho, \quad \tilde{T}_{ij} = p\delta_{ij}, \quad \tilde{T}_{0i} = \tilde{T}_{i0} = 0,$$

其中 p 是标量压强。将其切换到我们的参考系 x^μ 中来，那么流体在给定点就是以 \vec{v} 运动，则我们像这样 boost 一下：

$$T_{\mu\nu} = \Lambda^\gamma{}_\mu \Lambda^\sigma{}_\nu \tilde{T}_{\gamma\sigma},$$

其中我们得先求出任意方向的 Lorentz 变换，空间坐标中与 \vec{v} 垂直的分量应当在变换下不变，而平行的分量应当与时间混合起来 boost：

$$x^0 = \gamma \tilde{x}^0 - \gamma v_i \tilde{x}^i, \quad x^i = \tilde{x}^i + \frac{\gamma - 1}{v^2} (v_j \tilde{x}^j) v^i - \gamma v^i \tilde{x}^0, \quad (1.4.7)$$

这样写成矩阵就是

$$\Lambda^0{}_0 = \gamma, \quad \Lambda^i{}_0 = -\gamma v^i, \quad \Lambda^0{}_i = -\gamma v_i, \quad \Lambda^i{}_j = \delta^i_j + v^i v_j \frac{\gamma - 1}{v^2}. \quad (1.4.8)$$

这里不区分 δ^i_j 指标的左右是考虑到恒等变换 I 是对称矩阵。这样我们就有

$$T_{00} = \gamma^2 (\rho + v^2 p), \quad T_{i0} = \gamma^2 (\rho + p) v_i, \quad T_{ij} = p \delta_{ij} + \gamma^2 (\rho + p) v_i v_j,$$

或者说

$$T_{\mu\nu} = p \eta_{\mu\nu} + (\rho + p) U_\mu U_\nu. \quad (1.4.9)$$

这就是理想流体的能动张量 (*energy-momentum tensor*)。虽然目前是第一次正经地看到张量 (*tensor*) 这个词，但读者对此并不陌生：因为先前我们讨论的所有具有协变性的指标物理量都叫张量！从直观上讲，这个词最开始是由 Cauchy 等人研究连续介质中的应力理论而提出的。所谓的应力 (*stress*)，就是压强的推广，Cauchy 用以表示形变的连续介质内各个部分的相互作用力，但连续介质、流体当然不同于刚体，外力作用是通过作用于各个部分之间的“力”来表现的，其作用于相邻部分的界面上的。所以，应力应当依赖于两个方向：一是作用力的方向，一是界面的方向（用其法线方向表示）。如果某作用力和某面垂直，这样就成了该方向的正压强，降低了自由度就成为了矢量；而一般的应力不是矢量，因为其自由度由两个独立的矢量表示，所以叫做 2 阶 (*second order*) 张量。

这样根据连续介质力学的思想，也可以说，这些物理量就随流体而“流动”，而 T 正好能够描述物理量如何沿不同坐标流动⁵⁶。比如说 T_{00} 就是 ρ ，描述多少能量经

⁵⁶ 量的流动观点不止在宏观流体中出现。谈及与 Newton 力学不同的一些经典体系，人们可能会首先想到 Hamilton 体系等。历史上有另一种纯粹物理角度上的有趣观点，同熟悉的理论一样也抛弃了“力”，不同的是，取而代之的是所谓的动流，它企图以一些电路知识解释经典力学。动流是动量的定向流动，定义一个矢量为动流强度，也便替代了力即 $\mathbf{F} = \dot{\mathbf{p}}$ 。区别在于，任意的静力学系统便可视作一个动流回路。取向右为正，那么小明向右试图拖动静止的大木箱的静力学系统便可视为，由小明作为动流泵，或者动源，不断地从地面汲取动量沿着手臂传给木箱，最终又回到大地。由于许多物体都有摩擦性质，该系统涉及到的物质都是动量的良导体。这里面还有许多更有趣的理论和概念，读者可以自行搜索相关的文献，如 [13]。

过这一点向未来流动，能量的流动成了能流 (*energy flux*)； T_{0i} 就是动量密度，也即能流密度，表示能量随时间在空间中运动的速度； T_{ij} 就是应力也即动流密度，描述多少能量沿该空间坐标运动，即从时间向空间挤压的趋势；非正压强的部分包含了坐标间的耦合，表述能量把其运动传递给周围的趋势。这些通量大小显然与参考系相关：事件在空间上静止，但不断地往未来迁移，能量不断经时间移动；如果在空间上运动，则有一部分能量会流经空间方向——在同一个时空上，静能可以转化为空间动能，这就是空间运动的相对性。将物质的运动、影响等均视为是能量的流动，因为质量就是能量！时空的各种内容物：普通质量、电磁场等等，当然就影响能流。各种形式的能量本身就是穿梭时空的流体。

考虑时空中的某一点，能量在该点的流动方向可归结为每个坐标的正负方向。但对于理想流体而言，由于忽略真实情况里的那些粘性、压缩和热传导⁵⁷，一个点或者一个曲面等流经能量，必定是从这儿进多少，从这儿就得出多少。这就是连续性 (*continuity*)。因此我们只需考察四个坐标即可：不论所涉及的物质场⁵⁸本性如何，这就是在说 $T_{\mu\nu}$ 的散度为零：

$$\nabla^\mu T_{\mu\nu} = 0, \quad (1.4.10)$$

其中协变导数在普通坐标下 $\nabla^\mu = \eta^{\mu\nu} \frac{\partial}{\partial x^\nu}$ ，而推广到任意时空任意坐标系就需要协变性的介入，像这样缩并一下就称作协变散度 (*covariant divergence*)。上式体现了在微分水平上能动张量的守恒 (*conservation*)，在第二章中会知道，可利用积分求得整体上能动张量的平衡关系 (Gauss 定理)。

1.4.2 广义协变性

对张量有这样一种更为严格的定义，使得其坐标变换直接成为推论。对于时空中的一点 q ，比如说，一个具备两个下指标的张量 T 其实就是个线性函数

$$T : \mathbf{R}_q^4 \times \mathbf{R}_q^4 \rightarrow \mathbf{R} \quad (1.4.11)$$

而已，其中 \mathbf{R}_q^4 表示 q 处的矢量空间。我们知道，狭相中可不强调任意点矢量空间同时空 \mathbf{R}^4 本身的区别，但弯曲时空的任意坐标系中，有必要强调这个矢量空间只是 q 的。 T 作用在两个矢量上，而给出一个实数

$$T(U, \tilde{U}) = T_{\mu\nu} U^\mu \tilde{U}^\nu \quad (1.4.12)$$

⁵⁷ 这样一些条件在连续介质力学中由一系列状态方程 (*equations of state*) 表示。一般来说能动张量应从物质场的整体由它的构成状态得出，而这些状态方程则依赖于其本性与之间的相互关系。由于考虑的是理想流体，我们在此不去为这些事操心。

⁵⁸ 一个点的能动张量在忽略自身的影响下，只考虑所处环境的物质构成。比如忽略月球自身，那么它在地球外的能动张量为零，但如果地球是气态星球，月球比较小而处于地球内部，假如气体为无粘性均衡的理想球状，则能动张量与地球质量、指向外侧以抵抗自身坍塌的压强有关。

作为其像. 因此, 选择保持度规的坐标变换, 当然也就能推出 T 具有 Lorentz 协变性 (那一串偏微分乘积)! 其实在知道张量定义后, 很容易发现, 除了 T 以外, 度规当然也是个张量; 不仅如此, 它与 Ric, T 都是 2 阶对称张量.

很清楚, 想在一般坐标系里表示内积, 度规的分量和它们的导数就会出现在方程中. 所以我们所企图寻求的物质和引力场的方程, 要使得它们在任意坐标系下都有同样的形式, 则按照他的理解, 这就意味着所有出现的对象都应该按照张量那样变换, 故应该看成是先验地存在着但是“尚未算出的”对象. 这个原理称为“广义协变性 (*general covariance*)”. 从狭相到广相, 应该把 η 换成一个尚未算出的 2 阶对称张量, Lorentz 度规 g . 可以写出“尚未算出的” g 的方程, 而这个方程就会给出“已经算出的” η . 这样, 广义协变性本身并不是强迫我们抛开 η 另起炉灶, 而是注意到 g 和 T 有一样的分量, 所以, 把 g 就看作是引力场的体现, 并且努力去找一个把 g 和 T 直接连接起来的方程就是很自然的了.

对于一般的度规, 能动张量就是

$$T_{\mu\nu} = pg_{\mu\nu} + (\rho + p)U_\mu U_\nu, \quad (1.4.13)$$

那么怎样只用 g 来类似地陈述能动守恒呢? 那还是之前所说的协变散度了, 借助它就可以对任意 g 写出能动守恒.

甚至可以预料这一点: 如果考虑质量场集中在一点的极限情况, 或者说能动张量只在一条世界线上非零的极限情况, 此时可以证明, 此守恒流将是 g 下的测地线. 在这个极限下, 不论能动张量的本性如何, 物质的运动都是由 g 的几何学所决定的. 这样, 所有的对象都以同样的方式“下落”, 就如 Galileo 实验中重物的下落与质量无关一样. 这些考虑使得等效原理有了具体的体现, 并且正如 Einstein 所想的, 一旦写出场方程, 甚至可以抛弃掉测地线方程, 即运动学已经蕴藏于场方程之中.

此外, 对于一般的 g , $\nabla^\mu T_{\mu\nu} = 0$ 很可能不给出能量或动量的守恒. 只有当 g 还有一些对称性时才会是这样. 基本的守恒律一般地只在无穷小层次上仍然保留, 这件事使我们对这些物理定律的本性又有了新的洞察. 我们将在第三章对此继续详述.

1.4.3 场方程

在此前潮汐力的类比中已给出了关于 g 的曲率同 T 的联系. 由于希望方程能以尽可能简单的形式体现“广义协变性”, 即其中除了 g 和 T 以外别无它物, 而类比中已给出一个垂手而得的张量 Ric , 因此可首先猜测场方程是

$$R_{\mu\nu} = 4\pi G T_{\mu\nu},$$

但另一方面, 我们所要求的限制是, 场方程必须有推论 $\nabla^\mu T_{\mu\nu} = 0$, 这即要求

$$\nabla^\mu R_{\mu\nu} = 0,$$

虽然大可以直接拆开定义计算，但不妨先学习这个结论以简化：可从定义证明，对任意 g 有 *Bianchi* 恒等式 (*identity*)

$$\nabla_\alpha R^\mu{}_{\beta\gamma\delta} + \nabla_\beta R^\mu{}_{\gamma\alpha\delta} + \nabla_\gamma R^\mu{}_{\alpha\beta\delta} = 0, \quad (1.4.14)$$

令 $\mu = \alpha$ 得

$$\nabla_\alpha R^\alpha{}_{\beta\gamma\delta} + \nabla_\beta R_\gamma{}^\alpha{}_\delta - \nabla_\gamma R_\beta{}^\alpha{}_\delta = 0,$$

用 $g^{\beta\delta}$ 给其升指标有

$$\nabla_\alpha R^\alpha{}_\gamma + \nabla^\delta R_\gamma{}^\delta - \nabla_\gamma R = 2\nabla^\mu R_{\mu\gamma} - \nabla_\gamma R = 0,$$

此即得到

$$\nabla^\mu R_{\mu\nu} = \frac{1}{2}\nabla_\nu R, \quad (1.4.15)$$

于是 $\nabla^\mu T_{\mu\nu} = 0$ 要求

$$\nabla_\nu R = 0,$$

这会导致这样一种荒谬结论：令 T 的迹 $T = g^{\mu\nu}T_{\mu\nu}$ ，这样 $R_{\mu\nu} = 4\pi GT_{\mu\nu}$ 就是

$$R = 4\pi GT,$$

因此

$$\nabla_\nu T = 0,$$

即 T 在整个时空里是个常数。但另一方面，取低速情况使得 p 相对于 ρ 而言忽略不计，则

$$T \approx \rho U_\mu U^\mu = -\rho, \quad (1.4.16)$$

推出 ρ 为常数，这当然是不一定成立的！可见还需要在 $R_{\mu\nu}$ 基础上再尽可能简单地凑出新张量 $G_{\mu\nu}$ 使得 $\nabla^\mu G_{\mu\nu} = 0$ 且与 $T_{\mu\nu}$ 有线性关系，但其实只要稍微移项就有

$$\nabla^\mu \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \right) = 0, \quad (1.4.17)$$

括号内称为 *Einstein* 张量。不妨再试试它的可能性，假设比例系数是 κ ，即

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa T_{\mu\nu}, \quad (1.4.18)$$

现在来检查其是否能给出潮汐力近似

$$R_{\mu\nu}U^\mu U^\nu \sim 4\pi GT_{\mu\nu}U^\mu U^\nu.$$

一方面

$$\kappa T = R - \frac{1}{2}\delta_\mu^\mu R = -R, \quad (1.4.19)$$

因此

$$R_{\mu\nu} = \kappa T_{\mu\nu} + \frac{1}{2}g_{\mu\nu}(-\kappa T) = \kappa \left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T \right), \quad (1.4.20)$$

则

$$\begin{aligned} R_{\mu\nu}U^\mu U^\nu &= \kappa \left(T_{\mu\nu}U^\mu U^\nu - \frac{1}{2}g_{\mu\nu}U^\mu U^\nu T \right) = \kappa \left(\rho + \frac{T}{2} \right) \\ &\approx \kappa \left(\rho - \frac{\rho}{2} \right) = \frac{\kappa}{2}\rho = \frac{\kappa}{2}T_{\mu\nu}U^\mu U^\nu, \end{aligned}$$

可见确实能导出，并且还能知晓 $\kappa = 8\pi G$ ，称作引力密度。综上，经过一系列充满灵感的思维飞跃，我们终于得到了场方程 (*field equation*) 最简单的形式

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu}, \quad (1.4.21)$$

因为在符合了潮汐力近似后，场方程中就再没有其他可以调整的参数了，余下要做的就是用各种实验和应用去验证它，但并不排除在某些特殊情况做小的修补。

1.4.4 作用量

令 $T_{\mu\nu} = 0$ ，当然也就有 $T = 0$ ，则立即得到这样一种特例——真空方程 (*vacuum field equation*)

$$R_{\mu\nu} = 0. \quad (1.4.22)$$

它也可以对应于一个变分原理，其泛函是

$$S(g) = \int R\sqrt{-g} d^4x, \quad (1.4.23)$$

这里以后会知道 $\sqrt{-g} d^4x$ 表示时空体元 (*volume element*)，见第二章； g 是度规的行列式。先暂时接受结论： $\delta R_{\mu\nu} = 0$ 和代数学的 Jacobi 公式

$$\frac{d|A|}{dt} = |A| \text{Tr} \left\{ A^{-1} \frac{dA}{dt} \right\}, \quad (1.4.24)$$

这里 $|A|$ 表示行列式， Tr 表示求迹，则

$$\delta g = gg^{\mu\nu}\delta g_{\mu\nu} = gg_{\mu\nu}\delta g^{\mu\nu}. \quad (1.4.25)$$

下面开始变分：

$$\begin{aligned} \delta S &= \int (\sqrt{-g} \delta(g^{\mu\nu}R_{\mu\nu}) + R \delta\sqrt{-g}) d^4x \\ &= \int \left(\sqrt{-g}R_{\mu\nu}\delta g^{\mu\nu} - \frac{R}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu} \right) d^4x \\ &= \int \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \right) \delta g^{\mu\nu} \sqrt{-g} d^4x = 0, \end{aligned}$$

而读者易证 $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 0$ 等价于 $R_{\mu\nu} = 0$ 。看来的确是可行的。从场方程得到的许多有趣现象，其实在真空情况中就已经出现了：这多少有点讽刺，因为正是有了物质及其守恒方程才给出了场方程。

Hilbert 一直在紧随着 Einstein 构建具有动态 g 的理论，他往泛函中加入了关于物质的部分：

$$S(g) = \int (R + 2\kappa L_M) \sqrt{-g} d^4x, \quad (1.4.26)$$

于是我们发现可从另一角度定义能动张量

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta}{\delta g^{\mu\nu}} (L_M \sqrt{-g}), \quad (1.4.27)$$

使得

$$\delta S = \int \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \kappa T_{\mu\nu} \right) \delta g^{\mu\nu} \sqrt{-g} d^4x = 0,$$

便从不同的思路得到场方程⁵⁹，这个泛函也就称为 *Einstein-Hilbert 作用量 (action)*。

1.4.5 时空的拓扑结构

走了这么远，我们还一直没想过“度规究竟定义在哪儿”这个问题。当然我们知道，从 η 走向更一般的 g 时，确实应该考虑把 \mathbf{R}^4 换成别的背景。当然，这在理论构建之初无需过多考虑，然而在后期发展中，人们逐渐意识到，引力除体现在用局部坐标写的度规上，还与时空的整体拓扑性质 (*global topology*) 密不可分——而这不是由场方程先验决定的。

这一点在几何学中看得很清楚：原来，时空连续统的数学模型应该是一个 *Lorentz 流形 (manifold)* (M, g) ，其中 M 是微分流形 (*differentiable manifold*)——为了把 g 用局部坐标写出来，需要在最底层的拓扑对象 (*topological space*) 上处处 (光滑地) 指定坐标，办法就是用若干个坐标系 $\{x^\mu\}$ 去覆盖整个拓扑对象，这样我们可以说流形处处局部地同胚于 (*be homeomorphic to*) \mathbf{R}^4 ；而这若干个坐标系相容地粘贴 (*compatibly stick to*) 在一起，“拼”出了一个光滑的 M ，从而保证可以处处定义 g ——说明我们其实允许时空可以不再是 \mathbf{R}^4 ，而是一个更一般的流形。

这样，就可以囊括其拓扑性质了，即场方程的未知量不只是 g ，而是整个 (M, g) 。

1.4.6 这套理论有哪些应用？

以下我们将不赘述具体细节，而只对其应用做一个概览即可。

⁵⁹ 但后来的文献表示其实 Hilbert 并没有先于 Einstein。且 Einstein 在原论文中亦使用的是作用量手段，但也不代表他是用作用量凑出场方程的。

水星近日点进动

通过求解在所谓的线性近似 (*linear approximation*) 下的场方程，就可以算出水星近日点进动的正确值。只要能够取得这个进动的正确值，就表示广相的正确性已经通过了该实验的检验。

星光偏折

Sir Arthur Stanley Eddington(1882-1944, 英国天体物理学家) 在 1919 年的一次日食观测中摄下了光线被引力“弯曲”的图片。这一点已经在几何光学近似中从理论上计算过。从那以后，用太阳系的种种检验对场方程作了所谓的后 *Newton* 理论 (*post-Newtonian*)，这些预测和实验都在这个物理条件下以很高的精度证实了广相。

黑洞

除了闵氏度规，真空中另一著名的解 (solution) 是 Schwarzschild⁶⁰ 度规，它描述的是一个小型均匀静态球体的引力场。球坐标下是

$$-A dt^2 + A^{-1} dr^2 + r^2 d\Omega^2, \quad (1.4.28)$$

其中 $A = 1 - \frac{2Gm}{r}$ ，这里 m 是球体质量。我们注意当距离渐小到 $r = 2Gm$ 时，此处的时间在外界看来似乎停滞流逝（请观察度规的时间项）。我们称这个曲面 \mathcal{H}^+ 为黑洞 (*black hole*) 的视界 (*horizon*)。以后会知道，若对其做简单的坐标变换，则可把这个度规越过视界而延拓到内部去，而只有在视界上需要用正则坐标 (*normal coordinate*) 重写。视界定义了一个时空区域的边缘，而从这个区域可以把信号传送到“无穷远 (*infinity*)” \mathcal{I}^+ 。不能把信号传送到远处的点之集合称为黑洞区域 (*region*)，进而视界就成了黑洞表面。人们现在普遍认为星体引力坍缩 (*gravitational collapse*) 的末日归宿就是黑洞。如果我们考虑黑洞附近的光线，一路走向测地线方程，所描述的轨迹曲线可以作为引力透镜 (*gravitational lensing*) 的定量解释。

这些问题花了很多时间才理清，因为整体微分几何的流形语言是在场方程之后才发展完整的。

在 $r = 0$ 处会发生什么？首先会发现曲率⁶¹发散，而时空又没办法“越过”这个点而改写坐标，因此这个点叫做奇点 (*singularity*)。进入黑洞的测地线会在有限固有时内到达奇点，不可无限制地延伸下去，称作不可延测地线 (*inextendible geodesic*)。接近奇点的宏观观者会被差距强烈的潮汐力撕碎。

⁶⁰Karl Schwarzschild, 1873-1916, 德国物理学家，他给出的解是第一个除了平坦真空以外的解析解。1915 年论文发表的数月后，他在一战的俄罗斯战场完成了证明，又于数月后病逝。

⁶¹具体是 Kretschmann 标量 $R_{\mu\nu\sigma\lambda}R^{\mu\nu\sigma\lambda}$ ，这里 $R_{\mu\nu\sigma\lambda} = g_{\mu\alpha}R^\alpha{}_{\nu\sigma\lambda}$ 也可叫做 Riemann 曲率。

这种病态性质与高度的球对称性有关。Penrose 在 1965 年给出的奇性定理 (*singularity theorem*) 中指出，耦合适当的物质并附加某些条件后，一定会出现不可延测地线，因而就产生了这么个奇点。奇性定理给出了两个重要猜想 (*conjecture*) 或者假设 (*hypothesis*)。第一个是弱宇宙监督 (*weak cosmic censorship*) 假设：对于适当的物质方程组和一般物理意义的初始条件，奇点一定限制在黑洞区域内而“看不见”；第二个是强 (*strong*) 宇宙监督假设：解的奇性一定与其延拓遇到局部阻碍有关。后一个假设保证了动力学问题的唯一解必定是初始数据产生的经典时空，即经典决定性原理 (*Classical Newtonian determinism*)。如果抛弃条件的一般性，则这两个假设都不成立。Christodoulou 做出过标量场方程组的球对称解，其中的奇点就不在黑洞区域内，这样的时空就称作包含了裸 (*naked*) 奇点。裸奇点很容易构造出来，比如直接令 $m < 0$ ，但此时就不允许渐近平直 (*asymptotically flat*) 的 *Cauchy* 面了，因而可能不太符合咱们的物理常识。这件事与所谓的正能量定理 (*positive mass theorem*) 有关。

宇宙学

之前讨论的时空都是理想的孤立系统。但我们有更大的野心，用时空流形来代表整个宇宙会怎样呢？观测结果建议，极大尺度上，宇宙是均匀 (*homogeneous*) 且各向同性的 (*isotropic*)。这使得不能用 \mathbf{R}^4 上的常数 $\nabla\phi$ 和非零 ρ 解出 Poisson 方程，这样 Newton 的宇宙学就从未成为理性科学。但另一方面，广相确实容许均匀的各向同性的解及其扰动。

宇宙学的早期发展毫无疑问是 Einstein 带头。当时占统治地位的观点是静态宇宙。这使得 Einstein 为了强行凑出静态解，而在场方程左边加上一项 $\Lambda g_{\mu\nu}$ （因为适配性 $\nabla^\lambda g_{\mu\nu} = 0$ ），即

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}, \quad (1.4.29)$$

这里 Λ 称为宇宙学常数 (*cosmological constant*)，假设静态的情况下测出来是个小值。但现代观点自 Hubble 开始普遍认为宇宙是膨胀的 (*expanding*)。比如，FLRW 解就是膨胀宇宙的一阶近似。向时间过去回溯，这些解具有奇点，称为宇宙大爆炸 (*the big bang*)。不过，1990 年代部分天文观测显示，只有将 Λ 重新写进来才能解释所谓暗能量 (*dark energy*) 的贡献，但具体数值却与粒子物理的预言矛盾。

前景

场方程的解析解固然有很多，但它们都比较简单，我们只能对简单解邻域中的一般解有定性了解。前面提及的宇宙监督假设、一般条件的奇点等问题都未得到普

适解答. 而这些问题恰恰对于广相的物理意义和适用范围都有基础衡量. 我们只能期望场方程会继续揭示出美丽的数学结构, 帮助人类进一步地认识物理世界.

Draft

Chapter 2

平直时空

现代物理学中，闵氏时空通常能作为除引力外其他定律的背景，包括量子理论，因此有必要用严格的术语陈述其数学框架。随后将阐明一些已被经验证明的事实，其对相对性理论做出了正确的解释。此外，亦要通过实例讨论物理学性质，如电磁学和流体力学等。

2.1 代数学拾遗

前文谈及矢量空间 \mathbf{R}^4 可理解为所有矢量所构成的集合，但读者大抵是不太喜欢这般搪塞的，毕竟什么又是矢量呢？“满足叠加原理的箭头”这样的说法恐怕还不够。

2.1.1 集合与映射

既然是个集合，那咱就先说回一开始的集合。在集合论定义了集合 (*set*) 和元素 (*element*) 此二概念后，可进一步定义如下四种常见运算（其中承认逻辑学的自然语言表达）¹：

$$\begin{aligned} A \cup B &= \{x|x \in A \text{ 或 } x \in B\} && \text{并集,} \\ A \cap B &= \{x|x \in A \text{ 且 } x \in B\} && \text{交集,} \\ A \setminus B &= \{x|x \in A \text{ 且 } x \notin B\} && \text{余集、差集或补集,} \\ A \times B &= \{(x,y)|x \in A, y \in B\} && \text{卡氏积.} \end{aligned}$$

这里读者可能不熟悉的是第 4 种——*Descartes* 积，简称卡氏积，它不外乎是将两个集合的元素打包在一起，组成一个括号，比如你甚至能把 (M, g) 这种打包行为看成

¹对于结构，诸如集合、群等，通常能以直接的方式定义“子结构”，例如子集、子群等。故将不赘述这一点。

卡氏积，只不过 M 可能属于某种“流形空间”而 g 属于某种“度规空间”。这种写法当然不是唯一的，实际上应当按照 x, y 各自的所有可能“相乘”，因为它们之间“独立”，就好像我们在排列组合学中看到的那样。于是看到，在用分析学定义了实数集 \mathbf{R} 后，就可以像这样复合出一个“2 维”集合： $\mathbf{R} \times \mathbf{R} = \mathbf{R}^2$ 。

定义 2.1 (等价关系). 集合 A 上的一个关系是一个子集 $R \subseteq A \times A$ 。我们说 $x \sim y$ ，如果 $(x, y) \in R$ 。一般只对等价关系感兴趣。等价关系是指对 $\forall x, y, z \in A$ 满足如下定义的关系：

- (反身性) $x \sim x$;
- (对称性) $x \sim y \Rightarrow y \sim x$;
- (传递性) $x \sim y, y \sim z \Rightarrow x \sim z$.

若“ \sim ”是 A 上的一个等价关系，则对 $\forall x \in A$ ，都可定义其等价类 $[x] = \{y \in A : y \sim x\}$ 。可以证明等价类非空。任意两个等价类要么相等，要么不相交。所有等价类构成 A 的一个分解。等价类中的一个元素称为该类的一个代表元。

以后凡提及“等价的”“等价类”均指集合上的等价关系。

在学习集合后，随之而来的便是映射 (*map*) 这么一个概念，而它与函数 (*function*) 又总混为一谈。其实，映射并非一个玄乎的概念，如果按自然语言讲：

定义 2.2 (映射). 映射 $f : A \rightarrow B$ 对 A 中的 $\forall x$ 指定 B 中的一个元素 $f(x)$ 。这样 A 就称为定义域， B 称为上域， x 称为变量， $f(x)$ 称为 x 的像。

可见，映射只是一个指定行为。人类自古以来就寻求万事万物的变量间的对应关系，因此映射只是对人类因果认知规律的描述而已。于是经常有人将映射比喻为一台机器，其输入自变量而输出因变量；又或是一个法则，用以约束二者的关系。

定义 2.3 (复合映射). 映射 $g : A \rightarrow B$ 和 $f : B \rightarrow C$ 可构成复合映射 $f \circ g : A \rightarrow C$ ，定义为 $f \circ g(x) = f(g(x))$, $\forall x \in A$ 。

定义 2.4 (对易). 若 $f(g(x)) = g(f(x))$ ，则称 f, g 是对易的。称 $[f, g](x) = f(g(x)) - g(f(x))$ 的映射 $[f, g]$ 为 f, g 的对易子，则 f, g 对易即说 $[f, g](x) = 0$ ，简写为 $[f, g] = 0$ 或 $f \circ g = g \circ f$ 。

有些细节要注意。映射 f 必须要求对 $\forall x \in A$ 有定义，且只给其指定 $f(x) \in B$ ，可称单值映射以强调。与 $f(x)$ 对应的 x 一般可不止一个。如果 $f(x)$ 只有一个 x 与之对应，就定义了一个特殊的映射：

定义 2.5 (单射). 映射 f 称为单射 \Leftrightarrow 若 $f(x) = f(y)$ ，则 $x = y$ 。

上域不是值域 $f[A]$ ，而是包含它的。如果 B 等于值域，就定义了一个特殊的映射：

定义 2.6 (满射). 映射 f 称为满射 \iff 对 $\forall y \in B$, 都 $\exists x \in A$ 使得 $y = f(x)$. 亦称“到上”映射.

两个结合起来可以使元素一一对应, 即

定义 2.7 (双射). 映射 f 称为双射, 若其既是单射又是满射².

进而可以讨论逆的概念:

定义 2.8 (逆). f 的逆映射 $f^{-1} : B \rightarrow A$ 给 $\forall f(x) \in B$ 指定了 $!x \in A$, 这样 x 就称为 $f(x)$ 的逆像或逆. 因此 f^{-1} 有定义 $\iff f$ 是双射. 一般情况下“元素的逆”不唯一, 但对 $\forall f$ 及 $\forall C \subseteq B$, 总可定义其“集合的逆” $f^{-1}[C] = \{x | f(x) \in C\} \subseteq A$, 尽管它可以是 \emptyset . 元素的逆要求双射, 但集合的逆则随意.

函数与映射的区别在于

定义 2.9 (函数). 上域含于 \mathbf{R} (或 \mathbf{C}) 的映射称为函数. 目前只用到 \mathbf{R} .

更准确地说是含于某些常见数域, 但“域”这个概念还需要一些铺垫, 目前先姑且理解为 \mathbf{R} 或 \mathbf{C} 未尝不可.

2.1.2 矢量空间

说了这么多, 那矢量空间是什么呢? 我们以前就学过矢量应该满足“线性叠加”和“数乘”这两个运算结构. 矢量空间也就是按此思路定义的:

定义 2.10 (实矢量空间). 实矢量空间是指 \mathbf{R} 上的一个矢量空间 (又称矢量空间), 即一个集合 V , 其上定义了两个映射:

$V \times V \rightarrow V$ 表示“加法运算”, 使得对 $\forall x, y, z \in V$ 有

(结合律) $(x + y) + z = x + (y + z)$,

(零元存在性) $\exists \bar{0} \in V$ 使 $x + \bar{0} = x$,

(负元存在性) $\exists \bar{x} \in V$ 使 $x + \bar{x} = \bar{0}$;

$\mathbf{R} \times V \rightarrow V$ 表示“数乘运算”, 使得对 $\forall x, y \in V, \alpha, \beta \in \mathbf{R}$ 有

(分配律 a) $\alpha(x + y) = \alpha x + \alpha y$,

(分配律 b) $(\alpha + \beta)x = \alpha x + \beta x$,

(结合律) $(\alpha\beta)x = \alpha(\beta x)$,

(单位元数乘) $1x = x$.

其中 \mathbf{R} 上的加法和乘法满足代数学定义. V 中的元素称为矢量, \mathbf{R} 中的元素相对而言就称为标量.

²由于一一对应, 它有时也被称为 1-1 映射 (one-to-one map), 但这个术语在部分文献里指单射, 因此这是一个应该避免使用的术语.

所谓的复矢量空间只要把 \mathbf{R} 拓展成 \mathbf{C} 即可. 部分书籍选择将“加法交换律”放入定义中，但下面证明可从此 7 条推出.

定理 2.1.1 (加法交换律). $\forall x, y \in V, x + y = y + x$.

证. 设 z 是 x 的一个负元, 先证 $z + x = \bar{0}$: 首先 $(z+x) + z = z + (x+z) = z + \bar{0} = z$, 其次设 w 为 z 的一个负元, 则 $z + x = (z+x) + \bar{0} = (z+x) + (z+w) = ((z+x) + z) + w = z + w = \bar{0}$; 再证 $\bar{0} + x = x$: $\bar{0} + x = (x+z) + x = x + (z+x) = x + \bar{0} = x$. 由此两条, 构造 $(1+1)(x+y)$, 由分配律可展开为两种: $x + x + y + y = x + y + x + y$, 两边右加 y 的负元 \bar{y} , 即 $((x+x+y)+y)+\bar{y} = (x+x+y)+(y+\bar{y}) = (x+x+y)+\bar{0} = x+x+y$ 和 $x+y+x+y+\bar{y} = x+y+x$, 二者相等; 同理再左加 \bar{x} 得 $x+y=y+x$. \square

负元、零元的唯一性可以独立于交换律而得到证明.

定理 2.1.2 (唯一性). 设 $\bar{0}_1$ 是一个零元, 则 $\bar{0}_1 = \bar{0}$; 设 z 是 x 的一个负元, 则 $z = \bar{x}$.

证. $\bar{0}_1 = \bar{0}_1 + \bar{0} = \bar{0}$; $z = z + \bar{0} = z + (x + \bar{x}) = (z + x) + \bar{x} = \bar{0} + \bar{x} = \bar{x}$. \square

因此以后可只用 \bar{x} 或者干脆就用 $-x$ 表示 x 的负元. 由此可证

定理 2.1.3. $0x = \bar{0}$; $(-1)x = -x$; $\forall \alpha \in \mathbf{R}, \alpha\bar{0} = \bar{0}$.

证. 首先 $0x + 0x = (0+0)x = 0x$, 最后一步用到 \mathbf{R} 的加法定义, 进而 $0x = 0x + \bar{0} = 0x + (0x + (-0x)) = (0x + 0x) + (-0x) = 0x + (-0x) = \bar{0}$; $x + (-1)x = (1-1)x = 0x = \bar{0}$; 首先 $\alpha\bar{0} + \alpha\bar{0} = \alpha(\bar{0} + \bar{0}) = \alpha\bar{0}$, 进而 $\alpha\bar{0} = \alpha\bar{0} + \bar{0} = \alpha\bar{0} + (\alpha\bar{0} + (-\alpha\bar{0})) = (\alpha\bar{0} + \alpha\bar{0}) + (-\alpha\bar{0}) = \alpha\bar{0} + (-\alpha\bar{0}) = \bar{0}$. \square

由这几条亦可证明加法交换律, 只需先证 $(x+y) + ((-x-y)) = \bar{0}$ 即可. 此外, 今后将零元 $\bar{0}$ 简记为 0, 不必担心混淆问题, 结合语境是很容易分清的.

读者应该发现其实 \mathbf{R}^n 自己就是个矢量空间. 这是因为 \mathbf{R}^n 的元素之间自带了 \mathbf{R} 的加法, 配上 \mathbf{R} 乘法又可构造 \mathbf{R}^n 的任意元素, 以满足那 7 条定义. 所以此前让读者先默认时室内一点 q 处的矢量空间 V_q 就是该处的所有矢量所构成的集合, 并且 4 维矢量构成 \mathbf{R}_q^4 .

由于作为矢量空间的 \mathbf{R}^n 的元素是“箭头”, 而一方面它又应当是集合的“点”元素, 尤其是作为普通集合的 \mathbf{R}^n 的点. 这种观点, 无非就是在一矢量空间内“默认”矢量起于原点 $(0, 0)$, 这样一个矢量只对应一个终点, 那“箭头”和“点”可不作区分. 总之, 在讨论抽象集合或矢量空间时, 读者可选择放弃对箭头的几何直观, 直接将矢量想象为一个点³.

下面介绍基的概念, 以严格地叙述之.

³当讨论不同起点或者箭头模型时, 实际上是在讨论不同“处”的矢量空间. 本节末对此详解.

定义 2.11 (线性无关). 设 $\{x_\mu\}$ 是 V 的一组矢量, 任给 s 个 $k^\mu \in \mathbf{R}$, 矢量 $k^\mu x_\mu$ 称为 $\{x_\mu\}$ 的一个线性组合, $\{k^\mu\}$ 称为该组合的系数组; 对 $y \in V$, 若 $\exists k^\mu \in \mathbf{R}$ 使 $y = k^\mu x_\mu$, 则称 y 可由 $\{x_\mu\}$ 线性表出或表出; $k^\mu x_\mu = 0 \implies k^\mu = 0$, 则称 $\{x_\mu\}$ 是线性无关的; 否则是线性相关的, 即存在不全为零的 $\{k^\mu\}$ 使得 $k^\mu x_\mu = 0$.

包含零元的矢量组必线性相关, 因为 $\forall \alpha \in \mathbf{R}, \alpha 0 = 0$, 因此线性无关组一定由非零矢量构成 (但非零矢量组不一定线性无关); 进而若单个矢量 x 构成了线性相关的矢量组, 则 $x = 0$.

定义 2.12 (有限维矢量空间的基). 设 $n \in \mathbb{N}$, 若 V 中存在一个 n 元线性无关组 $\{e_\mu\}$ 使 V 中任意矢量都可由其表出, 则称 $\{e_\mu\}$ 是 V 的一个基底或一组基矢, 简称基. n 称为 V 的维数 $\dim V$.

注意这里有两个条件: “线性无关”、“所有矢量均可表出”, 这样才能说是基, 且维数等于基的个数. 如果只满足前者, 只能说明 $\dim V \geq n$, 因为一旦 $\dim V < n$, V 中任意 n 元组必线性相关; 只满足后者, 只能说明 $\dim V \leq n$, 且表示不唯一, 因为一旦 $\dim V > n$, V 中任意 n 元组不可能把 V 的所有矢量表出. 综上可以说说明一个矢量空间不可能有两组个数不同的基, 由此定义的维数是良定义的⁴.

定义 2.13 (矢量的分量). 对一个 V 中的任意一组基 $\{e_\mu\}$, V 中的每个矢量 x 可唯一地表示为 $\{e_\mu\}$ 的线性组合 (作差, 利用线性无关性证明). 则称其系数组 $\{k^\mu\}$ 为 x 由 $\{e_\mu\}$ 表出的坐标, 或称在其下的 (坐标) 分量, 写作 $(k^1, k^2, \dots, k^n)^T$.

欲谈及同一矢量在不同基底间的分量变换, 先要给定基变换.

定理 2.1.4. 设 $\{e_\mu\}, \{e'_\mu\}$ 是矢量空间 \mathbf{R}^n 上的两组不同的基, 而矢量 U 在两组基下的坐标分别是 $\{U^\mu\}, \{U'^\mu\}$. 设

$$e'_\mu = \Lambda^\nu{}_\mu e_\nu, \quad (2.1.1)$$

这里 $\Lambda^\nu{}_\mu$ 无非是新基矢 e'_μ 用原基底展开的第 ν 分量. 设以 $\Lambda^\nu{}_\mu$ 为元素排成的 (非退化) 方阵记作 Λ , 求证

$$U'^\mu = (\Lambda^{-1})_\nu{}^\mu U^\nu. \quad (2.1.2)$$

可见矢量的基变换和分量变换是互逆的.

证. 注意 $U'^\mu e'_\mu = U^\nu e_\nu$, 则 $U'^\mu \Lambda^\nu{}_\mu e_\nu = U^\nu e_\nu$, 同一矢量在同一基底下分量相同, 即 $U'^\mu \Lambda^\nu{}_\mu = U^\nu$, 得 $U'^\mu = (\Lambda^{-1})_\nu{}^\mu U^\nu$. \square

为更全面地研究变换 (映射) 并学习以后的张量, 下面介绍对偶空间.

⁴当然, 并非每个空间都有一个有限元组作为基, 若一个矢量空间中任意有限元组都不构成其基, 则称这个矢量空间是无限维的. 有限维里有很多结论不能简单推广至无限维, 否定这些结论带来了更多“反直觉”的性质.

2.1.3 对偶空间

定义 2.14 (线性映射). 设 $f : V \rightarrow W$ 是矢量空间之间的映射. 若 f 满足对 $\forall v, u \in V, \alpha \in \mathbf{R}$ 有 $f(\alpha v + u) = \alpha f(v) + f(u)$, 则称 f 是线性映射⁵. $V = W$ 时称为线性变换; $W = \mathbf{R}$ 时称为线性函数.

定理 2.1.5. 易得 $f(0) = 0 \in W; f(-v) = -f(v); f(k^\mu x_\mu) = k^\mu f(x_\mu)$. 进而设有限维 V 的一组基 $\{e_\mu\}$, 对 V 中任意矢量 $k^\mu e_\mu$ 有

$$f(k^\mu e_\mu) = k^\mu f(e_\mu). \quad (2.1.3)$$

因此只要知道了 V 的一组基 $\{e_\mu\}$ 在 f 下的像组 $\{f(e_\mu)\}$, 则 V 中任意矢量在 f 下的像就确定了. 故我们说线性映射可完全由其在 V 上一组基的作用决定.

定理 2.1.6 (复合映射亦线性). 设 A, B 分别是 $U \rightarrow V, V \rightarrow W$ 的线性映射, 易知 $B \circ A : U \rightarrow W$ 也是线性映射.

据此定义如下:

定义 2.15 (同构). 设存在 A 是 $U \rightarrow V$ 的线性双射, 可证 $A^{-1} : V \rightarrow U$ 也是线性的, 称 A 是 U, V 的一个线性同构或同构映射, 两矢量空间互为线性同构的, 记作 $U \cong V$.

定理 2.1.7. $U \cong V \iff \dim U = \dim V$. 证明略.

下面我们来看看线性函数和矢量的关系.

定义 2.16. V 上全体线性映射 $f : V \rightarrow W$ 之集记作 $\text{hom}(V, W)$.

定理 2.1.8. $\text{hom}(V, W)$ 是矢量空间.

证. 对 $\text{hom}(V, W)$ 可定义如下 (逐点) 加法、数乘、零元. $\forall f, g \in \text{hom}(V, W), v \in V, \alpha \in \mathbf{R}$ 有 $(f + g)(v) := f(v) + g(v), (\alpha f)(v) := \alpha f(v), 0(v) := 0$. \square

这样就可以称 $\text{hom}(V, \mathbf{R})$ 为对偶矢量空间, 简称对偶空间 V^* , 其元素即线性函数就称为对偶矢量.

可以给线性映射定义如下很自然的概念.

定义 2.17 (核、像、秩). 设 $A : V \rightarrow V'$ 是线性映射, 分别称 A 的核、像为

$$\ker A = \{\alpha \in V | A(\alpha) = 0\}, \quad \text{Im } A = \{A(\alpha) | \alpha \in V\}; \quad (2.1.4)$$

⁵这一线性性质契合了 V 上的加法与数乘运算.

易知 A 是单射 $\iff \ker A = \{0\}$; A 是满射 $\iff \text{Im } A = V'$. 并且一般地有

$$\dim \ker A + \dim \text{Im } A = \dim V, \quad (2.1.5)$$

一般称 $\dim \text{Im } A$ 为 A 的秩.

接下来讨论线性映射的矩阵表示. 设 V, W 分别是 n, m 维矢量空间, $A : V \rightarrow W$ 是线性映射, 在 V 中取一组基 $\{\alpha_\mu\}$, 在 W 中取一组基 $\{\beta_\nu\}$, 因 $A(\alpha_\mu) \in W$, 故 $A(\alpha_\mu)$ 可由 W 的基 $\{\beta_\nu\}$ 唯一地表为 $A(\alpha_\mu) = a^\nu{}_\mu \beta_\nu$, 或形式地表示为

$$A \begin{pmatrix} \alpha_1 & \cdots & \alpha_n \end{pmatrix} = \begin{pmatrix} \beta_1 & \cdots & \beta_m \end{pmatrix} \begin{pmatrix} a^1_1 & \cdots & a^1_n \\ \vdots & \ddots & \vdots \\ a^m_1 & \cdots & a^m_n \end{pmatrix},$$

将上式右端的 $m \times n$ 矩阵记作 $A_{\alpha;\beta}$, 称 M_A 是 A 在 V 的基 $\{\alpha_\mu\}$ 和 W 的基 $\{\beta_\nu\}$ 下的矩阵. $A_{\alpha;\beta}$ 的第 j 列是 $A(\alpha_j)$ 在 $\{\beta_\nu\}$ 下的坐标 ($j = [1, n] \cap \mathbb{N}$). 因此 $A_{\alpha;\beta}$ 被 A 唯一决定. 事实上在两组基给定的情况下, 每个 $m \times n$ 矩阵 $(a^i{}_j)_{m \times n}$ 也按上式唯一决定了一个 $V \rightarrow W$ 的线性映射 A , 这样就建立了一个全体 $m \times n$ 实矩阵组成的集合 $M_{m \times n}$ 到 $\text{hom}(V, W)$ 的同构 $M_{m \times n} \cong \text{hom}(V, W)$, 从而

$$\dim \text{hom}(V, W) = mn. \quad (2.1.6)$$

定理 2.1.9. $V \cong V^*$.

证. $\dim V^* = \dim \text{hom}(V, \mathbf{R}) = \dim V \cdot \dim \mathbf{R} = \dim V \iff V \cong V^*$. □

定义 2.18 (对偶基). 设 $\{e_\nu\}$ 是 V 的一组基, 则在 V^* 中可取这样一组基 $\{e^\mu\}$ 使 $e^\mu(e_\nu) = \delta_\nu^\mu$, 称为 $\{e_\nu\}$ 的对偶基.

定理 2.1.10. 对偶基确为 V^* 的一组基.

证. 容易验证 $\{e^\mu\}$ 是 V^* 中一组线性无关的矢量. 由于 e^μ 的作用是线性的, 这实际上也定义了 e^μ 对 $\forall v \in V$ 的作用. 进而 V^* 中任意元素可由 $\{e^\mu\}$ 表出, 只需 $\forall \omega \in V^*$, 令 $\omega_\mu = \omega(e_\mu)$, 必有 $\omega = \omega_\mu e^\mu$, 因为 $\omega(e_\nu) = \omega_\mu e^\mu(e_\nu) = \omega_\nu$. □

下面尝试在对偶空间上“套娃”, 无非是将 V^{**} 定义为 V^* 上全体线性函数之集 $\text{hom}(V^*, \mathbf{R})$. 但有这样一个问题, V, V^* 间的同构是不唯一的. 注意到 $\{e_\mu\}$ 到其对偶基 $\{e^\mu\}$ 的线性映射就是同构, 那这会随着 $\{e_\mu\}$ 的不同选取而产生不同的同构. 如果要有充分理由挑选出“唯一”的特殊同构, 则需要另加结构以限制之. 比如度规结构就建立了 V, V^* 的一个较为特殊的同构 (按 $v_\mu = g_{\mu\nu} v^\nu$ 一一对应).

有别于 V, V^* 间的同构, V, V^{**} 间存在一个不依赖于其它结构的很自然的同构. 这甚至比较明显, 读者不妨这样设想: $\omega \in V^*$ 是 $v \in V$ 的线性函数, 得到 $\omega(v)$, 总

可选取基及其对偶基展开，进而 $\omega(v) = \omega_\mu v^\mu$ 是与基无关的实数，这说明对偶矢量和矢量有天生的乘法，那就是对偶矢量作用在矢量上所得实数。这个“乘法”可否交换呢？我的意思是，矢量可否看作“作用在对偶矢量上的线性函数”，因此该实数 $\omega_\mu v^\mu = v^\mu \omega_\mu$ 是矢量作用在对偶矢量上所得的“ $v(\omega)$ ”？当然是可以的，因为实数结果只是“绑定”了二者，使之有一个对应关系，如果给矢量指配一个对偶矢量产生一个实数，当然也就可以看作是给对偶矢量指配一个矢量以产生了相同实数。此即把矢量看作“对偶对偶”矢量，毕竟对偶矢量的线性函数那么多，矢量本身就已满足定义，故而是最简单的选择。

定义 2.19 (自然同构). 设 $v^{**} \in V^{**}$ 是矢量 $v \in V$ 在某种同构下的像，对 $\omega \in V^*$ 作用产生实数 $v^{**}(\omega)$ 。定义 v^{**} 满足

$$v^{**}(\omega) := \omega(v), \quad \forall \omega \in V^*, \quad (2.1.7)$$

易证该映射 $V \rightarrow V^{**}$ 是同构映射，称为自然同构。

在代数上这相当于把 V 和 V^{**} 视为同一空间（因为把每一 $v \in V$ 与 $v^{**} \in V^{**}$ 认同），进而在自然同构的意义下可不再继续“套娃”，而只是“两面镜子的相互反射”。故真正有用的是 V 和 V^* ，不论再取对偶多少次，也得不到更多“有用”的空间⁶。

这样的话，从基按照 $e^\mu(e_\nu) = \delta_\nu^\mu$ 诱导对偶基，也可以看作对偶基诱导原来的基 $e_\nu(e^\mu) = \delta_\nu^\mu$ 。当然也可以把 e^μ 就视作基而 e_μ 是其对偶基。这些概念本身就是互相对偶的，只是遵从惯例，矢量的基使用下标，其对偶基是对偶矢量的基，用上标。设 $v = v^\mu e_\mu, \omega = \omega_\mu e^\mu$ ，当然 $v(\omega) = \omega(v) = v^\mu \omega_\mu$ ，自然有

$$\omega_\mu = \omega(e_\mu), v^\mu = v(e^\mu). \quad (2.1.8)$$

最后来研究各种分量、基变换的关系。

定理 2.1.11. 若 V 中有一基变换 $e'_\mu = \Lambda^\nu{}_\mu e_\nu$ ，则相应的对偶基变换为

$$e'^\mu = (\Lambda^{-\top})^\mu{}_\nu e^\nu, \quad (2.1.9)$$

其中 $\Lambda^{-\top}$ 是 Λ^\top 之逆。

证. 只须证明等式两边作用于 e'_α 所得结果相同，证明如下：

$$\begin{aligned} (\Lambda^{-\top})^\mu{}_\nu e^\nu (e'_\alpha) &= (\Lambda^{-\top})^\mu{}_\nu e^\nu (A^\beta{}_\alpha e_\beta) = A^\beta{}_\alpha (\Lambda^{-\top})^\mu{}_\nu e^\nu (e_\beta) \\ &= (\Lambda^\top)_\alpha{}^\beta (\Lambda^{-\top})^\mu{}_\nu \delta^\nu{}_\beta = (\Lambda^\top)_\alpha{}^\nu (\Lambda^{-\top})^\mu{}_\nu = \delta_\alpha{}^\mu = e'^\mu (e'_\alpha), \end{aligned}$$

其中第二个等号用到对偶矢量对矢量的作用的线性性，第三、五个等号分别用到转置、逆矩阵的定义，第六个等号用到对偶基的定义式。□

⁶ 自然同构的准确含义见 Spivak(1979) 第 7 章习题 6。

但另一方面，将新对偶基 e'^μ 在旧对偶基 $\{e^\nu\}$ 下展开是

$$e'^\mu = e'^\mu(e_\nu)e^\nu, \quad (2.1.10)$$

可见这系数就是那 $(\Lambda^{-\top})^\mu_\nu$. 这种变换形式称为逆变 (*contravariance*), 新量 = 矩阵 \times 旧量. 而另一方面

$$e'_\mu(e^\nu) = e'_\mu((\Lambda^\top)_\sigma^\nu e'^\sigma) = (\Lambda^\top)_\sigma^\nu e'_\mu(e'^\sigma) = (\Lambda^\top)_\mu^\nu,$$

于是原来的基变换是 $e'_\mu = (\Lambda^\top)_\mu^\nu e_\nu$, 即

$$e_\nu = (\Lambda^{-\top})^\mu_\nu e'_\mu, \quad (2.1.11)$$

这种变换形式称为协变 (*covariance*), 旧量 = 矩阵 \times 新量. 这说明 V 的基是协变的, 其对偶基是逆变的. 进而可见矢量分量变换是逆变形式, 因此也称 V 中的矢量为逆变矢量 (*contravector*); 同理对偶矢量称为协变矢量 (*covector*). 这两种形式的变换之所以称为协变和逆变, 其实是与矢量空间基底的变换相对比而来的. 与矢量基底相同形式变换的称为协变, 与矢量基底相反形式变换的称为逆变⁷.

2.1.4 张量

按此前理解, 张量是个对应矩阵表示的某种无坐标依赖的“深刻事物”. 在几何直观上, 比如应力张量, 欲确定其“观测值”, 比如想要确定某方向或某个面上的压力, 必定要先给定一个面法矢, 这样就产生对应的一个作用力 (或反之用力确定面法矢). 这样应力张量就可以看作从某个矢量空间到另一个矢量空间的映射. 不仅如此, 还可以再进一步地输入一个矢量去作用. 鉴于第一步已输出一个矢量, 第二步就选其对偶的矢量与之作用, 最终得到了实数. 第一步输出矢量和最后的输入矢量作用得到实数, 因此前者是后者的一个对偶矢量, 那么这个映射当然是线性的; 且第一步反之所得到亦为第二步输入矢量的对偶矢量, 则亦为线性映射, 故先后输入无所谓. 因此总地来看, 应力张量可以看作输入两个矢量而输出实数的“一台售货机”, 且对每个矢量都是线性映射, 故我们常说张量就是线性映射的推广.

这种做法无非是将定义域扩展到多个矢量空间、对偶空间甚至混合起来的卡氏积上, 可以类比从单元函数拓展到多元函数. 我们循序渐进, 对此可先介绍简单的“二元函数”:

定义 2.20 (双矢量线性映射). 设 $f : V \times V \rightarrow W$ 对每个 V 而言都是线性映射, 即

⁷请注意区分此前的“协变性”这个词, 虽然都叫“covariance”, 但我们之前所讨论的更多是“无坐标依赖”这个概念, 即强调“Lorentz 协变性”.

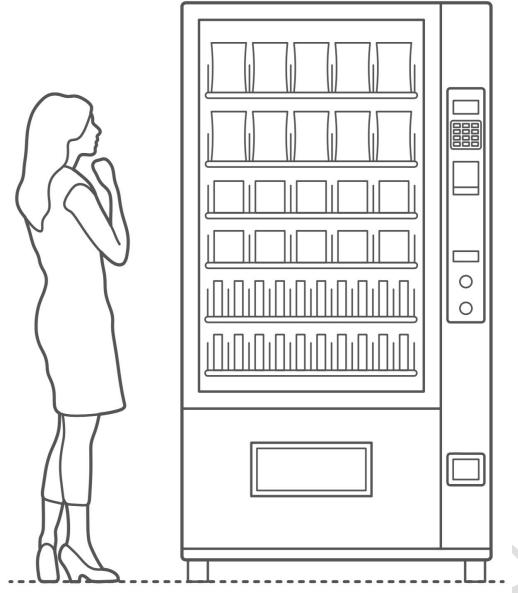


图 2.1: 往售货机的槽 (slat) 里投入足够数量的纸币或硬币可获得一件商品。选择一听可乐，投入过程中，显示屏会不断改变内容，以提示用户待支付的剩余金额，直至最终获得商品。这个过程类似于: $f(\cdot, \cdot) \Rightarrow f(v, \cdot) \in V^* \Rightarrow f(v, w) \in \mathbf{R}$ 。

$\forall \alpha_1, \alpha_2, \beta_1, \beta_2 \in V, k_1, k_2 \in \mathbf{R}$ 有

$$f(k_1\alpha_1 + k_2\alpha_2, \beta_1) = k_1f(\alpha_1, \beta_1) + k_2f(\alpha_2, \beta_1),$$

$$f(\alpha_1, k_1\beta_1 + k_2\beta_2) = k_1f(\alpha_1, \beta_1) + k_2f(\alpha_1, \beta_2),$$

称 f 是 V 上的一个双矢量线性映射，目前暂不混淆，可简称双线性映射。 $W = V$ 时为双线性变换； $W = \mathbf{R}$ 时为双线性函数。

上述表明若固定某一输入的矢量，则 f 就成为另一矢量的“单元”线性映射，当然，由于二者都隶属于同一矢量空间 V ，则哪怕 f 是“双元”的，它当然也是 V 上的“单元”线性映射。

定义 2.21 (降映射). 固定 α ， f 相当于把 $\beta \in V$ 映到 $f(\alpha, \beta) \in W$ 的线性映射，记作 $f_b(\alpha)$ ，即⁸

$$f_b(\alpha)(\beta) = f(\alpha, \beta). \quad (2.1.12)$$

以下专注于 $W = \mathbf{R}$ 的情况。由于 $f_b(\alpha)$ 是 V 的线性函数，因而是一个对偶矢量，则 f_b 就是一个 V 到 V^* 的线性映射。

⁸ “ b ”是五线谱里的降记号，读作“降”(flat)。

在 V 中取一组基 $\{e_\mu\}$, 任取两个 V 中的矢量 x, y , 设 $x = x^i e_i, y = y^j e_j$, 则

$$\begin{aligned} f(x, y) &= f(x^i e_i, y^j e_j) \\ &= x^i y^j f(e_i, e_j), \end{aligned}$$

把上式的系数 $f(e_i, e_j)$ 作为矩阵 A 的第 i 行第 j 列元素, 即令

$$A = \begin{pmatrix} f(e_1, e_1) & \cdots & f(e_1, e_n) \\ \vdots & \ddots & \vdots \\ f(e_n, e_1) & \cdots & f(e_n, e_n) \end{pmatrix},$$

称上述 A 为双线性函数 f 在基 $\{e_\mu\}$ 下的 (分量) 矩阵, 它由 f 和基 $\{e_\mu\}$ 唯一决定. 这样有

$$f(x, y) = X^\top A Y. \quad (2.1.13)$$

其中 $X = (x^1, \dots, x^n)^\top, Y = (y^1, \dots, y^n)^\top$ 分别是 x, y 在 $\{e_\mu\}$ 下的坐标. 可以证明, 设两个 $n \times n$ 方阵 A, B , 若对任意两个 n 维列矢量 X, Y 都有 $X^\top AY = X^\top BY$, 则 $A = B$, 因此, 在基 $\{e_\mu\}$ 下, 任何一个 $n \times n$ 矩阵也按照 $x^i y^j f(e_i, e_j)$ 的方式唯一决定了一个线性映射 f .

在 V 上任取两组基 $\{\alpha_\mu\}$ 和 $\{\beta_\mu\}$, 若 $\{\alpha_\mu\}$ 到 $\{\beta_\mu\}$ 的变换矩阵是 P , f 在 $\{\alpha_\mu\}$ 下和 $\{\beta_\mu\}$ 下的矩阵分别为 A, B , 则由 (2.1.13) 不难导出

$$B = P^\top A P. \quad (2.1.14)$$

从而同一双线性函数在不同的基底下的矩阵是合同的⁹. 由于任意两个合同矩阵有相同的秩, 一个双线性函数 f 的所有可能矩阵都只有一个固定的秩, 将其定义为 f 本身的秩, 记作 $\text{rank } f$. 事实上, 双线性函数 f 的秩与 f_b 作为线性映射的秩相等.

定义 2.22 (非退化). 如果 f 在 V 的某一组基下的矩阵是满秩的, 那么它在 V 的其它任何基底下也是满秩的, 此时我们说 f 是非退化的.

大家学过其充要条件是行列式非零 (矩阵非奇异), 或者说

$$\forall w \in V, f(v, w) = 0 \iff v = 0. \quad (2.1.15)$$

因而 f 是非退化的 $\iff f_b$ 是 $V \rightarrow V^*$ 的同构. 也就是说, 我们可以借助 f 这个 V 上的新结构建立 V, V^* 间 “最特殊” 的同构映射.

⁹ 合同变换 (*matrix congruence*) 即 $A \mapsto X^\top A X$ 这类变换, 其中 X 是可逆的; 若两个对称矩阵合同, 则他们有相同的正负惯性指数, 即有相同的正、负特征值个数.

定义 2.23 (对称与反称). 若 f 对 $\forall \alpha, \beta \in V$ 满足

$$f(\alpha, \beta) = f(\beta, \alpha), \quad (2.1.16)$$

则称 f 是对称的; 而若

$$f(\alpha, \beta) = -f(\beta, \alpha), \quad (2.1.17)$$

则称 f 是反称的.

容易验证, 对称双线性函数的矩阵全都是对称的, 反称双线性函数的矩阵也全都是反称的.

定义 2.24 (度规). 矢量空间 V 上的一个对称、非退化的双线性函数 $g : V \times V \rightarrow \mathbf{R}$ 称为 V 上的一个度规. 另一方面 g_b 是 V, V^* 的同构映射, 称作音乐同构 (*canonical isomorphism*, 或译为典范同构).

在 V 上规定一个度规后, $g_b(\alpha) \in V^*$ 就可看作专门与 α 对应的对偶矢量, 记为 α_b , 这样就与另一矢量 β 产生天生的乘法 $\alpha_b(\beta)$, 因此度规相当于把矢量和对偶矢量的天生乘法引入到两个矢量之间, 即 $g(\alpha, \beta)$, 可记为 $\alpha \cdot \beta$ 或 $\langle \alpha, \beta \rangle$, 对称双线性运算 $\langle \cdot, \cdot \rangle$ 称为 V 上的一个内积¹⁰. 定义了内积的矢量空间称为内积空间. 内积空间上可以定义长度、夹角等复合直观的几何概念, 这各位高中就学过, 比如矢量 $\alpha \in V$ 的模或范数就是

$$|\alpha| := \sqrt{|\langle \alpha, \alpha \rangle|}, \quad (2.1.18)$$

有时为与行列式区分, 写作 $\|\alpha\|$. 但若不追求虚实性, 最常用的写法是 $\alpha^2 = \langle \alpha, \alpha \rangle$, 这样允许“模方”是负的. 注意, 不要求内积正定性时, 非零矢量的内积或者模长也是可以为零的, 下一节立即就会学到.

定理 2.1.12 (Cauchy-Schwartz 不等式). 可以证明

$$|\langle \alpha, \beta \rangle| \leq |\alpha| |\beta|. \quad (2.1.19)$$

有限维矢量空间定义了正定内积后, 即有限维的正定内积空间, 就称为欧氏矢量空间, 不混淆时可称欧氏空间.

定义 2.25 (正交基). 基 $\{e_\mu\}$ 在 g 的衡量下是正交的, 若 $g(e_i, e_j) = 0, i \neq j$; 在其衡量下为归一的, 若 $g(e_i, e_i) = \pm 1$. 正交且归一的基底称为正交基, 不会混淆时可简称正交基.

比如在欧氏度规的正交基下, 对 $\forall x = x^i e_i, y = y^i e_i \in V$ 有 $\langle x, y \rangle = \delta_{ij} x^i y^j$.

¹⁰注意, 不涉及物理学的数学教材会强调内积应当满足正定性: $\forall \alpha \in V, \langle \alpha, \alpha \rangle > 0$, 但我们会讨论不定度规, 因此不作这个要求. 此外, 正定内积必然是非退化的, 因此若只定义了正定性, 就可略去非退化这个限制.

定理 2.1.13. 任意内积空间总存在正交基，且度规在此基的对角矩阵中对角元正负个数与所选正交基无关。证明见 Schutz(1980) P.65-66.

由此，欧氏空间总存在正交基，故也可把正定内积空间 V 就看成配有“勾股定理”的 \mathbf{R}^n .

定义 2.26 (多重矢量线性映射). 设 $f : \underbrace{V \times \cdots \times V}_k \rightarrow W$ 对每个 V 都是线性映射，即 $\forall v_1, \dots, v_k, v \in V, \lambda \in \mathbf{R}$ 有

$$f(v_1, \dots, v_i + \lambda v, \dots, v_k) = f(v_1, \dots, v_k) + \lambda f(v_1, \dots, v, \dots, v_k), \quad (2.1.20)$$

其中 v 处于第 i 个分量的位置， $i \in [1, k] \cap \mathbb{N}$ ，称 f 是 V 上的一个 k 重矢量线性映射。 $W = V$ 时为 k 重矢量线性变换； $W = \mathbf{R}$ 时为 k 重矢量线性函数。

定义 2.27 (张量). 设

$$T : \underbrace{V^* \times \cdots \times V^*}_k \times \underbrace{V \times \cdots \times V}_l \rightarrow W \quad (2.1.21)$$

对每个 V^*, V 都是线性映射，称 T 是 V 上的 (k, l) 型的多重线性映射。 V 上一个 (k, l) 型张量就是 (k, l) 型的多重线性函数。 $k + l$ 叫作张量的阶数。

以后考虑 $W = \mathbf{R}$ 时，就不再使用多重线性函数这个词，而统一叫作张量。这样，矢量、对偶矢量就视作一种 1 阶张量¹¹。此前所谓的“多重矢量线性函数”其实是强调只与矢量作用，其身份就相当于协变矢量（对偶矢量），因此只与矢量作用的张量又称作协变张量。比如说度规就应是一个 $(0, 2)$ 型协变张量，或者 2 阶协变张量。

定义 2.28 (张量的对称). 张量 T 关于某两个下槽对称，若

$$T(\dots, v, \dots, w, \dots) = T(\dots, w, \dots, v, \dots), \quad \forall v, w \in V.$$

在分量上，这意味着交换此二指标结果不变。若

$$T(\dots, v, \dots, w, \dots) = -T(\dots, w, \dots, v, \dots), \quad \forall v, w \in V.$$

则称其关于此二下槽反称，这样交换指标会使得分量反号。为上槽时同理。称多个槽对称或反称，若其中任意两槽是对称或反称的。

这样，度规和 Ric, T 都是 2 阶协变对称张量。类似地，度规的逆就应当看作 2 阶逆变对称张量，而 Riemann 曲率 $R^\lambda_{\mu\nu\sigma}$ 就是个 $(1, 3)$ 型张量（的分量）。

¹¹当然，张量构成矢量空间，因此张量也是一种矢量。看起来有点怪，其实是自然语言混淆了这个地方的矢量和先前所说的普通矢量空间的逆变矢量。

读者应该看出，其实可用矢量、对偶矢量（一元线性函数）构造出张量来。比如，设 f_1, f_2 是 2 个 V 上的线性函数，那么定义其这样一种用以复合的“乘法”： $f_1 \otimes f_2 : V \times V \rightarrow \mathbf{R}$ ，使得 $f_1 \otimes f_2(v_1, v_2) = f_1(v_1)f_2(v_2)$ ，而此当然对每个 V 都是线性的，这便从两个对偶矢量构造出了一个 2 阶协变张量。这就是电动力学、量子力学所说的“并矢”的实际含义，也就是将两个矢量排成一个矩阵出来，但是一般不满足交换律。

定义 2.29 (张量积). $(k, l), (k', l')$ 型张量 T, T' 的张量积 $T \otimes T'$ 是一个 $(k+k', l+l')$ 型张量，使得

$$\begin{aligned} & T \otimes T'(\omega^1, \dots, \omega^k, \omega^{k+1}, \dots, \omega^{k+k'}; v_1, \dots, v_l, v_{l+1}, \dots, v_{l+l'}) \\ &= T(\omega^1, \dots, \omega^k; v_1, \dots, v_l) T'(\omega^{k+1}, \dots, \omega^{k+k'}; v_{l+1}, \dots, v_{l+l'}). \end{aligned}$$

定理 2.1.14. 设 $\mathcal{T}_V(k, l)$ 表示 V 上全体 (k, l) 型张量之集。 $\mathcal{T}_V(k, l)$ 是矢量空间， $\dim \mathcal{T}_V(k, l) = (\dim V)^{k+l}$ 。

证。仿照 V^* 的逐点加法、数乘、零元的定义来构造矢量空间 $\mathcal{T}_V(k, l)$ 。下只证其一组基矢的个数为 $(\dim V)^{k+l}$ 。设 $\{e_\mu\}$ 及对偶基 $\{e^\mu\}$ 。可以预料，张量的基矢应是这些基矢的张量积组合，即

$$e_{\mu_1} \otimes \cdots \otimes e_{\mu_k} \otimes e^{\sigma_1} \otimes \cdots \otimes e^{\sigma_l},$$

而每个指标取便 $\dim V$ ，共有 $(\dim V)^{k+l}$ 种。下证它们构成 $\mathcal{T}_V(k, l)$ 的一组基。容易验证其线性无关，再证 $\forall T \in \mathcal{T}_V(k, l)$ 可表为

$$T = T^{\mu_1 \cdots \mu_k}_{\sigma_1 \cdots \sigma_l} e_{\mu_1} \otimes \cdots \otimes e_{\mu_k} \otimes e^{\sigma_1} \otimes \cdots \otimes e^{\sigma_l}. \quad (2.1.22)$$

只需作用于对应基失为

$$e_{\mu_1} \otimes \cdots \otimes e_{\mu_k} \otimes e^{\sigma_1} \otimes \cdots \otimes e^{\sigma_l}(e^{\mu'_1}, \dots, e^{\mu'_k}; e_{\sigma'_1}, \dots, e_{\sigma'_l}) = \delta_{\mu_1}^{\mu'_1} \cdots \delta_{\mu_k}^{\mu'_k} \delta_{\sigma'_1}^{\sigma_1} \cdots \delta_{\sigma'_l}^{\sigma_l},$$

说明其在这组基下的分量为

$$T^{\mu_1 \cdots \mu_k}_{\sigma_1 \cdots \sigma_l} = T(e^{\mu_1}, \dots, e^{\mu_k}; e_{\sigma_1}, \dots, e_{\sigma_l}), \quad (2.1.23)$$

故总是可表。 \square

下面用映射语言准确定义张量的缩并。回顾一下， $(1, 1)$ 型张量 T 可看作从 V 到 V 的线性映射，这正是线性代数所讲的线性变换。 T 在任一基底 $\{e_\mu \otimes e^\nu\}$ 的分量排成的矩阵 (T^μ_ν) 显然与基底有关，不难证明同一 T 在任意两个基底的分量对应的两个矩阵 (T^μ_ν) 和 (T'^μ_ν) 互为相似矩阵，证明拆到基底即可：

$$\begin{aligned} T'^\mu_\nu &= T(e'^\mu; e'_\nu) = T((A^{-\top})_\rho^\mu e^\rho; A^\sigma_\nu e_\sigma) = (A^{-\top})_\rho^\mu A^\sigma_\nu T(e^\rho; e_\sigma) \\ &= (A^{-\top})_\rho^\mu A^\sigma_\nu T^\rho_\sigma = (A^{-1})_\rho^\mu T^\rho_\sigma A^\sigma_\nu = (A^{-1}TA)^\mu_\nu, \end{aligned}$$

于是有矩阵等式 $T' = A^{-1}TA$ (其中 T', A, T 都代表矩阵¹²), 可见 T' 与 T 互为相似矩阵. 以 T'^{μ}_{μ} ($\sum_{\mu=1}^n T'^{\mu}_{\mu}$ 的简写) 及 T^{ρ}_{ρ} 分别代表矩阵 T' 和 T 的迹, 则易得

$$T'^{\mu}_{\mu} = (A^{-1})^{\mu}_{\rho} T^{\rho}_{\sigma} A^{\sigma}_{\mu} = A^{\sigma}_{\mu} (A^{-1})^{\mu}_{\rho} T^{\rho}_{\sigma} = \delta^{\sigma}_{\rho} T^{\rho}_{\sigma} = T^{\rho}_{\rho}.$$

这就证明了同一 $(1, 1)$ 型张量在不同基底的矩阵有相同的迹. 在关心张量时, 应该抓住其与基底无关的性质, $(1, 1)$ 型张量 T 的迹 T^{μ}_{μ} 就是这样一种性质, 通常称为 T 的缩并 (contraction), 记作 CT , 即

$$CT := T^{\mu}_{\mu} = T(e^{\mu}; e_{\mu}).$$

再讨论 $(2, 1)$ 型张量 T 的缩并. T 可记作 $T(\cdot, \cdot; \cdot)$, 它有两个上槽和一个下槽, 故有两种可能缩并: (1) 第一上槽与下槽的缩并 $C_1^1 T := T(e^{\mu}, \cdot; e_{\mu})$; (2) 第二上槽与下槽的缩并 $C_1^2 T := T(\cdot, e^{\mu}; e_{\mu})$. 若改用另一基底 $\{e'_\rho\}$ 定义这两种缩并, 分别记作 $(C_1^1 T)'$ 和 $(C_1^2 T)'$, 则易证 $(C_1^1 T)' = C_1^1 T$, $(C_1^2 T)' = C_1^2 T$. 由映射语言可知 $C_1^1 T$ 和 $C_1^2 T$ 都是 $(1, 0)$ 型张量, 它们在任一基底的分量可用 T 在该基底的分量表为 $(C_1^1 T)^{\nu} = T(e^{\mu}, e^{\nu}; e_{\mu}) = T^{\mu\nu}_{\mu}$, $(C_1^2 T)^{\nu} = T^{\nu\mu}_{\mu}$ (已略去求和号). 不难推广上述讨论而得出 (k, l) 型张量的缩并定义如下:

定义 2.30 (缩并). $T \in \mathcal{T}_V(k, l)$ 的第 i 上标 ($i \leq k$) 与第 j 下标 ($j \leq l$) 的缩并定义为

$$C_j^i T := \sum_{\mu} T \left(\cdot, \dots, \underset{\text{第 } i \text{ 上槽}}{\color{red} e^{\mu}}, \cdot, \dots; \cdot, \dots, \underset{\text{第 } j \text{ 下槽}}{\color{blue} e_{\mu}}, \cdot, \dots \right), \quad (2.1.24)$$

缩并结果属于 $\mathcal{T}_V(k-1, l-1)$.

注意, $C_j^i T$ 与基底选择无关. 联合使用张量积和缩并运算可从原有张量得到各种类型的新张量. 例如, 设 $v \in V, \omega \in V^*$, 则 $v \otimes \omega$ 是 $(1, 1)$ 型张量, 而 $C(v \otimes \omega)$ 则是 $(0, 0)$ 型张量 (标量).

后面经常遇到先求张量积再做缩并的运算, 其结果可看作张量对矢量 (或对偶矢量) 的作用.

例 2.1. 用 v^{μ}, ω_{μ} 表示 v, ω 在同一基底的分量, 有

- (a) $C(v \otimes \omega) = \omega_{\mu} v^{\mu} = \omega(v) = v(\omega), \quad \forall v \in V, \omega \in V^*$;
- (b) $C_2^1(T \otimes v) = T(\cdot, v), \quad \forall v \in V, T \in \mathcal{T}_V(0, 2)$;
- (c) $C_2^2(T \otimes \omega) = T(\cdot, \omega; \cdot), \quad \forall \omega \in V^*, T \in \mathcal{T}_V(2, 1)$.

下面只给出 (b) 的证明, 其他结论的证明由此类推.

¹² T 有时代表张量有时代表矩阵, 读者应能根据上下文识别.

证. 待证的等式 (b) 左边的 $T \otimes v$ 是 $(1, 2)$ 型张量, 是一部有 1 个上槽、2 个下槽的售货机, 可表为 $T \otimes v(\cdot; \cdot, \cdot)$, 故

$$C_2^1(T \otimes v) = T \otimes v(e^\mu; \cdot, e_\mu),$$

所以欲证等式 (b) 只须证明下式

$$T \otimes v(e^\mu; \cdot, e_\mu) = T(\cdot, v),$$

而上式是对偶矢量等式, 欲证上式只须证明两边作用于任一 $u \in V$ 给出相同实数.

$$\begin{aligned} \text{左边作用于 } u &= T \otimes v(e^\mu; u, e_\mu) = T(u, e_\mu)v(e^\mu) \\ &= T(u, e_\mu)e^\mu(v) = T(u, e_\mu)v^\mu = T(u, v) = \text{右边作用于 } u. \end{aligned}$$

可见式 (b) 成立. \square

除以上三式外还有许多类似等式. 这些等式是如下规律的表现: “ T 对 ω (或 v) 作用就是先求 T 与 ω (或 v) 的张量积再缩并”, 或者粗略地说, “作用就是先积后并”. 对两个张量先求张量积再缩并的操作常又简称为对它们做缩并, 因此上述粗略提法还可简化为“作用就是缩并”.

定义 2.31 (升降指标). 张量的升降指标操作就是利用音乐同构, 故是相对于某度规而言的, 比如对 $(1, 1)$ 型张量 T 降指标即 $C_1^1(g \otimes T) = T_b$, 而升指标用度规的逆¹³, 即 $C_1^1(g^{-1} \otimes T) = T^\sharp$. 由度规非退化性知, 其矩阵是可逆的, 因而一定存在对应的度规的逆.

当运算过程中使用的度规较为单一时, 可略去对度规的声明. 使用音乐同构记号默认是升降全部指标.

2.1.5 坐标系

集合 \mathbf{R}^n 的元素就是“数字”, 而我们期望坐标 (*coordinate*) 用实数表示, 那很明显这里元素自己就是坐标, 于是形容 \mathbf{R}^n 天生就具有坐标.

定义 2.32 (自然坐标). 作为集合的 \mathbf{R}^n 的元素称为自然坐标.

当然, 另一方面, 我们也知道 \mathbf{R}^n 应可具有多种不同坐标, 比如 2 维白纸 \mathbf{R}^2 的某点 p 可被赋予正交坐标系的 $(0, 1)$, 但也可被指派极坐标系的 $(1, \pi/2)$. 这是怎么一回事呢? 原来应该要这样厘清关系: 首先应当有的是一个几何背景 \mathbf{R}^2 , 视它“没有坐标”, 而其上可覆盖两种用以存放坐标的坐标空间 \mathbf{R}^2 , 一种对应着正交坐标系, 一种对应着极坐标系.

¹³这里五线谱符号 \sharp 读作“升” (*sharp*).

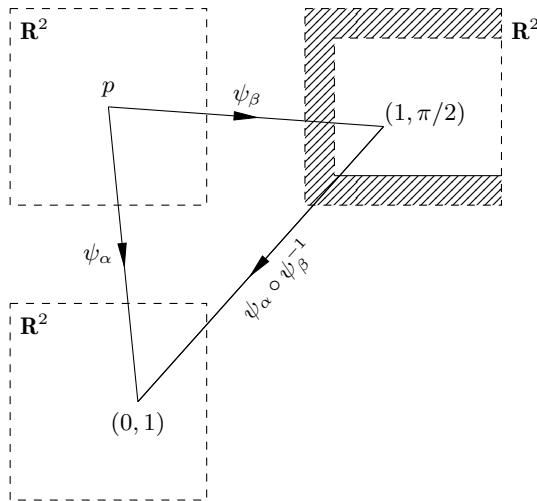


图 2.2: 坐标变换

如图, 假设有一背景对象 \mathbf{R}^2 , 其上的点 p 可以按照双射 $\psi_\alpha : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ 映射到 $\psi_\alpha(p) = (x, y)$, 而亦可按另一 $\psi_\beta : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ 映射到 $\psi_\beta(p) = (r, \theta)$, 但 ψ_β 可以不是满射. 当然, 二者一定都是单射¹⁴. 若给定极坐标 $\psi_\beta(p) = (r, \theta)$, 则其正交坐标就应是 $\psi_\alpha(p) = (r \cos \theta, r \sin \theta)$, 而它也是 $\psi_\alpha(\psi_\beta^{-1}(r, \theta))$, 这里把 ψ_β^{-1} 的定义域限制为 ψ_β 的值域. 这即是说, “从极坐标到正交坐标”的坐标变换定义为双射

$$\psi_\alpha \circ \psi_\beta^{-1} : \{(0, 0)\} \cap (0, \infty) \times [0, 2\pi) \rightarrow \mathbf{R}^2,$$

使得

$$x = r \cos \theta, \quad y = r \sin \theta,$$

其中 x, y 是 \mathbf{R}^2 的坐标, r, θ 是定义域的坐标. 当然, 这个范围只是个人为规定, 但可见定义域并非整个坐标空间 \mathbf{R}^2 , 因此为作区分, 应当单独视作其子集 V_β , 简称坐标子集. 且为了确保坐标映射是个(单值)映射, 还不能将其规定为闭集, 否则就成了多值映射. 图中将 V_β 以外的部分涂了斜线. 坐标子集也可以是整个 \mathbf{R}^2 , 比如正交坐标系的 V_α 就是这样.

只谈 V_β 作为 ψ_β 的上域时, ψ_β 当然就一定是双射了, 因此 V_β 可利用逆映射 ψ_β^{-1} 完全覆盖到背景上. 不过, 这里所举的例子是想给背景 \mathbf{R}^2 的所有点都“一下子”赋予坐标, 而无遗漏, 但一般的情况下, 一个坐标系的坐标子集的逆像可以不是整个背景, 为作区分, 可称坐标域 (*coordinate patch*). 这时坐标映射的定义域就不是整个背景 M 了, 而应该是坐标域 $O \subset M$. 这种情况是有的, 一个坐标系的“能力”可能有限, 而不能“管控”背景所有点的坐标指配工作. 但为了圆满完成所有指

¹⁴不能给两个点分配同一坐标. 否则面对某个坐标, 我怎么知道它到底是指北京还是纽约? 这样的歧义是不希望的.

派任务，就只能再借助其它的坐标系。当然，不要求其它的某一坐标系能够直接揽下全局，而只需要互相分担对方力所不能及之地。为了确保能够覆盖完整，必然要求所有坐标域间有交集，否则就有“漏网之鱼”¹⁵。交集的存在保证了交集内的点可进行坐标变换。

如果不人为规定坐标子集，而将其周期性地拓展到整个 \mathbf{R}^2 ，这样极坐标映射看起来像是个多值映射，但准确而言，应当看作由若干具有不同坐标子集的单值映射组合而来，使之看起来像是单独的一个多值映射。比如可以规定极坐标映射由这样两个映射组成，其一是

$$\psi_{\beta_1} : \mathbf{R}^2 \rightarrow (0, \infty) \times (0, 2\pi),$$

那剩下一条极轴 $[0, \infty) \times \{0\}$ 该怎么办？我们只要再用一个坐标系的坐标子集覆盖即可，比如

$$\psi_{\beta_2} : \mathbf{R}^2 \rightarrow (-\varepsilon, \infty) \times (-\varepsilon, \varepsilon), \quad \forall \varepsilon > 0.$$

这样做的优点是甚至能把坐标子集都设为开集（进而坐标域是开集），只不过要规定 $\theta \in (0, 2\pi)$ 用 ψ_{β_1} ，而 $[0, \infty) \times \{0\}$ 用 ψ_{β_2} 。因此，极坐标映射可以看作一个多值映射，也可以看作多个单值映射之组，特点是只要求坐标域是开集。撇开极坐标这种原点具有奇性的特例，一般来讲我们可以获得性质良好的坐标系，即坐标域覆盖完整且为连通开集¹⁶、坐标映射单值。故不妨定义：

定义 2.33 (坐标系). 集合 M 上的一个坐标映射 ψ_α 就是从开集 $O_\alpha \subset M$ 到开集 $V_\alpha \subset \mathbf{R}^n$ 的双射，这样 (O_α, ψ_α) 称为一个坐标系。要求 M 需要由坐标系完整覆盖，即 $M = \bigcup_\gamma O_\gamma$ ，这样 M 上可具备若干坐标系，且总存在两坐标域满足 $O_\alpha \cap O_\beta \neq \emptyset$ ，则交集内的点可定义映射 $\psi_\alpha \circ \psi_\beta^{-1}$ ，称为从 (O_β, ψ_β) 到 (O_α, ψ_α) 的坐标变换，具体形式是 n 个 n 元函数。如果该函数按分析学定义是 r 阶导连续的，则称集合 M 为 C^r 流形。

C^r 流形 M 想要成为真正有用的“背景”，还需另加限制。准确术语放入第 4 章。至少看来，应该希望这个背景的点“多”“密”“连续统”，目前确切知道， \mathbf{R}^n 是可以做这样一种几何背景的，甚至可以用一个坐标系的坐标域完整覆盖。（ \mathbf{R}^n 是开集！）以下默认上述 M 可充当“背景”。

这样的坐标系存在且有无限种，比如正交坐标系按任意角度旋转、任意长度平移，甚至各种缩放、倾斜、扭曲等等。但很多时候，我们不方便每次都画坐标空间，因而只能在一个背景上体现不同坐标系，应该怎样区分它们呢？

我们在中学的做法是很朴实的，即把正交坐标系画成相互垂直的坐标轴，一条条网格亦相互垂直，而极坐标系更多是一些同心圆。可究竟凭什么把正交坐标网格

¹⁵为确保坐标映射单值，坐标子集规定为非闭集，这样坐标域也是非闭的（具体证明需要我们以后学习“背景”的具体定义），故所有坐标域仅仅“相靠”会漏掉某些边界上的背景点，因此必须有交集。

¹⁶连通形容集合只有一整块而不分开，稍微更准确说：其中任意点间都存在处于集合内的路径。

画成欧氏几何的直线呢？原来，正交坐标系和极坐标系的区分不能肤浅地看作两个先验绝对的映射，而应只体现在相对不同的坐标变换上。实际上我们并不要求正交坐标网格是“直”的！可以设想到一个背景及一种画法，其上极坐标系是“直”的，而正交坐标系是“弯”的！于是画成垂直直线这件事只能理解为是习惯，如同你在想象集合 \mathbf{R} 时，通常不会将其看作曲线。如此，一旦规定某一坐标系，尤其是正交坐标系是平直网格，那么区分其它任意坐标网格就只需观察相对差异即可。

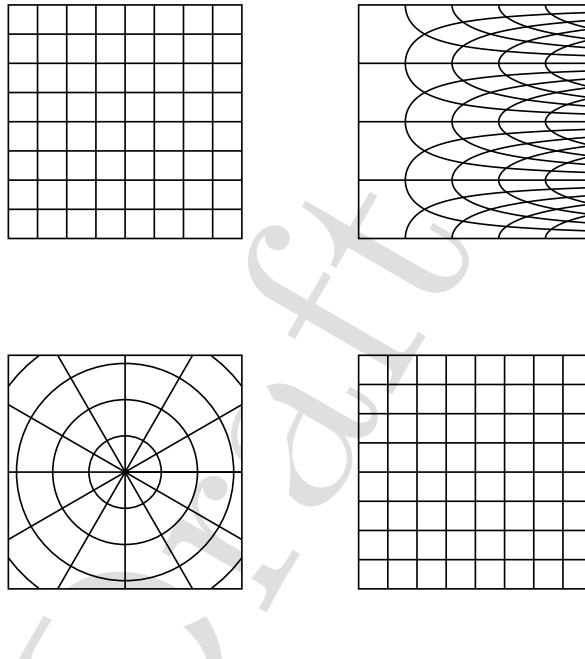


图 2.3: 坐标系间的差异是相对而言的

如上说法说明有必要定义什么叫“坐标网格”。为允许我们开始在背景上“作画”，先给出曲线之定义。

定义 2.34 (曲线、路径). 设 I 为 \mathbf{R} 的一个区间，则连续映射 $C : I \rightarrow X$ 就称为集合 X 上的一条曲线 (*curve*)。自变量 $t \in I$ 称为曲线的参数，映射的像 $C[I] \subset X$ 称为路径 (*path*)，不混淆时亦可称曲线。

更准确地，要求集合 X 是拓扑空间 (*topological space*)，则映射的连续性有确切含义，这在第 4 章会立即解释。目前讨论的是背景 M 上的曲线，则至少可以用区间 I 及坐标系上的分析学来衡量函数 $\psi \circ C$ 的连续性。给定两点 x 和 y ，一条从 x 到 y 的曲线 f 总可选择参数使得 $f(0) = x, f(1) = y$ 。可称路径 $f : [0, 1] \rightarrow X$ 连接 x 和 y 。

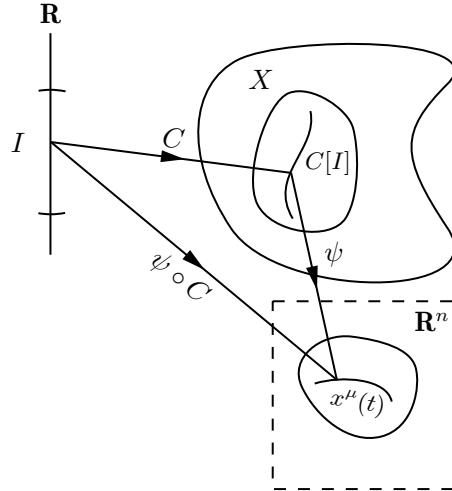


图 2.4: 曲线映射

坐标域上某点在坐标系下的坐标为 $(0, 0)$, 则称该点是该坐标系的原点. 不过我们可以不只是在原点画坐标轴, 也可以到处都画坐标线, 取等距坐标线便可“织”成一张坐标网格.

定义 2.35 (坐标线). 给定背景及一坐标系 (O, ψ) , 坐标为 x^μ . O 的子集

$$\{p \in O \mid x^2(p) = C_2, \dots, x^n(p) = C_n\}$$

可看成以 x^1 为参数的曲线之像, 其中 C_2, \dots, C_n 为可改动的常数. 称其是一条 x^1 坐标线. 其它情况以此类推.

为给以后做铺垫, 尤其是马上对本章的学习, 下一节的主角是所谓的微分形式 (*differential form*), 这样就可引入一套特别漂亮的语言, 来阐明张量和坐标系之联系.

2.1.6 微分形式

“形式” (*form*), 用自然语言讲就是 Riemann 积分“所积的内容”, 微分一词便是强调那些微元. 比如这样一些多元积分式

$$\int P \, dx + Q \, dy + R \, dz, \quad \int P \, dx \, dy + Q \, dy \, dz + R \, dz \, dx, \quad \dots$$

可看成这样的表达:

$$\int \omega_i \, dx^i, \quad \int \sum_{i < j} \omega_{ij} \, dx^i \, dx^j, \quad \int \sum_{i < j < k} \omega_{ijk} \, dx^i \, dx^j \, dx^k, \quad \dots$$



图 2.5: 地球的局部坐标网格, 图示为欧洲

求和下面像这样排序, 其实就相当于只对这些指标的组合求和. 考虑到不存在任意指标相等的项, 干脆就让 $\omega_{i_1 \dots i_l}$ 是个反称张量的分量, 这样任意指标相等时分量就是零, 因为 $\omega_{\dots i_i \dots i_l} = -\omega_{i_i \dots \dots}$.

定义 2.36 (形式). V 上的一个 l 阶反称协变张量 ω 称为一个 l 次微分形式, 简称 l -形式.

坐标基

不过还暂不清楚微分在这套语言下有何含义. 不妨回顾一下 *Newton-Leibniz* 公式 (线积分定理)

$$\int df = \int \frac{\partial f}{\partial x^\mu} dx^\mu, \quad (2.1.25)$$

可见函数 f 的微分 df 应当视作 1-形式, 也就是一个对偶矢量. 这样 $df \in V^*$ 就可以作用于 $\forall v \in V$ 出一个实数 $df(v)$. 该出怎样的实数好呢? 如果取某个 $\{e_\mu\}$ 及其对偶基 $\{e^\mu\}$ 下展开, 应该有 $df(e_\mu) = (df)_\mu$, 那么 $(df)_\mu$ 应该选择什么? 这里便已经给了我们答案: 如果 d 将函数 f 映射成对偶矢量 df , 那么也应将坐标函数 x^μ 映射成对偶矢量 dx^μ , 而这就给出了 V^* 的一组基 $\{dx^\mu\}$! 而上式告诉我们, df 在 dx^μ 上的展开系数 $(df)_\mu = \frac{\partial f}{\partial x^\mu}$, 因此

定义 2.37 (函数的微分). 设背景 M 上一点处全体函数构成的集合 \mathcal{F}_M . 若该点在某坐标系 $\{x^\mu\}$ 的坐标域内, 函数 $f: M \rightarrow \mathbf{R}$ 可用坐标写为 n 元函数 $f: \mathbf{R}^n \rightarrow \mathbf{R}$. 函数的微分是映射 $d: \mathcal{F}_M \rightarrow V^*$. 给定坐标系时, 坐标函数按 d 映射成对偶矢量 dx^μ 后, $\{dx^\mu\}$ 可作为一组基. 定义

$$df = \frac{\partial f}{\partial x^\mu} dx^\mu, \quad (2.1.26)$$

且可见, 按此式进一步给出 dx^ν 在 $\{dx^\mu\}$ 下第 μ 分量是 δ_μ^ν , 称作对偶坐标基 (*dual coordinate basis*).

这样便成功用线性代数的语言自然地引进了分析学的代数结构. 设展开系数 $v(dx^\mu) = v^\mu$, 有

$$df(v) = v^\mu \frac{\partial f}{\partial x^\mu}, \quad \forall v \in V, \quad (2.1.27)$$

式右就是分析学的方向导数. 可见 $dx^\mu(v) = v(dx^\mu)$. 利用 $\{dx^\mu\}$ 就可给出张量在坐标系变换中分量的变换公式. 设另一坐标系所产生的另一组对偶坐标基 $\{dx'^\mu\}$, 为使

$$\frac{\partial f}{\partial x^\nu} dx^\nu = \frac{\partial f}{\partial x'^\mu} dx'^\mu, \quad \forall f \in \mathcal{F}_M,$$

应有对偶坐标基间的变换

$$dx'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} dx^\nu, \quad (2.1.28)$$

这遍契合了分析学中对坐标变换函数求全微分. 根据先前的讨论, $\{dx^\mu\}$ 的对偶也就是原来的基 $\{e_\mu\}$ (相应就称作坐标基) 应该满足

$$e'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} e_\nu. \quad (2.1.29)$$

定理 2.1.15 (张量变换律).

$$T'^{\mu'_1 \cdots \mu'_k}_{\sigma'_1 \cdots \sigma'_l} = T^{\mu_1 \cdots \mu_k}_{\sigma_1 \cdots \sigma_l} \frac{\partial x'^{\mu'_1}}{\partial x^{\mu_1}} \cdots \frac{\partial x'^{\mu'_k}}{\partial x^{\mu_k}} \frac{\partial x^{\sigma_1}}{\partial x'^{\sigma'_1}} \cdots \frac{\partial x^{\sigma_l}}{\partial x'^{\sigma'_l}}. \quad (2.1.30)$$

证. 注意

$$T'^{\mu'_1 \cdots \mu'_k}_{\sigma'_1 \cdots \sigma'_l} = T(e'^{\mu'_1}, \dots, e'^{\mu'_k}; e'_{\sigma'_1}, \dots, e'_{\sigma'_l}),$$

其中 $e'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} e_\nu, e'^\mu = \frac{\partial x^\nu}{\partial x^\mu} e^\nu$ 即可. \square

当然, 坐标基其实有它对应的符号. 注意每个对偶矢量都可以看作一个映射, 那么其实矢量也可以看作一个映射, 不过这次是对函数 f 的作用, 即 $v: \mathcal{F}_M \rightarrow \mathbf{R}$. 对此式

$$df(v) = v^\mu \frac{\partial}{\partial x^\mu}(f),$$

若右边看作 $v(f)$, 这就意味着 $v = v^\mu \frac{\partial}{\partial x^\mu}$, 因此坐标基的符号其实就是 $\frac{\partial}{\partial x^\mu}$. 这是自然的, 因为 $dx^\mu(\frac{\partial}{\partial x^\nu}) = \frac{\partial}{\partial x^\nu}(x^\mu) = \delta_\nu^\mu$. 这样实质上可给出脱离于微分形式、对偶坐标基的定义:

定义 2.38 (坐标基). 坐标基是映射 $e_\mu : \mathcal{F}_M \rightarrow \mathbf{R}$ 使得

$$e_\mu(f) = \frac{\partial f}{\partial x^\mu}. \quad (2.1.31)$$

这相当于想从函数和坐标系构造出其相适配的矢量空间¹⁷, 只是目前还不太需要像这样做. 但是, 读者也不应对此过于陌生, 因为我们当然学过将算符理解为矢量的思想, 比如 ∇f . 依此语言便可书写曲线切矢的符号. 按朴素的理解¹⁸, 切矢分量无非就是曲线上点的坐标求导所得实函数 $\frac{dx^\mu}{dt}$, 那么可以预料到 $\frac{d}{dt}(\psi \circ C(t)) = (\frac{dx^1}{dt}, \dots, \frac{dx^n}{dt})$. 这其实就是切矢在坐标基下的分量, 即

$$\frac{d}{dt} = \frac{dx^\mu}{dt} \frac{\partial}{\partial x^\mu}, \quad (2.1.32)$$

将其作用于 f 就得到分析学中 f 沿曲线的导数

$$\frac{df}{dt} = \frac{dx^\mu}{dt} \frac{\partial f}{\partial x^\mu},$$

所以切矢的符号就是 $\frac{d}{dt}$, 有时记作 $\frac{\partial}{\partial t}$ 也无伤大雅. 用这套语言就很容易看出, 其实坐标基就是坐标线的切矢!

例 2.2. 在 \mathbf{R}^n 上确定某坐标系后, 对某点 (a^1, a^2, \dots, a^n) 取 x^1 坐标线, 则其上任意点坐标为 (x^1, a^2, \dots, a^n) , 因此该点矢量空间 V 中对应 x^1 的坐标基是 $(1, 0, \dots, 0)$. 同理对应 x^2 的为 $(0, 1, \dots, 0)$ 、 \dots 对应 x^n 的为 $(0, 0, \dots, 1)$. 而常数 a^μ 是任意的, 因此任意点的坐标基在该系都有相同坐标, 进而是一致的. 因而我们说 \mathbf{R}^n 中可以“随意平移矢量”. 容易发现这使得矢量的分量将等于坐标系的坐标.

背景上可讨论场的含义。

定义 2.39 (张量场). 一般就讨论张量场足矣。场是背景上像为张量的映射, 即给背景 (或其子集) 的每一点都赋予一个张量。

从而可借助在背景上给定度规场以定义一段曲线的弧长:

定义 2.40 (线长). 背景 M 上 (至少分段可导的) 曲线 $C : [t_1, t_2] \rightarrow M$ 的线长定义为 Riemann 积分

$$L(C) = \int_{t_1}^{t_2} \sqrt{|g(\partial_t, \partial_t)|} dt, \quad (2.1.33)$$

这里 ∂_t 是曲线的切矢。因此我们常象特征地将线元表示为 $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$ 。

这样我们便将矢量空间上的“长度”搬到了背景上来。因而我们说度规场赋予了背景以几何结构。

¹⁷以后会知道, 如此构造的矢量空间称作背景 (流形) 的切空间 (*tangent space*).

¹⁸切矢当然可以有很多, 只是一般约定是曲线 (坐标) 对参数的求导, 这样很可能在许多情况使得模长归一. 这样便可把曲线一点的所有切矢与该矢量本身等同。

楔积

说回微分形式. V 上全体 l -形式之集记作 $\Lambda_l(V)$, 它当然是 $\mathcal{F}_V(0, l)$ 的一个子集, 但更进一步, 容易验证它还是线性子空间, 只需按逐点运算构造出矢量空间. 其维数的计算可仿照 $\mathcal{F}_V(k, l)$ 那样, 用张量积构造基底. 然而两形式的张量积并非全反称, 故所得不再是形式. 于是我们应尝试用张量积构造出一种专用于形式的乘积“ \wedge ”, 要求对结果的全体指标施行“反称化”, 即可得到形式. 首先, 预期想得到:

$$\omega = \omega_{\mu_1 \dots \mu_l} dx^{\mu_1} \otimes \dots \otimes dx^{\mu_l} = \sum_{\mu_1 < \dots < \mu_l} \omega_{\mu_1 \dots \mu_l} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_l}, \quad (2.1.34)$$

这样积分就可以写成

$$\int \omega,$$

如此, 积分中的微元就不应理解为普通乘法, 而是楔积. 比如 $dx dy = -dy dx, dx dx = 0$ 这样的式子¹⁹, 其实表明积分微元间只是略去了楔积符. 由于 ω 反称性, 式左非零项下标各不相同; 单个组合的全排列可分为偶、奇排列, 而奇排列总可保持偶排列的系数下标, 而只给后面的基添个负号, 于是可提出单个组合的唯一系数, 其下标从小到大:

$$\begin{aligned} & \sum_{\mu_1 \neq \dots \neq \mu_l} \omega_{\mu_1 \dots \mu_l} dx^{\mu_1} \otimes \dots \otimes dx^{\mu_l} \\ &= \sum_{\mu_1 < \dots < \mu_l} \omega_{\mu_1 \dots \mu_l} \sum_{\sigma \in S_l} \operatorname{sgn} \sigma dx^{\mu_{\sigma(1)}} \otimes \dots \otimes dx^{\mu_{\sigma(l)}}, \end{aligned}$$

这里 S_l 是全体 $\{1, \dots, l\}$ 到自身双射组成的集合; σ 为偶排列时 $\operatorname{sgn} \sigma = 1$, 否则为 -1 ; $\sigma(k)$ 为 σ 的第 k 项. 说明

$$dx^{\mu_1} \wedge \dots \wedge dx^{\mu_l} = \sum_{\sigma \in S_l} \operatorname{sgn} \sigma dx^{\mu_{\sigma(1)}} \otimes \dots \otimes dx^{\mu_{\sigma(l)}},$$

这就可看作是对 $dx^{\mu_1} \otimes \dots \otimes dx^{\mu_l}$ 全体指标的“反称化”, 因为

$$\begin{aligned} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_l}(v_1, \dots, v_l) &= \sum_{\sigma \in S_l} \operatorname{sgn} \sigma dx^{\mu_{\sigma(1)}} \otimes \dots \otimes dx^{\mu_{\sigma(l)}}(v_1, \dots, v_l) \\ &= \sum_{\sigma \in S_l} \operatorname{sgn} \sigma dx^{\mu_{\sigma(1)}}(v_1) \cdots dx^{\mu_{\sigma(l)}}(v_l) \\ &= \det(dx^{\mu_i}(v_j)) \end{aligned}$$

确为反称函数。

¹⁹这涉及“定向”这一概念, 见本小节末尾.

定义 2.41. 全反称化算子 $\text{alt} : \mathcal{T}_V(0, k) \rightarrow \mathcal{T}_V(0, k)$ 将一个 k 阶协变张量 T 变为一个反称的协变张量：

$$\text{alt } T = \sum_{\sigma \in S_k} \text{sgn } \sigma \sigma T, \quad (2.1.35)$$

其中

$$\sigma T(v_1, \dots, v_k) = T(v_{\sigma(1)}, \dots, v_{\sigma(k)}) \quad (2.1.36)$$

是 T 按某一 σ 给定的一个分量。对其它类型张量的指标可采取类似方法定义。

全对称化算子 sym 就按

$$\text{sym } T = \sum_{\sigma \in S_k} \sigma T \quad (2.1.37)$$

定义。

这样我们可以改写为

$$dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_l} = \text{alt}(dx^{\mu_1} \otimes \cdots \otimes dx^{\mu_l}), \quad (2.1.38)$$

而另一方面，我们希望楔积有结合性，这样任意两形式的楔积就可以提出系数，而把它们的基“凑在一起”：

$$(dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}) \wedge (dx^{\mu_{k+1}} \wedge \cdots \wedge dx^{\mu_{k+l}}) = dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_{k+l}}, \quad (2.1.39)$$

而式右就是

$$\text{alt}(dx^{\mu_1} \otimes \cdots \otimes dx^{\mu_{k+l}}) = \text{alt}((dx^{\mu_1} \otimes \cdots \otimes dx^{\mu_k}) \otimes (dx^{\mu_{k+1}} \otimes \cdots \otimes dx^{\mu_{k+l}})),$$

这里利用张量积的结合性拆开了二者，以便于我们寻找一般楔积和一般张量积的关系。但我们希望里面应是两个形式之张量积，因此观察

$$\begin{aligned} & \text{alt}((dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}) \otimes (dx^{\mu_{k+1}} \wedge \cdots \wedge dx^{\mu_{k+l}})) \\ &= \text{alt}(\text{alt}(dx^{\mu_1} \otimes \cdots \otimes dx^{\mu_k}) \otimes \text{alt}(dx^{\mu_{k+1}} \otimes \cdots \otimes dx^{\mu_{k+l}})), \end{aligned}$$

其形如 $\text{alt}(\text{alt } f \otimes \text{alt } g)$ ，尝试化简它。考虑到 alt 的线性，不妨先考虑一边一边地找。

定理 2.1.16. 设 $f \in \mathcal{T}_V(0, k)$ 和 $g \in \mathcal{T}_V(0, l)$ ，则 $\text{alt}(\text{alt } f \otimes g) = k! \text{alt}(f \otimes g)$ ，
 $\text{alt}(f \otimes \text{alt } g) = l! \text{alt}(f \otimes g)$ 。

证。根据定义，

$$\text{alt}(\text{alt } f \otimes g) = \sum_{\sigma \in S_{k+l}} \text{sgn } \sigma \sigma \left(\sum_{\tau \in S_k} \text{sgn } \tau (\tau f) \otimes g \right),$$

可以发现 $\tau \in S_k$ 相当于固定 $k+1, \dots, k+l$ 的 $\sigma \in S_{k+l}$, 这样 τ 就满足

$$(\tau f) \otimes g = \tau(f \otimes g),$$

进而

$$\text{alt}(\text{alt } f \otimes g) = \sum_{\sigma \in S_{k+l}} \sum_{\tau \in S_k} \text{sgn } \sigma \text{ sgn } \tau \sigma(\tau(f \otimes g)),$$

这里注意, 若 $\sigma, \tau \in S_k$, 则对 $\forall v_i \in V$ 有

$$\begin{aligned} \sigma(\tau f)(v_1, \dots, v_k) &= (\tau f)(v_{\sigma(1)}, \dots, v_{\sigma(k)}) \\ &= f(v_{\sigma(\tau(1))}, \dots, v_{\sigma(\tau(k))}) \\ &= f(v_{(\sigma\tau)(1)}, \dots, v_{(\sigma\tau)(k)}) \\ &= (\sigma\tau)f(v_1, \dots, v_k), \end{aligned}$$

故 $\sigma(\tau f) = (\sigma\tau)f$, 进而

$$\text{alt}(\text{alt } f \otimes g) = \sum_{\sigma \in S_{k+l}} \sum_{\tau \in S_k} \text{sgn } \sigma \text{ sgn } \tau (\sigma\tau)(f \otimes g), \quad (2.1.40)$$

对 $\forall \mu \in S_{k+l}, \tau \in S_k$, $\exists! \sigma = \mu\tau^{-1} \in S_{k+l}$ 使 $\mu = \sigma\tau$ (从而 $\text{sgn } \mu = \text{sgn } \sigma \text{ sgn } \tau$), 故对 $\forall \tau$, 每个 μ 在上式中都只出现一次, 因而总共有 $k!$ 次, 即

$$\text{alt}(\text{alt } f \otimes g) = k! \sum_{\mu \in S_{k+l}} \text{sgn } \mu \mu(f \otimes g) = k! \text{alt}(f \otimes g).$$

同理可证 $\text{alt}(f \otimes \text{alt } g) = l! \text{alt}(f \otimes g)$. □

综上, $\text{alt}(\text{alt } f \otimes \text{alt } g) = k! \text{alt}(f \otimes \text{alt } g) = k!l! \text{alt}(f \otimes g)$, 即

$$\begin{aligned} &\text{alt}((dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}) \otimes (dx^{\mu_{k+1}} \wedge \cdots \wedge dx^{\mu_{k+l}})) \\ &= k!l!(dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_k}) \wedge (dx^{\mu_{k+1}} \wedge \cdots \wedge dx^{\mu_{k+l}}), \end{aligned}$$

这说明

定义 2.42 (楔积). 任意 k -形式 ω 和 l -形式 μ 的楔积为

$$\omega \wedge \mu = \frac{1}{k!l!} \text{alt}(\omega \otimes \mu). \quad (2.1.41)$$

至于为什么改为“定义”, 这是因为从这一条出发即可证得全部结论. 所以其实我们相当于花了较大篇幅, 以解释该定义尤其是那个系数是怎么来的.

定义 2.43. $(0, k)$ 型张量 T 的反称 (*alternating*) 部分为

$$AT = \frac{1}{k!} \sum_{\sigma \in S_k} \operatorname{sgn} \sigma \sigma T, \quad (2.1.42)$$

除以 $k!$ 表示与原张量平均。分量写法上，会在指标上标注方括号，比如 $T^{\lambda}_{[\mu\nu]} = \frac{1}{2}(T^{\lambda}_{\mu\nu} - T^{\lambda}_{\nu\mu})$ ；

其对称 (*symmetric*) 部分是指

$$ST = \frac{1}{k!} \sum_{\sigma \in S_k} \sigma T, \quad (2.1.43)$$

分量写法上，会在指标上标注圆括号，比如 $T^{\lambda}_{(\mu\nu)} = \frac{1}{2}(T^{\lambda}_{\mu\nu} + T^{\lambda}_{\nu\mu})$ 。

二者相加即为 T ，因此称为“部分”(*part*)。

请注意，有的书会将 $\operatorname{alt} T$ 定义为 T 的反称部分，导致系数仿佛“变了”：

$$\omega \wedge \mu = \frac{(k+l)!}{k!l!} A(\omega \otimes \mu), \quad (2.1.44)$$

更有甚者，直接将 $\omega \wedge \mu$ 定义为 $\omega \otimes \mu$ 的反称部分，使得系数“归一化”。因此查阅不同书籍用以辅助时，请注意分辨。回到 (2.1.34)，由于 ω 非零分量至多只有指标组合的情况，式右可把求和换成 $\frac{1}{l!}$ ，表示除序。利用本节知识， $\frac{1}{l!} \operatorname{alt} = A$ ，故此式即是说

$$\omega = A\omega, \quad (2.1.45)$$

这等价于 ω 为全反称张量。

说回 $\Lambda_l(V)$ ，已经看到，楔积可构造其基底，设 $\dim V = n$ ，注意任意 ω 表为分量时是按组合求和，则显然维度为组合数 $\dim \Lambda_l(V) = \frac{n!}{l!(n-l)!}$ ，当然这只是 $l \leq n$ 时，如果 $l > n$ ，等你排完 n 个指标后，余下指标肯定有与之重复的（抽屉原理），因此所有分量都是零，即此时 $\Lambda_l(V) = \{0\}$ 。

定理 2.1.17. 楔积的运算律：

（结合律） $(\omega \wedge \mu) \wedge \nu = \omega \wedge (\mu \wedge \nu)$ ；

（分配律） $\omega \wedge (\mu + \nu) = \omega \wedge \mu + \omega \wedge \nu$ ；

（交换律） $\omega \wedge \mu = (-1)^{kl} \mu \wedge \omega$ 。

均拆到基底证明即可。

外微分

接下来探究形式与微分相关的性质。首先是函数微分 d 的推广：

定义 2.44 (外微分). 外微分是映射 $d : \Lambda_l(V) \rightarrow \Lambda_{l+1}(V)$, 使得

$$\begin{aligned} d\omega &= \sum_{\mu_1 < \dots < \mu_l} d\omega_{\mu_1, \dots, \mu_l} \wedge dx^{\mu_1} \wedge \dots \wedge dx^{\mu_l} \\ &= \sum_{\sigma < \mu_1 < \dots < \mu_l} \frac{\partial}{\partial x^\sigma} (\omega_{\mu_1, \dots, \mu_l}) dx^\sigma \wedge dx^{\mu_1} \wedge \dots \wedge dx^{\mu_l}, \end{aligned}$$

即作用于形式的系数并添上楔积.

定理 2.1.18 (Poincaré 引理). $d^2 = 0$.

证. 注意

$$dd\omega = \sum_{\nu < \sigma < \mu_1 < \dots < \mu_l} \frac{\partial^2 \omega_{\mu_1, \dots, \mu_l}}{\partial x^\nu \partial x^\sigma} dx^\nu \wedge dx^\sigma \wedge dx^{\mu_1} \wedge \dots \wedge dx^{\mu_l},$$

由函数的偏导对易性知 $(dd\omega)_{\sigma\nu\mu_1\dots\mu_l} = (dd\omega)_{\nu\sigma\mu_1\dots\mu_l}$, 由反称性知 $(dd\omega)_{\sigma\nu\mu_1\dots\mu_l} = -(dd\omega)_{\nu\sigma\mu_1\dots\mu_l}$, 则 $(dd\omega)_{\nu\sigma\mu_1\dots\mu_l} = 0$, 故 $dd\omega = 0$. \square

定义 2.45 (闭、恰当). 给定背景 M 及坐标系 (X, ψ) , 设至少 X 上有 l -形式场 ω . ω 称闭的 $\iff d\omega = 0$; ω 称恰当的 $\iff \exists (l-1)$ -形式场 μ 使 $\omega = d\mu$.

根据 Poincaré 引理, 恰当形式一定为闭形式. 反之不一定, 欲成立则需对背景 (坐标域) 另设要求, 但暂不展开²⁰, 只需强调 $X = \mathbf{R}^n$ 上闭形式一定为恰当形式 (可直接用分析学证明). 对一般的背景而言, 任意局域 X 内亦成立, 即逆命题至少在处处点的邻域上成立.

2.1.7 积分

上节一开始引入这套语言, 便旨在改写甚至推广积分. 首先是有关楔积交换律, 很明显交换对偶坐标基会导致反号. 但为何有时交换积分次序并不影响符号呢? 其实是因为譬如二重积分这样, 意义明显的积分在交换楔积时, 亦“偷偷”地调换了上下限, 导致抵消. 但对更高维的积分, 这样做易出错, 因为没有几何图像可根据, 不知道应在哪一步交换上下限. 为此, 引入定向 (*orient*) 的概念, 有了这根弦, 便不再怕高维情况了.

定向

以欧氏背景 \mathbf{R}^3 为例, 设 \vec{v} 为矢量场, l 为光滑曲线, σ 为光滑曲面. 在指定 l 的方向和 σ 的法向之前, 积分 $\int_l \vec{v} \cdot d\vec{l}$ 和 $\int_\sigma \vec{v} \cdot d\vec{S}$ 都只唯一确定到差一个负号的程度.

²⁰剧透: 条件为 $X \subset M$ 比较“平凡”, 例如 \mathbf{R}^n , 具体指其第二种 de Rham 上同调 $H^2(X; \mathbf{R})$ 平凡. 例子还包括 \mathbf{R}^n 的开球及其更一般的“星形式”开区域等.



图 2.6: 曲线、曲面的定向

度. 欲完全确定此二积分就要指定 l 的方向和 σ 的法向. 推而广之, 计算任意积分域上的积分前应指定该区域的“定向”.

定义 2.46 (可定向的). 背景内 n 维区域称为可定向的 (*orientable*), 若其上存在一个非退化的 n -形式场 ε . 也可称这种形式场是非退化的.

例 2.3. \mathbf{R}^3 可定向, 其上存在光滑的 3-形式场 $\varepsilon = dx \wedge dy \wedge dz$, 这里 x, y, z 指其自然坐标.

例 2.4. Möbius 环就是不可定向的一例 (图 2.7), 因为若尝试给其连续地指定非零定向, 则沿途回到原处后, 一定会在该处断开, “突然” 反号, 这不满足连续要求. 从直观上讲, 这意味着它身为一张 2 维曲面却只有一个面, 不能区分正反面!

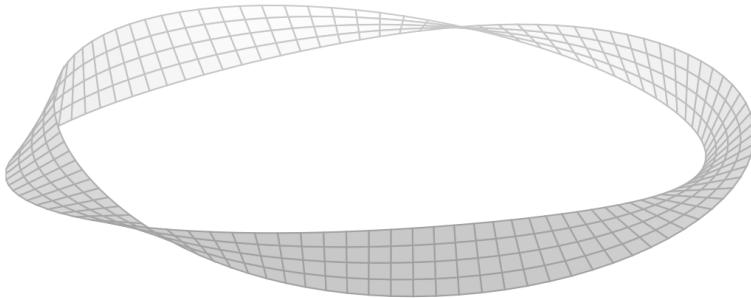


图 2.7: Möbius 环

定义 2.47 (相同定向). 在 n 维可定向区域 M 上选定了一个非退化 n -形式场 ε , 就说 M 是 (已经) 定向的 (*oriented*). 设任意两个非退化 n -形式场 ε_1 和 ε_2 , 若 $\varepsilon_1/\varepsilon_2$ 处处恒正, 就说 ε_1 和 ε_2 给出相同定向; 否则为相反定向.

从给出定向这个角度看, 满足 $\varepsilon_1 = h\varepsilon_2 (h > 0)$ 的 ε_1 和 ε_2 是等价的. 全体非退化 n -形式之集按此标准可分为两个等价类, 每个等价类称为 M 的一个定向. 某个非退化最高阶形式 ε 所在的等价类称为 M 上由 ε 决定的定向.

n 维区域上每点的全体 n -形式之集合相当于 1 维矢量空间 \mathbf{R} , 因此任意两个 n -形式场 ε_1 和 ε_2 必存在 $\varepsilon_1 = h\varepsilon_2$, 其中 h 是 M 上的函数. 若 ε_1 和 ε_2 处处非零, 则 h 处处非零; 若 ε_1 和 ε_2 连续, 则 h 连续. 因此 h 非退化. 就连通区域而言, 非退化函数只能处处为正或处处为负. 可见连通区域至多有两种定向.

定义 2.48. M 上选好以 ε 为代表的定向后 (规定称作正定向), 开域 $O \subset M$ 上的基底场 $\{e_\mu\}$ 称为在 ε 衡量下为右手的 (*right handed*), 若 O 上存在函数 $h > 0$ 使 $\varepsilon = he^1 \wedge \cdots \wedge e^n$, 其中 $\{e^\mu\}$ 是其对偶基; 反之称为左手的. 若坐标系的坐标基是右 (左) 手的, 则称为右 (左) 手系.

\mathbf{R}^n 通常取将 $dx^1 \wedge \cdots \wedge dx^n$ 衡量为右手系的定向为正定.

积分

欲定义形式的积分, 同分析学一样, 只能要求区域是与其同维的. 在分析学中, 通常只说“对函数 f 进行积分”, 但实际上积分对象是一个函数 f 与基底的乘积, 组成一个微分形式 ω .

比如这样一个重积分

$$\int_{\Omega} f(x, y) \, dx \, dy,$$

看上去是对函数 $f(x, y)$ 积分, 但其实应当看作对“带上微元的整个形式”积分.

$$\omega = f(x, y) \, dx \wedge dy,$$

这里 Ω 是 2 维, 则 ω 就是其上的 2-形式. 如果换成在极坐标下积分, 那这个积分应该写为

$$\int_{\Omega} f(r \cos \theta, r \sin \theta) r \, dr \, d\theta,$$

被积函数的形式完全变了, 但被积分的微分形式没变: 这是因为可以证明形式的分量坐标变换满足

$$\omega'_{1 \dots n} = \omega_{1 \dots n} \det \left(\frac{\partial x^\mu}{\partial x'^\nu} \right). \quad (2.1.46)$$

因此, 分析学中对函数的重积分的自然拓展实际上是对区域上的最高阶形式的积分, 而对函数的积分可以看作乘以最高阶基底后再积分. 当然, 如果这个区域相较于背景是低维度的, 则还会有 $dy \, dz, dz \, dx$, 但不论长成什么样, 它总可以拆到单独的基底上. 这样在定义的时候, 积分总可看作对 n 维定向区域 M 上 n -形式场 ω 的积分. 所谓的微元即基底本身也是形式, 考虑到它与积分域 (背景) 是同维的, 可称体积形式, 简称体元.

定义 2.49 (体元). n 维可定向背景 M 上的任意非退化 n -形式场 ε 称为体元.

体元和定向长得有点像，不过区别在于：若 ε_1 和 ε_2 是两个非退化的 n -形式场，而且有处处为正的函数 h 使 $\varepsilon_1 = h\varepsilon_2$ ，则 ε_1 和 ε_2 代表同一定向，但只要 $\varepsilon_1 \neq \varepsilon_2$ ，它们就是不同体元。对可定向连通背景，定向只有两个，而体元却有无限多个。

有了这些概念后，就可尝试严格叙述。 ω 可用对偶坐标基展开为

$$\omega = \omega_{1\dots n} dx^1 \wedge \dots \wedge dx^n,$$

更准确地讲，系数应表为其在坐标域上给出的 n 元函数 $\omega_{1\dots n} \circ \psi^{-1}(x^1, \dots, x^n)$ ，简记为 $\omega_{1\dots n}(x^\mu)$ 。这样就可将任意背景上的积分严格地表示为

定义 2.50 (积分)。设 (O, ψ) 是 n 维定向背景 M 上的右手系。 ω 是开集 $G \subset O$ 上的连续 n -形式场，则 ω 在 G 上的积分为

$$\int_G \omega = \int_{\psi[G]} \omega_{1\dots n}(x^\mu) dx^1 \dots dx^n, \quad (2.1.47)$$

式右是函数在坐标空间 \mathbf{R}^n 开子集 $\psi[G]$ 上的 Riemann 积分²¹，而已在分析学中有定义。注意，若积分域甚至不连通，都可看作连通集之并，将各部分累加即可，此即说 $\int_{G \cup D} \omega = \int_G \omega + \int_D \omega$ ，其中注意 ω 取各自分支的情况。此即不妨默认连通之原因。

式右与坐标系选取无关，因为另选一右手系 $(O', \psi'; x'^\mu)$ 并使得 $G \subset O \cap O'$ ，则有

$$\omega = \omega_{1\dots n} dx^1 \wedge \dots \wedge dx^n = \omega'_{1\dots n} dx'^1 \wedge \dots \wedge dx'^n,$$

进而

$$\begin{aligned} \int_{\psi[G]} \omega_{1\dots n} dx^1 \dots dx^n &= \int_{\psi'[G]} \omega_{1\dots n} \det\left(\frac{\partial x^\mu}{\partial x'^\nu}\right) dx'^1 \dots dx'^n \\ &= \int_{\psi'[G]} \omega'_{1\dots n} dx'^1 \dots dx'^n, \end{aligned}$$

其中用到了分析学中有关 Jacobi 变换矩阵的结论。当然，亦可看作从 ω 作为反称张量的张量变换律以逆序证明（即默认积分与坐标无关），但为避免循环论证之嫌，目前先默认分析学已自成派系，以此从 \mathbf{R}^n 推广至任意背景²²。注意，若 $\{x'^\mu\}$ 选为左手系，则 $\det(\partial x^\mu / \partial x'^\nu) < 0$ ，因此在分析学中的规范写法应取其绝对值，即 $-\det(\partial x^\mu / \partial x'^\nu)$ 。这样

$$\int_{\psi[G]} \omega_{1\dots n} dx^1 \dots dx^n = - \int_{\psi'[G]} \omega'_{1\dots n} dx'^1 \dots dx'^n,$$

²¹以后介绍 Lebesgue 积分的情况，它们的故事很微妙。

²²当然，考虑到本章只讨论平直时空 \mathbf{R}^4 ，甚至可说无推广之需，视本节所讲代数皆属分析学背景。

因此为定出同一积分，约定左手系时积分定义为

$$\int_G \omega = - \int_{\psi[G]} \omega_{1\dots n}(x^\mu) dx^1 \cdots dx^n.$$

故我们说积分符号取决于左右手系，其根源上是取决于区域的定向。欲将积分域推广至全部背景，需要由每一个局部积分“缝合”而成，于第四章再对此详述。

回到积分上。前面提到了将函数的积分视作形式的积分，事实上也可作如下定义：

定义 2.51 (函数的积分). 设 ε 为背景 M 选定的任意体元， f 为其上的连续函数，则 f 在其上的积分就是 n -形式场 $f\varepsilon$ 在其上的积分

$$\int_M f := \int_M f\varepsilon. \quad (2.1.48)$$

为说清究竟 ε 能否代表体积，下面介绍与度规相适配的体元。

适配体元

谈及定向背景的积分和体元时不要求其上选定度规，这时体元的选择十分任意²³，没有一个与众不同的体元。不过，给定度规场 g 后，便存在选择特定体元的自然方法。考虑 n 维背景，设 ε 为任一体元，则 $\varepsilon^\sharp \equiv g^\sharp(\varepsilon)$ 有意义，且 $\varepsilon^\sharp(\varepsilon)$ 是标量场，可借任一基底计算。在正交基下，设 g 为度规，则

$$\begin{aligned} \varepsilon^\sharp(\varepsilon) &= g^{\mu_1\nu_1} \cdots g^{\mu_n\nu_n} \varepsilon_{\mu_1 \dots \mu_n} \varepsilon_{\nu_1 \dots \nu_n} \\ &= \sum_{\sigma} g^{1\sigma(1)} \cdots g^{n\sigma(n)} \varepsilon_{\sigma(1) \dots \sigma(n)} \varepsilon_{1 \dots n} \\ &= n! g^{11} \cdots g^{nn} (\varepsilon_{1 \dots n})^2 \\ &= (-1)^s n! (\varepsilon_{1 \dots n})^2, \end{aligned}$$

其中 $\varepsilon_{1 \dots n}$ 是 ε 在正交基的分量， s 是 g 在正交基的分量中 -1 的个数，例如正定度规有 $s = 0$ ，号差 $+2$ 的 Lorentz 度规有 $s = 1$ 。所谓借用度规选择一个特定的体元，是指规定体元 ε 在正交基 $\{e^\mu\}$ 的分量满足如下的简单性要求 $\varepsilon_{1 \dots n} = \pm 1$ 或者说 $\varepsilon = \pm e^1 \wedge \cdots \wedge e^n$ ，相当于

$$\varepsilon^\sharp(\varepsilon) = (-1)^s n!. \quad (2.1.49)$$

满足上式的 ε 称为与度规 g 相适配（相容）的体元。上式只把体元确定到差一个负号的程度，加上“体元与定向相容”的要求就可确定唯一体元。于是 $+, -$ 号分别对应于右手和左手正交基。

²³ 只有一个要求，就是体元与定向相容，即代表体元的 ε 与代表定向的 ε 之间的乘子为正。

首先选好一个定向使 M 成为定向背景. 任一基底的右(左)手性由所选定向规定. 没有度规 g 这类几何结构时, 体元相当任意, 但要求与定向相容. 指定 g 后, 体元 ε 由 g 以及“体元与定向相容”的要求唯一确定, 简称适配体元. 在欧氏 \mathbf{R}^3 中任取符合直观含义的右手正交系 $\{x, y, z\}$ 并指定 3-形式场 $\varepsilon = dx \wedge dy \wedge dz$ 为定向, 则 $\{x, y, z\}$ 就是以 ε 衡量的右手系, 并且还是适配体元. 设 G 是 \mathbf{R}^3 的开集且 Riemann 积分 $\int_G dx dy dz$ 存在, 则此积分自然代表 G 的体积²⁴. 另一方面, 由定义知 3-形式场 ε 在 $G \subset \mathbf{R}^3$ 上的积分 $\int_G \varepsilon$ 正是 $\int_G dx dy dz$, 可见 $\int_G \varepsilon$ 就是 G 的体积. 推广至任意带正定度规 g 的定向背景 N , 设 ε 为适配体元, 若 $\int_N \varepsilon$ 存在, 就称它为 N 用 g 衡量的体积(1、2 维时又分别叫长度和面积). 这便是为何把 ε 称为“体积微元”.

定理 2.1.19. 设 ε 为适配体元, $\{e_\mu\}$ 及 $\{e^\mu\}$ 为基底及其对偶基底, g 为 g 在此基底的矩阵行列式, $|g|$ 为 g 的绝对值, 则(式中 +, - 号分别适用于右手和左手基底)

$$\varepsilon = \pm \sqrt{|g|} e^1 \wedge \cdots \wedge e^n. \quad (2.1.50)$$

证. 由式(2.1.49)知 ε 及 g 在所给基底的分量满足

$$(-1)^{s!} = \varepsilon^{\mu_1 \cdots \mu_n} \varepsilon_{\mu_1 \cdots \mu_n} = g^{\mu_1 \nu_1} \cdots g^{\mu_n \nu_n} \varepsilon_{\nu_1 \cdots \nu_n} \varepsilon_{\mu_1 \cdots \mu_n}, \quad (2.1.51)$$

考虑到 ε 的全反称性, 上述求和简化为对排列的求和. 以 \sum_π 代表对 $1, 2, \dots, n$ 所有排列求和, 则

$$\begin{aligned} \text{式右} &= \sum_{\pi(\mu_1 \cdots \mu_n)} \sum_{\pi(\nu_1 \cdots \nu_n)} g^{\mu_1 \nu_1} \cdots g^{\mu_n \nu_n} \varepsilon_{\nu_1 \cdots \nu_n} \varepsilon_{\mu_1 \cdots \mu_n} \\ &= \sum_{\pi(\mu_1 \cdots \mu_n)} g^{\mu_1 1} g^{\mu_2 2} g^{\mu_3 3} \cdots g^{\mu_n n} \varepsilon_{123 \cdots n} \varepsilon_{\mu_1 \cdots \mu_n} \\ &\quad + \sum_{\pi(\mu_1 \cdots \mu_n)} g^{\mu_1 2} g^{\mu_2 1} g^{\mu_3 3} \cdots g^{\mu_n n} \varepsilon_{213 \cdots n} \varepsilon_{\mu_1 \cdots \mu_n} + \cdots. \end{aligned}$$

上式右边共有 $n!$ 项. 以 $\hat{\varepsilon}_{\mu_1 \cdots \mu_n}$ 代表 Levi-Civita 记号: 若 π 是个排列, 则 $\hat{\varepsilon}_\pi \equiv \text{sgn}(\pi)$; 否则为零. 有 $\varepsilon_{\mu_1 \cdots \mu_n} = \varepsilon_{123 \cdots n} \hat{\varepsilon}_{\mu_1 \cdots \mu_n}$. 再把 $\sum_{\pi(\mu_1 \cdots \mu_n)}$ 简记为 \sum_π , 则上式右边第一项为

$$(\varepsilon_{123 \cdots n})^2 \sum_\pi g^{\mu_1 1} g^{\mu_2 2} g^{\mu_3 3} \cdots g^{\mu_n n} \hat{\varepsilon}_{\mu_1 \mu_2 \mu_3 \cdots \mu_n} = (\varepsilon_{1 \cdots n})^2 \det(g^{\mu\nu}),$$

²⁴按普通分析学的体积定义(Riemann 积分)而非测度学的(Lebesgue 积分).

其中 $\det(g^{\mu\nu})$ 代表矩阵 $(g^{\mu\nu})$ 的行列式 (最末一步用到行列式的定义). 而第二项为

$$\begin{aligned}
 & - \sum_{\pi} g^{\mu_1 2} g^{\mu_2 1} g^{\mu_3 3} \cdots g^{\mu_n n} \varepsilon_{123 \dots n} \varepsilon_{\mu_1 \mu_2 \mu_3 \cdots \mu_n} \\
 &= - (\varepsilon_{123 \dots n})^2 \sum_{\pi} g^{\mu_1 2} g^{\mu_2 1} g^{\mu_3 3} \cdots g^{\mu_n n} \hat{\varepsilon}_{\mu_1 \mu_2 \mu_3 \cdots \mu_n} \\
 &= - (\varepsilon_{1 \dots n})^2 \sum_{\pi} g^{\mu_2 2} g^{\mu_1 1} g^{\mu_3 3} \cdots g^{\mu_n n} \hat{\varepsilon}_{\mu_2 \mu_1 \mu_3 \cdots \mu_n} \\
 &= (\varepsilon_{1 \dots n})^2 \sum_{\pi} g^{\mu_1 1} g^{\mu_2 2} g^{\mu_3 3} \cdots g^{\mu_n n} \hat{\varepsilon}_{\mu_1 \mu_2 \mu_3 \cdots \mu_n} \\
 &= (\varepsilon_{1 \dots n})^2 \det(g^{\mu\nu}).
 \end{aligned}$$

同理可证上式右边每一项都等于 $(\varepsilon_{1 \dots n})^2 \det(g^{\mu\nu})$. 注意到该式右边共有 $n!$ 项, 代回式 (2.1.51) 便得 $(-1)^s n! = (n!) (\varepsilon_{1 \dots n})^2 \det(g^{\mu\nu})$, 或 $(-1)^s = (\varepsilon_{1 \dots n})^2 \det(g^{\mu\nu})$. 矩阵 $(g_{\mu\nu})$ 与 $(g^{\mu\nu})$ 互逆导致 $\det(g^{\mu\nu}) = 1 / \det(g_{\mu\nu}) \equiv 1/g$. 代入上式得

$$(-1)^s g = (\varepsilon_{1 \dots n})^2, \quad \varepsilon_{1 \dots n} = \pm \sqrt{|g|}. \quad (2.1.52)$$

此即结果. 对正交基有 $|g| = 1$, 故 $\varepsilon = \pm \sqrt{|g|} e^1 \wedge \cdots \wedge e^n$ 回到 $\varepsilon = \pm e^1 \wedge \cdots \wedge e^n$. 如此, 适配体元与坐标选取无关. \square

函数的积分与体元的选择有关, 故只要给定度规, 约定总是用适配体元来定义函数的积分. 以欧氏 \mathbf{R}^3 为例. 设 $\{x, y, z\}$ 为右手正交系, 则 $\varepsilon = dx \wedge dy \wedge dz$ 是适配体元, 于是 \mathbf{R}^3 上的函数 $f : \mathbf{R}^3 \rightarrow \mathbf{R}$ 的积分为 $\int_{\mathbf{R}^3} f = \int_{\mathbf{R}^3} f \varepsilon$, 而右边无非是 3-形式 $\omega \equiv f \varepsilon$ 的积分, 按定义应把 ω 用右手对偶坐标基展开. 设 f 在正交系 $\{x, y, z\}$ 下表达为 3 元函数 $F(x, y, z)$, 则 $\omega = F(x, y, z) dx \wedge dy \wedge dz$, 故 $\int f = \int f \varepsilon = \int \omega = \int F(x, y, z) dx dy dz$. 若愿意, 亦可用 (右手) 球坐标系 $\{r, \theta, \varphi\}$ 计算. 度规在球坐标下写成 $ds^2 = dr^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2)$ 可知 $g = r^4 \sin^2 \theta$, 故由 (2.1.50) 知 $\varepsilon = r^2 \sin \theta dr \wedge d\theta \wedge d\varphi$, 于是现在的情况是 $\omega \equiv f \varepsilon = \hat{F}(r, \theta, \varphi) r^2 \sin \theta dr \wedge d\theta \wedge d\varphi$, 其中 $\hat{F}(r, \theta, \varphi)$ 是由 f 在 $\{r, \theta, \varphi\}$ 下的表达, 故

$$\int f = \int f \varepsilon = \int \omega = \int \hat{F}(r, \theta, \varphi) r^2 \sin \theta dr d\theta d\varphi.$$

可见, 在带度规的定向区域上, 函数给定后其积分也就确定.

Hodge 对偶

最后花简短篇幅讲解形式的对偶, 利于完善这套语言. 我们知道, 在给定 n 维定向背景一点处 V 上, 若 $l \leq n$, 则 $\dim \Lambda_l(V) = \frac{n!}{l!(n-l)!}$, 而另一方面这个组合数也等于 $\dim \Lambda_{n-l}(V)$, 则二者同构. 借助度规及其适配体元, 就可在二者之间建立最自

然的同构，称为 *Hodge 对偶*，以符号 \star 代表 *Hodge 星算子 (star)*， $\star\omega$ 称为 ω 的对偶形式 (*dual form*)。

一种定义是：对任意 k -形式 ω, μ 有

$$\mu \wedge \star\omega = \langle \mu, \omega \rangle \varepsilon, \quad (2.1.53)$$

这里 $\langle \mu, \omega \rangle$ 表示两个 k -形式的 Gram 行列式：将二者拆到基底，所谓的 Gram 矩阵由其基底的内积排成，故

$$\langle dx^1 \wedge \cdots \wedge dx^k, dx^1 \wedge \cdots \wedge dx^k \rangle = \det \langle dx^i, dx^j \rangle = \det(g^{ij}) = (-1)^s, \quad (2.1.54)$$

这说明在正交系下

$$dx^1 \wedge \cdots \wedge dx^k \wedge \star(dx^1 \wedge \cdots \wedge dx^k) = (-1)^s dx^1 \wedge \cdots \wedge dx^n. \quad (2.1.55)$$

于是形象地说，基底楔积的对偶形式，就是“其它”基底的楔积，正负号取决于度规。从上节的知识，可知反称张量缩并同 Gram 行列式的关系： $\varepsilon^\sharp(\varepsilon) = (-1)^s k!$ ，这说明 Gram 行列式是缩并再除序，于是也有如下常用的缩并表达：

定义 2.52 (*Hodge 对偶*). 对任意 $\omega \in \Lambda_l(V)$ ，其 *Hodge 对偶* $\star\omega \in \dim \Lambda_{n-l}(V)$ 与原形式的关系为

$$\star\omega = \frac{1}{l!} C_1^1 \cdots C_l^l (\omega^\sharp \otimes \varepsilon), \quad (2.1.56)$$

这里 ε 是适配体元。

部分书籍选择用 Gram 行列式表达，因而平均系数 $\frac{1}{l!}$ 看上去好似“归一化”了，请注意分辨。

定理 2.1.20. 如果 $f \in \Lambda_0(V)$ ，则 $\star f = f\varepsilon$ ，这说明

$$\int f = \int \star f. \quad (2.1.57)$$

定理 2.1.21. 容易证明 $\star\star\omega = (-1)^{s+l(n-l)}\omega$ 。

上式在 +2 号差的 Lorentz 度规及 4 维时空中就简化为了 $\star\star\omega = (-1)^{l+1}\omega$ 。

例 2.5. \mathbf{R}^3 中的叉乘可用对偶形式表达。采取正交系分量写法。用欧氏度规降下矢量 A^i, B^j 的指标，令 $\omega_{ij} = A_i \wedge B_j = 2A_{[i}B_{j]}$ ，则

$$\star\omega_k = \frac{1}{2} \omega^{ij} \varepsilon_{ijk} = \varepsilon_{ijk} A^i B^j, \quad (2.1.58)$$

这里 ε 是欧氏度规的适配体元，可见其在右手正交基下的分量 ε_{ijk} 正是 Levi-Civita 符号。因此对两矢量求叉乘，实质上是按顺序做了：降指标、楔积、*Hodge 对偶*和升指标，忽略升降过程可记作 $\times = \star\wedge$ 。

例 2.6 (矢量分析). 容易验证如下公式的外微分语言改写:

$$\begin{aligned} (\text{grad } f)_b &= \frac{\partial f}{\partial x^i} dx^i = df, \\ \text{curl } A &= \varepsilon^{ijk} \partial_i A_j \frac{\partial}{\partial x^k} = (\star dA_b)^\sharp, \\ \text{div } A &= \partial_i A^i = \delta A_b, \\ \Delta &= \nabla^2 = \delta^{ij} \partial_i \partial_j = d\delta + \delta d. \end{aligned}$$

这里 $\delta = \star d \star$, 有时也记作 d^* . ∇^2 便是读者耳熟能详的 Laplace 算子²⁵.

注意, 若认为旋度必须将矢量映射为矢量, 则只能在 3 维空间定义; Laplace 算子用到了 \mathbf{R}^3 的欧氏度规, 用闵氏时空 \mathbf{R}^4 的 η 定义出的就是电动力学中的 d'Alembert 算子 $\square = \eta^{\mu\nu} \partial_\mu \partial_\nu$.

例 2.7. 易从 $d^2 = 0$ 证明 $\nabla \times \nabla f = 0$ 和 $\nabla \cdot \nabla \times A = 0$. 此即 3 维矢量分析中的“梯度无旋, 旋度无散”.

Stokes 定理

利用这套语言, 可将分析学中的积分公式, 如 Green 公式、Gauss 公式、Stokes 公式等统一成(广义) Stokes 定理. 这些定理的共性在于描述了区域边界与区域内部的积分关系. 对此需要先规范何为“边界”.

定义 2.53. 设 n 维背景 M 的一个子集 D , 和一点 p , p 可分为三类:

(内点) 存在一个开集 U 包含 p (称 p 的邻域), 使得 $U \subset D$, 则称 p 为 D 的内点.

(外点) 存在一个 p 的邻域 U 使得 $U \not\subset D$, 则称 p 为 D 的外点.

(边界点) 非内外的点称为边界点, 即任意 p 的邻域都同时包含 D 的内外点. 边界点之集称作边界 ∂D .

注意, 边界点的存在并不要求 D 要包含它, 实际上 D 为开集时可称作 $D \cup \partial D$ 的内部, 而为闭集时就称为带边区域, 内部就应写为 $i(D) = D \setminus \partial D$. 现借助坐标系叙述之. 在 M 中有一包含 p 的局部坐标系 $(O, \psi; x^i)$, 使得 $\psi(p)$ 是坐标原点且 $U \cap D = \{q \in U | x^1(q) \leq 0\}$, 则称 D 是 M 上的一个带边区域, 其边界点 x^1 坐标为零. 坐标空间里这个区域就称作“半空间 (half-space)” \mathbb{H}^n , 有时可记作 $\frac{1}{2}\mathbf{R}^n$. 由于边界点失去了一个坐标自由度, 因此我们很容易发现 ∂D 是相较于 D 而言“余 1 维 (1-codimension)”. 考虑到背景和内部都是同维开集, 因此说带边区域不能直接认同为普通开集, 因其在实质上“强行”拼凑了两个不同维度的东西, 即内部和边界.

²⁵ 它还有其它各种名称: Laplace-Beltrami 算子 (强调在高维背景上)、Laplace-de Rham 算子 (同左), 又或是 Hodge-Laplace 算子 (强调在闵氏时空上).

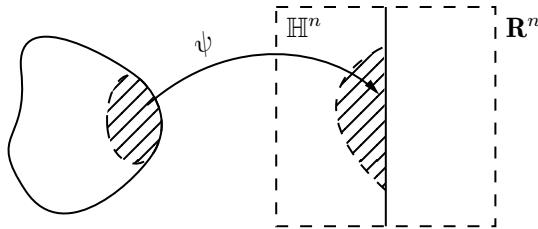


图 2.8: 带边区域

区域定向可在其边界上也诱导出与之相应的定向，为此先说明一些很直观的概念。

定义 2.54 (切矢和法矢). 设 S 是背景 M 上可具有不同维度的子集。将 S 上一点 q 处的矢量称为切矢 w ，若 S 上存在过 q 的曲线与其相切。用坐标基易验证， q 处所有矢量构成的矢量空间 V_q 与 M 同维，而所有 w 所构成的集合称作切空间，记作 $T_q S$ ，是 V_q 的矢量子空间。

可以定义 q 处的法矢 n ，若对任意切矢 w 有 $\langle n, w \rangle = 0$ ，这里内积取决于 M 所选度规。

定理 2.1.22. 当 S 是 M 的余 1 维子空间时，又可以形象地叫做一张超曲面 (*hypersurface*)，其上一点的任意法矢间只差倍数关系。

证。证明设 $\{e_2, \dots, e_n\}$ 为 $T_q S$ 任一坐标基，因 $\dim V_q = n$, V_q 必有与 $\{e_2, \dots, e_n\}$ 线性无关的元素，任取其一并记作 e_1 ，则 $\{e_\mu \mid \mu = 1, \dots, n\}$ 为 V_q 的基底。令 $n = e_1$ ，则 $\langle n, e_\tau \rangle = g(e_1, e_\tau) = 0$ ，这里 $\tau = 2, \dots, n$ ，可见 n 为法矢。若存在 m 满足 $\langle m, e_\tau \rangle = 0$ ，则其在 $\{e_\mu\}$ 的分量 $m^\tau = 0$ ，因而只剩下分量 m^1 ，即 $m = m^1 (e_1)^a = m^1 n$ ，即 m 与 n 只差一因子 m^1 。□

注意，余维数太多的话，比如三维空间的曲线，其上法矢就没有该性质了，因此选取更为任意。对于超曲面，约定其模长可以使得差别仅体现在“内外方向”上。以后默认取单位法矢，即要求 $\langle n, n \rangle = \pm 1$ ，正负号来自度规的正负定性。设 $p \in S$ 上的一个局部坐标系 $(U; x^1, \dots, x^n)$ 是 M 正定的右手系使得域内点 $x^n \leq 0$ ，则 $(U \cap S; x^1, \dots, x^{n-1})$ 也是 S 的局部坐标系，且 $\frac{\partial}{\partial x^n}$ 就是法矢而且是朝外的，又可强调为外指法矢。

对一点 q ，度规的作用范围是其整个矢量空间 V_q ，因此类似于切矢，度规也有“切于” S 的部分，那就是只作用于 $T_q S$ 的部分。

定义 2.55 (诱导度规). 设 S 是背景 M 的子集而别无他求，其上一点 q 处的切空间

$T_q S$ 的度规 \tilde{g} 称为 V_q 上 g 的诱导度规 (*induced metric*), 若对任意 $w_1, w_2 \in T_q S$ 有

$$\tilde{g}(w_1, w_2) = g(w_1, w_2), \quad (2.1.59)$$

可见诱导度规只是将度规的作用对象从 V_q 限制于 $T_q S$, 因此又称作度规在 S 上的限制 (*restriction*). g 定义在 V_q 上, 而 \tilde{g} 可视作定义在 $T_q S$ 上.

满足上式的诱导度规有很多, 下面介绍一种用于超曲面的惯例. 在规定其单位法矢后, 将 v 沿着法向和切向分解, 可得垂直的 $v_n = \pm \langle v, n \rangle n$ (正负跟随 $\langle n, n \rangle = \pm 1$), 切向就是 $v_t = v - v_n$, 即

$$v = v_t \pm \langle v, n \rangle n,$$

如果将 v_t 视作 v 的投影 (*projection*) 或限制, 则应当有 (1,1) 型张量 h 满足 $v_t = C_1^2(h \otimes v)$, 即

$$v_t = h(\cdot; v) = v \mp n_b(v)n = I(\cdot; v) \mp n \otimes n_b(\cdot; v),$$

说明 $h = I \mp n \otimes n_b$, 用 g 降下来,

$$h_b = C_1^1(g \otimes I) \mp C_1^1(g \otimes n \otimes n_b) = g \mp n_b \otimes n_b, \quad (2.1.60)$$

可以验证这就是一个诱导度规. 下面从缩并角度解释其合理性. 为便于陈述, 设 V_q 为 4 维, 因而 $T_q S$ 为 3 维. 作为诱导度规, \tilde{g} 是 $T_q S$ 上的 3 维张量, 即 $\tilde{g} \in \mathcal{T}_{T_q S}(0, 2)$ (它不能作用于 $V_q \setminus T_q S$ 的元素). 但为了便于用 4 维等式演算, 我们希望找到一个 4 维的 (0,2) 型张量 (即 $\mathcal{T}_{V_q}(0, 2)$ 的元素), 它能代表 3 维张量 \tilde{g} . 而 $h_b \equiv g \mp n_b \otimes n_b$ 就是这样的 4 维张量等式. 可以证明 $\mathcal{T}_{V_q}(0, 2)$ 的子集 $\mathcal{S}_{V_q}(0, 2) \equiv \{T \in \mathcal{T}_{V_q}(0, 2) \mid T(n, \cdot) = 0, T(\cdot, n) = 0\}$ 与 $\mathcal{T}_{T_q S}(0, 2)$ 自然同构, 因而可作认同²⁶. 易见 $g \notin \mathcal{S}_{V_q}(0, 2)$ 而 $h_b \in \mathcal{S}_{V_q}(0, 2)$, 且 $h_b(w_1, w_2) = g(w_1, w_2), \forall w_1, w_2 \in T_q S$, 所以可把 h_b 认同为 \tilde{g} . 容易验证 $\mathcal{S}_{V_q}(0, 2)$ 中满足诱导度规定义 (因而可充当 \tilde{g}) 的元素只有 h_b , 这就是把 4 维张量 $h_b \equiv g \mp n_b \otimes n_b$ 看作诱导度规的理由. 以后将不在符号上区分 h_b 和 \tilde{g} .

以上关于 (0,2) 型张量的结论还可推广: 由 $T \in \mathcal{T}_{V_q}(0, k)$ 满足其任意槽与 n 缩并为零, 所构成的集合与 $\mathcal{T}_{T_q S}(0, l)$ 有自然同构关系. 这使我们在讨论和书写公式时可用前者的元素代替后者的元素 (写成 4 维而非 3 维张量等式), 从而带来许多方便.

此外, 任意 l -形式也能按相应思路定义其诱导形式, 这是有必要的, 因为一旦把 S 看作一个独立区域, 而非从背景的嵌入视角观察, 那么“不切于”的形式都是没有意义的. 我们在讨论形式的积分时, 总是默认对相应的诱导形式积分, 即

$$\int_S \omega = \int_S \tilde{\omega}. \quad (2.1.61)$$

²⁶详见第 3 章参考系一节.

体元、定向也是形式，自然也有其诱导，当然，我们也要从中选择。先谈体元，考虑到适配体元与度规有联系，这意味着上述超曲面的诱导度规可给出唯一惯例，即该体元 $\tilde{\varepsilon}$ 应与 \tilde{g} 适配，故

$$\tilde{\varepsilon}^\sharp(\tilde{\varepsilon}) = (-1)^{\tilde{s}}(n-1)!, \quad (2.1.62)$$

这里 $\tilde{\varepsilon}^\sharp$ 用 \tilde{g} 升，而 \tilde{s} 表示 \tilde{g} 对角元负项个数。这就称为诱导体元。诱导体元在闭合超曲面上的形式，其实就包含了上文所说的缩并技巧。为此先介绍矢量和形式的缩并。

定义 2.56 (内缩)。线性运算 $i_{(\cdot)}(\cdot) : V \times \Lambda_k(V) \rightarrow \Lambda_{k-1}(V)$ 称为内缩、内乘或内导数，用以指代矢量和形式以此顺序缩并：

$$i_v \omega = C_1^1(v \otimes \omega), \quad (2.1.63)$$

即对任意 $(k-1)$ 个矢量 Y_i 有

$$i_v \omega(Y_1, \dots, Y_{k-1}) = \omega(v, Y_1, \dots, Y_{k-1}).$$

规定函数的内缩为零。

定理 2.1.23. 设 k 个对偶矢量 a^i ，则对任意 $v \in V$ 有

$$i_v(a^1 \wedge \cdots \wedge a^k) = \sum_{i=1}^k (-1)^{i-1} a^i(v) a^1 \wedge \cdots \wedge \widehat{a^i} \wedge \cdots \wedge a^k, \quad (2.1.64)$$

这里 a^i 头顶上的 $\widehat{}$ 表示将 a^i 从楔积中删去²⁷。

证。作用于任意 $v_2, \dots, v_k \in V$ ，不妨记 $v_1 = v$ ，有

$$\begin{aligned} i_{v_1}(a^1 \wedge \cdots \wedge a^k)(v_2, \dots, v_k) &= (a^1 \wedge \cdots \wedge a^k)(v_1, v_2, \dots, v_k) \\ &= \text{alt}(a^1 \otimes \cdots \otimes a^k)(v_1, \dots, v_k) \\ &= \sum_{\sigma \in S_k} \text{sgn } \sigma a^1(v_{\sigma(1)}) \cdots a^k(v_{\sigma(k)}) \\ &= \det(a^i(v_j)) \\ &= \sum_{i=1}^k (-1)^{i+1} a^i(v_1) \det(a^\mu(v_j)), \end{aligned}$$

这里 $\det(a^\mu(v_j))$ 表示余子式，要求 $1 \leq \mu \leq k, \mu \neq i$ 且 $2 \leq \mu \leq k$ 。因此余子式就是 $(a^1 \wedge \cdots \wedge \widehat{a^i} \wedge \cdots \wedge a^k)(v_2, \dots, v_k)$ ，而 $(-1)^{i+1} = (-1)^{i-1}$ 。 \square

²⁷ 读作脱字号 (caret)。

定理 2.1.24 (诱导体元). 设 D 是背景 M 的 n 维积分域, 则 M (进而 D) 的体元 ε 在 ∂D 上的诱导体元为

$$\tilde{\varepsilon} = i_n \varepsilon, \quad (2.1.65)$$

这里 n 是外指 (*outward-pointing*) 法矢场 (以 $i(D)$ 为内部).

证. $\forall q \in \partial D$, 设 $\{e_\mu\}$ 是 q 的正交基, 满足 $e_1 = n$, 则

$$\varepsilon = \pm n_b \wedge e^2 \wedge \cdots \wedge e^n,$$

其中 $+, -$ 号对应于右手、左手基. 注意 $\{e^2, \dots, e^n\}$ 可以看作 ∂D 上 q 点的局部坐标系的坐标基, 规定诱导定向应将坐标系衡量为右手系, 则可以取 $\bar{\varepsilon} = a e^2 \wedge \cdots \wedge e^n$. 但由于其只给出“正负”方向, 因此系数选择更为任意, 然而为保证诱导定向和诱导体元相容 (满足适配体元定义), 只差一个正系数, 干脆选择 $a = 1$, 则

$$\bar{\varepsilon} = e^2 \wedge \cdots \wedge e^n,$$

故

$$\varepsilon = \pm n_b \wedge \bar{\varepsilon},$$

由此可证 $\bar{\varepsilon} = i_n \varepsilon$, 可见系数选择使得定向、体元直接“相等”. 按照之前的思路, 可以验证 $\tilde{\varepsilon}^\sharp(\tilde{\varepsilon}) = (-1)^{\tilde{s}}(n-1)!$. 注意, 这只能把 $\tilde{\varepsilon}$ 确定到差正负号的程度, 只有在与诱导定向相容后, 才能完全确定. \square

以后可不再区分诱导定向、体元的记号.

例 2.8. 以 $\{(x^1, \dots, x^n) | x^n \leq 0\} \subset \mathbf{R}^n$ 举例. 则 $dx^1 \wedge \cdots \wedge dx^n$ 在超平面 $x^n = 0$ 上的一个诱导定向是

$$\begin{aligned} i_{\partial/\partial x^n}(dx^1 \wedge \cdots \wedge dx^n) &= (-1)^{n-1} i_{\partial/\partial x^n}(dx^n) dx^1 \wedge \cdots \wedge dx^{n-1} \\ &= (-1)^{n-1} dx^1 \wedge \cdots \wedge dx^{n-1}, \end{aligned}$$

例 2.9. 以 \mathbf{R}^3 中的 2 维单位球面 $S = \{(x^1, x^2, x^3) | \delta_{ij} x^i x^j = 1\}$ 举例, 显然存在一个法矢场 $x^i \frac{\partial}{\partial x^i}$. 则定向 3-形式 $\varepsilon = dx^1 \wedge dx^2 \wedge dx^3$ 的一个诱导 2-形式为

$$\begin{aligned} \tilde{\varepsilon} &= i_{x^i \frac{\partial}{\partial x^i}}(dx^1 \wedge dx^2 \wedge dx^3) \\ &= x^1 dx^2 \wedge dx^3 + x^2 dx^3 \wedge dx^1 + x^3 dx^1 \wedge dx^2, \end{aligned}$$

同理 \mathbf{R}^2 给回路的诱导正定就是绕 z 轴右手螺旋 (逆时针). 可见这些都与分析学中的习惯一致.

读者还可验证: 3 维欧氏度规用球坐标系的对偶坐标基展开后, 球面度规确为诱导度规, 只需注意球坐标的单位外法矢 $\frac{\partial}{\partial r}$.

定理 2.1.25 (Stokes). Ω 是背景 M 的一个 n 维带边区域, 则对 $\forall \omega \in \Lambda_{n-1}(M)$ 有

$$\int_{\Omega} d\omega = \int_{\partial\Omega} \omega, \quad (2.1.66)$$

证. 一般背景上需要处理一些技术性细节, 为此不妨模仿分析学的做法, 设想带边区域靠着一个个“微小单位”粘贴出来, 这样可供在整体的场上积分, 于是问题的关键在于内点邻域、边界点邻域, 此二情况分别可以近似于 \mathbf{R}^n 和 \mathbb{H}^n , 不妨设 $\{(x^1, \dots, x^n), x^n \geq 0\}$.

我们先从半空间开始. 微小邻域总是有界的 (理解为背景里存在一个开的球区域包含它), 边界点邻域内的坐标总能用 $A = [-R, R]^{n-1} \times [0, R]$ 囊括, 这里 R 是有限正数, 使得 $(n-1)$ -形式 ω 在 A 边缘及以外为零. 将 ω 展开为

$$\omega = \sum_{i=1}^n \omega_i dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n,$$

则

$$\begin{aligned} d\omega &= \sum_{i,j=1}^n \frac{\partial \omega_i}{\partial x^j} dx^j \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n \\ &= \sum_{i=1}^n (-1)^{i-1} \frac{\partial \omega_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^n. \end{aligned}$$

证明核心仍是读者最为熟知的 Newton-Leibniz 公式. 由分析学的 Fubini 定理,

$$\begin{aligned} \int_{\mathbb{H}^n} d\omega &= \sum_{i=1}^n (-1)^{i-1} \int_A \frac{\partial \omega_i}{\partial x^i} dx^1 \wedge \cdots \wedge dx^n \\ &= \sum_{i=1}^n (-1)^{i-1} \int_0^R \int_{-R}^R \cdots \int_{-R}^R \frac{\partial \omega_i}{\partial x^i} dx^1 \cdots dx^n, \end{aligned}$$

由于 ω 在边界处为零, 这意味着除 $i = n$ 外任意 i 有 $\int_{-R}^R \frac{\partial \omega_i}{\partial x^i} dx^i = 0$, 则最后只剩下第 n 项:

$$\begin{aligned} \int_{\mathbb{H}^n} d\omega &= (-1)^{n-1} \int_0^R \int_{-R}^R \cdots \int_{-R}^R \frac{\partial \omega_n}{\partial x^n} dx^1 \cdots dx^n \\ &= (-1)^{n-1} \int_{-R}^R \cdots \int_{-R}^R \int_0^R \frac{\partial \omega_n}{\partial x^n} dx^n dx^1 \cdots dx^{n-1} \\ &= (-1)^n \int_{-R}^R \cdots \int_{-R}^R \omega_n(x^1, \dots, x^{n-1}, 0) dx^1 \cdots dx^{n-1}; \end{aligned}$$

记 $\partial\mathbb{H}^n = \{(x^1, \dots, x^n), x^n = 0\}$, 则另一方面

$$\int_{\partial\mathbb{H}^n} \omega = \sum_{i=1}^n \int_{A \cap \partial\mathbb{H}^n} \omega_i(x^1, \dots, x^{n-1}, 0) dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n,$$

$x^n = 0$ 意味着没了 dx^n , 故上式也只剩下

$$\int_{\partial \mathbb{H}^n} \omega = \int_{A \cap \partial \mathbb{H}^n} \omega_i(x^1, \dots, x^{n-1}, 0) dx^1 \wedge \dots \wedge dx^{n-1},$$

考虑到 n 为偶数时, 坐标系 (x^1, \dots, x^{n-1}) 正定 (否则为负定), 二者便是相等的.

再考虑 \mathbf{R}^n , 这样 ω 的邻域就囊括于 $A = [-R, R]^n$ 中, 那任意 i 有 $\int_{-R}^R \frac{\partial \omega_i}{\partial x^i} dx^i = 0$, 故

$$\int_{\mathbf{R}^n} d\omega = 0;$$

由于 \mathbf{R}^n 没边界, 即 $\partial \mathbf{R}^n = \emptyset$, 故

$$\int_{\partial \mathbf{R}^n} \omega = 0.$$

每个邻域都有局部坐标系, 而内部邻域没有贡献, 这样总体的 $\int_{\Omega} d\omega$ 就只剩下表面的变化 $\int_{\partial \Omega} \omega$. 证明的严格术语将在第四章介绍. \square

积分域 Ω 亦可写作 $i(\Omega)$, 因为 $\int_{\partial \Omega} d\omega = \int_{\Omega} dd\omega = 0$, 这里 $dd\omega = 0$ 可看作任意阶形式. 当然, 我们只定义了同阶形式在同维区域的积分, 故此亦可视作规定不同维时积分为零. 可做对照: $\int_{\partial \partial \Omega} \omega = \int_{\partial \Omega} d\omega = 0$, 这里不妨设 ω 为 $(l-2)$ -形式场而 Ω 为 l 维. 这对任意 ω 成立, 因而可能性只有 $\partial^2 \Omega = \emptyset$. 如何从直观上理解这件事? 我们知道默认 Ω 为任意连通闭集, 这应当使连通开集 $i(\Omega)$ 有界, 因而 $\partial \Omega$ 有界, 且 $\partial^2 \Omega = \emptyset$ 还意味着它无边²⁸! 如此便可确定 $\partial \Omega$ 的内外部, 因而有外指法矢的明确含义, 故从背景 M 可立即诱导出边界的定向.

定理 2.1.26 (Green 公式). Green 公式是 2 维平面上的 Stokes 公式. 设 \vec{A} 是 \mathbf{R}^2 上的矢量场, L 是 \mathbf{R}^2 上的闭合回路, S 是 L 包围的开集. 建立正交坐标系 $\{x^1, x^2\}$, 则

$$\int_S \left(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2} \right) dx^1 dx^2 = \int_L A_1 dx^1 + A_2 dx^2. \quad (2.1.67)$$

证. $S \cup L$ 可充当 Stokes 定理的 Ω , 其中 S 和 L 分别充当 $i(\Omega)$ 和 $\partial \Omega$. 虽然这里计算积分并不需要度规, 但考虑到实际情形是个 \mathbf{R}^2 上的矢量场 \vec{A} , 不妨用欧氏度规把 \vec{A} 变为 1 形式场 $A \equiv I_b(\vec{A})$, 则 A 可充当定理中的 ω . 把 A 用对偶坐标基展开为 $A_\mu dx^\mu$, 则

$$\begin{aligned} d\omega &= dA_\mu \wedge dx^\mu = \frac{\partial A_\mu}{\partial x^\nu} dx^\nu \wedge dx^\mu = \frac{\partial A_1}{\partial x^2} dx^2 \wedge dx^1 + \frac{\partial A_2}{\partial x^1} dx^1 \wedge dx^2 \\ &= \left(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2} \right) dx^1 \wedge dx^2, \end{aligned}$$

²⁸有界无边的流形有时可称闭流形 (取闭合之意), 请注意同闭集等名称区分.

所以 $\int_S \left(\frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2} \right) dx^1 dx^2$ 可表为 $\int_{i(\Omega)} d\omega$; 另一方面, $\int_{\partial\Omega} \omega = \int_L A_1 dx^1 + A_2 dx^2$, 证毕.

下面再用诱导形式更严格地叙述之. 选线长 ℓ 为 L 的局部坐标, 把 $\tilde{\omega}$ 用坐标基展为 $\tilde{\omega} = \tilde{\omega}_1(\ell) d\ell$, 两边作用于 $\partial/\partial\ell$ 得

$$\tilde{\omega}_1(\ell) = \tilde{\omega}(\partial/\partial\ell) = \omega(\partial/\partial\ell) = A(\partial/\partial\ell) = A_\ell,$$

故 $\tilde{\omega} = A_\ell d\ell$, 于是亦可写为

$$\int_L A_\ell d\ell = \int_L \omega,$$

上式推导中有一点要讲清楚. $\int_L \tilde{\omega}$ 的积分域是闭合曲线 L , 至少要用两个坐标域覆盖, 因此应先对每个坐标域做局部积分再“缝合”. 幸好可做这样的简单处理: 设 L' 是 L 挖去一点所得, 则它用一个坐标域即可覆盖, 但挖去一点不影响积分值. \square

这些常见的多元微积分公式亦可用对偶形式的语言证明, 不再赘述.

以后多用到高维的 Gauss 定理. 回忆一下, 设 V 是 \mathbf{R}^3 的有界区域, 边界 S 是闭合曲面, \vec{A} 是空间内的矢量场, \vec{n} 是曲面的正定向法矢场, 则 Gauss 公式是指

$$\int_V \nabla \cdot \vec{A} dV = \int_S \vec{A} \cdot \vec{n} dS. \quad (2.1.68)$$

其特点是, 左右两边可分别形象地说成是“函数 $\nabla \cdot \vec{A}$ 与体元 dV 的乘积的积分”和“函数 $\vec{A} \cdot \vec{n}$ 与面元 (2 维体元) dS 的乘积的积分”. 下面从 Stokes 定理可导出专门相应于 Gauss 公式的推广.

定理 2.1.27 (Gauss). 设 M 是 n 维定向背景, N 是 M 中的 n 维带边区域, g 是 M 上的度规, ε 是适配体元, n 是 ∂N 的外法矢场, $\tilde{\varepsilon} = i_n \varepsilon$ 是诱导体元, v 是 M 上的可导矢量场, 则

$$\int_{i(N)} (\operatorname{div} v) \varepsilon = \int_{\partial N} i_v \varepsilon = \pm \int_{\partial N} \langle v, n \rangle \tilde{\varepsilon}, \quad (2.1.69)$$

证. 先证 $i_v \varepsilon = \pm \langle v, n \rangle \tilde{\varepsilon}$. 将 v 沿着法向和切向分解, 得

$$\begin{aligned} i_v \varepsilon &= i_{v_n} \varepsilon + i_{v_t} \varepsilon \\ &= \pm \langle v, n \rangle i_n \varepsilon + i_{v_t} \varepsilon, \end{aligned}$$

故只需证 $i_{v_t} \varepsilon = 0$, 这是显然的, 切于 $(n-1)$ 维 ∂N 的矢量构成 $(n-1)$ 维切空间, 其中任意矢量 X_1, \dots, X_{n-1} 再配上 v_t 总是线性相关的, 有

$$i_{v_t} \varepsilon(X_1, \dots, X_{n-1}) = \varepsilon(v_t, X_1, \dots, X_{n-1}) = 0,$$

这里用了 ε 的反称性. 再证 $di_v \varepsilon = \operatorname{div} v \varepsilon$. 不失一般性, 正交系下有

$$\begin{aligned} di_v(dx^1 \wedge \cdots \wedge dx^n) &= \sum_{i=1}^n (-1)^{i-1} d(dx^i(v)) \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n \\ &= \sum_{i=1}^n (-1)^{i-1} \frac{\partial v^i}{\partial x^i} dx^i \wedge dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n \\ &= \sum_{\mu=1}^n \frac{\partial v^\mu}{\partial x^\mu} dx^1 \wedge \cdots \wedge dx^n, \end{aligned}$$

此即 $d(i_v \varepsilon) = (\operatorname{div} v) \varepsilon$. \square

例 2.10. 整个平直时空 \mathbf{R}^4 能动守恒.

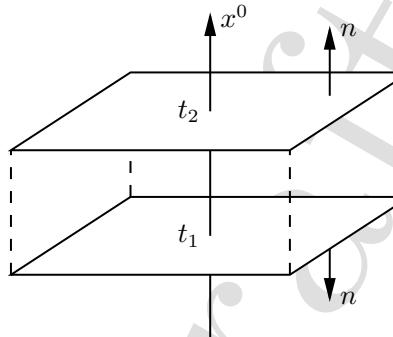


图 2.9: 世界管 (压缩一个空间维度)

证. 时空是 4 维的, “扔掉” 时间就得到了某时刻所有同时事件所构成的 3 维空间, 这张超曲面就称作同时面 (*simultaneity surface*). 将 $T_{\mu\nu}$ 在两个同时面上积分, 目标即证:

$$\int_{x^0=t_1} T_{0\alpha} d^3x = \int_{x^0=t_2} T_{0\alpha} d^3x, \quad (2.1.70)$$

该式在 $\alpha = 0$ 时就是能量守恒, 在其它情况下就是动量守恒. 于是应考虑对 t_1, t_2 间应用 Gauss 定理. 平直时空可定义全局的正交坐标系. 对任意选定的正交系, 协变散度取为普通偏导算符, 即 $\partial^\mu T_{\mu\alpha} = 0$, 由 Gauss 定理知

$$\int_{\partial N} T_{\mu\alpha} n^\mu \tilde{\varepsilon} = 0, \quad (2.1.71)$$

这里用 \mathbf{R}^3 代表若干同时面²⁹, 则 $N = [t_1, t_2] \times \mathbf{R}^3$. ∂N 类似于微分满足 Leibniz 律. 由于 $\partial \mathbf{R}^3 = \emptyset$ (直观上代表管没有侧壁, 管口无限大), 则只有 $\partial[t_1, t_2] \times \mathbf{R}^3$. 而

²⁹准确说是世界管上的 3 维截面, 但现在我们将世界管取为了整个时空

$\partial[t_1, t_2]$ 其实就是两个点 $\{t_2\}, \{t_1\}$, 和 \mathbf{R}^3 一乘就是各自的同时面了 (直观上代表管的两端). 此即

$$\int_{x^0=t_1} T_{\mu\alpha} n^\mu \tilde{\varepsilon} + \int_{x^0=t_2} T_{\mu\alpha} n^\mu \tilde{\varepsilon} = 0,$$

这里后者 $n = \frac{\partial}{\partial x^0}$, 因而分量只有 $n^0 = 1$, 且由 $\varepsilon = dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3$, 得 $\tilde{\varepsilon} = dx^1 \wedge dx^2 \wedge dx^3$; 前者是 $n = -\frac{\partial}{\partial x^0}$, 因而 $n^0 = -1, \tilde{\varepsilon} = -dx^1 \wedge dx^2 \wedge dx^3$, 故

$$\int_{x^0=t_1} T_{0\alpha} dx^1 \wedge dx^2 \wedge dx^3 + \int_{x^0=t_2} T_{0\alpha} dx^1 \wedge dx^2 \wedge dx^3 = 0,$$

但 t_1 同时面上 $dx^1 \wedge dx^2 \wedge dx^3$ 按其 $\tilde{\varepsilon} = -dx^1 \wedge dx^2 \wedge dx^3$ 衡量是左手系, 写成 Riemann 积分时要填上负号, 故有 (2.1.70). \square

2.2 时空结构

2.2.1 时空的几何学

熟悉这些概念后, 可给闵氏时空一个严格的数学表达:

定义 2.57 (闵氏时空). 背景 \mathbf{R}^4 中一点 q 处的闵氏度规是该处矢量空间 \mathbf{R}_q^4 上的 2 阶协变张量 $\eta : \mathbf{R}_q^4 \times \mathbf{R}_q^4 \rightarrow \mathbf{R}$, 如果 $\{e_\mu\}$ 是 \mathbf{R}_q^4 的坐标基, 则

$$\eta(e_\mu, e_\nu) = \eta_{\mu\nu}.$$

闵氏时空 \mathbf{R}^{3+1} 是背景 \mathbf{R}^4 再处处配上一个闵氏度规, 或者说直接配上了个闵氏度规场. 有时我们写作 (\mathbf{R}^4, η) .

可见坐标基在 η 的衡量下是正交基。

定义 2.58 (矢量分类与光锥). 取号差 +2. 如果 v 的“平方” $v \cdot v$ 为正、负或零, 那就分别称为类空 (spacelike)、类时 (timelike) 或类光 (null) 矢量. 所有类光矢量之集 C_N 称为矢量空间 \mathbf{R}^4 的光锥 (light cone). 号差 -2 不外乎正负定义颠倒之.

设 $v = v^\mu e_\mu, w = w^\nu e_\nu$ 为坐标基下的展开, 立即有 $v \cdot w = \eta_{\mu\nu} v^\mu w^\nu$. 因此类时矢量位于光锥“内部”, 类空矢量则位于光锥“外部”.

定理 2.2.1. 从 \mathbf{R}^3 的 Cauchy 不等式可以看到, 若 v 类时, 而 w 是非零的类时或类光矢量, 那么当且仅当在任意正交基中满足 $v^0 w^0 > 0$ 时, 有 $v \cdot w < 0$.

这种 v, w 可在光锥内部定义一种等价关系, 称 v, w 具有相同的时间指向. 存在两种等价方式, 一种是选择并指定指向未来 (future directed) 的矢量, 那另一种方式的类时矢量就称为指向过去 (past directed) 的. 这种分类可以推广到非零类光矢量,

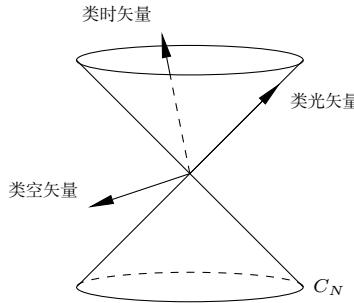


图 2.10: 光锥

但不能推广到类空矢量。任一时空点 p 的全体类时和类光矢量（除零元）可分为指向未来和指向过去两大类。无论弯曲还是平直时空，一点的切空间并无区别（无非是 4 维矢量空间配以度规 η ），因此弯曲时空中任一点的类时和类光矢量（除零元）的集合也可类似地分为两大部分。若只孤立地讨论一点 p ，可任意指定其中一个部分为指向未来部分 \tilde{F}_p ，其中每一元素称为指向未来的类时（或类光）矢量，另一部分 \tilde{P}_p 的元素则称为指向过去的类时（或类光）矢量。此二部分不仅是 V_p 的子集，更是对 C_N 及其内部的划分。当然，在讨论全时空时，物理上总希望这种指定在从一个时空点到另一时空点的过渡中是连续的（闵氏时空就如此）。但并非所有时空都能做到这一点。

例 2.11. 考虑一个“圆柱面时空” $S^1 \times \mathbf{R}$ ，设其上闵氏度规这样给定：沿着任意 S^1 一圈，其上光锥将均匀地旋转 180° 。这样便无法连续地指定 \tilde{F} ，因为若尝试按此指定，则一定存在某母线 \mathbf{R} 上的 \tilde{F} 有 180° 的突变。不妨认为这样的时空没有物理意义。

能连续地指定光锥未来部分的时空叫时间可定向时空 (*time orientable spacetime*)。今后谈到时空都指时间可定向时空，并认为每一时空都已作了这样的连续指定。设时空存在一个连续的类时矢量场，就可把其在每点的值所在部分指定为指向未来部分，进而使得指定是连续的。因此，存在连续类时矢量场的时空一定是时间可定向时空，反之亦然³⁰。

有确定指向的正交基称为时间定向的，若其类时矢量 e_0 指向未来——而后将始终只考虑这种定向的正交基。在矢量空间 \mathbf{R}^4_q 上一个具有确定指向的、时间定向的正交基称为惯性基矢 (*frame*)、惯性标架 (*tetrad*) 或者惯性观者，由于矢量空间和背景本身的等同，我们也可以把这些矢量都“搬”到时空里来。

定义 2.59 (正齐次 Lorentz 变换). 设 $\{e_\mu\}$ 和 $\{\hat{e}_\nu\}$ 是两个相同时间定向的惯性基矢，之间满足 $e_\mu = \Lambda_\mu^\nu \hat{e}_\nu$ ，容易验证，矩阵 Λ 满足：

³⁰ 证明需要第 4 章的知识，且可见 Penrose(1972)。

(保度规或正交性) $\Lambda^T \eta \Lambda = \eta$;

(时空定向或固有性) $\det \Lambda = 1$;

(时间定向或正时性) $\Lambda^0_0 \geq 1$.

把任意满足上述三个条件的四阶方阵 Λ 称为正齐次 *Lorentz* 变换，一般简称为 *Lorentz* 变换。

全体这样的矩阵之集结合矩阵乘法运算，就构成了闵氏几何学的所谓对称群，可称正齐次 *Lorentz* 群，或简称 *Lorentz* 群，记作 L ；如果只有保度规即正交的要求，相当于包含了时空反演，称为齐次 *Lorentz* 群³¹。“群”这个概念相比于集合稍微多了一种元素间的乘法运算，用以更方便地描述对称性。现严格叙述之：

定义 2.60 (群). 群是指集合 G 配上映射（称为群乘法或群法则） $G \times G \rightarrow G$ ，元素记作 $(x, y) \mapsto xy$ ，满足

(结合律) 对 $\forall x, y, z \in G$, $(xy)z = x(yz)$.

(单位元) $\exists 1$ 使得 $\forall x \in G$, $1x = x1 = x$.

(逆元) $\forall x \in G$ 对应有一个 x^{-1} ，即 $xx^{-1} = x^{-1}x = 1$.

这里把群法则表示为群内的乘法。如果群是 *Abel* 的，即对 $\forall x, y \in G$ 有交换律 $xy = yx$ ，则也可把群法则表示为通常意义上的加法。在这种情况下，可以把条件写成 $x + y = y + x$ ，并称单位元（或中性元）为 0，逆元即填个负号。

由保度规条件很容易证明，若 $\Lambda^0_0 = 1$ ，那 Λ 必须是如下形式：

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & (R^i_j) & \\ 0 & & & \end{pmatrix},$$

其中 (R^i_j) 是行列式为 1 的 3×3 正交矩阵，它所构成的集合 R （再配上矩阵乘法）是 L 的一个子群，称为空间转动群³²。应指出，在诸多领域中（如粒子物理）须考虑由 *Lorentz* 群和时空平移 ($x^\mu \rightarrow x^\mu + C^\mu$ ，其中 C^μ 是常矢量) 在闵氏时空中生成的更大的变换群，称为非齐次 *Lorentz* 群或 *Poincaré* 群。当然，这些概念可能对本节所述无用，不过为读者以后的学习考虑，分清这些也有必要。

2.2.2 时空的物理学

狭相讨论的是闵氏时空中的物理学。这种情形下，若想比较不同起点的矢量，可以选择平移至一点，从而不同点处的矢量空间可不作区分，这个时候所有的矢量空间

³¹又称广义正交群，在号差 +2 下记作 $O(3, 1)$ ，前者指度规正项个数，后者指负项个数。我们常说号差不改变实质，这可以体现在群理论中：可以证明其与 $O(1, 3)$ 是同构的。

³²行列式为 1 意味着旋转变换将保持对象的原样，也称作 $SO(3)$ 。类似地， L 有时记作 $SO(3, 1)$ 。

“都一样”，而且还可以和时空本身等同。但是，为从物理上想像闵氏时空和 Lorentz 群，最好还是在直观和术语上把矢量空间 \mathbf{R}_q^4 的“矢量”和背景 \mathbf{R}^4 的“点”（矢量一端）区分开来，这样我们还是要讲清矢量的起点在哪儿。

定义 2.61 (事件). \mathbf{R}^{3+1} 中的点称为事件。

虽然作为理想化的“点事件”没有空间尺度也没有时间流逝，但还是应尽量将其想像为真实“发生”的事件。例如，一种物理图像是瞬时碰撞，或瞬时爆炸，或某种点粒子或光子运动历程中的“瞬间”。

平直时空保证我们可将一点矢量空间的光锥搬到时空里来，因而可在时空全局地给任意时空点 p 分配其对应的“过去点”和“未来点”。但在更加任意的时空中，光锥只能帮助我们区分附近的过去和未来，要想确定更大范围的点是 p 的过去还是未来，我们需要研究它们之间的连线。

前面看到，粒子的事件在时空中划出一条曲线，称为相应粒子的世界线。回忆一下，若 $I \subseteq R$ 是一种参数区间，映射 $\alpha : I \rightarrow \mathbf{R}^{3+1}$ 是 \mathbf{R}^{3+1} 中的一条曲线。在惯性坐标下，对任意 $\xi \in I$ ，我们可以写下 $\alpha(\xi)$ 的坐标为 $x^\mu(\xi)$ 。要求曲线 α 至少可导，即 $x^\mu(\xi)$ 在 I 上是可导的，且速度矢量

$$\alpha'(\xi) = \frac{dx^\mu}{d\xi} e_\mu \quad (2.2.1)$$

对任意 $\xi \in I$ 不为零。这种可导性的定义与 \mathbf{R}^{3+1} 中惯性坐标基矢的选取无关。若对任意 $\xi \in I$ ，

$$\alpha'(\xi) \cdot \alpha'(\xi) = \eta_{\mu\nu} \frac{dx^\mu}{d\xi} \frac{dx^\nu}{d\xi} \quad (2.2.2)$$

是正、负或者零，则曲线 α 分别称为类空、类时或类光线。

定义 2.62 (类时世界线). 若一条类时线 α 在任意 $\xi \in I$ 上的 $\alpha'(\xi)$ 均指向未来，则这条曲线称为类时世界线（或指向未来类时线），而且它的像与某个质点历程中的全部事件之集等同。

由于是未来指向，即 $\langle d\alpha/d\xi, e_0 \rangle < 0$ ，那其实亦即 $dx^0/d\xi > 0$ ，我们默认参数是这样选择的。准确来说，还要规定参数是均匀的（仿射参数），甚至是归一的（固有时，即线长参数），这样才能成为物理意义上的观者。这涉及对钟的问题，下一节立即谈到。

定义 2.63 (因果线). 若曲线 γ 上任意点切矢是指向未来的类时或类光矢量（含零元），该线就成为了指向未来因果线 (*future directed causal curve*)。其他定义类似。

因果线之所以要包含类光情况，是因为因果关系亦可按光速传递并影响，比如在光缆中传递的信号。然而，要想借助具有质量的载体传递信息，速度就是小于光速的，这一结论称为相对论限制，其准确的几何语言为：

定理 2.2.2 (相对论限制). 有质粒子走类时世界线.

证. 对任意质点, 总存在与之瞬时共速的参考系 (见 2.4 节), 使得其中质点瞬时相对静止, 这样世界线在这一瞬间的切矢就是严格类时的。Lorentz 变换的保度规性使得矢量模长不会变号, 因而类时矢量在变换下仍是类时矢量, 不可能变成类光, 更不可能超出光锥。□

定义 2.64 (未来). 时空点 p 的时序未来 (*chronological future*) 定义为集合 $I^+(p)$, 其元素 q 满足: 存在从 p 到 q 的类时世界线。因果未来 (*causal future*) $J^+(p)$ 只需再包括类光情况 (含零元)。相应的过去集合符号将 + 改为 -。

定义 2.65 (因果菱形). 从 p 到 q 的因果菱形 (*causal diamond*) 或因果核 D_p^q 是指 $J^+(p) \cap J^-(q)$ 。

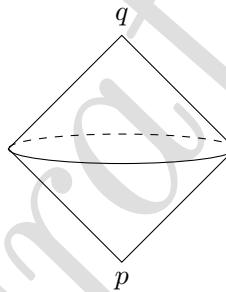


图 2.11: 3 维平直时空的因果菱形。可见其形似一颗“钻石”, 而其 2 维截面为“菱形”状 (准确说呈平行四边形)。英语世界皆以“diamond”代之。

定理 2.2.3. 从 p 到 q 的所有指向未来因果线在时空中扫出的区域含于 D_p^q 。

证. 假设 p, q 间存在并不完全落在 D_p^q 中的因果线, 不落在 D_p^q 的点仍同 p, q 间存在因果线, 因而又属于 $J^+(p), J^-(q)$, 矛盾。□

要想完全仅用可导曲线扫出 D_p^q 是不可能的, 因为很明显 ∂D_p^q 就是不可导的, 但它确实似乎是一条物理经验上“允许”的路径, 尽管这样拐角处加速度得“无穷大”才可使速度突变³³。在现实情况中, 实际刚体的分子微扰将反而使近似宏观现象可导, 甚至是光滑的, 因而很多情况下可直接假定模型具有极其优良的性质, 只不过这不太符合数学家的癖好。少数情况, 如讨论时空奇点时, 则有必要讨论许多极端情形。下面给出 Penrose 所发明的一个有用概念:

³³读者可证明这种类似于“碰撞”的理想模型的加速度符合 Dirac δ 分布, 过程只涉及普通物理学。

定义 2.66 (旅程). 一个旅程 (*trip*) 或旅途就是一条“分段类时测地线”，即由有限段的测地世界线连接而成的连续曲线。这样连接点处允许不可导。“分段因果测地线”称为因果旅程³⁴。

可以证明，如果用分段的类时测地线替代原来的类时线，给出的 $I^+(p), J^+(p)$ 等概念仍是等价的³⁵。此后要想讨论一般的类时世界线，就可直接替代为旅程（指向未来因果线对应替换为因果旅程），这将大大地简化各类话题的数学细节。总之，可导曲线只能扫出 D_p^q 的内部开集，替换为旅程概念后， D_p^q 就可完全扫出。

2.3 惯性运动

2.3.1 对钟

下面我们所感兴趣的，是由一类惯性观者以一种特别方式指定的事件坐标。惯性观者处于 3 维右手空间坐标系（坐标基由所选坐标系产生），光子相对于该坐标系在任意方向上沿直线运动。坐标系原点放有一只时钟，观者可通过如下的 *Fizeau* 对钟程序确定真空光速 c : 时钟读数为 t_1 时，从原点发射一个光子，然后由放置在 (x^1, x^2, x^3) 的镜子反射回来；时钟读数为 t_2 时，光子回到原点，测得

$$c = \frac{2\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}}{t_2 - t_1}.$$

目前观者只有原点一只时钟，因此只能观测原点处发生的事件。为了能够观察到一定范围内的情况，观者需要建立一个参考系 (*reference frame*)。这相当于在坐标系里每个空间点上，都放置一只走时全同的时钟。如何保证它们起点一致呢？可以通过从原点发射的球面电磁波（所有方向上都有光子传播）来同步各点时钟，即 (x^1, x^2, x^3) 点接收到这个电磁波时，该处时钟的读数就应调到 $\sqrt{\delta_{ij}x^i x^j}/c$ ，这样就对好了空间中所有的钟。于是观者现在可以在他的坐标系中，给发生在当地的事件指定 3 个空间坐标及 1 个时间坐标。

一个时钟就是一条类时世界线，其固有时就是此观者所配有时钟的读数。观者除了需要是一条可以定义固有时的世界线外，还需要配以标架，也就是那点矢量空间的基底，只不过 e^0 就是世界线的速度，而其余基矢一般选择与其正交，这样就成为了空间矢量。这样来看，所谓的惯性观者就是走 \mathbf{R}^{3+1} 测地线（直线）的标架场³⁶。空间中所有点的时钟都对应有观者标架，因此整个空间处处布满了观者，形成了线

³⁴ 细心的读者或许发现还未用严格术语重新定义测地线。这是因为协变导数等概念要在下一章严格给出，“矢量平行移动不变”这一定义不依赖于度规，而“线长泛函变分为零”则需借助度规，并且还使协变导数同度规适配。暂时认为平直时空中测地线就是平直几何的直线。

³⁵ 早期证明由 Penrose 给出，一旦在第 4 章补充少量知识后，证明将非常通俗。

³⁶ 还要求惯性标架无自转！请见 2.4 节。

江。可以看到, \mathbf{R}^{3+1} 中所有观者的整体线汇, 就可以称为一个参考系. 现严格叙述之:

定义 2.67 (参考系). 时空中的一个矢量场 U 称为一个参考系, 若其每一条积分曲线³⁷对应一个观者的世界线。

如果标架是坐标基, 就相当于选配了沿着线汇的坐标系, 这样坐标系就和参考系如此这般“紧密贴合”。尤其像狭相往往偏好研究(无论是否有相对运动)惯性观者³⁸。狭相中经常可以不区分参考系和坐标系, 毕竟当然不喜欢还给其折腾个旋转了的坐标系出来。一般而言, 一个惯性观者的世界线就视作其坐标系的 x^0 轴。

定义 2.68 (惯性系). 惯性参考系是 \mathbf{R}^{3+1} 的一个测地参考系。惯性坐标系一般就选择处处与参考系标架一致的坐标基。考虑到狭相中二者实在太相似, 笼统称为惯性系未尝不可。

一个实验事实是, 所有惯性系均承认相同的真空光速数值³⁹。观者起初在某种距离单位上测量时间(如时针的轨迹), 距离均匀地走过其时间坐标 t 乘以 c 。于是, 若定义时间坐标为 $x^0 = ct$, 则时空的四个坐标将具有一致量纲。进一步地, 选取单位制使得光速等于无量纲的 1。后文一般将默认这一点。

在时空的数学模型 \mathbf{R}^{3+1} 中, 惯性观者这种极为微妙的概念, 与惯性基矢 $\{e_\mu\}$ 的简单概念是等同的。若 $x \in \mathbf{R}^{3+1}$ 是事件且写成 $x = x^\mu e_\mu$, 这里把坐标基搬到了平直时空里来, 则 (x^1, x^2, x^3) 就是由相应观者确定的 x 的空间分量, 而 x^0 是时间分量。若 $\{\hat{e}_\nu\}$ 是由 Λ 同 $\{e_\mu\}$ 联系在一起的另一组基, 且设 $x = \hat{x}^\nu \hat{e}_\nu$, 则 $\hat{x}^\nu = \Lambda^\nu_\mu x^\mu$ 。因此, Lorentz 变换将两个惯性观者对任意给定事件的测量所分别得到的坐标联系了起来。若 $(\Lambda^\nu_\mu) \in \text{SO}(3)$, 那么这两个观者的差别只在于其各自空间坐标轴取向不同。

另一方面, 对 $\forall \theta \in \mathbf{R}$, 定义 L 的一个特殊元素:

$$\Lambda(\theta) = \begin{pmatrix} \cosh \theta & -\sinh \theta & 0 & 0 \\ -\sinh \theta & \cosh \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

³⁷如果一条曲线处处与该矢量场相切, 就称为该矢量场的一条积分曲线。一点只对应一条积分曲线, 因为倘若对应多条, 矢量场于该点应取哪一切矢呢?

³⁸尽量和 Newton 力学意义的惯性观者相照应(第一定律), 即在所引进的参考系中, 自由粒子之轨迹或是一点, 或是一条以不变速度运动的直线。

³⁹此处应当理解为双程平均真空光速, 而非单向真空光速。至今为止, 单向光速从未被实验测量过, 也不可能实验测量, 除非找到光信号之外的对钟方式。真空光速在各个方向上相同是假定性的原理。

且若两组惯性基由该 Lorentz 变换联系，则坐标变换具体写出来就是

$$\begin{aligned}\hat{x}^0 &= (\cosh \theta)x^0 - (\sinh \theta)x^1, \\ \hat{x}^1 &= -(\sinh \theta)x^0 + (\cosh \theta)x^1, \\ \hat{x}^2 &= x^2, \\ \hat{x}^3 &= x^3.\end{aligned}$$

令 $v = \tanh \theta$ (因而 $|v| < 1$) 得

$$\begin{aligned}\hat{x}^1 &= \frac{1}{\sqrt{1-v^2}}x^1 - \frac{v}{\sqrt{1-v^2}}x^0, \\ \hat{x}^0 &= -\frac{v}{\sqrt{1-v^2}}x^1 + \frac{1}{\sqrt{1-v^2}}x^0.\end{aligned}$$

这里容易看出上述 Lorentz 变换是线性的. 上面的公式，就是狭相中通常所说的 boost 变换，即

$$\begin{aligned}\hat{x} &= \gamma(x - vt), \\ \hat{t} &= \gamma(t - vx),\end{aligned}$$

其中 Lorentz 因子 $\gamma = \frac{1}{\sqrt{1-v^2}}$. 要想研究狭相运动学，只要考虑所谓的 boost 变换 $\Lambda(\theta)$ 足矣. 可以证明，如果 $\Lambda' \in L$ 是任意 Lorentz 变换，那就存在实数 θ 及两个转动 $R_1, R_2 \in SO(3)$ ，使得 $\Lambda' = R_1 \Lambda(\theta) R_2$. 因为 R_1, R_2 不包含相对速度，所以全部的运动学都包含在 $\Lambda(\theta)$ 之中.

现在假定， x, x_0 是 \mathbf{R}^{3+1} 中两个不同事件，并考虑从 x_0 到 x 的平移矢量 $x - x_0$ ，若 $\{e_\mu\}$ 是一个惯性基矢，且定义 $x = x^\mu e_\mu, x_0 = x_0^\nu e_\nu$ ，则 $x - x_0 = (x^\mu - x_0^\mu) e_\mu = \Delta x^\mu e_\mu$. 若 $x - x_0$ 类光，则

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 = (\Delta x^4)^2,$$

即事件 x_0 和 x 之间的空间间隔等于光在事件间的时间间隙中传播的距离. 这对其他任意惯性基矢一样成立，因为 Lorentz 变换是线性变换，它将保持 Lorentz 度规不变. 因而，所有惯性观者都会有一致结论，即 x_0 和 x 可以“由光子联系起来”. 甚至还会一致地看到两个事件中，哪一个是“发射”光子，哪一个是“接受”光子，因为可以证明，非零的类时或类光矢量在任意惯性基矢下，其 4 个坐标符号一致（因为 $\Lambda^0_0 \geq 1$ ）. 所以， $x^0 - x_0^0$ 要么对任意惯性观者为正（即 x_0 发生在 x 之前），要么对任意惯性观者为负（即 x_0 发生在 x 之后）. 由于光子在惯性坐标系中沿直线运动，所以可以定义如下概念：若 $x_0, x \in \mathbf{R}^{3+1}$ 之间的 $x - x_0$ 类光，则在 \mathbf{R}^{3+1} 中过 x 和 x_0 的直线称为光子世界线，这条世界线被看作光子历史中经历的所有事件（包括 x 和 x_0 ）之集. 现在另外假定 $x - x_0$ 是类时矢量. 那么，在任意惯性基矢下有

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 < (\Delta x^4)^2.$$

在这种情况下，可以证明，必存在惯性基矢 $\{\hat{e}_\mu\}$ 使得 $\Delta\hat{x}^i = 0$ ，即存在这样一个观者，在其看来，这两个事件一先一后地发生在空间中的同一位置。设想该位置被某一质点所占据，如放在此处的观者时钟，我们将发现，事件 x_0 和 x 都由该质点“经历”，且 $\sqrt{|g(x - x_0, x - x_0)|}$ 就是由该质点携带的钟所记录的事件时间间隔。对于其他任意惯性观者来说，这个质点看起来是自由而不受“力”的，因为其以不变速度沿直线运动。这使我们可以有如下定义：若 $x_0, x \in \mathbf{R}^{3+1}$ 之间的 $x - x_0$ 类时，则在 \mathbf{R}^{3+1} 中过 x_0 和 x 的直线称为自由质点世界线，且 $\sqrt{|g(x - x_0, x - x_0)|}$ (通常写作 $\tau(x - x_0)$ ，或简写 $\Delta\tau$) 是 x_0 和 x 间的固有时。于是，可以把 $\tau(x - x_0)$ 看成是 $x - x_0$ 的一种“长度”，衡量自身局部或内部演化的物理过程的时间，其由经历了 x_0 和 x 这两个事件的自由粒子所携带的时钟测量。然而，这是一种不同于欧氏几何的长度，因为其满足的不是欧氏三角不等式，而是如下“反转”形式：

定理 2.3.1. 设 x_0, x 和 y 是 \mathbf{R}^{3+1} 中的事件，且使 $y - x$ 和 $x - x_0$ 是具有相同时间取向的类时矢量。则 $y - x_0 = (y - x) + (x - x_0)$ 亦是类时的，且

$$\tau(y - x_0) \geq \tau(y - x) + \tau(x - x_0),$$

当且仅当 $y - x$ 和 $x - x_0$ 线性无关时取等。

这将导致很有意思的结果，如双生子佯谬，稍后将对此介绍。

最后，考虑 $x - x_0$ 类空的情况，这时有

$$(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2 > (\Delta x^4)^2.$$

现在就不存在那种惯性观者了，亦不存在自由粒子（或是光子）能够经历这两个事件。然而，可以证明，给定任意实数 T ，存在惯性基矢 $\{\hat{e}_\mu\}$ 使得 $\Delta\hat{x}^0 = T$ 。即，一些观者判定两个事件同时发生，另一些观者则断言 x_0 先于 x ，剩下的观者则判断相反。这种类型的两个事件的时间先后次序，即原因和结果，没有意义。对于能够断定两个事件同时发生 ($\Delta\hat{x}^0 = 0$) 的那些惯性观者而言， $\sqrt{|g(x - x_0, x - x_0)|}$ 这个量就是它们之间的距离，因而这个量称为 x_0 和 x 之间的固有空间距离（无论 $x - x_0$ 是否类空）。

对任意事件 $x_0, x \in \mathbf{R}^{3+1}$ ， $g(x - x_0, x - x_0)$ 在任意惯性基矢下由

$$\eta_{\mu\nu} \Delta x^\mu \Delta x^\nu$$

给出，并称为 x_0 和 x 之间的间隔。闵氏几何中的这种长度非常像欧氏几何中的勾股定理。然而，依据事件 x_0 和 x 之间的物理关系，这两个事件之间的间隔可以是任意实数。当然，历史上对上文所说的事件间隔有过各种物理解释，这种时空结构是以 Minkowski 的名字命名的，见 [4]。

2.3.2 运动学

狭相所有熟知的运动学效应 (速度叠加、同时的相对性、钟慢、尺缩) 均可以从以下讨论获得。鉴于前面已说过的原因，我们只需要讨论 boost 变换。

设 θ_1 和 θ_2 是两个实数，考虑 boost 变换 $\Lambda(\theta_1)$ 和 $\Lambda(\theta_2)$ 。根据双曲函数的性质，有 $\Lambda(\theta_1)\Lambda(\theta_2) = \Lambda(\theta_1 + \theta_2)$ 。定义 $v_i = \tanh \theta_i, i = 1, 2$ ，以及 $v = \tanh(\theta_1 + \theta_2)$ ，那么根据 $\tanh \theta$ 求和公式有

$$v = \frac{v_1 + v_2}{1 + v_1 v_2}. \quad (2.3.1)$$

其物理解释很简单。设存在 3 个惯性观者，其空间轴的设置如前文所述。若第二个观者相对于第一个的速度是 v_1 ，而第三个相对于第二个的速度是 v_2 ，那第三个相对于第一个的速度就不是经典力学的线性叠加所给出的 $v_1 + v_2$ ，而是 $(v_1 + v_2)/(1 + v_1 v_2)$ 。这就是相对论的速度叠加公式。

如前文所述，若 x_0 和 x 之间的间隔类空，那么这两个事件在某些惯性观者看来是同时的，但在其他观者看来不是。事实上，如果 $\Delta x^0 = 0$ 并且这些观者由 boost 联系在一起，那么 $\Delta \hat{x}^0 = -(v/\sqrt{1-v^2}) \Delta x^1 = -v \Delta \hat{x}^1 \neq 0$ (除非 $v = 0$ ，即没有相对运动)，毕竟 $\Delta \hat{x}^1$ 不可能是零，除非 $x = x_0$ 使 $\Delta \hat{x}^\mu = 0$ 。这种现象称为同时的相对性，我们现在可进一步用几何简单说明这一现象。

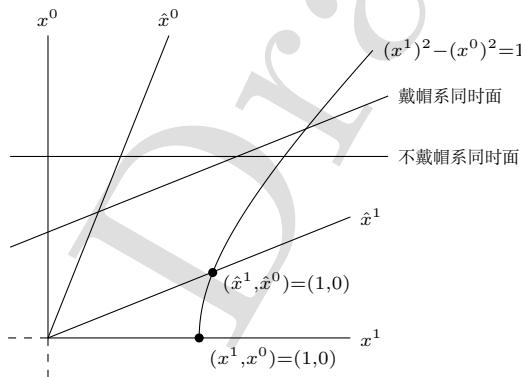


图 2.12: 同时的相对性

在平面中选取两条互相垂直的直线来代表 x^1 轴和 x^0 轴。当然，这两条线的欧氏垂直性没有任何物理上的必要，但更容易想象一些。而 \hat{x}^1 轴即 $\hat{x}^0 = 0$ ，按照 boost 变换，这即是 $x^0 = vx^1$ (不妨假定 $v > 0$)。同理， \hat{x}^0 轴就是 $x^0 = (1/v)x^1$ 。因为 Lorentz 变换保持度规不变，双曲线 $(\Delta x^1)^2 - (\Delta x^0)^2 = k$ 与 $(\Delta \hat{x}^1)^2 - (\Delta \hat{x}^0)^2 = k$ 都可在调整坐标轴标度后，使得比如 $(\Delta x^1)^2 - (\Delta x^0)^2 = 1$ (其中 $x^1 > 1$) 与 x^1 轴交于 $(x^1, x^0) = (1, 0)$ 。可证明在此标度下，所有坐标值均可以作平行于另一轴的直线来投影得到。例如，时间 x^0 坐标和 \hat{x}^0 坐标可以分别做平行于 x^1 轴和 \hat{x}^1 轴的直线

而截得.

因此, 戴帽 (或不戴帽) 坐标的同时性直线平行于 \hat{x}^1 轴 (或 x^1 轴), 所以一般情况下, 在一条同时性直线上的一对事件不会在另一条同时性直线上. 注意, 这些直线实际上是指 3 维超平面, 即所谓同时面, 故我们所说的交点实际是“相应的平面”, 其上任意两个事件都会被这两个观者看做是同时的.

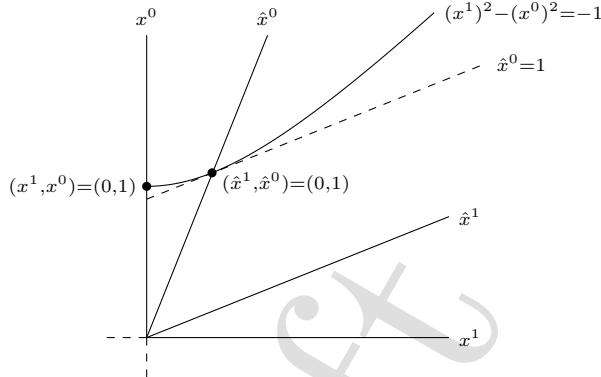


图 2.13: 钟慢效应

无论如何, 对任意两个事件, 戴帽坐标的事件间隔 $\Delta\hat{x}^0$ 与不戴帽坐标的事件间隔 Δx^0 之间的关系由 boost 给出

$$\Delta\hat{x}^0 = -\frac{v}{\sqrt{1-v^2}}\Delta x^1 + \frac{1}{\sqrt{1-v^2}}\Delta x^0,$$

故一般情况下两个时间间隔不等. 例如, 考虑静止在不戴帽坐标系中的原点, 其世界线上的两个事件, 由同一时钟分别测得两个读数. 这种情况 $\Delta x^1 = 0$, 所以

$$\Delta\hat{x}^0 = \frac{1}{\sqrt{1-v^2}}\Delta x^0 > \Delta x^0. \quad (2.3.2)$$

这种效应完全对称, 因为如果 $\Delta\hat{x}^1 = 0$, 那么 boost 给出

$$\Delta x^0 = \frac{1}{\sqrt{1-v^2}}\Delta\hat{x}^0 > \Delta\hat{x}^0.$$

每个观者都认为其他的相对运动的时钟走时慢. 这种现象称为钟慢效应或时间膨胀. 这在几何上一目了然. 例如, 两个观者都看到不戴帽坐标系原点上的时钟读数“0”, 但直线 $\Delta\hat{x}^0 = 1$ 与该时钟世界线, 即 x^0 轴, 在 $(x^1, x^0) = (0, 1)$ 的下方相交.

须强调, 这种现象的物理意义是完全真实的. 例如, 在宇宙射线内发现的介子的寿命是如此之短, 以至于即使以光速运动, 它们穿透大气层所需时间也要比其通常寿命大几十倍. 按理来说, 其不可能到达地面, 但其实还是可以到达, 因为钟慢效应“使其保持年轻”, 即在我们看来, 其实际寿命比通常寿命长很多.

最后, 由于不同的惯性观者一般来说, 对事件的同时性各抒己见, 而测量一运动物体“长度”的唯一方法, 就是要利用一把量尺“同时”测量物体两端; 所以我们不应当诧异, 不同的惯性观者所测量的长度和时间的确会不同. 考虑一把沿戴帽坐标系

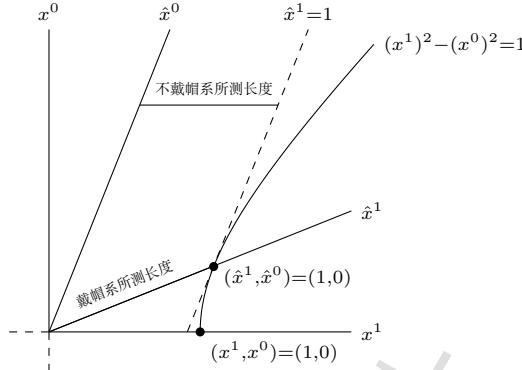


图 2.14: 尺缩效应

\hat{x}^1 轴静止放置的测量尺子. 在这个坐标系中它的“长度”是 $\Delta\hat{x}^1$. 尺子两端的世界线是平行于 \hat{x}^0 轴的两条直线. 若不戴帽观者在这些世界线上同时观测两个事件, 那么他们的坐标将满足 $\Delta x^0 = 0$, 且由 boost 有 $\Delta\hat{x}^1 = (1/\sqrt{1-v^2}) \Delta x^1$, 则

$$\Delta x^1 = \sqrt{1-v^2} \Delta\hat{x}^1 < \Delta\hat{x}^1, \quad (2.3.3)$$

即量尺在其运动方向上缩短了 $\sqrt{1-v^2}$ 倍. 就像钟慢效应那样, 这种被称为尺缩效应或长度收缩的现象完全对称, 且实在, 在几何上亦一目了然.

我们已经看到, 不同的惯性观者对一些特别的事情可以有不同的结论: 两个事件是否同时、对于各个观者来说均不同的两个事件间的时间间隔, 及量尺长度等. 不必关心在任何给定的情况下, 我们是否能够确定“谁”对. 当然, 两个事件要么同时, 要么不是, 我们只需分出哪些惯性观者的观点是正确的. 遗憾的是, 或者侥幸地按某种观点, 不同惯性观者所给出的不同结论在物理上是不允许的.

由于度规在 Lorentz 变换中不变, 所以闵氏时空里的一切几何量在其下也不变, 从而其中的物理学也不变. 利用相对性原理的几何学表述(见 1.3.4 节), 会更容易发现同时的相对性. 在通过 \mathbf{R}^{3+1} 的一点的所有超平面中, 当然有一类对某参考系坐标时 t 为常数的特殊超平面, 也就是同时面. 在一定条件下, 一定有保度规的 Lorentz 变换把这个同时面变成过此点的其他同时面, 所以在这个度规下找不到什么东西能把特定的同时面区别出来. 这与之前复杂的代数计算相一致, 可见几何语言的优越性.

必须清楚, 这并非准确的数学陈述. 如果要对这世界做更准确的陈述, 我们还要清楚, 如何从现象中提炼出描写它的数学方程. 这些现象中, 一些乃纷繁复杂的, 而

另一些则乃平凡的（例如，在相对于地面以不变速度平稳飞行的飞机上，乘客不可能“感觉到”他相对于地球的运动）。相对性原理对我们构造相对论性物理学是一种强有力地指导，但更彻底的观念是，我们应当避免，把任何一种特殊的惯性观者对宇宙的看法当成权威。比如下面这些问题：“两颗超新星是同时爆发的吗？”“介子存活了多长时间？”“蟹状星云和半人马星座 α 星之间有多远？”这些都极其困难，但将来须要面对。

2.3.3 力学

回忆一下，设 $I = [\xi_0, \xi_1]$ ，且 $\alpha : [\xi_0, \xi_1] \rightarrow M$ 是一条类时世界线，则 α 的这一段线长定义为

$$\begin{aligned} L(\alpha) &= \int_{\xi_0}^{\xi_1} \sqrt{|g(\alpha'(\xi), \alpha'(\xi))|} d\xi \\ &= \int_{\xi_0}^{\xi_1} \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\xi} \frac{dx^\nu}{d\xi}} d\xi, \end{aligned}$$

可以理解为 $\alpha(\xi_0)$ 和 $\alpha(\xi_1)$ 的时间间隔，由 α 携带的钟记录。因为积分是 Riemann 求和极限，这与前文的固有时 $\Delta\tau = \sqrt{-\eta_{\mu\nu} \Delta x^\mu \Delta x^\nu}$ 的物理解释一致，但在数学和物理上更加精湛。数学解释由如下结果给出，其类似于三角不等式：

定理 2.3.2. 设 x_0 和 x 是 \mathbf{R}^{3+1} 中的事件， $x - x_0$ 类时且指向未来，当且仅当在 \mathbf{R}^{3+1} 中存在一条类时世界线 $\alpha : [\xi_0, \xi_1] \rightarrow M$ ，其中 $\alpha(\xi_0) = x_0$ 和 $\alpha(\xi_1) = x$ ，且有

$$L(\alpha) \leq \tau(x - x_0),$$

当且仅当 α 是类时直线时取等。

这说明，若两个质点都经历了 x_0 和 x ，则自由的那一个（也即静止在某惯性坐标系中）会需要更长的时间等候第二个事件发生，因为做加速运动的另一时钟整体走时较慢。这一明显的事实在称为“双生子佯谬”。

一条类时世界线具有如下定义的固有时，也即线长参数。

定义 2.69 (固有时). 对 $[\xi_0, \xi_1]$ 中的任意 ξ ，设

$$\tau(\xi) = \int_{\xi_0}^{\xi} \sqrt{|g(\alpha'(\zeta), \alpha'(\zeta))|} d\zeta, \tag{2.3.4}$$

此即 α 从 $\alpha(\xi_0)$ 到 $\alpha(\xi)$ 所经历的固有时。

$\tau = \tau(\xi)$ 存在逆函数 $\xi = \xi(\alpha)$ ，所以 α 可用 τ 来参数化。我将稍微混用记号为

$$\alpha(\tau) = x^\mu(\tau) e_\mu.$$

具有这种参数的速度矢量定义为

$$U = U(\tau) = \frac{dx^\mu}{d\tau} e_\mu,$$

此即世界线的 4-速，而且它还是 α 的单位切矢，即对任意 τ 有归一化

$$U(\tau) \cdot U(\tau) = -1. \quad (2.3.5)$$

定义 2.70 (测地线). 在平直时空任意惯性坐标系中，4-速满足

$$U^\nu \frac{\partial U^\mu}{\partial x^\nu} = 0 \quad (2.3.6)$$

的世界线为（仿射参数化的）测地线。只需代入 $U^\mu = \frac{dx^\mu}{d\tau}$ ，则立即有 $\frac{d^2 x^\mu}{d\tau^2} = 0$ ，说明其为平直时空的直线。结合 $U \cdot U = -1$ 可将仿射参数唯一确定到线长参数。

这里偏导实质是与 η 适配 ($\partial_\sigma \eta_{\mu\nu} = 0$) 的惯性坐标系的协变导数（由于克氏符为零，其成为普通导数算符）。容易验证，在平直时空中，可取任意非惯性坐标系使得测地线表为 $U^\nu \nabla_\nu U^\mu = 0$ ，这里 ∇_ν 是该坐标系相应的协变导数。在弯曲时空中协变导数仍保持其协变性（且其上坐标系一般不能使克氏符全局为零），因此在推广到任意时空时会直接用协变导数定义测地线，这样可绕开线长的定义。

定理 2.3.3 (一点一矢确定). 一点及一矢量唯一决定（仿射参数化的）测地线。

证. $U^\nu \frac{\partial U^\mu}{\partial x^\nu} = 0$ 是二阶常微分方程，由分析学知结论成立。 \square

一个惯性观者更可能用他自己的时间坐标 x^0 对世界线进行参数化，即

$$\alpha'(x^0) = \frac{dx^\mu}{dx^0} e_\mu = \frac{dx^i}{dx^0} e_i + e_0,$$

所以

$$|g(\alpha'(x^0), \alpha'(x^0))| = 1 - v^2,$$

其中速率

$$v = |\vec{v}| = \sqrt{\left(\frac{dx^1}{dx^0}\right)^2 + \left(\frac{dx^2}{dx^0}\right)^2 + \left(\frac{dx^3}{dx^0}\right)^2} \quad (2.3.7)$$

是粒子的相对速度

$$\vec{v} = \vec{v}(x^0) = \frac{dx^i}{dx^0} e_i = v^i e_i$$

在给定的惯性系中所测值。则我们发现 4-速和 3-速的关系是

$$U = \gamma(\vec{v} + e_0). \quad (2.3.8)$$

把 \mathbf{R}^{3+1} 中一个质点用 (α, m) 来代表，其中 α 是一条类时世界线， $m > 0$ 是这个粒子的质量。若 $dx^\mu/d\xi$ 是常数，则 (α, m) 就是一个具有质量 m 的自由质点。

质点 (α, m) 的 4-动量可自然地定义为 $P = mU$. 因此

$$P \cdot P = -m^2. \quad (2.3.9)$$

在任意惯性基矢下，有

$$\begin{aligned} P &= P^\mu e_\mu = mU^\mu e_\mu = m \frac{dx^\mu}{d\tau} e_\mu \\ &= \gamma m (\vec{v} + e_0), \end{aligned}$$

即惯性坐标中 P 的“空间”部分，即 β -动量是

$$\vec{p} = \gamma m \vec{v}, \quad (2.3.10)$$

如果 $|\vec{v}| \ll 1$, 那么这个量近似于 $m\vec{v}$. 若把 m 等同于 Newton 力学中的惯性质量 (这是由相对于该粒子的速度很小的观者测量的)，则此即经典动量. 如果把 $1/\sqrt{1-v^2}$ 按照广义二项式定理展开，则得到

$$p^i = \frac{m}{\sqrt{1-v^2}} v^i = mv^i + \frac{1}{2}mv^i v^2 + \dots,$$

这给出了经典动量的分量及其“相对论性修正”. Einstein 为保持在形式上与 Newton 力学类似，会把 γm 称为“动质量”，但此属历史遗留问题，现代不怎么用这个概念. 再来看看 P 的时间分量，即

$$P^0 = -P \cdot e_0 = \frac{m}{\sqrt{1-v^2}} = m + \frac{1}{2}mv^2 + \dots,$$

相应于 Newton 动能项 $(1/2)mv^2$ 的出现表明， P^0 是给定惯性观者对其测量的相对论性(总)能量，记作 E . 须知道，物理学中的“能量”概念是一个微妙的概念，把 $-P \cdot e_0$ 称为能量没有带来任何物理内容. 这个名字是否合适只能由实验决定. 我们尤其应该问，上式出现的 m 这一项是否同 P^0 代表“能量”的观点相符. 观察 $|\vec{v}| = 0$ 的情况 (即粒子相对于观者静止)，则有质能方程

$$E_0 = m,$$

此即表明，粒子即使静止也有能量. 若这种静能真的是物理意义上的“能量”，那它应该能释放出来并得到应用. 当然，事实上这也是可能的，并且已经被令人信服地证明了. 实际例子是家喻户晓的，比如核能.

定理 2.3.4 (狭相能动量公式). $E^2 = m^2 + \vec{p}^2$.

证. 该恒等式是 4-动量模长恒定的体现. 取其模长平方，一方面 $P^\mu P_\mu = mU^\mu mU_\mu = -m^2$ ，另一方面由 $P = Ee_0 + \vec{p}$ 和 $\vec{p} \cdot e_0 = 0$ 可知 $P^\mu P_\mu = -E^2 + \vec{p}^2$. \square

2.3.4 Döppler 效应

类光世界线对应于电磁理论的几何光学极限. 不仅有质粒子有“动量”和“能量”，光子也有，因而也应当有 4-动量（例如光电效应，在这种效应中光子与逃离原子中电子轨道的电子之间的碰撞）. 然而，不像有质粒子，光子的特性不是质量而是频率 ν （或者波长 $\lambda = 1/\nu$ ），它与能量 ε 的关系是 $\varepsilon = h\nu$ （ h 是 Planck 常数），且与观者有密切关联（Döppler 效应）. 而且，我们没有相应类比出“质量”的“固有频率”，因为不存在光子在其中静止的惯性系，因此宁愿说光子没有质量. 为了尝试刻画这种特性，考虑一点 $x_0 \in \mathbf{R}^{3+1}$ 、一条指向未来的类光矢量 N 和一个间隔 $I \subseteq R$. 由

$$\alpha(\xi) = x_0 + \xi N$$

定义的曲线 $\alpha : I \rightarrow M$ 是光子经过 x_0 的参数化世界线， ξ 是仿射参数. 对于类光世界线， N 可以在任意惯性基矢下写成

$$N = (-N \cdot e_0) (\vec{d} + e_0),$$

其中

$$\vec{d} = \frac{(N \cdot e_1) e_1 + (N \cdot e_2) e_2 + (N \cdot e_3) e_3}{\sqrt{(N \cdot e_1)^2 + (N \cdot e_2)^2 + (N \cdot e_3)^2}}$$

是这条世界线在相应空间坐标系中的方向矢量. 现在，通过类比能量定义，我们把 \mathbf{R}^{3+1} 中的光子定义成 \mathbf{R}^{3+1} 中的一条曲线 α ，把 N 取成是光子的 4-动量，且光子的能量 ε 在惯性基矢 $\{e_\mu\}$ 下写成

$$\varepsilon = -N \cdot e_0,$$

则

$$N = \varepsilon (\vec{d} + e_0).$$

相应的频率 ν 和波长 λ 则定义为 $\nu = \varepsilon/h$ 和 $\lambda = 1/\nu$. 在另一组惯性基矢下，有 $N = \hat{\varepsilon}(\hat{d} + \hat{e}_0)$ ，其中 \hat{d} 和 $\hat{\varepsilon}$ 是相应物理量的戴帽. 可以看出

$$\frac{\hat{\varepsilon}}{\varepsilon} = \frac{\hat{\nu}}{\nu} \frac{1 - v \cos \theta}{\sqrt{1 - v^2}} = (1 - v \cos \theta) + \frac{1}{2} v^2 (1 - v \cos \theta) + \dots,$$

其中 v 是两个空间坐标系的相对速率， θ 是（不戴帽坐标系中）光子的方向 d 与戴帽坐标系运动方向之间的夹角. 此即相对论 Döppler 效应公式，其中展开级数第一项是经典 Döppler 效应公式.

2.3.5 接触作用

下面来考察上述定义的粒子间简单的相互作用，为此只假设 4-动量在这些相互作用中守恒。出于方便，将用自由粒子这个词表示自由的有质粒子或光子。若 \mathcal{A} 是自由粒子的一个有限集，则 \mathcal{A} 中每一元素都有唯一的 4-动量，或是指向未来的类时矢量，或是类光矢量。任何这种集合中的矢量求和都是类时且指向未来的，除非所有的矢量全是互相平行的类光矢量，这时矢量求和是指向未来的类光矢量。我们把这样的求和称为 \mathcal{A} 的总 4-动量。下面把自由粒子有限集的定义公式化，该集在某个事件中同（或许是新的）另一自由粒子的有限集相碰撞，而从碰撞中产生新的集合，例如，一个电子和一个质子碰撞而产生一个中子和一个中微子。

定义 2.71 (接触作用). \mathcal{A} 中的一个接触作用是一个三元组 $(\mathcal{A}, x, \tilde{\mathcal{A}})$ ，其中 \mathcal{A} 和 $\tilde{\mathcal{A}}$ 是自由粒子的两个有限集，它们都不包含具有线性相关的 4-动量的粒子对（它们或许在物理上不可区分），且 $x \in \mathbf{R}^{3+1}$ 是这样的事件，使得

(接触性) x 是 \mathcal{A} 中所有粒子的终点。即对 \mathcal{A} 中一个粒子的每条世界线 $\alpha : [\xi_0, \xi_1] \rightarrow M$ ，有 $\alpha(\xi_1) = x$ ；

(分散性) x 是 $\tilde{\mathcal{A}}$ 中所有粒子的起点；

(4-动量守恒) \mathcal{A} 的总 4-动量等于 $\tilde{\mathcal{A}}$ 的总 4-动量。

如果 \mathcal{A} 只包含单个自由粒子，那么 $(\mathcal{A}, x, \tilde{\mathcal{A}})$ 称为衰变，例如，一个中子衰变到一个质子、一个电子和一个中微子。

考虑一个相互作用 $(\mathcal{A}, x, \tilde{\mathcal{A}})$ ，其中 $\tilde{\mathcal{A}}$ 包含单个光子。 $\tilde{\mathcal{A}}$ 的总 4-动量是类光矢量，所以对于 \mathcal{A} 来说也是如此。因为 \mathcal{A} 中各粒子的 4-动量有类时、类光的，且是未来指向的，故只有其全是平行的类光矢量时，总和才可能是类光矢量。因为 \mathcal{A} 不能包含有平行 4-动量的不同光子，故它必定由单个光子组成，根据 4-动量守恒，这个光子的 4-动量与 $\tilde{\mathcal{A}}$ 中光子的相同。因此从本质上讲，在 x 处仿佛“什么也没发生”。我们得出结论，不存在我们所限定的那种非平凡相互作用，能够在单个光子上发生，而不发生其它事件。把 \mathcal{A} 和 $\tilde{\mathcal{A}}$ 的颠倒后表明，若 4-动量守恒，则单个光子不能衰变。

下面，我们考虑单个有质粒子衰变到两个有质粒子，例如一个原子核自发的 α 衰变。考虑一个有接触的相互作用 $(\mathcal{A}, x, \tilde{\mathcal{A}})$ ，其中 \mathcal{A} 包含一个质量 m_0 的单个自由有质粒子，而 $\tilde{\mathcal{A}}$ 包含两个自由有质粒子（一个质量 m_1 ，另一个 m_2 ）。设 P_0, P_1 和 P_2 分别是质量 m_0, m_1 和 m_2 的有质粒子的 4-动量。则 $P_0 = P_1 + P_2$ 。按照“反转三角不等式”，由于 P_1 和 P_2 是线性无关且指向未来的，结合 $P \cdot P = -m^2$ ，有

$$m_0 > m_1 + m_2.$$

初始粒子的额外质量 $m_0 - (m_1 + m_2)$ 按质能方程，可认为是把 m_0 分成两部分所要求的能量的一种度量。稍微不同的说法是， $\tilde{\mathcal{A}}$ 中的两个粒子结合而形成 \mathcal{A} 中的一个粒子时，给后面粒子的质量贡献了“结合能”。此即质量亏损。

这个例子里，颠倒 \mathcal{A} 和 $\tilde{\mathcal{A}}$ 就成了完全非弹性碰撞 (两个质量 m_1 和 m_2 的自由粒子碰撞并结合成第三个质量 m_0 的自由粒子) 中所含有的那种接触作用。质量亏损仍然成立。更详细的分析将给出 $m_0 - (m_1 + m_2)$ 的近似公式，这可与该碰撞中动能损失的 Newton 公式作比较 (经典上把能量看做是从结合粒子中获得热量的形式)。对含有一个电子和一个光子的 \mathcal{A} 和 $\tilde{\mathcal{A}}$ 的相互作用分析可得所谓的 Compton 效应公式⁴⁰。

2.4 非惯性运动

2.4.1 瞬时观者

狭相建立初期对狭义相对性原理做了自然语言叙述，即规定惯性系平权，但却没有对非惯性系做类似规定。这导致的问题是，仅基于狭义相对性原理，不能直接在非惯性系中讨论物理。现代理论认为狭相的研究背景就是平直时空（无引力），因此也可讨论非惯性系，但却需要一点“弯曲时空”的数学知识。本节考虑到读者接受能力，暂只介绍在惯性系下非惯性运动的性质。在非惯性系中讨论物理见 3.3 节。

设 $\alpha(\tau)$ 是质点的世界线，观测其运动需借助任一参考系 R 。对 $\alpha(\tau)$ 上任一点 p ，总有 R 中的一条类时世界线（观者） G 经过，即 $G \cap \alpha(\tau) = p$ ， G 便可对事件 p 的相关性质进行测量。以 Z 和 U 分别代表 G 和 $\alpha(\tau)$ 在 p 的 4-速。若 $Z = U$ ， G 会认为 α 在 p 时刻静止。反之， G 会认为 α 在 p 时刻有 3 维速度。

为给质点在时刻 p 相对于 G 的 3-速下定义，先做如下铺垫。设想你自己就是观者 G 。首先，你能直接观测发生在你身上的事件。如果事件发生在你身外，你当然也可能听见或看见，但这涉及从该事件到你的信号传递（如用声或光），要花费一定时间，讨论起来较复杂。这就叫间接观测⁴¹。理论上最简单、明确、基本的观测是直接观测，即对发生在观者世界线上的事件的观测，亦称当时当地观测 (*local measurement*)。不过，参考系由无处不在的观者组成，发生在别处的事件由别处的观者观测便是，只不过涉及到大家对标准钟零点的统一，暂不详述，细节见 3.3 中对任意时空任意参考系的对钟问题。

此外，你在观测发生在你世界线上 p 点的事件时，在某些情况下重要的不是你的整条世界线而只是你在 p 点的 4-速。这时没必要强调观者的世界线如何，只须给定该世界线在 p 点的切矢 Z 。于是可提炼一个更为抽象的概念：

定义 2.72 (瞬时观者). 瞬时观者 (*instantaneous observer*) 由 p 点及 p 点的一个指向未来类时单位矢 Z 构成，记作 (p, Z) 。

⁴⁰这个例子在 [5] 第 6 章, §14 中有详细分析。

⁴¹理论方面，狭相中高速物体所呈形象的课题就属这一范畴，见 Part II 相对论图形学一章；实用方面，天文观测都属间接观测。

最后，你作为观者，除有时间感（用标准钟）外还有空间方向感。假定你手握一支笔，其笔尖的任一指向就代表你能感到的一个空间方向。你在时刻 p （世界线 G

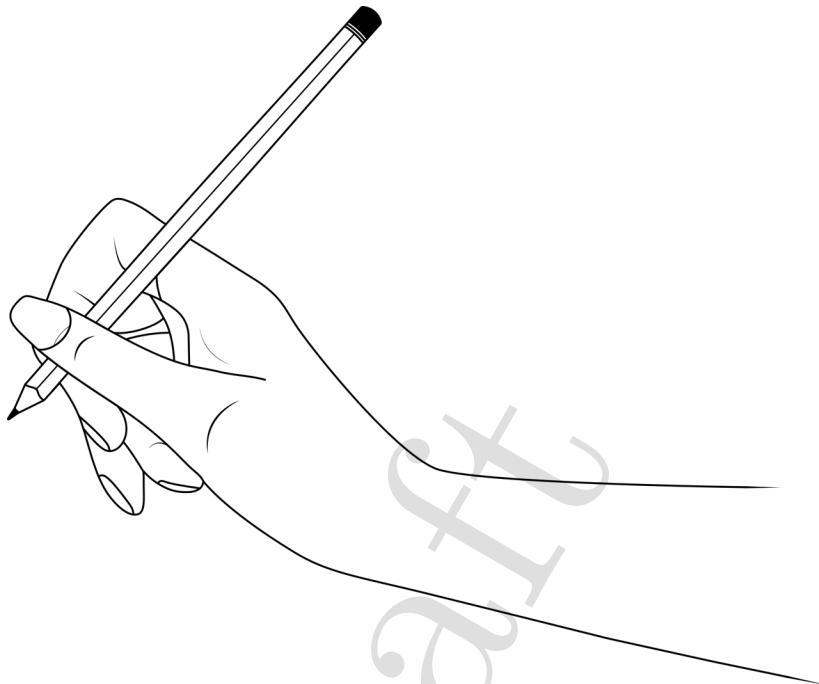


图 2.15: 空间矢量

的一点）能感到的所有空间矢量的集合 W_p 是 3 维的，而 p 作为 \mathbf{R}^{3+1} 的一点，其矢量空间 V_p 是 4 维的。 W_p, V_p 的关系类似于此前所讨论的超曲面。回忆一下，如果 R 是惯性系，而你是其中的惯性观者。 R 的同时面就是 R 在某时刻的 3 维空间，而同时面与该系的所有惯性观者世界线正交，所以你在交点 p 的全部空间矢量都与你在 p 点的 4-速 Z 正交，故 W_p 对应于 V_p 中与 Z 正交的 3 维子矢量空间。即

$$W_p = \{\vec{w} \in V_p | \eta(\vec{w}, Z) = 0\}.$$

这一对应也可用于非惯性观者，因为我们关心的只是观者世界线上一点 p 的情况。可用一“无限小”平面⁴²代表 W_p 。

定义 2.73 (空间矢量). 设 $w \in V_p$, 当 $w \in W_p$ 时, 就说 w 对观者 G 而言是空间矢量 (*spatial vector*)。

非零的空间矢量一定是类空矢量，反之不然。由定义知类空矢量是绝对的，即不

⁴²对 W_p 最准确的理解还是 p 点矢量空间的子空间，而非时空中的子集，但借助“弯曲时空局部类似平直时空”这一思想，可模仿平直时空将矢量空间的东西搬到时空来，只不过在图上画得小一点。

依赖于观者、参考系或坐标系等人为因素，但空间矢量是相对的，其取决于观者 4-速 Z 。

设研究对象 $\alpha(\tau)$ 同某观者世界线 G 交于 $p = \alpha(\tau_1)$ ，则在此前讨论中已经知道，若二者都做惯性运动，以 G 所在惯性系建立坐标系（取其世界线为 $x^0 = t$ 坐标线，坐标基 $e_0 = Z$ ），则 $U = \gamma(\vec{u} + Z)$ 。取 α 上临近点 $q = \alpha(\tau_1 + d\tau)$ ，则直线 pq 代表质点在微小固有时内所经历的事件。对 G 而言这个过程的时间测量值为 dt ，空间位移测量值则为 $dx^i e_i$ ，故 α 相对于 G 的 3-速就是

$$\vec{u} = \frac{dx^i}{dt} e_i = \frac{U^i}{U^0} e_i.$$

回顾一下， p 点的 η 在 W_p 上的诱导度规是 $\tilde{\eta} = \eta + Z_b \otimes Z_b$ ，其对应 V_p 到 W_p 的投影映射 $h = I + Z \otimes Z_b$ 。可见， $\gamma\vec{u}$ 正是 U 在 W_p 上的空间投影，因此 \vec{u} 又可改写为 $h(U)/\gamma$ ，而 γ 也可表为 $-\eta(U, Z)$ （因为 $-\eta(U, Z) = -\eta_{00} U^0 Z^0 = U^0 = dt/d\tau$ ）。由于现在 G 是惯性系 R 中的观者， α 相对于 G 的 3-速也就是相对于 R 的。

上文的关键结论与坐标系无关，因此在非惯性运动下就可定义：

定义 2.74 (3-速). 任意质点 α 相对于任意瞬时观者 (p, Z) 的 3-速为

$$\vec{u} = \frac{h(U)}{\gamma} = -\frac{U + \eta(U, Z)Z}{\eta(U, Z)} = -\frac{U}{\eta(U, Z)} - Z. \quad (2.4.1)$$

容易发现 \vec{u} 作为 4-矢时 $u^0 = 0$ 。3-速率为其模长

$$u = \sqrt{\eta_{\mu\nu} u^\mu u^\nu} = \sqrt{\tilde{\eta}_{\mu\nu} u^\mu u^\nu} = \sqrt{\delta_{ij} u^i u^j}.$$

定理 2.4.1 (相对论限制). “质点走类时世界线”要求“亚光速”，即 $u < 1$ 。

证. 考虑速率平方： $u^2 = \frac{\eta(U, U)}{\eta^2(U, Z)} + \eta(Z, Z) + \frac{2\eta(U, Z)}{\eta(U, Z)} = \frac{1}{\gamma^2} \eta(U, U) + 1$ 。质点取固有时参数，则 $u^2 = 1 - \frac{1}{\gamma^2} < 1$ 。这与此前 $\gamma = \frac{1}{\sqrt{1-u^2}}$ 的结论是一致的。□

定义 2.75 (瞬时静止观者). 若瞬时观者 (p, Z) 恰好与被测对象世界线相切，则称为该对象的瞬时静止观者或瞬时共动观者，在其看来对象的 3-速瞬时为零。由 (p, Z) 决定的（线长参数化）测地线 G 称为对象在 p 时的瞬时静止惯性观者。

定义 2.76 (瞬时静止惯性系). 上述 G 所隶属的惯性参考系（线汇均为测地线且相互无交点， G 是其中一条）称为瞬时静止惯性参考系。该参考系中任意惯性坐标系称为瞬时静止惯性坐标系。一般取 $x^0 = t$ 坐标线为 G 世界线，则可统称为瞬时静止惯性系。

此前在惯性运动中讨论的结论在瞬时观者中也适用，如

定理 2.4.2. 4-速可借 (p, Z) 做 $3+1$ 分解： $U = \gamma(\vec{u} + Z)$ 。

证. 由定义知 $\gamma\vec{u} = U - \gamma Z$ 。 □

4-动量的分解类似。

平直时空中惯性观者走直线, 这可用 4 维加速度刻画。

定义 2.77 (4-加速). 在平直时空的任意惯性坐标系中, 可定义一个质点的 4-加速 (*4-acceleration*) 为

$$A^\mu = \frac{dU^\mu}{d\tau} = \frac{d^2x^\mu}{d\tau^2}, \quad (2.4.2)$$

这里 U 为 4-速, τ 为其固有时。

因此我们说测地线满足 $A^\mu = U^\nu \partial_\nu U^\mu = 0$ 。

定理 2.4.3. A, U 正交。

证. 即证明 $U_\mu A^\mu = U_\mu U^\nu \partial_\nu U^\mu = 0$ 。此即 $\frac{1}{2}U^\nu \partial_\nu (U_\mu U^\mu) = 0$, 显然成立。 □

与测量 3-速不同, 只用一个观者 G 不足以测量 α 的 3-加速, 因为要确定其 p 处 3-加速需要比较 p 与 α 上另一邻近点 q 的 3-速。而观者 G 不能直接观测 q 点⁴³。这一困难一般借坐标系克服: 可以定义 α 在其任意点相对于任意坐标系的所谓坐标 3-加速。常取惯性坐标系。

定义 2.78. 质点世界线在惯性坐标系的参数表达式为 $x^\mu = x^\mu(\tau)$, 记 $x^0 = t$, 则其相对于该坐标系的 3-加速定义为

$$\vec{a} = \frac{d\vec{u}}{dt} = \frac{d^2x^i(t)}{dt^2} e_i. \quad (2.4.3)$$

定理 2.4.4. 质点的 4-加速等于它相对于瞬时静止惯性坐标系的 3-加速。因此 $A = 0$ 的质点相对于任意惯性系的 3-加速为零。

定义 2.79 (4-力). 平直时空惯性坐标系中, 质点所受的 4-力 (*4-force*) 定义为

$$F^\mu = \frac{dP^\mu}{d\tau} = U^\nu \frac{\partial P^\mu}{\partial x^\nu}, \quad (2.4.4)$$

这里 P 是 4-动量。容易验证 $F^0 = dE/d\tau$, $F^i = \gamma f^i$, 其中 3-力 $\vec{f} := d\vec{p}/dt$ 。

定理 2.4.5. 若质量在运动中保持常数 ($dm/d\tau = 0$), 则 $F^\mu = mA^\mu$, $\vec{f} = m\vec{a}$, $\vec{f} \cdot \vec{u} = dE/dt$ 。

⁴³除非某邻域内与 G, α 重合, 这样 G 必然瞬时共动。但“邻域”具有有限大小, 这意味着重合是“可观的”(曲线差距至少是高二阶无穷小), 因而 3-加速在邻域内为零。当然, 由于“有限大小”可以任意小, 我们不妨放宽“直接观测”的限制, 这样便不要求严格相交(只要求曲线差距高一阶), 仍可用瞬时静止观者观测非零 3-加速。但其它瞬时观者仍然测不了, 因为曲线差距仍然同阶。

2.4.2 无自转观者

此前提及过，为了观测，观者除需标准钟外，往往还要配备一个标架。在生活中，3-标架 (frame) 可以是由三根单位长的短直杆或刻度尺焊成的、两两互相正交的架子，每根直杆代表一个观测方向，它们在每一时刻的指向由该观者选定。当然，读者亦可尝试用右手大拇指、食指及中指按右手螺旋比划出 3-标架。

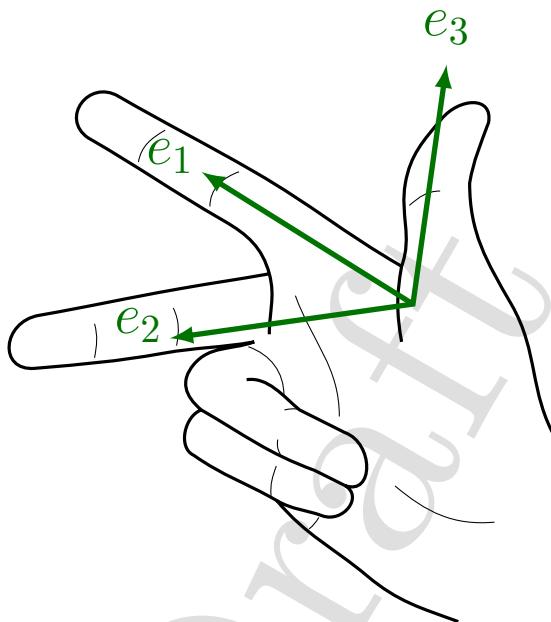


图 2.16: 观者的空间 3-标架

在数学上，3-标架被抽象为观者世界线上的空间正交基矢场 $\{e_i\}$ ，“空间”是指它们都与观者的 4-速 Z 正交。于是，连同 $e_0 = Z$ 在内，观者世界线上就有四个正交归一矢量场，称为观者的 4-标架场 (tetrad)。今后谈及 4-标架场时如无声明都规定为右手标架。因为参考系由充斥于时空 (或其开子集) 的无数观者组成，过每一时空点有且仅有一条观者世界线，所以，给定一个参考系后，全时空 (或其开子集) 就有一个 4-标架场。任何时空点的任何张量都可用该点的 4-标架作为基底表出。

定义 2.80 (观者). 观者时指的是一条定义了 4-标架场的类时世界线。

世界线为测地线的运动就称为惯性运动，则对惯性观者的准确定义是：

定义 2.81 (惯性观者). 惯性观者是做惯性运动的无自转观者。

“无自转”则是对线上的 4-标架场的要求。粗略地说，设 Alice、Bob 两人坐在地面的两把椅子上，Alice 坐普通椅子而 Bob 坐转椅 (但底座固定于地面) 且不停转

动，则 Alice 可视为惯性观者而 Bob 不能（他有自转）。请注意，虽然观者概念本身已要求把此二人看成没有大小的点（于是由一世界线代表，且是测地线），但自转涉及的仅是线上各点每一空间基矢的方向沿线是否改变，故仍有明确意义（下面立即讨论其内涵）。例如，设 R 为惯性坐标系，把其中每一 $x^0 = t$ 坐标线看作一个观者的世界线，并选惯性坐标基为线上的 4-标架场，则直观看来（或按下面给出的严格定义）他就是无自转的。平常谈及惯性系中的惯性观者时，一般默认他们以惯性坐标基为 4-标架。相应地，决定一个瞬时观者的要素除 p, Z 外，有时还要辅以 3-标架，它们与 Z 组成 p 点的正交 4-标架 $\{e_\mu\}$ ，在此情况下一个瞬时观者应表为

$$(p, \{e_\mu\}),$$

其中 $e_0 = Z$ 。不必强调标架时，瞬时观者仍可表为 (p, Z) 。

下面一步步地引出无自转的严格定义。描述 3 维矢量沿着世界线转没转，自然是要比较微小固有时两端的 3 维矢量，而这需要我们确定两点一致的“不变方向”。Newton 力学中明确规定了“陀螺仪自转轴”代表不变的方向⁴⁴。因此假设读者手持一部陀螺仪（譬如读者手机中所配备的），若一支笔在 τ_1 时同陀螺仪轴平行而在 τ_2 却不平行，便可肯定其在这段固有时 $\Delta\tau$ 内发生了转动。这一判据可延伸至相对论中。以 $G(\tau)$ 代表观者世界线，欲研究 $p_1 = G(\tau_1)$ 处一单位空间矢量 \vec{w} ，其与 τ_1 时的 4-速 $Z|_{p_1}$ 正交。随固有时流逝至 τ_2 ，则自然产生了 $G(\tau)$ 上的单位空间矢量场 \vec{w} 。

定义 2.82 (Fermi-Walker 导数).

定义 2.83 (Fermi-Walker 移动).

Einstein 转盘

2.5 经典电磁理论

2.5.1 回顾

18 世纪末，人们已经知道四类电磁现象，但还不知晓它们之间的联系：磁 (*magnetism*)，希腊语关于“来自 Magnesia 之石”转意而来的词；静电 (*electrostatic*)，琥珀与毛皮摩擦产生，“电” (*electricity*) 这个词由 Gilbert 取自于希腊语的“琥珀” ($\eta\lambda\epsilon\kappa\tau\rho\nu\omega\zeta$)；光 (*light*)；Galvani 状态或“动物电”，Galvani 发现的电池所产生的电。进入 19 世纪后，人们对电磁学的认识实质上并非一蹴而就：

1800 年 Volta 演示了 Galvani 状态与静电的联系。

1820 年 Ørsted 显示了电池的电流引发对磁体的作用力。

⁴⁴ 相对论领域有关的实验证明了陀螺仪轴在环球旅行的过程中会受到扰动，因而我们暂且只考虑在惯性系中讨论转动。

1822 年 Ampère 建议光是两类电流体组成的“传光以太”中的波动. 同年, Galileo 的《关于两大世界体系的对话》从被禁书中解禁.

1831 年 Faraday 显示运动磁体可感应电流.

1846 年 Faraday 建议光是磁力线中的振动.

1863 年 Maxwell 发表了描述电场和磁场动力学的方程组.

1905 年 Einstein 的论文《论动体的电动力学》.

这场统一事业是漫长而艰苦的. 现在回顾起来, 似乎各种古怪的概念阻挠了人们对电磁学的理解, 如“电流体”、“磁涡旋”的力学模型等等, 直到 Einstein 的论文为止, 这个时期的“遗产”都仍有许多在现代术语库里继续误导的, 也有部分已不通用, 比如“磁通量”、“力线”、“电位移”等. Maxwell 的贡献是决定性的, 之所以称为“Maxwell 理论”是其继承者 (Lorentz、Hertz、Einstein 等人) 的缘故; 正如我们看到, 在 Maxwell 自己描述的电磁学中的关键成分是“电磁以太”, 一种被假定传播电磁波的弥漫介质, 但后来就被 Einstein 摈弃了.

值得注意的是虽然在 Maxwell 方程组的命名上他的贡献普遍被承认, 但是很少见到引证“Maxwell 理论”. 这是大家认同的说法. Maxwell 在他电磁学的早期研究中依据精心制作的力学模型工作, 他把它看成类比物, 而不是根本的物理实在的文字描述. 在他的后期工作中, 力学模型, 特别是传播光波的“传光以太”的力学性质作为电磁理论的基础更为文字化地提出来了. 现代理论中方程组幸存了下来, 但 Maxwell、Faraday 以及其他人为之奋斗的力学模型只得在如“力线”、“磁通量”等陈旧的术语中残存.

Maxwell 的遗产不是他的理论, 而是他的方程组: 一个和谐一致的偏微分方程组体系, 它描述了电场和磁场与运动电荷间已知相互作用的整个领域. 当然, Maxwell 没有认识到其紧凑的矢量表达, 更不用说它们在狭相中的张量形式. 通过第一次揭示电场和磁场之间充分的对偶性, 来统一处理电学和磁学. 方程组连同无穷远处的边界条件 (*boundary condition*) 确定了源所激发的场. 这些方程可导出静电学和静磁学熟知的各种预言, 以及场和电荷的动力学行为, 特别是电磁波的激发和传播. 这组方程证实了几乎不可想像的各种各样的物理过程, 从光通过星系磁场的行为而越过宇宙学距离的传播, 电机和实验室实验的日常应用, 一直到个别电子间交换光子.

Maxwell 方程组的过去以许多人的蜿蜒曲折旋绕着. Maxwell 本人在他 1865 年论文《电磁场的动力理论》中对第一方程中的 ρ 以不同符号, 写下方程组的不协调形式. 而协调的形式迟至他的著作《电磁通论》(1873) 才出现, 参见 [14].

本节将不按照历史的脉络来谈电磁学. 上面某些评述中提到有关进展的复杂故事可在论文 [14]、[15] 和 [16] 中找到. 我也将不按照许多教科书通过研究基本电磁现象建立充分的动力方程组这样一条传统的适宜于教学法的路线. 替代之, 我将沿着这样一条路径到达 Maxwell 方程组, 认识它最关键的特征, 即 Lorentz 协变性, 得

出 Maxwell 方程组. 当然, Maxwell 对此一无所知. 我们将从有关电荷在电场和磁场中行为的基本事实概要出发, 然后从不同运动参考系考察这一行为来建立充分的动力学构架. 当然, 自从 Maxwell 方程组与 Galileo 相对性不和谐以来, 在时空的经典概念的构架内一致地这样做是不可能的. 但是理解某些方程组的关键特征至少是可能的, 特别是在 Maxwell 方程组第三方程中需要包含 \mathbf{E} 的时间导数项, 即所谓“位移电流”.

电荷、场和 Lorentz 力定律

经典电磁理论的基本对象是荷电粒子, 以及依赖于位置和时间的电场 \mathbf{E} 、磁场 \mathbf{B} . 粒子电荷 q 可正可负, 是类似于引力质量的固有内禀量, 其确定了电场和磁场中粒子相互作用的强度, 正如粒子的质量确定了引力场中粒子相互作用的强度.

这一相互作用有两方面. 第一, 电场和磁场对荷电粒子施加一依赖于电荷值、粒子速度及粒子所在处的 \mathbf{E} 和 \mathbf{B} 值的作用力. 这个力由 Lorentz 力定律给出

$$\mathbf{f} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}), \quad (2.5.1)$$

式中 \mathbf{v} 是速度. 它类似于引力场 \mathbf{g} 中作用于质量为 m 的粒子的引力

$$\mathbf{f} = m\mathbf{g}.$$

原则上借助于测定作用在具有已知速度的标准电荷上的作用力, 通过力学定律, 观者可测量电场和磁场.

第二, 运动电荷激发电场和磁场. 但我们将不详细研究运动电荷激发场的方法, 而只是陈述下列原理:

电磁基本原理 1 场线性依赖于电荷. 这意味着如果我们叠加两个电荷分布, 则总的 \mathbf{E} 场和 \mathbf{B} 场是此二分布分别激发各自的场之和.

电磁基本原理 2 一个固定不动的点电荷 q 激发电场而不激发磁场, 电场由下式给出:

$$\mathbf{E} = \frac{kqr}{r^3}, \quad (2.5.2)$$

式中 \mathbf{r} 是起于电荷的位矢, $r = |\mathbf{r}|$, k 是类似于引力常数的正常量.

联合 (2.5.2) 和 (2.5.1), 我们得到两个固定不动电荷之间的静电力平方反比律, 即 Coulomb 定律

$$\mathbf{f} = \frac{kqq'}{r^3}\mathbf{r}. \quad (2.5.3)$$

与重力不同, 当电荷取相同符号时为斥力.

电磁基本原理 3 以速度 \mathbf{v} 运动的点电荷激发磁场

$$\mathbf{B} = \frac{k'qv \times \mathbf{r}}{r^3}, \quad (2.5.4)$$

式中, k' 是另一个正常量.

此式是根据测定电路中流动电流所激发的磁场外推得来的 (*Biot-Savart 定律*).

电磁基本原理 2 和 3 中的常量 k 和 k' 确定了电磁相互作用的强度, 它们通常表示为

$$k = \frac{1}{4\pi\epsilon_0}, \quad k' = \frac{\mu_0}{4\pi}, \quad (2.5.5)$$

其中常量在国际单位制下为

$$\epsilon_0 = 8.9 \times 10^{-12}, \quad \mu_0 = 1.3 \times 10^{-6}. \quad (2.5.6)$$

电荷 q 单位用 Coulomb 量度, $|\mathbf{B}|$ 用 Tesla, $|\mathbf{E}|$ 用 Volt/m。对于这些单位可有如下概念: 电子元电荷为 $e = -1.6 \times 10^{-19}$ C, 而通过电炉的电流是 $5 \sim 10$ C·s⁻¹; 地球磁场约是 4×10^{-5} T; 一根磁棒的磁场大约 1 T; Oxford 的 Clarendon 实验室三楼大约有 50 T 的磁场; 中子星表面磁场约是 10^8 T; 虽然我们在日常生活中更熟悉重力, 但它比静电力弱得多, 两个质子之间的静电斥力比它们之间的引力吸引大 1.2×10^{36} 倍 (任何间隔上两种力都遵从平方反比律).

我们的目标是从电磁基本原理 1 ~ 3 出发, 用微分方程来替换 (2.5.2) 和 (2.5.4) 以获得 Maxwell 方程组, 而其将场强同电荷连续分布的电荷密度 ρ 和电流密度 \mathbf{J} 联系起来, 这些密度定义为

$$\rho = \frac{dQ}{dV}, \quad \mathbf{J} = \rho\mathbf{v}, \quad (2.5.7)$$

式中 dV 是该点的测度微元, dQ 是遍及 dV 内众多电荷所对应的测度微元, \mathbf{v} 是速度场在该点的取值.

固定不动的电荷分布

下面着眼于固定不动电荷分布的电场, 尝试把基本原理改写成偏微分方程, 其中采用 Gauss 定理重新陈述平方反比律来完成连续极限这一步骤, 以获得 Gauss 定律, 其将电场沿闭合曲面的积分同其所包围的总电荷联系起来。对于单个点电荷, 电场由电磁基本原理 2 给出:

$$\mathbf{E}(\mathbf{r}) = \frac{q\mathbf{r}}{4\pi\epsilon_0 r^3},$$

则

$$\nabla \cdot \mathbf{E}(\mathbf{r}) = \frac{q}{4\pi\epsilon_0} \nabla \cdot \frac{\mathbf{r}}{r^3} = \frac{q\delta(\mathbf{r})}{\epsilon_0},$$

则由 Gauss 定理知

$$\int_{\partial V} \mathbf{E} \cdot d\mathbf{S} = \frac{q}{\epsilon_0} \int_V \delta(\mathbf{r}) dV, \quad (2.5.8)$$

因此

$$\int_{\partial V} \mathbf{E} \cdot d\mathbf{S} = \begin{cases} q/\epsilon_0, & \text{如果电荷在 } \partial V \text{ 包围的体积内} \\ 0, & \text{其他情形} \end{cases}$$

当我们遍及电荷分布求和，公式左边的积分选择 S 内的总电荷，于是我们有：

Gauss 定律 对于包围体积 V 的任意闭合曲面 ∂V ,

$$\int_{\partial V} \mathbf{E} \cdot d\mathbf{S} = \frac{Q}{\epsilon_0}, \quad (2.5.9)$$

式中 \mathbf{E} 是总电场， Q 是 V 内的总电荷.

现在我们可以研究连续极限. 假定 \mathbf{E} 被具有密度 ρ (单位体积内的电荷) 的电荷分布所激发，则由 Gauss 定理，对任意体积 V 有

$$\int_{\partial V} \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\epsilon_0} \int_V \rho dV, \quad (2.5.10)$$

但另一方面由 Gauss 定理有

$$\int_V \left(\nabla \cdot \mathbf{E} - \frac{\rho}{\epsilon_0} \right) dV = 0,$$

由于对任意体积 V 都成立，从而

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0}. \quad (2.5.11)$$

按照类似思路，也可以证明固定不动的电荷分布的电场是保守的，其意义是当一电荷沿一闭合回路移动，场所做的总功为零，即对任意闭合路径有

$$\int_{\partial S} \mathbf{E} \cdot d\mathbf{r} = 0, \quad (2.5.12)$$

这等价于

$$\nabla \times \mathbf{E} = 0, \quad (2.5.13)$$

因为根据 Stokes 定理

$$\int_{\partial S} \mathbf{E} \cdot d\mathbf{r} = \int_S \nabla \times \mathbf{E} \cdot d\mathbf{S}, \quad (2.5.14)$$

式中 S 是以路径为边界的任意曲面，(2.5.13)成立，当且仅当对一切可能的路径和一切可能的 S 积分为零. 单个固定不动电荷的场是保守的，因为

$$\mathbf{E} = -\nabla\phi, \quad \phi = \frac{q}{4\pi\epsilon_0 r}, \quad (2.5.15)$$

由于梯度无旋，因此 $\nabla \times \mathbf{E} = 0$. 对于连续分布，式中

$$\phi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \int_V \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3x', \quad (2.5.16)$$

此积分中 \mathbf{r} 是固定的，积分遍及各电荷位置 \mathbf{r}' ，除开 $\mathbf{r} = \mathbf{r}'$ 的奇异性，积分是良定义的，所以 (2.5.13) 对固定不动电荷的连续分布也成立.

磁场的散度

可采用相同论证建立低速运动电荷的磁场 *Gauss* 定律，观察

$$\mathbf{B} = \frac{\mu_0 q \mathbf{v} \times \mathbf{r}}{4\pi r^3}, \quad (2.5.17)$$

式中 \mathbf{r} 是从电荷到所测场点的矢量。因为 $\mathbf{r}/r^3 = -\nabla(1/r)$ ，我们有

$$\nabla \cdot \left(\mathbf{v} \times \frac{\mathbf{r}}{r^3} \right) = \mathbf{v} \cdot \nabla \times \nabla \frac{1}{r} = 0,$$

所以如同电场情形，除 $r = 0$ 外， $\nabla \cdot \mathbf{B} = 0$ 。然而在磁场的情形，场遍及包围电荷的曲面所做的积分也为零，因为假设 S_R 是中心在电荷处半径为 R 的球，则

$$\int_{S_R} \mathbf{B} \cdot d\mathbf{S} = \frac{\mu_0 q}{4\pi} \int_{S_R} \frac{\mathbf{v} \times \mathbf{r}}{r^3} \cdot \frac{\mathbf{r}}{r} dS = 0,$$

根据 *Gauss* 定理，对于包围电荷的任意曲面同样是正确的，因此如果磁场仅被运动电荷所激发，则对任意体积 V 有

$$\int_{\partial V} \mathbf{B} \cdot d\mathbf{S} = 0, \quad (2.5.18)$$

从而

$$\nabla \cdot \mathbf{B} = 0, \quad (2.5.19)$$

当然，如同电荷激发电场，如果存在激发磁场的自由“磁极”，此式将不成立；类似于 (2.5.11) 中的电荷密度，此方程的右边应有“磁极密度”。

Galileo 不变性的限度

我们主要关心的是电磁学定律与相对性原理的相容性。正如 Einstein 观察到的，简单的电磁相互作用确实仅依赖于相对运动；一个导体通过磁体的场运动所感应的电流等同于磁体以相同的相对速度运动，越过在固定不动的导体上所激发的电流。很遗憾，这一对称性没有反映在我们的基本原理中。如果假设其在各惯性系中都成立，则我们可很快消除许多矛盾。如下，一个观者 O 可以测定相对于电荷的速度为 \mathbf{v} 的标准电荷粒子 q 所受的力 \mathbf{f} ，依照 Lorentz 力定律， O 可以测量一点的 \mathbf{B} 和 \mathbf{E} 值。第二个观者 O' 相对于第一个观者以速度 \mathbf{v} 运动，将观察到相同的力 \mathbf{f} ，但现在是作用在一个静止粒子上。所以 O' 将测量到电场为 $\mathbf{E}' = \mathbf{f}/q$ 。由此得出结论，以速度 \mathbf{v} 通过磁场 \mathbf{B} 和电场 \mathbf{E} 运动的观者看到的电场满足关系

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B},$$

交换两个观者的角色有

$$\mathbf{E} = \mathbf{E}' - \mathbf{v} \times \mathbf{B}',$$

式中 \mathbf{B}' 是 O' 所测磁场。如果两者都成立，则 $\mathbf{B} - \mathbf{B}'$ 应等于 \mathbf{v} 的标量倍数。但这与电磁基本原理 3 不相容，如果场是静止点电荷相对于第一观者的场，则 \mathbf{E} 由 (2.5.2) 给出且

$$\mathbf{B} = 0;$$

另一方面，第二观者看到点电荷以速度 $-\mathbf{v}$ 运动的场，所以

$$\mathbf{B}' = -\frac{\mu_0 q \mathbf{v} \times \mathbf{r}}{4\pi r^3},$$

因此 $\mathbf{B} - \mathbf{B}'$ 是垂直于 \mathbf{v} ，而非与其平行。这一显而易见的矛盾可以部分地解决，只要我们认识到其实电磁基本原理 3 是不严格的，其仅当速度小到两粒子间的磁力同静电力比较时可忽略才成立。如果 \mathbf{v} 是某个这种典型速度，则条件是 $v^2 \mu_0$ 应远小于 $1/\epsilon_0$ ，即涉及的速度必须远小于 $\frac{1}{\sqrt{\epsilon_0 \mu_0}}$ 。当然，这就是光速。

现在必须认为电磁基本原理 1~3 是近似的，它们所描述的粒子和场的相互作用，对应于粒子相对于彼此运动在速率上远小于光速，特别地应该强调，不能预期对粒子以可与 c 比拟的速率运动时电磁基本原理 3 仍成立，必须代之以：

电磁基本原理 3' 一个以速度 \mathbf{v} 运动的电荷，当 $v \ll c$ 时，其激发的磁场可视作

$$\mathbf{B} = \frac{\mu_0 q \mathbf{v} \times \mathbf{r}}{4\pi r^3} + O(v^2/c^2), \quad (2.5.20)$$

电荷系统的磁场在一般运动中满足

$$\nabla \cdot \mathbf{B} = 0. \quad (2.5.21)$$

我们保留了上式作为的微分表述，即不存在自由磁极；磁场仅由电荷的运动激发。做了这一变更，只要我们忽略 v^2/c^2 级项，理论就与相对性原理和谐一致。以电磁基本原理 3' 替代电磁基本原理 3 解决了矛盾；只要速度远小于光速，Einstein 所指出的导体在磁场中运动激发的电流和磁体越过导体的运动所激发的电流之间的对称性就可被解释。然而主要的问题仍存在，电磁方程组在速度与 c 可比拟的 Galileo 变换下不协变。这一矛盾仍存在但比原先出现的更加微妙。有三条可能的出路：(1) 不协变性是实在的，有可观察的效应（须有 v^2/c^2 级或更小级），(2) Maxwell 理论是错的，(3) Galileo 变换是错的。物理学采取最后一条路。

Maxwell 方程组将严格保留其经典表述，但做相对运动的参考系的时空坐标间的变换却并不如此。当与光速相比较，所包含的速度不是无关紧要时，这种差别才显示出来，只要可忽略 u^2/c^2 级的项，Maxwell 方程组与 Galileo 相对性原理就是协调的。

电磁感应

低速运动电荷的磁场总小于其相关电场 (即使我们用 $c\mathbf{B}$ 代替 \mathbf{B} 并使用与 \mathbf{E} 相同的单位). 然而电路中电流所激发的磁场不会被强大的电场所支配. 这是因为在低速下电流是由电子的流动建立起来的, 而总地说来, 导线中的物质大致是电中性的, 带正电的核的电场与带负电的电子的电场相抵消.

下面是从速度远小于 c 的 Galileo 变换导出 Faraday 感应定律应铭记的脉络, 涉及到沿电路的电动势或“电压”、遍及电路所围曲面上磁场 \mathbf{B} 的变化率. 其微分表述是 Maxwell 方程组之一.

首先假设场被电荷激发, 电荷全部相对于给定的惯性系 R 并以相同的速度 \mathbf{v} 运动, 于是在相对于 R 以速度 \mathbf{v} 运动的第二参考系 R' 中存在电荷固定不动的分布. 若速度远小于光速, 则 R' 内测得的电场 \mathbf{E}' 与 R 内测得的电场 \mathbf{E} 、磁场 \mathbf{B} 有关:

$$\mathbf{E}' = \mathbf{E} + \mathbf{v} \times \mathbf{B}, \quad (2.5.22)$$

由于 R' 内测得的场是固定不动的电荷分布的场, 有

$$\nabla \times \mathbf{E}' = 0, \quad (2.5.23)$$

在 R 中, 电荷全以速度 \mathbf{v} 运动, 因此其位形从时刻 t 点 \mathbf{r} 到时刻 $t + \tau$ 点 $\mathbf{r} + \mathbf{v}\tau$ 严格相同, 因此

$$\begin{aligned} \mathbf{B}(\mathbf{r} + \mathbf{v}\tau, t + \tau) &= \mathbf{B}(\mathbf{r}, t), \\ \mathbf{E}(\mathbf{r} + \mathbf{v}\tau, t + \tau) &= \mathbf{E}(\mathbf{r}, t). \end{aligned}$$

于是取关于 τ 在 $\tau = 0$ 的导数

$$(\mathbf{v} \cdot \nabla) \mathbf{B} + \frac{\partial \mathbf{B}}{\partial t} = 0, \quad (2.5.24)$$

$$(\mathbf{v} \cdot \nabla) \mathbf{E} + \frac{\partial \mathbf{E}}{\partial t} = 0. \quad (2.5.25)$$

故必有

$$\begin{aligned} 0 &= \nabla \times \mathbf{E}' = \nabla \times \mathbf{E} + \nabla \times (\mathbf{v} \times \mathbf{B}) \\ &= \nabla \times \mathbf{E} + (\nabla \cdot \mathbf{B}) \mathbf{v} - (\mathbf{v} \cdot \nabla) \mathbf{B} \\ &= \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t}, \end{aligned}$$

式中用到 $\nabla \cdot \mathbf{B} = 0$. 由此可见

$$\nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} = 0, \quad (2.5.26)$$

上式对 \mathbf{B} 和 \mathbf{E} 是线性的；相对于 R 以不同速度运动的不同电荷流的磁场和电场相加起来，就得到对于运动电荷所激发的电场和磁场，此式普遍成立。它描述了磁场变化如何激发电流，在静态情形下

$$\frac{\partial \mathbf{B}}{\partial t} = 0,$$

化为 $\nabla \times \mathbf{E} = 0$ ，此即静电场保守条件；也即是说，当电荷沿一闭合路径运动，不存在净功。更普遍地，考虑一个闭合曲线 ∂S 形状的导线回路，令 S 是以 ∂S 为界的固定曲面，则根据 Stokes 定理，

$$\int_{\partial S} \mathbf{E} \cdot d\mathbf{r} = \int_S \nabla \times \mathbf{E} \cdot d\mathbf{S} = - \int_S \frac{\partial \mathbf{B}}{\partial t} \cdot d\mathbf{S} = - \frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S}, \quad (2.5.27)$$

最后一步是因为 S 固定。若磁场变化导致 \mathbf{B} 遍及 S 的积分非常量，则 \mathbf{E} 沿回路的积分不为零，将有一非零的电场沿着导线，它将对导线内的电子施以作用力，从而引起电流。这里

$$\int_{\partial S} \mathbf{E} \cdot d\mathbf{r}$$

用 V 量度，即单位电荷绕导线一周电场所做的功，称为沿电路的电动势，与电路的磁通量变化率相联系。

匀速运动电荷的场

根据电磁基本原理 3'，速度为 \mathbf{v} 的单一电荷 q 激发电场 \mathbf{E} 和磁场

$$\mathbf{B} = \frac{\mu_0 q \mathbf{v} \times \mathbf{r}}{4\pi r^3} + O(v^2/c^2),$$

式中 r 是从电荷到被测场点的矢量，在电荷静止的参考系 R' 中，它的电场是

$$\mathbf{E}' = \frac{q\mathbf{r}}{4\pi\epsilon_0 r^3},$$

在电荷以速度 \mathbf{v} 运动的参考系中 $\mathbf{E} = \mathbf{E}' + O(\mathbf{v}/c)$ ，因此

$$c\mathbf{B} = \frac{\mathbf{v} \times \mathbf{E}'}{c} = \frac{\mathbf{v} \times \mathbf{E}}{c} + O\left(\frac{v^2}{c^2}\right),$$

两边取旋度，省略 v^2/c^2 级的项：

$$\nabla \times (c\mathbf{B}) = \nabla \times \left(\frac{\mathbf{v} \times \mathbf{E}}{c} \right) = \frac{1}{c} ((\nabla \cdot \mathbf{E})\mathbf{v} - (\mathbf{v} \cdot \nabla)\mathbf{E}),$$

而由 (2.5.24)，

$$\nabla \cdot \mathbf{E} = \rho/\epsilon_0, \quad (\mathbf{v} \cdot \nabla)\mathbf{E} = -\frac{\partial \mathbf{E}}{\partial t},$$

因此

$$\nabla \times (c\mathbf{B}) - \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} = \frac{1}{c\epsilon_0} \mathbf{J} = c\mu_0 \mathbf{J},$$

式中 $\mathbf{J} = \rho \mathbf{v}$, 遍及各粒子速度求和得

$$\nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \mathbf{J}, \quad (2.5.28)$$

对任意电荷分布均成立, 条件是它们的速度均远小于光速.

Maxwell 方程组

基本原理连同速度远小于光速下的 Galileo 变换假设, 允许我们导出连续分布的运动电荷在不同的真空中激发的电场和磁场, 满足

$$\begin{aligned}\nabla \cdot \mathbf{E} &= \rho/\epsilon_0, \\ \nabla \cdot \mathbf{B} &= 0, \\ \nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} &= \mu_0 \mathbf{J}, \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{B}}{\partial t} &= 0.\end{aligned}$$

此即 Maxwell 方程组, 连同 Lorentz 力定律描述了电荷和电磁场的动力学.

我们得到它们是考虑基本电磁过程如何出现在运动参考系中, 这是一条不能令人满意的路线, 因为按这种方法作为演绎基础的原理, 同速度可与光速相比较下的 Galileo 变换是不相容的. Maxwell 导出它们是分析精心制作的电场和磁场的力学模型, 被看成是传光以太的位移. 这也是不能令人满意的, 因为这一模型已经抛弃很久了. 它们今天被接受作为电磁学理论和实际应用基础的理由很少与此两种论证有关. 首先它们是自洽的, 其次除开高频情况外, 它们以超乎寻常的精度描述了真实的场的行为.

连续性方程

方程组的自洽性并不明显, 给定 ρ 和 \mathbf{J} 作为坐标和时间的函数, 方程组是 \mathbf{E}, \mathbf{B} 未知分量的两个标量方程和两个矢量方程, 对于 6 个末知量总共有 8 个方程, 方程数多于末知量数, 因此有可能它们实际上是不相容的. 如果取方程组第四式的散度,

$$\frac{\partial}{\partial t} (\nabla \cdot \mathbf{B}) = 0,$$

这与方程 $\nabla \cdot \mathbf{B} = 0$ 是相容的, 因此这里没有问题. 然而取方程组第三式的散度, 代入 $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$ 得

$$\begin{aligned}0 &= \nabla \cdot \nabla \times \mathbf{B} = \frac{1}{c^2} \frac{\partial}{\partial t} (\nabla \cdot \mathbf{E}) + \mu_0 \nabla \cdot \mathbf{J} \\ &= \mu_0 \left(\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} \right),\end{aligned}$$

这给出一个矛盾，除非

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0, \quad (2.5.29)$$

所以 ρ 和 \mathbf{J} 的选择并非没有约束，它们必须通过这个连续性方程所关联，这一点只对物理上适当的电荷分布才成立，它是电荷不能创造也不能消灭的微分表述。

电荷守恒

为看出连续性方程和电荷守恒之间的联系，考察以曲面 S 为边界的固定体积 V 内的总电荷。如果电荷是守恒的，则在短时间内任何增加或减少必须与电荷通过 S 流入或流出精确地平衡掉。考虑 S 的面元 dS 有外法线，并考虑在时间 t 具有特定电荷 q 和特定速度 \mathbf{v} 的全部粒子，假定单位体积的粒子数是 n (n 是位置的函数)，在 t 和 $t + dt$ 之间穿过面元的粒子数等于在时间 t 位于体积区 $|\mathbf{v} \cdot dS dt|$ 内的粒子数，如图所示。它们对通过面元流出电荷的贡献为 $nq\mathbf{v} \cdot dS dt$ ，而 \mathbf{J} 在面元处的值

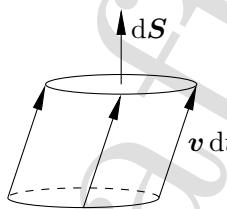


图 2.17: 通过面元的流出量

是 $nq\mathbf{v}$ 遍及 \mathbf{v} 和 q 所有可能值之和。因此，遍及 \mathbf{v}, e 和曲面的面元求和，并通过对连续分布取极限，总的流出率是

$$\int_S \mathbf{J} \cdot d\mathbf{S},$$

电荷守恒意味着流出率应等于 V 内总电荷的减少率，即

$$\frac{d}{dt} \int_V \rho dV + \int_S \mathbf{J} \cdot d\mathbf{S} = 0, \quad (2.5.30)$$

由 Gauss 定理，

$$\int_V \left(\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} \right) dV = 0, \quad (2.5.31)$$

如果此式对于任意选择的 V 成立，则 ρ 和 \mathbf{J} 必须满足连续性方程；反过来，连续性方程意味着电荷守恒。

位移电流

Maxwell 方程组第三式可写成

$$\nabla \times \mathbf{B} = \mu_0 \left(\mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right), \quad (2.5.32)$$

在这一表述中，可看成未知磁场 \mathbf{B} 依据已知电流分布 \mathbf{J} 和电场 \mathbf{E} 的方程，当 \mathbf{E}, \mathbf{J} 与 t 无关，则化为

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J},$$

它确定恒定电流的磁场，在某种程度上与 Maxwell 同时代的人已经熟知。但是右边的第二项 $\epsilon_0 \frac{\partial \mathbf{E}}{\partial t}$ 是新的，它加到 \mathbf{J} 上称为真空间移电流，名称来自类似于绝缘物质中电荷的行为。这里非恒定电流可能流动，但是物质内的电荷分布被外电场扭曲了。场改变，扭曲也改变，结果显示一种电流，即位移电流，其在改变期间流动。Maxwell 的远见卓识就是这一相同的项即使在真空中也应存在。结论是深远的，它允许 Maxwell 进而把光的传播解释为电磁现象。

真空方程

在远离激发电场和磁场的电荷的真空中可看作无源，Maxwell 方程组就成为了真空方程。在第一章中我们已经知道其可导出波动方程：

$$\frac{1}{c^2} \frac{\partial^2 \mathbf{B}}{\partial t^2} = \nabla^2 \mathbf{B}, \quad \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = \nabla^2 \mathbf{E}. \quad (2.5.33)$$

因此在真空中 \mathbf{E}, \mathbf{B} 的 3 个分量满足标量方程

$$\square E^i = 0, \quad \square B^i = 0, \quad (2.5.34)$$

这里 \square 是 *d'Alembert* 算符，定义为

$$\square = -\frac{1}{c^2} \frac{\partial^2}{\partial t^2} + \nabla^2. \quad (2.5.35)$$

单色平面波 *

\mathbf{E}, \mathbf{B} 是真空中波动方程矢值解的事实表明：我们找到了 Maxwell 方程组的“平面波”解，其中

$$\mathbf{E} = \boldsymbol{\alpha} \cos \Omega + \boldsymbol{\beta} \sin \Omega, \quad (2.5.36)$$

式中 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 是常矢量，而

$$\Omega = \frac{\omega}{c}(ct - \mathbf{r} \cdot \mathbf{e}) \quad \mathbf{e} \cdot \mathbf{e} = 1, \quad (2.5.37)$$

$\omega > 0, \boldsymbol{\alpha}, \boldsymbol{\beta}$ 和 \mathbf{e} 是常量； ω 是频率， \mathbf{e} 是给出传播方向的单位矢量 (t 增加 τ 且 \mathbf{r} 增加 $c\tau\mathbf{e}$ 保持 Ω 不变)。这满足波动方程，如果常量不可普遍选取，将不可能找到使真空方程也成立的 \mathbf{B} 。取上式的散度得

$$\nabla \cdot \mathbf{E} = \frac{\omega}{c}(\mathbf{e} \cdot \boldsymbol{\alpha} \sin \Omega - \mathbf{e} \cdot \boldsymbol{\beta} \cos \Omega) = 0, \quad (2.5.38)$$

即必须选择 α, β 使其垂直于 e . 而另一方面, 必须找到 B 使

$$\nabla \times \mathbf{E} = \frac{\omega}{c}(\mathbf{e} \times \boldsymbol{\alpha} \sin \Omega - \mathbf{e} \times \boldsymbol{\beta} \cos \Omega) = -\frac{\partial \mathbf{B}}{\partial t}, \quad (2.5.39)$$

一种可能的选择是

$$\mathbf{B} = \frac{\mathbf{e} \times \mathbf{E}}{c} = \frac{1}{c}(\mathbf{e} \times \boldsymbol{\alpha} \cos \Omega + \mathbf{e} \times \boldsymbol{\beta} \sin \Omega). \quad (2.5.40)$$

不难看出 \mathbf{E}, \mathbf{B} 同样满足真空方程. 这种方法得到的解称为单色平面电磁波. 注意, \mathbf{E} 和 \mathbf{B} 垂直于传播方向, 在此意义下这样的波是横波, \mathbf{E} 的定义可更简要地写成

$$\mathbf{E} = \operatorname{Re}\{(\boldsymbol{\alpha} + i\boldsymbol{\beta})e^{-i\Omega}\}, \quad (2.5.41)$$

用 Fourier 分析可证明真空中每个解都是单色平面波的组合. 如果 $\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 成正比, 平面波是“平面”偏振的或“线”偏振的; 如果 $\boldsymbol{\alpha} \cdot \boldsymbol{\alpha} = \boldsymbol{\beta} \cdot \boldsymbol{\beta}, \boldsymbol{\alpha} \cdot \boldsymbol{\beta} = 0$, 平面波则是“圆”偏振的.

Maxwell 理论的核心概念是具有确定频率或一定色的光波可用 Maxwell 方程组的单色平面解来表示.

势

对真空中 Maxwell 方程组的每个解, \mathbf{E}, \mathbf{B} 的分量满足 3 维波动方程, 但反过来则是不对的, 也就是说如果

$$\square \mathbf{B} = 0, \quad \square \mathbf{E} = 0,$$

则说 \mathbf{E}, \mathbf{B} 满足 Maxwell 方程组, 一般是不成立的. 要想发生这种情形, 两个场必须满足真空方程. 如果不用场而是用称为“势”的辅助量来处理, 这种附加的约束要稍许简单些. 势的定义依赖于矢量微积分里的标准可积条件⁴⁵: 假定 \mathbf{F} 是依赖于时间的矢量场, 若 $\nabla \times \mathbf{F} = 0$, 则存在函数 ϕ 使 $\mathbf{F} = \nabla \phi$; 若 $\nabla \cdot \mathbf{F} = 0$, 则存在矢量场 \mathbf{A} 使 $\mathbf{F} = \nabla \times \mathbf{A}$. ϕ 和 \mathbf{A} 都不能由 \mathbf{F} 唯一确定: 总可用 $\phi' = \phi + f$ 代替 ϕ , 这里 f 仅为时间的函数; 总可用 $\mathbf{A}' = \mathbf{A} + \nabla u$ 代替 \mathbf{A} , 这里 u 是位置和时间的任意标量场.

现假设给定场 \mathbf{E}, \mathbf{B} 满足用电荷密度 ρ 和电流密度 \mathbf{J} 代表的有源 Maxwell 方程组, 因为 $\nabla \cdot \mathbf{B} = 0$, 存在依赖于时间的矢量场 $\mathbf{A}(t, x, y, z)$ 使

$$\mathbf{B} = \nabla \times \mathbf{A}, \quad (2.5.42)$$

⁴⁵ 现在的陈述是局域的. 以后会知道, 如果 \mathbf{F} 在拓扑非平凡的区域 U 上有定义, 则不可能找到遍及整个 U 均合适的 ϕ 或 \mathbf{A} .

则

$$\nabla \times \left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} \right) = 0,$$

由此得出存在一个标量 $\phi(t, x, y, z)$ 使

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}. \quad (2.5.43)$$

这一矢量场 \mathbf{A} 称为磁矢势, 使上式成立的函数 ϕ 称为电标势. 反过来, 给定 t, x, y, z 的标量场 ϕ 和矢量场 \mathbf{A} , 可通过其定义 \mathbf{B}, \mathbf{E} , 于是因为 Maxwell 方程组中的两式自动成立, 剩下的一对方程转变为 \mathbf{A}, ϕ 的关系:

$$\nabla \cdot \mathbf{E} = -\nabla^2\phi - \frac{\partial}{\partial t}(\nabla \cdot \mathbf{A}) = \frac{\rho}{\epsilon_0}$$

和

$$\nabla \times \mathbf{B} - \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} = -\nabla^2 \mathbf{A} + \nabla(\nabla \cdot \mathbf{A}) + \frac{1}{c^2} \frac{\partial}{\partial t} \left(\nabla\phi + \frac{\partial \mathbf{A}}{\partial t} \right) = \mu_0 \mathbf{J},$$

如果令

$$(A^\mu) = (\phi/c, A^1, A^2, A^3), \quad (2.5.44)$$

则可将 \mathbf{A}, ϕ 的方程改写成更简练的

$$\begin{aligned} -\square\phi - \frac{\partial^2 A^\mu}{\partial t \partial x^\mu} &= \frac{\rho}{\epsilon_0}, \\ -\square\mathbf{A} + \nabla \frac{\partial A^\mu}{\partial x^\mu} &= \mu_0 \mathbf{J}. \end{aligned}$$

这里对 4 个未知量 (ϕ 和 \mathbf{A} 的分量) 有 4 个方程 (一个标量方程, 一个矢量方程). 任意一组解 ϕ, \mathbf{A} 通过 \mathbf{E}, \mathbf{B} 定义确定 Maxwell 方程组的一个解.

规范变换

给定 Maxwell 方程组的解 \mathbf{E}, \mathbf{B} , 在 \mathbf{A}, ϕ 的选择上还有什么自由度? 首先, \mathbf{A} 被 $\nabla \times \mathbf{A} = \mathbf{B}$ 确定到可用 $\mathbf{A}' = \mathbf{A} + \nabla u$ 替代 \mathbf{A} , 其中 u 是位置和时间的函数. 相应于 \mathbf{A}' 的标势 ϕ' 必须选择使

$$\begin{aligned} -\nabla\phi' &= \mathbf{E} + \frac{\partial \mathbf{A}'}{\partial t} = \mathbf{E} + \frac{\partial \mathbf{A}}{\partial t} + \nabla \left(\frac{\partial u}{\partial t} \right) \\ &= -\nabla \left(\phi - \frac{\partial u}{\partial t} \right), \end{aligned}$$

也就是说, $\phi' = \phi - \partial u / \partial t + f(t)$, 式中 f 仅为 t 的函数, 我们可以通过减去

$$\int f dt$$

把 f 吸收到 u 中 (这不改变 \mathbf{A}'), 所以在选择 \mathbf{A} 和 ϕ 的自由度是对任意的 $u = u(t, x, y, z)$ 取变换

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla u, \quad \phi \rightarrow \phi' = \phi - \frac{\partial u}{\partial t}, \quad (2.5.45)$$

称为规范变换. 其下

$$\frac{\partial A^\mu}{\partial x^\mu} \rightarrow \frac{\partial A'^\mu}{\partial x^\mu} = \frac{1}{c^2} \frac{\partial \phi'}{\partial t} + \nabla \cdot \mathbf{A}' = \frac{\partial A^\mu}{\partial x^\mu} - \square u,$$

在 $\frac{\partial A^\mu}{\partial x^\mu}$ 相当温和的条件下, 可证明非齐次波动方程

$$\square u = \frac{\partial A^\mu}{\partial x^\mu} \quad (2.5.46)$$

有解 $u = u(t, x, y, z)$. 如果选择 u 使上式成立, 则被变换的势 \mathbf{A}' 和 ϕ' 满足条件

$$\nabla \cdot \mathbf{A}' + \frac{1}{c^2} \frac{\partial \phi'}{\partial t} = 0, \quad (2.5.47)$$

此即 Lorenz 规范⁴⁶. 如果我们要求该规范, 则 \mathbf{A}, ϕ 的选择上唯一剩下的自由度是作规范变换, 其中 u 是波动方程 $\square u = 0$ 的解. 在 Lorenz 规范下, Maxwell 方程组取

$$\phi = \rho/\epsilon_0, \quad \square \mathbf{A} = \mu_0 \mathbf{J}. \quad (2.5.48)$$

与 Lorenz 规范的相容性可根据 ρ, \mathbf{J} 的连续性方程得出.

所以在无源情形下, 关于势的 Maxwell 方程组在 Lorenz 规范下化为

$$\square \phi = 0, \quad \square \mathbf{A} = 0, \quad (2.5.49)$$

例如, 我们可选择失势分量的标量波动方程的 3 个任意解, 从而用

$$\phi = c^2 \int \nabla \cdot \mathbf{A} dt \quad (2.5.50)$$

来定义 ϕ . 无论取怎样的选择, 我们将获得 Maxwell 方程组的一个解, 并且真空方程的每个解都是由某个这种选择所引起的.

2.5.2 几何语言

电磁张量 F

\mathbf{R}^{3+1} 中的一个荷电粒子由 (α, m, q) 代表, 其中 (α, m) 是一个有质粒子, 而非零实数 q 称为这个粒子的电荷. 关于荷电粒子有两件事情我们感兴趣: 荷电粒子的存在会产生电磁场, 而这些荷电粒子又受到其他电荷所产生的电磁场作用.

⁴⁶这个条件是由丹麦物理学家 Ludvig Valentin Lorenz(1829-1891) 给出的, 所以是 Lorenz 规范, 当然它又是 Lorentz 不变的, 所以也可叫做 Lorentz 规范. 二人名字的发音相同, 容易混淆, 需要注意 Lorenz 是荷兰的物理学家, 是相对论的奠基者之一.

荷电粒子通过改变它们的 4-动量来表达它们对外部电磁场的“反应”. 这种反应的定量表示, 也就是运动方程, 通常被看作是所谓的 4 维 Lorentz 力定律, 这个力作为 4-速的线性函数, 等于粒子的 4-动量在世界线上的每个点对固有时的变化率. 所以, 在世界线上的处处有

$$\frac{dP(\tau)}{d\tau} = q\tilde{F}_{\alpha(\tau)}(U(\tau)), \quad (2.5.51)$$

其中, $\tilde{F}_{\alpha(\tau)} : M \rightarrow M$ 是在惯性坐标系中由经典电场 E 和磁场 B 决定的线性变换 (假定 q 对周围电磁场的贡献忽略不计, 即 (α, m, q) 是一个“试验电荷”). 把上式简单写成

$$\tilde{F}(U) = \frac{m}{q} \frac{dU}{d\tau},$$

用 U 点乘得

$$\tilde{F}(U) \cdot U = \frac{m}{q} \frac{dU}{d\tau} \cdot U = \frac{m}{2q} \frac{d}{d\tau}(U \cdot U) = \frac{m}{2q} \frac{d}{d\tau}(-1) = 0.$$

因为任意指向未来的类时单位矢量 U 是某个荷电粒子的 4-速, 我们则发现对于任何这样的矢量, 都有 $\tilde{F}(U) \cdot U = 0$, 那么线性意味着对任何类时矢量都有 $\tilde{F}(V) \cdot V = 0$. 设 U 和 V 都是指向未来的类时矢量, 则 $U + V$ 就是类时的, 所以 $0 = \tilde{F}(U + V) \cdot (U + V) = \tilde{F}(U) \cdot V + U \cdot \tilde{F}(V)$, 即 $\tilde{F}(U) \cdot V = -U \cdot \tilde{F}(V)$. 在 \mathbf{R}^{3+1} 中具有未来指向的类时基矢下, 对所有的 $x, y \in \mathbf{R}^{3+1}$ 有

$$\tilde{F}(x) \cdot y = -x \cdot \tilde{F}(y), \quad (2.5.52)$$

因此, 在每一点, 线性变换 \tilde{F} 相对于 Lorentz 度规必须是反称的. 因此我们可以构造 \mathbf{R}^{3+1} 上的电磁场, 即对该反称线性变换的每一点制定一个荷电粒子的 4-动量, 这个粒子的世界线通过这一点, 而由于电磁场的存在, 这个粒子要经历 4-动量的改变. 另一种稍微不同的方法更方便. 注意到一个反称的线性变换 $\tilde{F} : M \rightarrow M$ 和 Lorentz 度规一起决定了下面给出的双线性形式 $F : M \times M \rightarrow \mathbf{R}$,

$$F(x, y) = \tilde{F}(x) \cdot y,$$

这显然是反称的, 即 $F(y, x) = -F(x, y)$, 而且反过来, 一个反称双线性形式唯一决定一个反称线性变换. 其实, 在 \mathbf{R}^{3+1} 的每一点指定一个反称双线性形式就是 \mathbf{R}^{3+1} 上的一个 2-形式, 因此, 可以使用微分形式的语言来表达经典电磁理论.

在自然界所给予的确定限制下, 2-形式能够合理地代表 \mathbf{R}^{3+1} 上的电磁场, 即 Maxwell 方程. 为了尝试微分形式语言的改写, 自然地定义如下的场源 1-形式 J . 如果 $\{x^\mu\}$ 是 \mathbf{R}^{3+1} 上的任意惯性坐标系, 那么

$$J_\mu = \rho \eta_{\mu\nu} U^\nu, \quad (2.5.53)$$

即

$$J = \rho U_b = -\rho dx^0 + j_i dx^i, \quad (2.5.54)$$

其中, $\rho : M \rightarrow \mathbf{R}$ 是电荷密度函数, U_b 是速度按 η 的对偶, $j = j^i e_i$ 是电流密度矢量场 (它们通常是光滑函数, 分别是相应的惯性观者所逐点测量的“单位体积电荷”和“单位面积单位时间的荷流”). 现在, 给出如下的正式定义: 由 \mathbf{R}^{3+1} 上的源的 1-形式 J 所决定的 \mathbf{R}^{3+1} 上的电磁场是 \mathbf{R}^{3+1} 上的 2-形式 F , 满足 Maxwell 方程

$$dF = 0 \quad (2.5.55)$$

和

$$\star d\star F = J. \quad (2.5.56)$$

下面依次给出几个讨论. 以下所选的单位制使得: 光速和 Maxwell 方程中的各种其他常数都是 1, 而且在以上 Maxwell 的第二式中 4π 的因子“归一化”了. 我们发现 Maxwell 的第二式等价于

$$d\star F = \star J,$$

在没有电荷的区域, 即 $J = 0$, 得到无源的真空 Maxwell 方程

$$dF = 0, \quad d\star F = 0.$$

这就是说, F 和 $\star F$ 都是闭 2-形式. \mathbf{R}^{3+1} 上的任何 2-形式 F 在任意惯性坐标系中可以写成 $F = (1/2)F_{\mu\nu} dx^\mu \wedge dx^\nu$, 其中 $(F_{\mu\nu})$ 是 F 的反称矩阵. 为了同物理学中使用的记号进行联系, 我们引入如下定义:

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & -E^1 & -E^2 & -E^3 \\ E^1 & 0 & B^3 & -B^2 \\ E^2 & -B^3 & 0 & B^1 \\ E^3 & B^2 & -B^1 & 0 \end{pmatrix},$$

因此,

$$\begin{aligned} F = & E^1 dx^1 \wedge dx^0 + E^2 dx^2 \wedge dx^0 + E^3 dx^3 \wedge dx^0 \\ & + B^3 dx^1 \wedge dx^2 + B^2 dx^3 \wedge dx^1 + B^1 dx^2 \wedge dx^3. \end{aligned}$$

计算 $\star F, dF, d\star F$ 和 $\star d\star F$ 并且记 $E = E^i e_i$ 和 $B = B^i e_i$, 我们发现 $dF = 0$ 等价于

$$\operatorname{div} B = 0, \quad \operatorname{curl} E + \frac{\partial B}{\partial t} = 0,$$

而 $\star d\star F = J$ 等价于

$$\operatorname{div} E = \rho, \quad \operatorname{curl} B - \frac{\partial E}{\partial t} = J.$$

此即 Maxwell 方程组的通俗表达.

在 \mathbf{R}^{3+1} 上的另一惯性坐标系 $\{\hat{x}^\alpha\}$ 下, 2-形式 F 写成 $F = (1/2)\tilde{F}_{\alpha\beta} dx^\alpha \wedge d\hat{x}^\beta$. 代入 $\hat{x}^\alpha = \Lambda^\alpha_\mu x^\mu$ 有 $F = (1/2) (\Lambda^\alpha_\mu \Lambda^\beta_\nu \tilde{F}_{\alpha\beta}) dx^\mu \wedge dx^\nu$, 所以有

$$F_{\mu\nu} = \Lambda^\alpha_\mu \Lambda^\beta_\nu \tilde{F}_{\alpha\beta},$$

现在假定一个匀速运动电荷的电磁场. 按照相对性原理, 究竟是认为电荷相对于一个“固定的”惯性观者运动, 还是认为这个观者相对于一个“静止的”电荷运动, 完全无关紧要. 因此, 固定在戴帽坐标系原点的一个电荷产生的电场 (“Coulomb 定律”), 由上式变换到相对于这个戴帽坐标系运动的不戴帽坐标系. 在 $\{\hat{x}^\alpha\}$ 下, 在空间原点放置的点电荷 q 由平方反比定律给出 $\hat{B} = 0$ 和 $\hat{E} = (q/\hat{r}^3)\hat{r}$, 其中 $\hat{r} = \hat{x}^i \hat{e}_i$ 且 $\hat{r} = ((\hat{x}^1)^2 + (\hat{x}^2)^2 + (\hat{x}^3)^2)^{1/2}$ (注意, \hat{E} 只是在 \mathbf{R}^{3+1} 上由 \hat{e}_0 测量的). 因此,

$$(\hat{F}_{\mu\nu}) = \frac{q}{\hat{r}^3} \begin{pmatrix} 0 & -\hat{x}^1 & -\hat{x}^2 & -\hat{x}^3 \\ \hat{x}^1 & 0 & 0 & 0 \\ \hat{x}^2 & 0 & 0 & 0 \\ \hat{x}^3 & 0 & 0 & 0 \end{pmatrix}.$$

很容易在其范围内证实, $(\hat{F}_{\mu\nu})$ 满足无源 Maxwell 方程. 把 Λ 取成 boost 变换, 并把 $F_{\mu\nu}$ 用上式给出的 $(\hat{F}_{\mu\nu})$ 写出来, 得

$$\begin{aligned} E^1 &= q \frac{\hat{x}^1}{\hat{r}^3}, & B^1 &= 0, \\ E^2 &= \frac{q}{\sqrt{1-v^2}} \frac{\hat{x}^2}{\hat{r}^3}, & B^2 &= \frac{-qv}{\sqrt{1-v^2}} \frac{\hat{x}^3}{\hat{r}^3}, \\ E^3 &= \frac{q}{\sqrt{1-v^2}} \frac{\hat{x}^3}{\hat{r}^3}, & B^3 &= \frac{qv}{\sqrt{1-v^2}} \frac{\hat{x}^2}{\hat{r}^3}. \end{aligned}$$

可以使用不戴帽观者, 在电荷通过其空间原点时所做的测量来表达这些式子. 在 boost 中代入 $x^0 = 0$ 后得到

$$\hat{x}^1 = \frac{1}{\sqrt{1-v^2}} x^1, \hat{x}^2 = x^2, \hat{x}^3 = x^3,$$

所以

$$\hat{r}^2 = \frac{1}{1-v^2} (x^1)^2 + (x^2)^2 + (x^3)^2.$$

为了方便写成 \hat{r}_v^2 . 在 $(\hat{F}_{\mu\nu})$ 做了这些代替后有

$$\begin{aligned} E &= \frac{q}{\sqrt{1-v^2}} \left(\frac{1}{r_v^3} \right) x^i e_i \\ &= \frac{q}{\sqrt{1-v^2}} \left(\frac{1}{r_v^3} \right) r \end{aligned}$$

和

$$\begin{aligned} B &= \frac{q}{\sqrt{1-v^2}} \left(\frac{1}{r_v^3} \right) (-vx^3 e_2 + vx^2 e_3) \\ &= \frac{q}{\sqrt{1-v^2}} \left(\frac{1}{r_v^3} \right) ((ve_1) \times r), \end{aligned}$$

这是以速度 ve_1 匀速运动的电荷，在其通过坐标原点时所产生的电磁场。当 $v \ll 1$, $r_v \approx r$, 那么上式就说明，缓慢运动电荷的电场近似于 Coulomb 场；同理，磁场也退化为 Biot-Savart 定律。

我们来考虑另一简单应用：在某惯性观者下，电磁场是恒稳的纯磁场，荷电粒子 (α, m, q) 会作何反应。为了简单，假设对该观者来说， $E = 0$ 且 $B = be_3$ ，其中 b 是非零常数。相应的 2-形式 F 具有分量

$$(F_{\mu\nu}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & -b & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

相应的线性变换 \hat{F} 在这组基矢下具有同样的矩阵，所以如果 $\alpha(\tau) = x^\mu(\tau)e_\mu$ 且 $U(\tau) = U^\mu(\tau)e_\mu$ ，那么 4-Lorentz 力定律退化为线性微分方程组

$$\begin{aligned} \frac{dU^1}{d\tau} &= \frac{bq}{m} U^2 & \frac{dU^2}{d\tau} &= -\frac{bq}{m} U^1 \\ \frac{dU^3}{d\tau} &= 0 & \frac{dU^0}{d\tau} &= 0. \end{aligned}$$

可以解得

$$\begin{aligned} \alpha(\tau) &= x_0 + a \sin \left(\frac{bq\tau}{m} + \phi \right) e_1 + a \cos \left(\frac{bq\tau}{m} + \phi \right) e_2 \\ &\quad + c\tau e_3 + \left(1 + \frac{a^2 b^2 q^2}{m^2} + c^2 \right) \tau e_0, \end{aligned}$$

其中， $x_0 = x_0^\mu e_\mu \in \mathbf{R}^{3+1}$ 是常数， a, ϕ 和 c 是常实数且 $a > 0$ (已用 $U \cdot U = -1$ 排除负数)。注意， α 上处处 $(x^1 - x_0^1)^2 + (x^2 - x_0^2)^2 = a^2$ 。因此，若 $c \neq 0$ ，那么在这个坐标系中的空间轨迹是一条沿 e_3 方向 (即磁力线方向) 的螺旋线；若 $c = 0$ ，那么这条轨迹就是一个在 x^1 - x^2 平面中的圆。在气泡室中引入恒稳磁场，就可以看到这种粒子沿一个圆轨道运动。可以利用这一点测量粒子的荷质比。上式取 $c = 0$ 并计算 $U(\tau)$ ，利用 $U = \gamma(v + e_0)$ 求解粒子的坐标速度得

$$v = \frac{abq/m}{\sqrt{1-v^2}} \left(\cos \left(\frac{bq\tau}{m} + \phi \right) e_1 + \sin \left(\frac{bq\tau}{m} + \phi \right) e_2 \right),$$

由此计算得常数

$$v^2 = \left(1 + \frac{m^2}{a^2 b^2 q^2}\right)^{-1}.$$

求解 q/m (为了方便, 假定 $q > 0$) 得

$$\frac{q}{m} = \frac{1}{a|b|} \frac{|v|}{\sqrt{1 - v^2}}.$$

因为 a, b 和 $|v|$ 都是可测量的, 所以能得到所要求的荷质比.

电磁 4-势 A

接下来简单讨论存在性并利用电磁场的“势”概念. 设 F 是定义在 \mathbf{R}^{3+1} 中的某块区域 X 上的电磁场, 那么 F 就是 X 上的 2-形式, 由 Maxwell 方程 ($dF = 0$), 其是闭的. 由于 $X \subset \mathbf{R}^4$, 其上每个闭 2-形式都是恰当的, 故存在 X 上的 1-形式 A 满足

$$F = dA.$$

A 当然不是唯一的, 因为 $d^2 = 0$, 如果 A 满足, 那 $A + df$ 也满足, 其中 0-形式 f 是定义在 X 上的任意光滑函数. 只有当选取一定规范⁴⁷使其确定时, A 才能看成是基本的物理对象. 这种 1-形式 A 称为 F 的(规范)位势, 简称 4-势. 对于某个 f , 平移 $A \rightarrow A + df$ 称为势的规范变换, 平移的自由度称为规范自由度. 可以看出, 给定 F , 对于按照 0-形式 $\star d \star A$ 定义的 A , 总能够局域求解 $dA = F$. 更严格地说, 如果 F 满足 $dF = 0$ 且 g 是任意的 0-形式, 那么在任意点的邻域上, 存在一个 1-形式 A 满足

$$dA = F \quad \text{和} \quad \star d \star A = g.$$

上式第二个条件可以如下解释⁴⁸. 设给出有关电荷和电流的某种(物理)构成, 即源的 1-形式 J , 我们想要求出相应的电磁场 F , 就必须求解 Maxwell 方程 $dF = 0$ 和 $\star d \star F = J$ (依边界条件是否合适而定). 至少在局域上可以寻求替代相应的势 A (所以 $F = dA$), 那么 Maxwell 方程组的第一个方程自动满足 ($dF = d(dA) = 0$), 故而只需求解方程 $\star d \star (dA) = J$. 有

$$\Delta A = d(\delta A) + \delta(dA) = d(\star d \star A) + \star d \star (dA).$$

按照上面给出的结果, 利用 $\star d \star A = 0$ 的条件简化前面的讨论, 于是

$$\delta A = 0,$$

⁴⁷ 该词首先由 Weyl 使用.

⁴⁸ [8] 的附录二中证明了更一般的结果, 而且在 2.9 节有更一般的情况

此即 Lorenz 规范。该规范下，自动有 $\Delta A = \star d\star(dA)$ ，那么第二个 Maxwell 方程就是

$$\Delta A = J.$$

因此，我们看到（局域）求解给定源 J 的 Maxwell 方程，简化为对势 A 求解 Lorentz 规范和上式。为了了解问题的求解是如何简化的，在惯性坐标系中计算，会发现算子 Δ 简化为在函数上定义的 d'Alembert 算子。因此， $\Delta A = J$ 退耦为波动表达 $\square A_\mu = J_\mu$ 。

电磁场的能动张量

常见的电动力学教材已使用 3 维语言定义了电磁场的能量密度、能流密度（动量密度）、动量密度、动流密度（应力张量）这些概念。这些 3 维量可统一表为一个 4 维的能动张量

$$T_{\mu\nu} = \frac{1}{4\pi} \left(F_{\mu\sigma} F_\nu^\sigma - \frac{1}{4} \eta_{\mu\nu} F_{\sigma\lambda} F^{\sigma\lambda} \right), \quad (2.5.57)$$

利用对偶形式语言可改写为

$$T_{\mu\nu} = \frac{1}{8\pi} (F_{\mu\sigma} F_\nu^\sigma + \star F_{\mu\sigma} \star F_\nu^\sigma), \quad (2.5.58)$$

仍选一惯性系可得

$$T_{00} = \frac{1}{8\pi} (E^2 + B^2), \quad -T_{i0} = \frac{1}{4\pi} (\vec{E} \times \vec{B})_i, \quad (2.5.59)$$

这正是该系惯性观者测电磁场所得能量密度、能流密度。

当 $J = 0$ 时，由 Maxwell 方程可证 $\partial^\mu T_{\mu\nu} = 0$ ，这里 $\partial^\mu = \eta^{\mu\nu} \frac{\partial}{\partial x^\nu}$ 。因此无源电磁场服从能量、动量和角动量守恒，但有源时上述 $T_{\mu\nu}$ 不满足 $\partial^\mu T_{\mu\nu} = 0$ 。从物理上，因为电磁场与带电质点间有相互作用，从而要交换能量、动量和角动量。然而，电磁场同荷电粒子的总能动张量仍是守恒的。

2.6 流体动力学

2.6.1 尘埃

2.6.2 连续介质的能动张量

定义 2.84. $W^\mu = -T^{\mu\nu} Z^\nu$ 称为瞬时观者 (p, Z) 所测的 4-动量密度。

本节主要讨论理想流体⁴⁹。

⁴⁹关于非理想流体，读者可提前参阅 Weinberg 及 Landau 等书。

Draft

Chapter 3

弯曲时空

3.1 几何学拾遗

3.1.1 协变微分

本节我们用严格术语重新定义协变微分。区别于外微分，协变微分要求背景附加额外结构——仿射联络 (*affine connection*)，以下简称联络¹。

先类比分析学的方向导数一步步地引出可作用于张量的联络概念。

定义 3.1 (方向导数). 背景上的一个联络 ∇ 是指一种法则，其给任意矢量场 X 赋以算子 ∇_X ，使得任意可导²矢量场 Y 映射为另一矢量场 $\nabla_X Y$ ，满足：

(张量性) $\nabla_X Y$ 是以 X 为变元的张量。即对任意函数 f, g 和可导矢量场 X, Y, Z 有 $\nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z$ 。此即要求 ∇_X 仅依赖于某点处 X 的方向；

(线性性) $\nabla_X Y$ 对 Y 是线性的。即 $\nabla_Z(\alpha X + \beta Y) = \alpha \nabla_Z X + \beta \nabla_Z Y$ ，这里 α, β 为实数；

(对函数) 对任意可导函数 f 和可导矢量场 Y 有 $\nabla_X(fY) = X(f)Y + f\nabla_X Y$ ，这里 $X(f)$ 在坐标基上表示为 f 的方向导数 $X^\mu \frac{\partial f}{\partial x^\mu}$ 。

如上 $\nabla_X Y$ 称为 Y 沿 X 方向关于 ∇ 的导数。

由张量性，还可直接定义 Y 关于 ∇ 的导数 ∇Y 定义为 $(1, 1)$ 型张量场，而它与 X 缩并后产生矢量 $\nabla_X Y$ 。于是亦可定义为

定义 3.2 (协变微分). C^{r+2} 流形上的一个 C^r 联络是指一种法则，其给任意点赋以协变微分 (取 *Lorentz* 协变性之意) ∇ ，使得任意 C^{r+1} 矢量场 Y 映射为 $(1, 1)$ 型

¹ 第 4 章会讲到，一般的联络可定义在各种“空间”上，定义在流形矢量丛上的称为仿射联络，此即本章所讨论的，且一律简称联络。

² 矢量场可导指矢量场按任意坐标基展开的分量为坐标域上的可导函数。

C^r 张量场 ∇Y , 满足:

(线性性) $\nabla(\alpha X + \beta Y) = \alpha \nabla X + \beta \nabla Y$, 这里 α, β 为实数;

(对函数) 对任意可导函数 f 和可导矢量场 Y 有 $\nabla(fY) = df \otimes Y + f \nabla Y$, 这里为绕开外微分概念, d 可仅理解为函数的微分。

在坐标域上, 给定任意 C^{r+1} 矢量基 $\{e_\mu\}$, 则 ∇Y 的展开就写为

$$\nabla Y = \nabla_\mu Y^\nu e^\mu \otimes e_\nu, \quad (3.1.1)$$

这里 ∇_μ 称为协变导数; 而方向导数就在相同基下写作

$$\nabla_X Y = X^\mu \nabla_\mu Y^\nu e_\nu. \quad (3.1.2)$$

可见 ∇ 实际上就相当于给其作用对象添加一个下标, 于是再结合一些微分学的特性即可推广至任意张量:

定义 3.3 (联络). 联络 ∇ 是背景上的一个算符场, 其将 (k, l) 型 $C^r (r \geq 1)$ 张量场映为 $(k, l+1)$ 型 C^{r-1} 张量场, 满足

(线性性) $\nabla(\alpha T + \beta S) = \alpha \nabla T + \beta \nabla S$, 这里 α, β 为实数而 T, S 同型;

(对函数) 对任意可导函数 f 有 $\nabla f = df$ (注意此处非梯度);

(Leibniz 律) 对任意张量 T, S 有 $\nabla(T \otimes S) = T \otimes \nabla S + S \otimes \nabla T$;

(缩并对易性) $[\nabla, C] = 0$, 这里 C 代表缩并算子。

后两条保证我们可推导任意张量的方向导数或协变微分。一般的所谓“微分算符 (operator)”或其相应的“导数算符”并不要求其作用结果仍为张量场 (但满足联络其它要求), 因此可能算符自身就与坐标系选择 (或其它人为因素) 有关³。有时为区别于微分算符, 可将协变微分记作 D 。

定义 3.4 (沿曲线求导). 若 T 是沿 C^1 曲线 $\lambda(t)$ 定义的 C^1 张量场 (只在路径上定义), 则我们还可定义 T 沿 $\lambda(t)$ 关于所给联络的导数为

$$\frac{DT}{dt} = \nabla_{\partial_t} \bar{T}, \quad (3.1.3)$$

这里 \bar{T} 表示将 T 以任意方式延拓至路径某开邻域的 C^1 张量场, 即选取任意范围包含路径的张量场 \bar{T} , 但满足在路径上取值为 T 且可导。 $\frac{DT}{dt}$ 只在路径上定义。

定义 3.5 (平行移动). 若 $DT/dt = 0$, 则称 T 沿曲线关于所给联络平行移动。

³除非强行将结果理解为“坐标依赖的张量场”, 但只有少数书秉持这种思想, 如 Wald, 见第 4 章抽象指标简介一节。本书将张量场定义为“背景上像为张量的单值映射”, 在一个背景点只赋以唯一张量, 因此所谈张量场默认“坐标无关”。

给定具有端点 p, q 的曲线 λ , 将任意给定张量由 p 出发沿曲线平行移动, 即可得到 q 点的唯一张量。因此, 沿曲线 λ 的平行移动是从 p 处张量空间到 q 处同型张量空间的线性映射, 且保持所以张量积、缩并不变。因此, 在特殊情形下, 若沿给定曲线从 p 至 q 平行移动矢量基, 就可决定从 V_p 至 V_q 的一个同构 (若曲线自相交, 则 p, q 也可以重合), 这样便联系了两点的矢量空间, 这便是称 ∇ 为“联络”的由来。通常考虑这样的特殊情形: 沿曲线 λ 的切矢量自身关于联络的导数。

定义 3.6 (测地线). 若沿 $\lambda(t)$ 的 $\nabla_X X$ 平行于 $X = \frac{\partial}{\partial t}$, 即若存在函数 f 使任意坐标有 $X^\nu \nabla_\nu X^\mu = f X^\mu$, 则称 $\lambda(t)$ 为测地线。总可取一个新参数 ξ 使

$$X^\nu \nabla_\nu X^\mu = 0, \quad X = \frac{\partial}{\partial \xi}, \quad (3.1.4)$$

则该参数称为仿射参数。容易验证仿射参数可确定到线性变换 $\xi' = a\xi + b$, 此处 a, b 为常数。此后谈及测地线均默认仿射参数化。

这样, 只要沿着测地线做关于给定联络的平行移动, 就可确定一点附近矢量空间的一致同构⁴。

3.1.2 克氏符

可以证明 (见第 4 章), 任意流形必定存在满足定义的联络, 并且由如下讨论可知, 其数量无限多。为此先给出克氏符的严格定义:

定义 3.7 (克氏符). 给定背景上的联络 ∇ , 设矢量基 $\{e_\mu\}$ 及其对偶, 则克氏符定义为 $\Gamma^\lambda_{\mu\nu} = e^\lambda(\nabla_{e_\mu} e_\nu)$, 或亦可记作

$$\Gamma^\lambda_{\mu\nu} = \langle e^\lambda, \nabla_{e_\mu} e_\nu \rangle. \quad (3.1.5)$$

上式等价于

$$\nabla e_\nu = \Gamma^\lambda_\nu \otimes e_\lambda, \quad (3.1.6)$$

这里 $\Gamma^\lambda_\nu = \Gamma^\lambda_{\mu\nu} e^\mu$ 称为联络 1-形式, 因而 $\Gamma^\lambda_{\mu\nu}$ 也称作联络系数。

定理 3.1.1. ∇Y 在坐标基上的表达为

$$\nabla_\mu Y^\lambda = \frac{\partial Y^\lambda}{\partial x^\mu} + \Gamma^\lambda_{\mu\nu} Y^\nu.$$

证. 对任意矢量场 Y 有

$$\begin{aligned} \nabla Y &= \nabla(Y^\nu e_\nu) = dY^\nu \otimes e_\nu + Y^\nu \Gamma^\lambda_{\mu\nu} e^\mu \otimes e_\lambda \\ &= \frac{\partial Y^\lambda}{\partial x^\mu} dx^\mu \otimes e_\lambda + Y^\nu \Gamma^\lambda_{\mu\nu} e^\mu \otimes e_\lambda, \end{aligned}$$

取矢量基为坐标基即可得到分量。 □

⁴第 4 章证明小的邻域内两点间只存在一条测地线。

同理可求得对偶矢量及任意张量的协变导数。

定理 3.1.2. 任意张量的协变导数为

定理 3.1.3. 设 $e'_{\mu'} = A_{\mu'}{}^{\mu} e_{\mu}$, $e'^{\mu'} = A^{\mu'}{}_{\mu} e^{\mu}$, 则

$$\Gamma'^{\lambda'}{}_{\mu'\nu'} = A^{\lambda'}{}_{\lambda} (e'_{\mu'}(0)). \quad (3.1.7)$$

其在坐标基下就是此前所学的克氏符的坐标变换式。其相较于张量变换律所多出来的部分称为非协变部分。

联络是坐标无关的微分算符。下面给出“坐标依赖的微分算符”一例：

定义 3.8 (普通微分). 设 $\{x^{\mu}\}$ 的坐标基及其对偶分别为 $\{\partial_{\mu}\}, \{dx^{\mu}\}$, 定义普通(*ordinary*)微分是一个微分算符 d (不是外微分), 其将一点处的 (k, l) 型张量映为 $(k, l+1)$ 型张量, 使得

$$dT = \frac{\partial}{\partial x^{\sigma}} T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l} dx^{\sigma} \otimes \partial_{\mu_1} \otimes \cdots \otimes \partial_{\mu_k} \otimes dx^{\nu_1} \otimes \cdots \otimes dx^{\nu_l}. \quad (3.1.8)$$

普通微分只在对函数 f 作用时才等同于外微分。

定理 3.1.4. $d(\partial_{\nu}) = 0, d(dx^{\nu}) = 0$.

证. 注意 dT 的分量就是 T 分量对坐标求偏导数, 但坐标基在该系的分量都是常数 0 或 1, 当然结果也就是个零张量。 \square

定理 3.1.5. $ddT(u, v, \dots) = ddT(v, u, \dots), \forall$ 矢量场 u, v .

证. 注意 ddT 的分量就是对 T 分量求两次坐标偏导, 由分析学知偏导是对易的, 因此分量关于此二指标对称。 \square

不难验证, 其满足联络其它要求, 但偏偏就不满足“将张量场映为张量场”, 因为其自定义起就依赖于坐标系的微分算符, 且只在该坐标系的坐标域上有定义, 则张量场的普通微分一般不为张量场。因而 d 可算作一种微分算符, 但不属于联络。

下面说明联络无限多。从一般的微分算符谈起, 若 ∇ 是微分算符, 假设存在至少一个不同的微分算符 $\tilde{\nabla}$, 我们欲分辨其差异。二者作用于同一函数 f 的结果均为 df , 故不能考虑对函数作用。不妨考虑作用于矢量 v 或对偶矢量 ω 。

考虑点 p 处对偶矢量在邻域上的任意两种延拓场 ω', ω , 则场 $\nabla \omega'$ 在 p 处取值一般不等于 $\nabla \omega$ 的⁵。但下面证明任意两联络的差 $\tilde{\nabla} - \nabla$ 作用于任意延拓 ω 是相同的。

⁵可类比分析学中, 两一元函数在同一处取值相同但切线斜率一般不同。

定理 3.1.6. 设 $\omega_p = \omega'_p$, 则

$$(\tilde{\nabla} - \nabla)\omega|_p = (\tilde{\nabla} - \nabla)\omega'|_p,$$

这里 $(\tilde{\nabla} - \nabla)\omega$ 视作 $\tilde{\nabla}\omega - \nabla\omega$ 。

证. 只需证明

$$\tilde{\nabla}\Omega|_p = \nabla\Omega|_p,$$

其中 $\Omega = \omega' - \omega$ 。取邻域上任意坐标系, 将 Ω 表为 $\Omega_\mu dx^\mu$ 。则由 Leibniz 律有

$$\nabla\Omega = \Omega_\mu \nabla dx^\mu + dx^\mu \nabla\Omega_\mu,$$

由 $\Omega_\mu|_p = 0$, 其于 p 处取值为

$$\nabla\Omega|_p = dx^\mu \nabla\Omega_\mu|_p.$$

同理 $\tilde{\nabla}\Omega|_p = dx^\mu \tilde{\nabla}\Omega_\mu|_p$, 但 Ω_μ 是函数, 任意微分算符 ∇ 作用结果均为 $d\Omega_\mu$, 因此二者相等, 当然 p 处取值也相等。 \square

虽然 $\nabla\omega$ 和 $\tilde{\nabla}\omega$ 都依赖于 ω 在一点某邻域的延拓情况, 但 $(\tilde{\nabla} - \nabla)\omega$ 只依赖于 ω 在一点的值, 且 $(\tilde{\nabla} - \nabla)$ 将一点 ω 线性地映为该处 $(0, 2)$ 型张量, 故其是该处矢量空间上的 $(1, 2)$ 型三重线性映射, 写作分量为

$$(\tilde{\nabla}_\mu - \nabla_\mu)\omega_\nu = \left\{ \begin{matrix} \lambda \\ \mu\nu \end{matrix} \right\} \omega_\lambda,$$

此关系对任意点成立, 因此我们说任意联络之间在任意坐标系差一个 $(1, 2)$ 型矩阵场⁶, 其任意性导致微分算符无限多。取两微分算符都为联络, 则 $(1, 2)$ 型矩阵任意性导致联络无限多。

该矩阵场是否可称为张量场的分量呢? 答案是不一定, 注意, 我们在定义此矩阵时使用的是分量语言, 而非映射语言。使用映射语言在一点给出的一定为张量, 如

$$(\tilde{\nabla} - \nabla)\omega = C^1_3(C \otimes \omega),$$

其中 C 是该点处与之对应的 $(1, 2)$ 型张量, 但不一定构成张量场。假设二者都是联络, 则 C 是不变的, 因而可构成 $(1, 2)$ 型张量场。实际上, 由此前讨论已知, 任意联络之差相当于克氏符之差, 而克氏符之差刚好消去了非协变部分, 因此克氏符之差满足张量变换律!

若其一恰为坐标依赖的, 比如取 $\tilde{\nabla}$ 为某坐标系的普通微分 d , 则

⁶矩阵场由于给每一点赋以的是坐标分量数表, 因此不能把矩阵场直接理解为定义在背景上, 而是依附于坐标系上。注意同张量场相区别。

定理 3.1.7. d 和 ∇ 在此坐标系所差的 $(1, 2)$ 型矩阵即为 ∇ 在该系的克氏符。部分书将此作为克氏符定义。

即使可对微分算符定义克氏符，亦成立相关的协变导数公式，但最好只考虑能映出张量场的联络，这样就可以说克氏符修正了一定不协变的普通微分，从而得到协变微分 ∇ ，此时克氏符同普通微分的不协变性相抵消。

设 $\{x^\mu\}$ 系下有

$$(d - \nabla)\omega = C_3^1(\Gamma \otimes \omega),$$

取另一 $\{x'^\mu\}$ 可给出不同张量：

$$(d' - \nabla)\omega = C_3^1(\Gamma' \otimes \omega),$$

由此发现一点处 Γ 本身就依赖于坐标系，当然也就不构成张量场。分量语言所给出的克氏符在 $\{x^\mu\}$ 系下是 Γ 在 $\{x^\mu\}$ 的分量，而在 $\{x'^\mu\}$ 系下是 Γ' 在 $\{x'^\mu\}$ 的分量，因此克氏符不满足张量变换律。这与此前的讨论是一致的。

定理 3.1.8. $DT/dt = \nabla_{\partial_t} T$ 。在坐标系下可表为

$$\frac{D}{dt} T^{\mu_1 \cdots \mu_k}_{\nu_1 \cdots \nu_l} = \frac{dx^\sigma}{dt} \nabla_\sigma T^{\mu_1 \cdots \mu_k}_{\nu_1 \cdots \nu_l}. \quad (3.1.9)$$

证. ∇ 虽取决于曲线外 \bar{T} 的情况，但方向导数 ∇_{∂_t} 只沿曲线，于是与 \bar{T} 无关。 \square

定理 3.1.9. 切矢平行移动直接给出使用仿射参数的测地线。

证. 取 $X = \frac{\partial}{\partial \xi}$, $DX/d\xi = \nabla_X \bar{X} = 0$ 等价于 $\nabla_X X = 0$ 。 \square

3.1.3 Levi-Civita 联络

3.1.4 曲率

3.2 基本原理探讨

3.3 参考系

3.4 作用量理论

3.5 线性引力论

3.6 引力辐射

3.7 引力波简介

3.7.1 洞穴论据

如果把分量全部显式地展开，就可以看到它是关于 $g_{\mu\nu}$ 的非线性微分方程.

如果在任意坐标系下把场方程显式地写出来，就可以看到，它并不属于哪一种通常的类型的 PDE，诸如椭圆型方程（像 Poisson 方程那样）、抛物型方程（像热传导方程那样）和双曲型方程（像波动方程那样）。这与以下的事实有关，即给出了一个解，就可以把这个解与一个坐标变换复合起来，得到一个新解。我们可以对原来的坐标系作一个如下的坐标变换以得出一个新坐标系：这个变换只在一个球体内与恒等变换不同。这个事实，即所谓洞穴论据⁷，使得 Einstein 和 Grossmann 大为困惑，他们是在坐标下的方程“代数地”在思考，因而竟使得他们拒绝广义协变性。这样导致的思维滞后使得场方程的正确陈述推迟了两年之久。对于 Einstein 的理论的几何解释给出了摆脱困境的出路：这样的解应该看成是“相同”的解，因为从所有的几何量度的观点看来，它们都是相同的。用现代的语言来说，我们就说场方程的解（例如真空方程的解）就是时空的等价类：我们说两个时空是等价的，就是说它们之间有一个微分同胚 ϕ ，使得在任意开集合里，如果我们把用 ϕ 互相转换的局部坐标视为相同的，则这两个度规都看成是相同的。结果，当这些概念上的问题被克服以后，场方程就可以看成是双曲型的。

⁷又称洞论证，这是 Einstein 在发展广义相对论时提出的关于时空的一个深刻的论点，或者说是一个悖论。它有深刻的哲学含义，所以下面介绍一个哲学网站上的一篇论文，以备有兴趣的读者参阅：<http://plato.stanford.edu/entries/spacetime-holearg/>

看到这一点最容易的方法是作一个规范，就是对坐标系作一定的限制。具体地说，就是要求坐标函数 x^α 满足波动方程 $\square_g x^\alpha = 0$ ，这里 \square_g 是 d'Alembert 算子。这样的坐标总是局部存在的，称为谐和坐标，或者波坐标。于是场方程可以写为一个方程组

$$\square_g g_{\mu\nu} = N_{\mu\nu}(\{g_{\alpha\beta}\}, \{\partial_\lambda g_{\alpha\beta}\}),$$

这里的 $N_{\mu\nu}$ 是一个非线性表达式，而且对于 $\{\partial_\lambda g_{\alpha\beta}\}$ 是二次式。考虑到 Lorentz 度量的号差，上面的方程组就是一个二阶非线性⁸双曲方程组。于是方程组的波动性质就是显然的了⁹。

场方程的双曲性有两个重要收获。第一即是应该存在引力波，Einstein 本人在 1918 年就看到了这一点；第二便是适定的初值问题。所谓适定性，就是对于适当的初始数据，必有唯一的解。这里的所谓数据应该是一个 3 维 Riemann 流形 (Σ, \bar{g}) 再附带一个 2 阶对称协变张量 K 。从而 Einstein 理论就变成了经典问题，即一个动力学问题。

⁸实际上拟线性。

⁹后期的 Yang-Mills 方程也具有这样的规范对称性。

Chapter 4

现代语言

4.1 拓扑流形

4.1.1 什么是拓扑？

拓扑学¹，经常被描述成橡皮膜几何学 (*rubber-sheet geometry*)，也就是说，人们可以在不撕裂的情况下形变所研究的对象。这是关于连续性的最重要的概念。连续性是我们在这里所要研究的绝大部分问题的基础。

研究映射连续性以及拓扑结构在连续变换下不变的性质的学科。在拓扑学中，人们把具有相同开集结构 (同胚) 的空间等同起来，从而架空“距离”、“长度”、“面积”这些传统几何学中的概念，把注意力集中到诸如“连通性”、“曲面是否可定向”、“曲面上是否有洞”、“绳子上打了几个结”这类问题上的。点集拓扑研究的是最一般的集合上的拓扑，讨论开集、闭集、连续映射、同胚这些基本的拓扑学概念，是所有拓扑学分支的基础。

后续所学的微分流形是一个可供微积分的平台，这在很大程度上需要借助拓扑学定义，也就是用拓扑结构提供平台的连续性。

点集拓扑对我们而言主要用途是定义流形，而流形则可以说是我们整个故事的起点：我们会用流形来定义力学问题中的各种空间（如位形空间，相空间等）。当然，在拓扑学框架中，我们还只能定义所谓“拓扑流形”，即每一个点附近有一个邻域同胚于 \mathbf{R}^n 上开集的拓扑空间，而不能定义我们最终要用到的微分流形——那是更进

¹本节主旋律是所谓点集拓扑，即用集合论的语言先定义出拓扑学概念来，但为讨论更为深刻的话题，还需要借助所谓代数拓扑。本节暂不涉及，尽管它在本科数学系里作为高级课程以引入。代数拓扑这一专题将在本章关于上同调 (*cohomology*) 的部分中得以处理。相信这样基础性的一节对这部书的作用纯粹在于夯实和查阅。许多物理学家对这里绝大部分概念是熟悉的，但通常在相当不精确的意义下。本节将提供严格的定义和结果，其在参阅本书其余内容时会很有用。从应用于物理的角度出发，在假定对于未加定义的更高级概念熟悉的情况下，我们将引用大量的例子。论题的选择也是基于同样的理由。

一步的工作。事实上，我们讨论力学问题时一般不会特别深入地探讨所研究空间的拓扑结构，所以仅就我们的需要而言，最基本的点集拓扑学知识已经足够。微分流形中还有 de Rham 上同调等更深层次的知识会用到代数拓扑、同调论等，但我们所介绍的内容不会涉及到这些。

基本概念

拓扑的基础作用就是在一般的集合上定义开集，而不借助任何距离概念。为此可先回顾一下分析学中的开闭概念。分析学在定义实数集或实直线 (*real line*) \mathbf{R} 时已经借助了序结构，因此 $>$, $<$, \geq , \leq 之类的关系是良定义的。 \mathbf{R} 上的任意开区间可表为若干“连通开集”之并，所谓连通开集就是形如 (a, b) 的集合。闭区间就是其在 \mathbf{R} 上的补集，则边界上的括号就改为中括号，表示可取等。

欲在一般情况下定义开集，当然要研究一个集合的全体开集组成的集合有何性质，这个集合就叫做拓扑。其性质就是开集所具备的特征，我们可从分析学中抽象出来。

比如，全集可以看作开的，则空集就是闭的，但空集也可以看作开的，则全集就是闭的。从定义的推广上来看，其实没有理由排除一个既开又闭，或非开非闭的集合。不应该仅靠直觉就否定这种存在。不过，考虑到拓扑是先定义开集，不妨先规定空集、全集是开的。在分析学中，除了空集、全集外，一个区间就不会存在这种兼备的现象了。

其次，任意多甚至无限个开集的并集是开集，这从分析学的定义就可看出。但是，要使得开集的交集是开集，就只能有限多个：在 \mathbf{R} 上，所有开区间 $(-1/n, 1)$, $n \in \mathbb{N}_+$ 的交集是 $[0, 1]$ ，显然不是开的。区分“任意并”、“有限交”很重要。

上述三条性质不显含任何有关集合上距离的信息，因此可作推广之基础。因此，拓扑学所研究的对象可以看作脱胎于距离的一种更广义的空间。把所有仅涉及到开集的定义和命题给抽提出来，就得到了拓扑学中相应的定义和命题。可见，拓扑学并不玄乎。

定义 4.1 (拓扑). 设 X 为一集合。 X 的一族子集 \mathcal{T} 称为一个拓扑 (*topology*)，其中的一个成员称为 X 在拓扑 \mathcal{T} 下的一个开集 (*open set*)，如果以下性质成立：

a. $\emptyset, X \in \mathcal{T}$ ；

b. 开集的任意并是开集。令 \mathcal{I} 为一指标集 (理解为存放任意数字的集合)，应有

$$A_\alpha \in \mathcal{T}, \alpha \in \mathcal{I} \implies \bigcup_{\alpha \in \mathcal{I}} A_\alpha \in \mathcal{T}.$$

也可以不借助指标集： $\forall \sigma \subseteq \mathcal{T}, \bigcup_{U \in \sigma} U \in \mathcal{T}$ ；

c. 开集的有限交是开集. 即

$$A_i \in \mathcal{T}, i = 1, \dots, n \implies \bigcap_{i=1}^n A_i \in \mathcal{T}.$$

例 4.1. 实数集的通常拓扑 (*usual topology*) 暂且理解为分析学中开闭定义所对应的拓扑。

定义 4.2. 如果假定 \emptyset 和 X 是仅有的开子集, 则得到所谓的非离散拓扑 (*indiscrete topology*) 或最粗拓扑 (*coarsest topology*). 另一个极端情形是假定所有子集都是开集, 则我们得到离散拓扑 (*discrete topology*) 或最细拓扑 (*finest topology*).

若以实直线 \mathbf{R} 或平面 \mathbf{R}^2 为例来考虑, 这两个拓扑都不太自然. 但事实上如果我们从定义中明确地排除它们则会更不自然. 在某些方面它们显得非常有用.

定义 4.3 (闭集). X 的一个子集称为闭的 (*closed*), 若它在 X 中的补集是开的.

基于并集之补等于补集之交这一集合论的命题, 可以很容易地用闭集代替开集来构造一个拓扑. 因此, 也可以把拓扑看作定义闭集的过程.

定义 4.4 (拓扑空间). 一个被赋予了拓扑 \mathcal{T} 的集合 X 称为一个 (关于给定拓扑的) 拓扑空间 (*topological space*), 记作 (X, \mathcal{T}) . 拓扑空间中的元素有时被称为点 (*point*), 拓扑就称其拓扑结构. 当然, 即使没有拓扑结构的集合, 也可以把元素称作点.

此外, 还有一些看似需借助距离的分析学概念可用拓扑语言改写.

定义 4.5 (邻域). 令 $x \in X$. x 的一个邻域 (*neighbourhood*) 是 X 的一个子集, 该子集包含一个含有 x 的开集.

这个定义看起来有些不方便, 但实际上在一般情况下比通常那种仅限于开邻域的定义要有用.

对于一般的拓扑空间, 其开集结构往往会比较复杂, 这时我们会用所谓“拓扑基”方式来描述一个拓扑结构. 对于一个集合 X , 所谓拓扑基, 是指 X 的一个满足“有限交”性质并且能够“覆盖”整个 X 的子集族, 这个子集族通过做“任意并”操作能够生成 X 的一个拓扑结构.

定义 4.6 (拓扑基). 一个开集的子族 $\mathcal{B} \subseteq \mathcal{T}$ 称作拓扑 \mathcal{T} 的一个基 (*basis*), 如果每个开集都是 \mathcal{B} 中集合的并集. 也不妨考虑: $\bigcup_{U \in \mathcal{B}} U = X$; 对任意 $U_i \in \mathcal{B}$ 存在 $\mathcal{F} \subseteq \mathcal{B}$ 使得 $\bigcup_{U \in \mathcal{F}} U = \bigcap_{i=1}^n U_i$. \mathcal{B} 中成员的所有可能并集 $\mathcal{T} = \{\bigcup_{U \in \mathcal{F}} U\}$ 构成了 X 上的一个拓扑, 称为由拓扑基 \mathcal{B} 生成的拓扑.

定义 4.7 (拓扑子基). 一个开集的子族 $\mathcal{S} \subseteq \mathcal{T}$ 称作拓扑 \mathcal{T} 的一个子基 (*sub-basis*), 如果每个开集都是 \mathcal{S} 中集合的有限交集之并.

定义 4.8. X 的子集 A 的闭包 (*closure*) \bar{A} 是 X 中包含 A 的最小闭集;

X 的子集 A 的内部 (*interior*) $i(A)$ 是 X 中含于 A 的最大开集;

有时可定义 $\partial A = \bar{A} \setminus i(A)$ 为 A 的边界 (*boundary*).

定义 4.9 (极限点). 设 A 为拓扑空间 X 的子集. 点 $x \in X$ 称为 A 的极限点 (*limit point*), 如果每一个包含 x 的开集都包含 A 中异于 x 的点.

定义 4.10 (稠密). 称 X 的子集 A 在 X 中稠密 (*dense*), 如果 $\bar{A} = X$.

定义 4.11 (Hausdorff 空间). 拓扑空间 X 称为 *Hausdorff* 空间, 如果对任意不同的两点 $x, y \in X$, 存在 x 的开邻域 A 与 y 的开邻域 B , 使得 $A \cap B = \emptyset$.

这看起来更像我们所期待的. 但是, 某些略微非 Hausdorff 的空间会很有用.

例 4.2. 在扭量 (*twistor*) 理论中. 一个“口袋” (*pocket*) 提供了这样的一个例子. 考虑实平面的子集 X , 由实轴上的区间 $[-1, 1]$ 和直线 $y = 1$ 上的区间 $[0, 1]$ 构成, 并且等同如下的点: $(x, 0) \approx (x, 1), 0 < x \leq 1$. 这样, 点 $(0, 0)$ 与 $(0, 1)$ 就没有不相交的开邻域. 严格说来, 我们需要下文中商拓扑的概念.

例 4.3. 一个更为地道的非 *Hausdorff* 空间: 考虑正整数组成的空间 \mathbb{N}_+ , 开集取成 \emptyset, \mathbb{N}_+ 以及集合 $\{1, 2, 3, \dots, n\}$. 这个空间既非 *Hausdorff* 亦非紧致的 (见后面关于紧致的定义).

定义 4.12 (连续). 设 X, Y 为拓扑空间. 我们称映射 $f: X \rightarrow Y$ 是连续的 (*continuous*), 如果“开集的原像是开集”, 即对于 Y 中任意开集 A , $f^{-1}[A]$ 是 X 中的开集.

连续性是这里最为重要的概念. 该拓扑学定义也不显含任何关于距离的信息. 在这个一般的框架下, 它看起来与 ε - δ 定义略有不同. 但是后者只对于定义了距离的度量空间有效, 我们马上将要谈到.

定义 4.13 (同胚). 一个双射 $f: X \rightarrow Y$ 称作同胚 (*homeomorphism*), 如果它是连续的且逆映射 f^{-1} 也是连续的. 两个同胚的空间从拓扑学的观点来看是不可区分的. 在这个意义上, 同胚是拓扑空间之间自然的映射. 拓扑不变量就是拓扑空间在同胚下保持不变的性质.

定义 4.14 (诱导拓扑). 设 $B \subseteq A$, 可定义 B 的相对拓扑 (*relative topology*) 或诱导拓扑 (*induced topology*): 一个子集 $C \subseteq B$ 是开集当且仅当存在 A 的开集 D 使得 $C = D \cap B$. 一个赋予了诱导拓扑的子集 $B \subseteq A$ 称为 A 的拓扑子空间 (*topological subspace*).

这样, 对于实直线的子集, 如果我们令 $A = [0, 3], B = [0, 2]$, 则在由 \mathbf{R} 上通常拓扑的诱导拓扑下, $C = (1, 2]$ 是 B 中的开集.

定义 4.15 (乘积拓扑). 给定两拓扑空间 X, Y , 可定义乘积拓扑空间 (*product topological space*) $Z = X \times Y$. 这里 Z 作为集合是 X 和 Y 的卡氏积. 令 A 为 X 中开集, B 为 Y 中开集, 则形如 $A \times B$ 的集合构成 Z 的一个拓扑基.

当然, $X \times Y$ 中的开集并不一定都具有乘积 $(A \times B)$ 的形式.

定义 4.16 (商空间). 设对于指标集 \mathcal{I} , 有一个将 X 分成互不相交的子集 $A_\alpha, \alpha \in \mathcal{I}$ 的分解. 或者等价地说, 在 X 上定义了一个等价关系. 则我们可以按以下方式在等价类的集合 $\{A_\alpha, \alpha \in \mathcal{I}\}$ 上定义商拓扑 (*quotient topology*), 该等价类称作商空间 (*quotient space*) $X / \sim = Y$. 考虑正则投影映射 $\pi: X \rightarrow Y$, 它把 X 中的一个元素 x 映到它的等价类 $[x]$. 则 Y 中的子集 $U \subseteq Y$ 是开集当且仅当 $\pi^{-1}(U)$ 是开集.

定理 4.1.1. 令 \mathcal{T} 为商空间 Y 上的商拓扑. 设 \mathcal{T}' 为 Y 上另一个拓扑使得正则投影是连续映射, 则 $\mathcal{T}' \subseteq \mathcal{T}$.

定义 4.17 (开覆盖). X 的一个开覆盖 (*open cover*) 或覆盖 $\{U_\alpha : \alpha \in \mathcal{I}\}$ 是指 X 的一族开子集 $U_\alpha \subseteq X$, 它们的并集是 X . 一个覆盖的子覆盖是指这族集合的一个子族, 它本身也是 X 的一个覆盖.

定义 4.18 (紧致). 一个拓扑空间 X 称为紧致的 (*compact*) 或紧的, 如果 X 的每个覆盖都有有限的子覆盖.

对于紧致空间来说, 无论我们选择怎样的覆盖, 用有限个开子集就足够了. 这是开区间 (非紧致) 和闭区间 (紧致) 的根本区别. 前者实际上同胚于整个实直线.

定义 4.19 (连通). 一个拓扑空间 X 称为连通的 (*connected*), 如果它不能写成两个互不相交的非空开集的并. 有用的两条等价说法是: 任何一个从 X 到赋予离散拓扑的两点集合 $\{0, 1\}$ 的连续映射都不是满射; 连通拓扑空间只有两个既开又闭的子集.

定义 4.20 (路径连通). 一个拓扑空间 X 称为路径连通的 (*path-connected*) 或弧连通的 (*arcwise connected*), 如果 X 中的任意两点都可被一条完全在 X 中的路径连接. (路径一词已定义于 2.34.)

定理 4.1.2. 路径连通的拓扑空间一定是连通的. 反之不一定, 存在“擦边”的反例². \mathbf{R}^n 中的连通开子空间是路径连通的.

定义 4.21 (连通分支). 给定一拓扑空间 X , 定义一等价关系如下: $x \sim y$ 当且仅当 x 和 y 属于 X 的同一个连通子空间. 则每个等价类称为 X 的一个连通分支 (*connected component*) 或分支.

² 流形的路径连通与连通等价, 稍后叙述之.

例 4.4. 3×3 正交矩阵构成的群 $O(3)$ 具有两个连通分支. 单位连通分支是子群 $SO(3)$.

例 4.5. 闵氏时空中的 *Lorentz* 变换构成 *Lorentz* 群的单位连通分支.

度量空间

度量空间是具有距离定义的特殊集合. 在 \mathbf{R}^n 上用勾股定理定义距离, 就可从拓扑学转向分析学, 并一路走向微积分. 本节只简述很少的分析学, 可能对流形的学习暂时无用, 但对因果结构、泛函分析等的理解很有裨益.

度量空间和内积空间的区别在于, 后者是定义矢量空间上的双线性函数 (度规), 前者是直接定义任意集合中两元素的距离函数 (度量). 二者本无依赖, 但亦可有所关联.

定义 4.22 (度量). 一个度量空间 (*metric space*) 是一个集合 X 以及一个函数 $d : X \times X \rightarrow \mathbf{R}$, 满足

(对称) $d(x, y) = d(y, x)$

(正定) $d(x, y) \geq 0$, 当且仅当 $x = y$ 时取等;

(三角不等式) $d(x, z) \leq d(x, y) + d(y, z)$.

函数 d 称作两点间的度量函数或距离. 在普通几何中度量的概念一般被理解为欧氏度量. 显著的特征是正定性 (以及三角不等式). 可以引入具有不同符号的不定度量, 就像不定度规一样, 但是这些度量不被用来在所关心的空间上诱导出拓扑.

定义 4.23 (开球). 给定度量空间 X 中一点 $x \in X$, 我们定义以 x 为中心, 正实数 r 为半径的开球 (*open ball*) 为 $B_r(x) = \{y \in X : d(x, y) < r\}$.

可以验证, 给定度量空间 X , 取 X 中全体开球组成的集族 $B_\varepsilon(x) : x \in X, \varepsilon > 0$ 构成了 X 上的一个拓扑基, 它所生成的拓扑称为 X 上的给定度量所诱导的拓扑. 于是我们可以重新得到关于连续性的通常的 ε - δ 定义.

定理 4.1.3 (分析学的 ε - δ 语言). 令 f 为度量空间 X 到度量空间 Y 的映射. 则 f 在 $x \in X$ 处连续 (对所诱导的拓扑) 当且仅当对任意 $\varepsilon > 0$, 存在 $\delta > 0$, 使得

$$B_\delta(x) \subseteq f^{-1}[B_\varepsilon(f(x))].$$

即 $f^{-1}[B_\varepsilon(f(x))]$ 是 x 点邻域.

上述说法显然等价于: 对任意 $\varepsilon > 0$, 存在 $\delta > 0$, 使得若 $d_X(x, x') < \delta$, 则有 $d_Y(f(x), f(x')) < \varepsilon$. 注意, 这里给两个度量空间以各自不同的度量函数.

定理 4.1.4. 如果 f 在 X 上任意一点处连续, 则 f 是连续映射.

证. 改用拓扑语言就容易发现, 如果 f 在 X 上任意点连续, 那么对于任意 $x \in X, \varepsilon > 0$, $f^{-1}[B_\varepsilon(f(x))]$ 是 X 上的开集, 再结合开集的“任意并”性质和原像的保并运算性质, 容易得到: 对于任意 Y 中的开集 U , $f^{-1}[U]$ 是 X 中的开集。此即连续的拓扑定义。□

拓扑定义的好处就是脱离了距离、某点连续性的概念而直接定义一个映射整体的连续性, 并且可以对更为任意的拓扑空间、映射定义相应的连续, 尽管可能会更反直觉。

定理 4.1.5. 容易证明度量空间是 *Hausdorff* 的。

定义 4.24 (一致连续). 一个度量空间之间的映射 $f : X \rightarrow Y$ 是一致连续的 (*uniformly continuous*), 如果对任意 $\varepsilon > 0$, 存在 $\delta > 0$, 使得对于所有 $x_1, x_2 \in X$, 若 $d(x_1, x_2) < \delta$, 则有 $d(f(x_1), f(x_2)) < \varepsilon$.

注意连续和一致连续的差别: 后者要更强而且对整个空间要求一致的 δ .

定义 4.25. 在 X 上定义的两个度量 d_1 和 d_2 是等价 (*equivalent*) 的, 如果存在正常数 a 和 b , 使得对任意两点 $x, y \in X$, 有 $ad_1(x, y) \leq d_2(x, y) \leq bd_1(x, y)$ 。

这显然是一个等价关系. 两个等价的度量诱导出相同的拓扑。

例 4.6. 给定集合 X , 我们定义离散度量如下: $d_0(x, y) = 1$ 若 $x \neq y$. 它在 X 上诱导出离散拓扑. 这是一个十分方便的描述离散拓扑的方法。

例 4.7 (通常拓扑). 在 \mathbf{R} 上, 通常度量是 $d(x, y) = |x - y|$, \mathbf{R} 上的通常拓扑由这个度量所诱导。

例 4.8. 更一般地, 在 \mathbf{R}^n 上, 我们可以定义一个度量

$$d_p(x, y) = \left(\sum_{k=1}^n |x_k - y_k|^p \right)^{1/p}, \quad p \geq 1,$$

这里 $x = (x_1, x_2, \dots, x_n), y = (y_1, y_2, \dots, y_n)$. 特别地, 对于 $p = 2$, 我们得到欧氏度量, 但是其他情形也是有用的. 继续这个序列, 可以定义

$$d_\infty = \max_{1 \leq k \leq n} \{|x_k - y_k|\}.$$

所有这些度量都在 \mathbf{R}^n 上诱导出相同的拓扑, 也就是通常拓扑. 当然, 亦可视作 \mathbf{R} 通常拓扑的乘积拓扑。

例 4.9 (范数). 在一个实数或复数域上的矢量空间 V 上, 一个函数 $\|\cdot\| : V \rightarrow \mathbf{R}^+$ 称为一个范数 (*norm*), 如果它满足以下公理:

- (a) $\|x\| = 0$ 当且仅当 $x = 0$,
- (b) $\|\alpha x\| = |\alpha| \|x\|$,
- (c) $\|x + y\| \leq \|x\| + \|y\|$.

易见利用范数可以定义一个度量

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

在许多情况下, 例如对于上面例子 4.8 中定义的度量, 我们可以把一个矢量的范数定义成它到原点的距离 (一个显然的例外是离散度量). 其合理性自明, 这样, 勾股定理所定义的距离, 就可看作欧氏内积 (度规) 诱导的度量: $|x - y| = \langle x - y, x - y \rangle$. 于是在 \mathbf{R}^n 的通常拓扑下, 度规和距离就仿若同一事物了。

定义 4.26 (Cauchy 列). 给定一个度量空间 X , 一个点列 $\{x_1, x_2, \dots\}$ 称为基本列 (*fundamental sequence*) 或 *Cauchy* 列, 如果对任意 $\varepsilon > 0$, 存在一个正整数 N , 使得对任意 $k, l > N$, 有 $d(x_k, x_l) < \varepsilon$ 。

定义 4.27 (极限). 在度量空间 X 中给定一个点列 $\{x_1, x_2, \dots\}$, 一点 $x \in X$ 称为这个序列的极限 (*limit*), 如果对任意 $\varepsilon > 0$, 存在一个正整数 N , 使得对任意 $n > N$, 有 $d(x, x_n) < \varepsilon$. 我们称这个序列收敛 (*converge*) 到 x .

定义 4.28 (完备). 一个度量空间 X 称为完备的 (*complete*), 如果其中每个 *Cauchy* 列都在 X 中有极限.

例 4.10. 实直线上的闭区间 $[0, 1]$ 是完备的, 而开区间 $(0, 1)$ 则不是. 例如, *Cauchy* 列 $\{1/n, n = 2, 3, \dots\}$ 在这个开区间上没有极限点 (看成实直线上的序列, 它当然有极限点 0).

例 4.11. 空间 \mathbf{R}^n 是完备的.

例 4.12. 由所有使得求和 $\sum_{k=1}^{\infty} x_k^2$ 收敛的实数序列 $\{x_1, x_2, \dots\}$ 构成的 *Hilbert* 空间 ℓ^2 在前面定义的 d_2 的无穷维推广所得度量下是完备的. 对任意 $p \geq 1$, 可以类似地定义 ℓ^p , 它们也是完备的, 从而是 *Banach* 空间.

完备性不是拓扑不变量. 例如, 开区间 $(0, 1)$ 与整个实直线同胚 (在通常拓扑下), 但是前者不完备而后者完备. 同胚可以方便地由正切函数给出.

定义 4.29 (有界). 度量空间 X 的一个子集 B 称为有界的, 如果存在一个半径为正实数 R 的球完全包含 B .

定理 4.1.6 (Heine-Borel). \mathbf{R}^n 的任何有界闭子集是紧致集. 其逆命题亦成立.

从而我们有一个非常好的关于 \mathbf{R}^n 的紧致子集的刻画: 有界闭集.

定理 4.1.7 (列紧性). \mathbf{R}^n 中的任何有界序列有一个收敛子列.

这些是分析学中老生常谈的定理, 故证明不再赘述 (包括下文), 感兴趣的读者可自行查阅。

定义 4.30 (函数逐点收敛). 考虑定义在 \mathbf{R} 的子集 A (通常为一区间) 上的一列实值函数 $\{f_n\}$. 我们说 $\{f_n\}$ 在 A 上逐点收敛 (*converge pointwise*), 如果对于每个 $x \in X$, 实数列 $\{f_n(x)\}$ 收敛. 从而我们可以定义函数 $f: A \rightarrow \mathbf{R}$ 为

$$f(x) = \lim_{n \rightarrow \infty} f_n(x),$$

并记成 $f_n \rightarrow f$.

定义 4.31 (函数一致收敛). 一列函数 $f_n: A \rightarrow \mathbf{R}, A \subseteq \mathbf{R}$ 称为一致收敛 (*converge uniformly*) 到函数 $f: A \rightarrow \mathbf{R}$, 如果对任意 $\varepsilon > 0$, 存在一个正整数 N , 使得对所有 $x \in X$, 只要 $n > N$, 就有 $|f_n(x) - f(x)| < \varepsilon$.

定理 4.1.8. 令 $f_n: (a, b) \rightarrow \mathbf{R}$ 为一列函数, 在 $c \in (a, b)$ 处连续. 假设 f_n 在 (a, b) 上一致收敛到 f , 则 f 在 c 处连续.

例 4.13. 连续函数的逐点极限不一定连续. 令 $f_n(x) = x^n, x \in [0, 1]$, 而极限函数 f 不连续.

定义 4.32 (压缩映射). 设 X 为一度量空间. 一个映射 $f: X \rightarrow X$ 称为压缩映射 (*contraction*), 如果存在 $c < 1$, 使得对所有 $x, y \in X$, 有 $d(f(x), f(y)) \leq cd(x, y)$.

定理 4.1.9 (Banach). 若 X 为一完备度量空间, f 为 X 上一压缩映射, 则 f 在 X 上有唯一不动点 x , 即 $f(x) = x$.

4.1.2 流形

Draft

参考文献

- [1] I. Newton, *Principia Mathematica Philosophia Naturalis*, 1687, English translation *The Principia: Mathematical Principles of Natural Philosophy* (City? University of California Press, 1999). 1.1.2, 25
- [2] R.Descartes, *Principia Philosophiae*(1644), Translated by V.R.Miller and R.P. Miller (Dordrecht: Reidel, 1983). 15
- [3] I. Newton, *De Gravitatione et Aequipondio Fluidorum*, translation in *Unpublished Papers of Isaac Newton*, ed. A.R. Hall and M.B. Hall (Cambridge: Cambridge University Press, 1962). 17
- [4] Einstein A et al. (1958) *The Principle of Relativity*. New York: Dover. 1.3.4, 2.3.1
- [5] Synge J L (1972) *Relativity: The Special Theory*. Amsterdam: North-Holland. 40
- [6] Spivak M (1965) *Calculus on Manifolds*. New York: W A Benjamin.
- [7] Choquet-Bruhat Y, De Witt-Morette C, and Dillard-Bleick M (1977) *Analysis, Manifolds and Physics*. Amsterdam: North-Holland.
- [8] Parrott S (1987) *Relativistic Electrodynamics and Differential Geometry*. Berlin: Springer. 48
- [9] Einstein A. *Zur Elektrodynamik bewegter Körper*(*On the Electrodynamics of Moving Bodies*)[J]. Annalen der Physik, 1905, 17(891): 50. 论文的翻译可在此书找到: *The Principle of Relativity* by Lorentz HA, Einstein A, Minkowski H, and Weyl H, 结合笔记: Sommerfeld A. New York: Dover, 1952. 28
- [10] Einstein A. *Does the inertia of a body depend upon its energy-content*[J]. Annalen der Physik, 1905, 18(13): 639-41. 1.2.4

- [11] H. Minkowski. *Raum und Zeit, Jahresberichte der Deutschen Mathematiker-Vereinigung*, 1909: 75-88. 34
- [12] A. Pais. *Subtle is the Lord*, 1982, Oxford University Press. (方在庆, 李勇的中译本《上帝难以捉摸》, 1998 年, 广东教育出版社出版) 44
- [13] Herrmann F., *New Physics Course For Senior High School*. (德国 KPK 物理课程教材第一分册《力学》) 56
- [14] Chalmers AF(1975) *Maxwell and the displacement current*. Physics Education January 1975: 45-49. 2.5.1
- [15] Siegel DM(1985) Mechanical image and reality in Maxwell's electromagnetic theory. In: Harman PM(ed.) *Wranglers and Physicists*. Manchester: Manchester University Press. 2.5.1
- [16] Roche J(1998) The present status of Maxwell's displacement current. European Journal of Physics 19:155-166. 2.5.1

索引

P

Poisson 方程, 7

Y

引力场强, 6

Draft