

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

First Edition

夏草

Beijing, PRC



Self Publishers Worldwide

NOT PUBLISHED

致永不褪色的青春.

Copyright © 2023 夏草 <https://chaoli.club/index.php/member/8927>

NOT PUBLISHED.

仅限教学科研使用，使用与传播应严格遵守相关法律规定.

Version: 2023.3.21, 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：

$$G_{im} = -\kappa \left(T_{im} - \frac{1}{2} g_{im} T \right)$$
$$\sum_{\sigma} g^{\sigma\sigma} T_{\sigma\sigma} = \sum_{\sigma} T_{\sigma}^{\sigma} = T$$

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

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。

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，深度上似乎要介绍玩微分拓扑，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。例如，集合（*set*）是具有某种特性的事物的整体。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不要不厌其烦地翻阅回顾定义。

Part II 将在其完成时再做序。

夏璇

2023 年 3 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	8
1.1.5 很大的引力	10
1.1.6 很小的引力 *	19
1.2 为何需要相对论?	19
1.2.1 经典电磁理论与经典力学的矛盾	19
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	27
1.2.5 Minkowski 对狭相的理解	27
1.3 怎么描述弯曲几何?	30
1.3.1 度规	30
1.3.2 闵氏几何	33
1.3.3 Einstein 对闵氏几何的理解	34
1.3.4 曲率	42
1.3.5 Lorentz 几何	45
1.3.6 测地偏离	57

1.4	广相应该是怎样的一套理论?	62
1.4.1	理想流体	62
1.4.2	广义协变性	65
1.4.3	场方程	66
1.4.4	作用量	68
1.4.5	时空的拓扑结构	69
1.4.6	这套理论有哪些应用?	69
2	平直时空	73
2.1	时空的几何学	73
2.1.1	闵氏时空	73
2.2	时空的物理学	75
2.3	运动学	79
2.4	力学	82
3	弯曲时空	89
3.1	引力同惯性的等效原理	89
3.2	广义协变性	89
3.3	弯曲几何	89
3.4	参考系	89
3.5	线性引力论	89
3.6	引力辐射	89
3.7	引力波简介	89
3.7.1	洞穴论据	89
4	现代语言	91
4.1	微分流形	91
4.2	切空间	91
4.3	微分算子、Lie 导数	91
4.4	微分形式和上同调论	91
	参考文献	93
	索引	95

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筐的职业球员；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，它自身就天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

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

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

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。

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。

1.1.2 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：

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

$\mathbf{F} = - \frac{GMm}{|\mathbf{r}|^3} \mathbf{r},$

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

$$\ddot{\mathbf{r}}(\mathbf{r}, t) = -G \sum_i \frac{\overset{\text{第 } i \text{ 个微元的质量}}{\Delta M_i} (\mathbf{r} - \mathbf{r}_i)}{|\mathbf{r} - \mathbf{r}_i|^3},$$

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

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

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

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 $\int_{\mathbf{R}^D} \delta(\mathbf{r}) d^D r = 1$ 。

起就是总密度 ρ ，再代入到 (1.1.3) 中就有

$$\begin{aligned}
 \ddot{\mathbf{r}}(\mathbf{r}, t) &= -G \int_M \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \sum_{i=1}^n M_i \delta(\mathbf{r}' - \mathbf{r}_i(t)) d^3 r' \\
 &= -G \sum_{i=1}^n M_i \int_M \frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} \delta(\mathbf{r}' - \mathbf{r}_i(t)) d^3 r' \\
 &= -G \sum_{i=1}^n M_i \frac{\mathbf{r} - \mathbf{r}_i(t)}{|\mathbf{r} - \mathbf{r}_i(t)|^3}, \quad \text{挑选性, } \delta \text{ 函数定义推论}
 \end{aligned}$$

因此这就又回到了质点的万有引力定律.

1.1.3 很远的引力

实际上，我们还可以继续扩大“高度”——或者应该说把视线放到“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 描述为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

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

⁶负号纯粹是习惯，但确实是个好习惯.

如果对其按 \mathbf{r} 取散度，则 关于 \mathbf{r}' 的积分和密度与散度无关 由 Gauss 定理得

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

即

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

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**。

⁷序言中的 Poisson 方程是 $G = 1$ 的版本，这涉及到**单位制选取**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）。

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

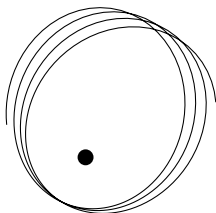


图 1.2: 轨道进动（离心率和进动角均夸大）

释进动，因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

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

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

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

⁹较近距离引力的传播速度确实可以视作瞬间。

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。

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

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

接下来处理有心力问题。通常从能量、角动量守恒出发：

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

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

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

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。

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.

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“ $43''$ /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, 并将其移用到引力场上. 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“调和和经

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.

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：作为一个实体或作为一种关系。

“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”：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”：如果空间是一种关系，运动只能被定义为从一个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

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。

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

¹⁶Aristotle 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。

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

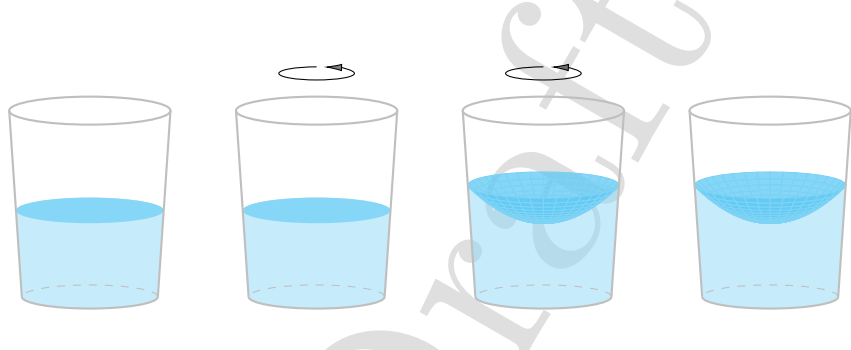


图 1.3: 水桶实验

还是挺大的: Newton 本人就声称“已经做过实验”, 并对该结果表示肯定.

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，即运动只能在与其它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”

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。或许哪朝哪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。

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**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:

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

而万有引力定律中定义的应叫引力质量:

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

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**. 我将在 1.2.2 以及 3.1 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²². 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

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?

²²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.

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.

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“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. 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. 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.

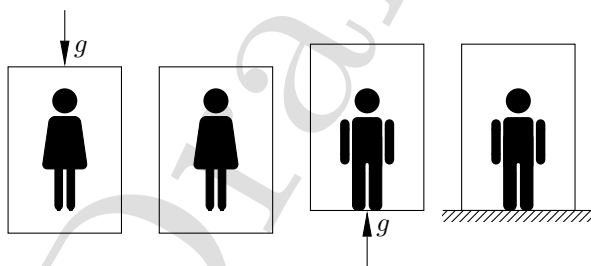


图 1.4: 电梯实验

设想一部封闭的电梯²³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

²³Einstein 本身是瑞士专利局职员, 其很多思想实验都与电驱机械有关. 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.

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 \blacksquare 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 $\mathbf{r}(t)$. 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

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

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他力 $\mathbf{F}_{\text{局域}}$ ，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！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²⁴。

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 所偏好的整体参考系，并认识到每个星系都有其自己的局部惯性系。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。

²⁴ 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度所推动，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. 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.

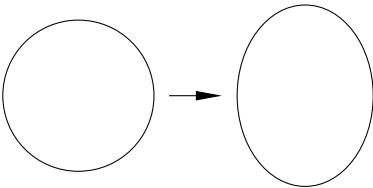


图 1.5: 潮汐效应

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²⁵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：

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

²⁵受到 Kaluza 理论的启发的 Teleparallel 引力，其数学语言采用 Weitzenböck 几何。

这里我还还原了国际单位制: μ_0 称为真空磁导率, ϵ_0 称为真空介电常数. 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, 他坚持 Galileo 的相对性原理, 将其精确地重述为机械系²⁶. 这可以表述为: 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. 因此我们说, Maxwell 方程组在 Galileo 变换下不具备协变性. 为了解决这一矛盾, 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

²⁶ 《原理》运动定律之推论 V.

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一样的”. 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²⁷.

1.2.2 Einstein 对狭相的理解

在 1905 年 6 月 30 日, Einstein 发表了他的原始论文《论动体的电动力学》²⁸, 这是狭义相对论 (下称狭相) 确立的重要标志.

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

²⁷ 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.

²⁸ [9], 第一部分主要解决运动学上的矛盾, 第二部分主要解释如何变换 Maxwell 方程组. 他没有完全使用初等数学, 而独立地得到了 Lorentz 的结果. 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”.

相对论时空观, 由相对性原理即可导出 *boost* 变换²⁹:

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

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.

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, 比如你想要测量一量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: 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时空图!

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

²⁹ 其含义取时间轴上的推动之意. 一个较好的翻译是“伪转动”, 我们将在本节末解释这一点.

知两个惯性系之间的时空变换是线性变换³⁰。假设这个线性变换是四阶方阵 Λ ，那么应该有

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

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，而叫做上指标，因为在给行矢量、矩阵分量编号的时候还可以用下指标，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。现在我们先暂时统一用 $\Lambda^\mu{}_\nu$ 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 $\Lambda^\mu{}_\nu$ 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 这个懒偷得非常巧妙但又十分深邃，所以后人给其取名为“求和约定”。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_ν^μ 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

³⁰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 2.1 节。

出现顺序混淆的问题³¹. 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 ds^2 , 这个等式实际上就是如下写法:

$$ds^2 = d(x^T)_\mu \eta^\mu_\nu dx^\nu = 0,$$

这里 η^μ_ν 是对角矩阵 $\text{diag}(-1, 1, 1, 1)$ 的矩阵元, 行矢量 $(x^T)_\mu$ 是列矢量 x^μ 的转置. 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.

关于矩阵的转置, 在目前暂时记号中, 比如对于 Λ^μ_σ , 我们像这样直接对调行列指标为 $(\Lambda^T)^\sigma_\mu$. 马上会看到, 这有助于我们理解转置、乘法的顺序问题. 但缺点也是明显的: 如果直接将二者划等号, 不能满足上下指标的平衡. 为了满足平衡而创设的新规则将在后续章节中讲解, 这可能涉及对协变性的探讨.

矢量怎么转置? 在当下记号中, 如果视列矢量为 x^σ_0 , 则对调看起来像是在“降指标”: $(x^T)_\sigma^0$, 这说明行矢量 $(x^T)_\sigma$ 的指标相当于矩阵列指标.

我们要求在 \hat{S} 下同样有

$$d\hat{s}^2 = d(\hat{x}^T)_\mu \eta^\mu_\nu d\hat{x}^\nu = 0,$$

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:

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

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³².

³¹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.

³²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, 更别提整个狭相理论, 比如所默认为“线性变换”就不能严格推出, 我们将在 2.1 节仔细探讨.

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

$$d\hat{s}^2 = d(\hat{x}^T)_\mu \eta^\mu_\nu \Lambda^\nu_\lambda dx^\lambda = 0,$$

可是 $(\hat{x}^T)_\mu$ 应该如何替换? 得先对 $\hat{x}^\mu = \Lambda^\mu_\sigma x^\sigma$ 整体进行转置, 在分量写法上就相当于“分别转”: $(\hat{x}^T)_\mu = (\Lambda^T)^\sigma_\mu (x^T)_\sigma$. 调换顺序后 σ 又可以再次相靠: $(x^T)_\sigma (\Lambda^T)^\sigma_\mu$, 即左乘再转置等于转置再右乘, 这与线性代数的知识是一致的. 现在我们有

$$d\hat{s}^2 = d(x^T)_\sigma (\Lambda^T)^\sigma_\mu \eta^\mu_\nu \Lambda^\nu_\lambda dx^\lambda = 0,$$

可见中间的 $(\Lambda^T)^\sigma_\mu \eta^\mu_\nu \Lambda^\nu_\lambda$ 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

$$(\Lambda^T)^\sigma_\mu \eta^\mu_\nu \Lambda^\nu_\lambda = k \eta^\sigma_\lambda.$$

为了方便, 不妨还原为矩阵写法

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

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

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

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. 如

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

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

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

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

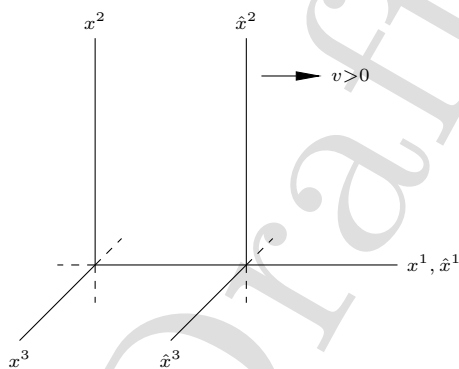


图 1.6: boost 变换示意图

轴亦重合, 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

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

此即

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

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治意义上可以定义为

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

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 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, 简称时空³³, 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³⁴.

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? 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, 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³⁵. 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 .

³³Minkowski 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].

³⁴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.

³⁵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.

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, \mathbf{x}) 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？那我们在推到 boost 时，这个矢量

$$(ct, x, y, z)^T$$

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：

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

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³⁶，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

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

上下同除 $d\tau$ 得

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

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 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。

³⁶对此，Einstein 曾表示相比于“相对论”这个名字，他更喜欢“绝对论”。

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

$$ds^2 = dx_\mu \eta^\mu{}_\nu dx^\nu$$

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

$$d\ell^2 = dx^2 + dy^2 + dz^2 = dx_i \delta^i_j dx^j,$$

其中 δ^i_j 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；并且为何这里又可以不区分 δ^i_j 指标的左右呢？因为这是个对称矩阵。

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”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 就是所谓的立体角了³⁷. 至此, 给定球面上的曲线, 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称作欧氏几何. 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:

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

³⁷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 $[0, 4\pi]$.

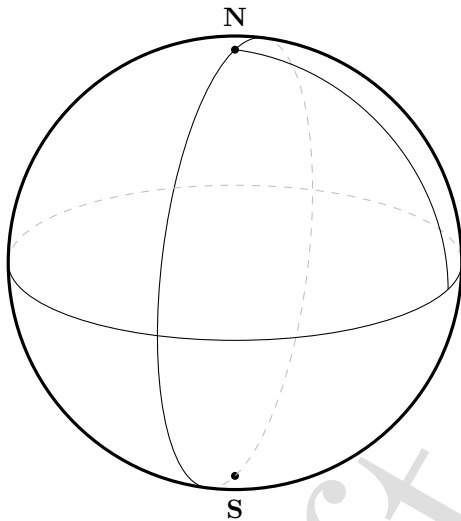


图 1.7: 球面

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某些光滑性.

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

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

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，即 $g_{\mu\nu} > 0$ ，可作强调地称其 *Riemann* 度规；否则就叫伪 *Riemann* 度规，即那些负定的或者不定的（正负都有）.

可我为什么不写成 $g^\mu{}_\nu$ 呢？原来，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

$$v^i \delta_{ij} u^j = (v^T)_i \delta^i_j u^j = (v^T)_j u^j.$$

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 闵氏几何

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，简称闵氏几何，与欧氏几何相对照. η 就叫做闵氏度规.

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，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

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

而一方面, $\Lambda^\mu_{\hat{\mu}} = \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}}$, 因此该式就是

$$\eta_{\hat{\mu}\hat{\nu}} = \Lambda_{\hat{\mu}}^\mu \eta_{\mu\nu} \Lambda^\nu_{\hat{\nu}}. \quad (1.3.6)$$

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. 这也暗示了一种新的平衡规则: 面对矩阵转置, 我们可以不直接沿对角线交换指标, 而只进行左右交换 (左右分别对应行列指标).

1.3.3 Einstein 对闵氏几何的理解

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³⁸, 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, 匀速运动时也一样. 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

³⁸ 此处语境只考虑引力场.

界线有个几何学里的名字——测地线，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。我们也可以说，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。

一般来讲，几何学将测地线定义为某种直线的推广。定义一条直线方程的方式有很多，我们不妨先来研究这样一种表述：矢量沿自身“平移”所得轨迹。如果我们设曲线是 α 的坐标表达式是 $x^\mu(\tau)$ ，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，那这不外乎是说沿着曲线速度矢量不变³⁹，即

$$\frac{d}{d\tau}(U^\mu e_\mu) = 0. \quad (1.3.7)$$

根据乘法法则，那么

$$\frac{dU^\mu}{d\tau} e_\mu + U^\mu \frac{de_\mu}{d\tau} = 0,$$

其中 $\frac{de_\mu}{d\tau}$ 是指基矢沿着曲线参数的变化率，可我们连曲线的坐标参数式都还不知道，怎么知道 $e_\mu(\tau)$ 呢？因此这不是我们想要的，但我们可以将其关于坐标展开成 $\frac{\partial e_\mu}{\partial x^\nu} \frac{dx^\nu}{d\tau}$ ，这样以来由于矢量 $\frac{\partial e_\mu}{\partial x^\nu}$ 是偏微分，当然只与坐标本身有关，而与曲线表达式无关了。可见 $\frac{\partial e_\mu}{\partial x^\nu}$ 是坐标系自己带有的信息。我们不妨将这个东西展开到基矢上：

$$\frac{\partial e_\mu}{\partial x^\nu} = \Gamma^\lambda_{\mu\nu} e_\lambda, \quad (1.3.8)$$

这里 $\Gamma^\lambda_{\mu\nu}$ 就表示 $\frac{\partial e_\mu}{\partial x^\nu}$ 的第 λ 分量，取名为 *Christoffel* 符号，简称克氏符。分母下方是一个上指标，而这相当于分子下指标，因此这样写是指标平衡的。现在让我们全代进去：

$$\frac{d}{d\tau} \left(\frac{dx^\mu}{d\tau} \right) e_\mu + \Gamma^\lambda_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} e_\lambda = 0.$$

看似无法消去基底，那是因为我们丢掉了求和号，从而忽略了指标的任意性！实际上，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！）左边因式中 μ 可以换

³⁹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，这种参数叫仿射参数。比如说，固有时就是一种仿射参数，我们马上会对此详述。

成 λ (必须两个一起换掉), 即

$$\frac{d}{d\tau} \left(\frac{dx^\lambda}{d\tau} \right) e_\lambda + \Gamma^\lambda_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} e_\lambda = 0,$$

这样便可提出基矢而合并同类项. 整个矢量为零当然意味着分量为零, 因此

$$\frac{d^2 x^\lambda}{d\tau^2} + \Gamma^\lambda_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0, \quad (1.3.9)$$

我们就得到了测地线方程. 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. 比如说, 在平面直角坐标系下, 每一处的基矢当然都是不变的, 因此 $\frac{\partial e_\mu}{\partial x^\nu} = 0$, 则克氏符处处为零, 这说明直线的微分方程就是

$$\frac{d^2 x^\lambda}{d\tau^2} = 0,$$

就像是 Newton 所意料的“加速度”为零那样. 那这是否意味着 $\Gamma^\lambda_{\mu\nu}$ 就是我们想要找的“引力”呢? 毕竟

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

长得像第二定律似的. 答案是否! 切记 $\Gamma^\lambda_{\mu\nu}$ 只是坐标系自身的信息 (每一处都有这 $4 \times 4 \times 4$ 个实数!), 很明显我们可以在平面上选取坐标系使得 $\frac{\partial e_\mu}{\partial x^\nu}$ 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?

这一切和等效原理有何关系呢? 不妨这样与前面所学联系: 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. 这样做就是要将 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, 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的局部惯性系 ξ^α , 那

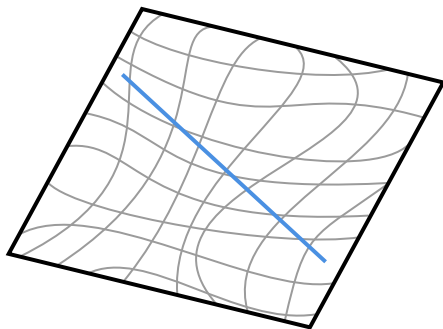


图 1.8: 一个完整的坐标系由处处的基矢决定, 而非仅仅只是原点处的坐标轴. 图中所示为一张平面上的曲线坐标网格, 蓝线为一条测地线.

么它在该坐标系里的世界线满足“不受力”的直线方程⁴⁰, 也就是:

$$\frac{d^2 \xi^\alpha}{d\tau^2} = 0, \quad (1.3.10)$$

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? 我的意思是, 其所定义的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:

$$\eta_{\alpha\beta} U^\alpha U^\beta = -1, \quad \eta_{\alpha\beta} d\xi^\alpha d\xi^\beta = -d\tau^2.$$

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, 因为根本不会有人问 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 2.1 中还会继续讲解.

⁴⁰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.

回到问题上, 现在我们看看把它从 ξ^α 换到别的任意坐标系 x^μ 会怎样:

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

可见有测地线的意味在里头, 但为了消除系数, 我们可以乘以一个 $\frac{\partial x^\lambda}{\partial \xi^\alpha}$, 则有

$$\frac{d^2 x^\lambda}{d\tau^2} + \frac{\partial x^\lambda}{\partial \xi^\alpha} \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0, \quad (1.3.11)$$

可见在局部坐标系中的直线方程就是一般坐标系的测地线方程, 这里说明

$$\Gamma_{\mu\nu}^\lambda = \frac{\partial x^\lambda}{\partial \xi^\alpha} \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu}, \quad (1.3.12)$$

这就称作与度规相适配的 $\Gamma_{\mu\nu}^\lambda$, 但为什么说它是与度规相适配的? 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

$$g_{\mu\nu} = \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{\partial \xi^\beta}{\partial x^\nu} \eta_{\alpha\beta}. \quad (1.3.13)$$

这称为合同变换. 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

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

即

$$\frac{\partial g_{\mu\nu}}{\partial x^\sigma} = g_{\mu\lambda} \Gamma_{\nu\sigma}^\lambda + g_{\nu\lambda} \Gamma_{\mu\sigma}^\lambda. \quad (1.3.14)$$

我们当然也可以从中“解”出 $\Gamma_{\mu\nu}^\lambda$: 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:

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

即可得到

$$\Gamma_{\mu\nu}^\lambda = \frac{1}{2} g^{\sigma\lambda} \left(\frac{\partial g_{\sigma\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \right). \quad (1.3.15)$$

因而称这个 $\Gamma^\lambda_{\mu\nu}$ 为与度规相适配的⁴¹，这里 $g^{\sigma\lambda}$ 是其逆矩阵的分量，此即规定 $g_{\mu\sigma}g^{\sigma\lambda} = \delta^\lambda_\mu$.

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 $116^\circ\text{E}, 40^\circ\text{N}$ 、 $74^\circ\text{W}, 40^\circ\text{N}$. 地球平均半径约 6371 km.



图 1.9: 从北京到纽约的最短航线

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 维欧氏空间中球面度规为 $g_{\theta\theta} = R^2, g_{\phi\phi} = R^2 \sin^2 \theta$ ，因此

$$g^{\theta\theta} = R^{-2}, \quad g^{\phi\phi} = R^{-2} \sin^{-2} \theta.$$

⁴¹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，它是一个无挠引力理论.

求导得

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

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

$$\Gamma_{\theta\theta}^{\sigma} = \frac{1}{2} g^{\sigma\rho} (g_{\rho\theta,\theta} + g_{\rho\theta,\theta} - g_{\theta\theta,\rho}) = 0,$$

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

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

求得非零分量如下:

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

因此测地线方程为:

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

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

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

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 $\theta'' = \frac{\pi}{2}$, $\phi'' = a\tau + b$:

$$\begin{aligned} \cos \theta &= \sin(a\tau + b + \beta) \sin \alpha = \sin \alpha \sin \varphi', \\ \tan \phi &= \tan(a\tau + b + \beta) \cos \alpha = \cos \alpha \tan \varphi', \end{aligned}$$

其中 $\varphi' = a\tau + b + \beta$, 即

$$\begin{aligned} \theta(\tau) &= \arccos(\sin \alpha \sin \varphi'), \\ \phi(\tau) &= \arctan(\cos \alpha \tan \varphi'). \end{aligned}$$

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:

$$\begin{aligned} \theta(\tau) &= \frac{\pi}{2} - \arccos(\sin \alpha \sin \varphi'), \\ \phi(\tau) &= \arctan(\cos \alpha \tan \varphi'). \end{aligned}$$

设北京 $(\phi(0), \theta(0))$ 、纽约 $(\phi(1), \theta(1))$. 设航线为曲线 $\mathbf{x} = (\phi(\tau), \theta(\tau))$, 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

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

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

1.3.4 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 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

$$R_{(x)} = \lim_{r \rightarrow 0} \frac{\pi r^2 - |B(x, r)|}{\pi r^4 / 24}. \quad (1.3.16)$$

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

$$\text{Ric}_{(x)}(v, v) = \lim_{r \rightarrow 0} \lim_{\theta \rightarrow 0} \frac{\frac{1}{2}r^2\theta - |A|}{r^4\theta/24}. \quad (1.3.17)$$

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⁴², 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⁴³!), 我们不妨来回看测地线方程中的这个 $\Gamma^\lambda_{\mu\nu}$, 它是矢量 $\frac{\partial e_\mu}{\partial x^\nu}$ 的分量. 这不就是 e_μ 沿着各个 x^ν 方向的变化么? 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 e_σ 沿 x^ν 走一小段以比较出 $\frac{\partial e_\sigma}{\partial x^\nu}$, 随后再将其沿 x^μ 走一小段以比较出 $\frac{\partial}{\partial x^\mu} \left(\frac{\partial e_\sigma}{\partial x^\nu} \right)$. 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), 再次求导必将带着括号里的基矢导数走, 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:

$$R = \frac{\partial}{\partial x^\mu} \left(\frac{\partial e_\sigma}{\partial x^\nu} \right) - \frac{\partial}{\partial x^\nu} \left(\frac{\partial e_\sigma}{\partial x^\mu} \right), \quad (1.3.18)$$

⁴² 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.

⁴³ 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.

代入分量结果，将其展开：

$$\begin{aligned}
 R &= \frac{\partial}{\partial x^\mu} (\Gamma^\lambda_{\sigma\nu} e_\lambda) - \frac{\partial}{\partial x^\nu} (\Gamma^\lambda_{\sigma\mu} e_\lambda) \\
 &= \frac{\partial \Gamma^\lambda_{\sigma\nu}}{\partial x^\mu} e_\lambda + \Gamma^\lambda_{\sigma\nu} \frac{\partial e_\lambda}{\partial x^\mu} - \frac{\partial \Gamma^\lambda_{\sigma\mu}}{\partial x^\nu} e_\lambda - \Gamma^\lambda_{\sigma\mu} \frac{\partial e_\lambda}{\partial x^\nu} \\
 &= \frac{\partial \Gamma^\lambda_{\sigma\nu}}{\partial x^\mu} e_\lambda + \Gamma^\lambda_{\sigma\nu} \Gamma^\alpha_{\lambda\mu} e_\alpha - \frac{\partial \Gamma^\lambda_{\sigma\mu}}{\partial x^\nu} e_\lambda - \Gamma^\lambda_{\sigma\mu} \Gamma^\alpha_{\lambda\nu} e_\alpha \\
 &= \frac{\partial \Gamma^\lambda_{\sigma\nu}}{\partial x^\mu} e_\lambda + \Gamma^\alpha_{\sigma\nu} \Gamma^\lambda_{\alpha\mu} e_\lambda - \frac{\partial \Gamma^\lambda_{\sigma\mu}}{\partial x^\nu} e_\lambda - \Gamma^\alpha_{\sigma\mu} \Gamma^\lambda_{\alpha\nu} e_\lambda,
 \end{aligned}$$

可见，其分量是

$$R^\lambda_{\mu\nu\sigma} = \frac{\partial \Gamma^\lambda_{\sigma\nu}}{\partial x^\mu} - \frac{\partial \Gamma^\lambda_{\sigma\mu}}{\partial x^\nu} + \Gamma^\alpha_{\sigma\nu} \Gamma^\lambda_{\alpha\mu} - \Gamma^\alpha_{\sigma\mu} \Gamma^\lambda_{\alpha\nu}, \quad (1.3.19)$$

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。为方便起见这样定义了一个新东西：

$$R_{\mu\nu} = R^\lambda_{\mu\lambda\nu}. \quad (1.3.20)$$

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。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⁴⁴，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 (1.3.17) 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

$$R = R^\nu_{\nu}. \quad (1.3.21)$$

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 (1.3.21) 又是

$$R = g^{\mu\nu} R_{\mu\nu}. \quad (1.3.22)$$

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

⁴⁴不过我们可以剧透其物理意义：“时空体积”沿不同方向的变化方式。

微分方程⁴⁵；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 $\Gamma^\lambda_{\mu\nu}$ 。但由于我们之前对 $\Gamma^\lambda_{\mu\nu}$ 只做了定性研究，而这里要想定量证明 $R^\lambda_{\mu\nu\sigma}$ 的不依赖性，就得先定量证明 $\Gamma^\lambda_{\mu\nu}$ 的依赖性，此即说要找到它们的坐标变换公式。

1.3.5 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)、Ricci⁴⁶ (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⁴⁷。

⁴⁵ $\Gamma^\lambda_{\mu\nu}$ 是 g 的一阶方程：克氏符就像力，而度规就像其势。

⁴⁶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。

⁴⁷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:

$$\hat{g}_{\hat{\mu}\hat{\nu}} = g_{\mu\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}}, \quad (1.3.23)$$

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

$$\hat{g}^{\hat{\sigma}\hat{\lambda}} = g^{\sigma\lambda} \frac{\partial \hat{x}^{\hat{\sigma}}}{\partial x^\sigma} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda}. \quad (1.3.24)$$

由于标量函数 f 就是个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. 但矢量会怎样呢? 我们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, 说明矢量的分量应当满足

$$\hat{U}^{\hat{\mu}} = \frac{\partial \hat{x}^{\hat{\mu}}}{\partial x^\mu} U^\mu, \quad (1.3.25)$$

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等同于时空本身了⁴⁸, 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:

$$\hat{e}_{\hat{\mu}} = \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} e_\mu, \quad (1.3.26)$$

可见这在狭相中也是说得通的.

再如更多指标的物理量, 比如曲率 $R^\lambda_{\mu\nu\sigma}$, 我们也可以按照类似思路构造标量, 比如说

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

因此

$$\hat{R}^{\hat{\lambda}}_{\hat{\mu}\hat{\nu}\hat{\sigma}} = R^\lambda_{\mu\nu\sigma} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial x^\sigma}{\partial \hat{x}^{\hat{\sigma}}}. \quad (1.3.27)$$

⁴⁸一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 .

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. 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 ∇_μ

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

$$\frac{\partial f}{\partial x^\nu} = \frac{\partial f}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\nu}}}{\partial x^\nu}, \quad \frac{\partial U}{\partial x^\nu} = \frac{\partial U}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\nu}}}{\partial x^\nu};$$

但二者的区别在于矢量还有分量和基矢之分. 若直接对其分量求的话, 一般情况下会不协变:

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

然而, 如果我们把 U 的偏导拆成分量写, 请看:

$$\frac{\partial}{\partial x^\nu} (U^\mu e_\mu) = \frac{\partial U^\mu}{\partial x^\nu} e_\mu + U^\mu \frac{\partial e_\mu}{\partial x^\nu} = \left(\frac{\partial U^\mu}{\partial x^\nu} + \Gamma^\mu_{\sigma\nu} U^\sigma \right) e_\mu.$$

可见由于 $\Gamma^\mu_{\sigma\nu}$ 的存在, 我们得以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. 我们称这个分量

$$\nabla_\nu U^\mu = \frac{\partial U^\mu}{\partial x^\nu} + \Gamma^\mu_{\sigma\nu} U^\sigma \quad (1.3.28)$$

就叫 U 的协变导数 (的分量), 可以将导数算符记作 ∇_ν . 实际上它是协变导数 $\frac{\partial U}{\partial x^\nu}$ 的第 μ 分量. 不过没关系, 我们多数使用分量语言. 而由于标量 f 的偏导数已经是协变的, 因此对 f 而言,

$$\nabla_\nu f = \frac{\partial f}{\partial x^\nu}. \quad (1.3.29)$$

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 $\Gamma^\mu_{\sigma\nu}$ 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**！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

$$\nabla_\sigma g_{\mu\nu} = \frac{\partial g_{\mu\nu}}{\partial x^\sigma} - g_{\mu\lambda} \Gamma^\lambda_{\sigma\nu} - g_{\nu\lambda} \Gamma^\lambda_{\sigma\mu}. \quad (1.3.30)$$

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：

$$\nabla_\sigma g_{\mu\nu} = 0. \quad (1.3.31)$$

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，所以我们可以称满足 (1.3.31) 式的是与度规相适配的协变导数算符。至此，我们就已经知道如何对标量、矢量以及度规求其对应的协变导数了！

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，也就是“直线”的“最短长度”的定义。这可以预料到，因为度规决定着距离，而 $\Gamma^\kappa_{\sigma\lambda}$ 最开始是用“直线”的平移定义导出的。当然，很明显球面上的测地线或许就不太能满足这件事，因为两点之间的测地线方程不仅给出了劣弧，它 also 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”。）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：

$$T(\alpha) = \int d\tau = \int_{\alpha} \sqrt{-g_{\mu\nu} dx^\mu dx^\nu} = \int_{\tau_0}^{\tau_1} \sqrt{-g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu} d\tau, \quad (1.3.32)$$

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，对其微分叫做变分 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

$$\delta T = 0, \quad (1.3.33)$$

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:

$$\frac{d^2 x^\lambda}{d\tau^2} + \frac{1}{2} g^{\sigma\lambda} \left(\frac{\partial g_{\sigma\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \right) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 0. \quad (1.3.34)$$

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

$$\frac{d}{d\tau} \frac{\partial L}{\partial \dot{x}^\mu} - \frac{\partial L}{\partial x^\mu} = 0, \quad (1.3.35)$$

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

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

$\Gamma^\lambda_{\mu\nu}, R^\lambda_{\mu\nu\sigma}$ 等的坐标变换

说了这么多, 到底还是没讲关于 Riemann 曲率的“几何意义”. 但只有我们在得到这些东西的坐标变换时, 才能够讲清楚这些事. 尤其是 $\Gamma^\lambda_{\mu\nu}$, 它是坐标系处处具备的自身信息, 我们可以从定义或度规计算它. 如果利用定义

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

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

$$\hat{\Gamma}^\lambda_{\hat{\mu}\hat{\nu}} = \Gamma^\lambda_{\mu\nu} \frac{\partial x^\mu}{\partial \hat{x}^{\hat{\mu}}} \frac{\partial x^\nu}{\partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda} + \frac{\partial^2 x^\lambda}{\partial \hat{x}^{\hat{\mu}} \partial \hat{x}^{\hat{\nu}}} \frac{\partial \hat{x}^{\hat{\lambda}}}{\partial x^\lambda}. \quad (1.3.36)$$

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.

接下来推 $R^\lambda_{\mu\nu\sigma}$ 的变换. 我们写下 \hat{S} 系的

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

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

从而只剩下协变部分. 同理

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

故

$$\hat{R}^{\hat{\lambda}}_{\hat{\mu}\hat{\nu}\hat{\sigma}} = R^{\lambda}_{\mu\nu\sigma} \frac{\partial\hat{x}^{\hat{\lambda}}}{\partial x^{\lambda}} \frac{\partial x^{\mu}}{\partial\hat{x}^{\hat{\mu}}} \frac{\partial x^{\nu}}{\partial\hat{x}^{\hat{\nu}}} \frac{\partial x^{\sigma}}{\partial\hat{x}^{\hat{\sigma}}}. \quad (1.3.37)$$

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

$$\hat{R}_{\hat{\mu}\hat{\nu}} = \hat{R}^{\hat{\lambda}}_{\hat{\mu}\hat{\lambda}\hat{\nu}} = R_{\mu\nu} \frac{\partial x^{\mu}}{\partial\hat{x}^{\hat{\mu}}} \frac{\partial x^{\nu}}{\partial\hat{x}^{\hat{\nu}}} \quad (1.3.38)$$

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.

$R^{\lambda}_{\mu\nu\sigma}$ 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. 由于我们是对基矢求偏导, 现在就可以用上协变导数的语言:

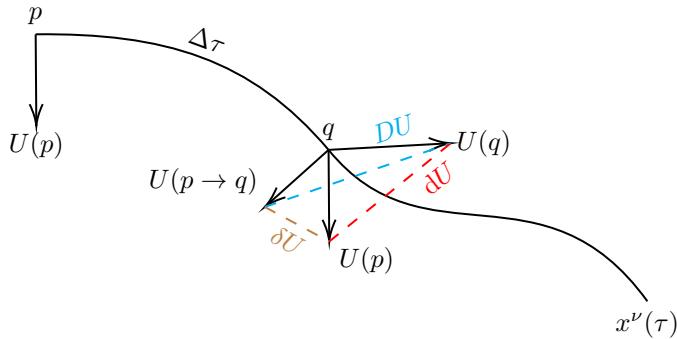
$$R^{\lambda}_{\mu\nu\sigma}e_{\lambda} = [\nabla_{\mu}, \nabla_{\nu}]e_{\sigma},$$

其中方括号称作对易子, 也就是交叉相减 $[\nabla_{\mu}, \nabla_{\nu}] = \nabla_{\mu}\nabla_{\nu} - \nabla_{\nu}\nabla_{\mu}$. 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 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:

$$\frac{\partial U^{\mu}}{\partial x^{\nu}} = \lim_{q \rightarrow p} \frac{U^{\mu}(q) - U^{\mu}(p)}{\Delta x^{\nu}}, \quad (1.3.39)$$

图 1.10: 图中 $\Delta\tau$ 等差值均夸大

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:

$$dU^\mu = \frac{\partial U^\mu}{\partial x^\nu} dx^\nu. \quad (1.3.40)$$

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:

$$\nabla_\nu U^\mu = \lim_{q \rightarrow p} \frac{U^\mu(q) - (U^\mu(p) - \Gamma^\mu_{\sigma\nu} U^\sigma(p) \Delta x^\nu)}{\Delta x^\nu}, \quad (1.3.41)$$

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

$$U^\mu(p \rightarrow q) = U^\mu(p) - \Gamma^\mu_{\sigma\nu} U^\sigma(p) \Delta x^\nu. \quad (1.3.42)$$

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:

$$DU^\mu = dx^\nu \nabla_\nu U^\mu. \quad (1.3.43)$$

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:

$$\delta U^\mu = U^\mu(p \rightarrow q) - U^\mu(p). \quad (1.3.44)$$

由 (1.3.42) 或者图示知

$$\delta U^\mu = dU^\mu - DU^\mu = -\Gamma^\mu_{\sigma\nu} U^\sigma dx^\nu. \quad (1.3.45)$$

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

$$\Gamma^\mu_{\sigma\nu} U^\sigma = -\frac{\delta U^\mu}{\partial x^\nu}, \quad (1.3.46)$$

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:

$$DU^\mu = dx^\nu \nabla_\nu U^\mu = 0. \quad (1.3.47)$$

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:

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

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

$$D(g_{\mu\nu} U^\mu U^\nu) = 0, \quad (1.3.48)$$

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⁴⁹. 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, 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

⁴⁹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?

不同. 在一张平直在白纸上, 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

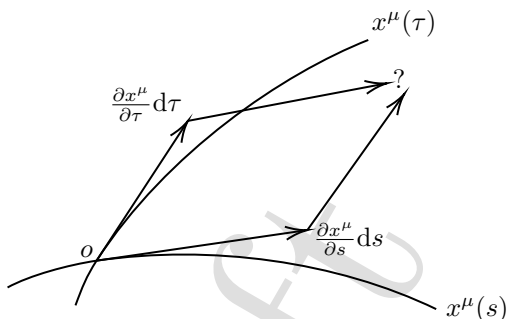


图 1.11: 无穷小测地回路

o , 但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

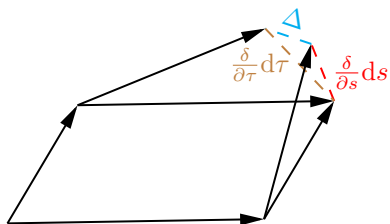


图 1.12: 挠率导致的差值

量加以区分就很有帮助了：

$$\Gamma^\mu_{[\lambda\nu]} = \frac{\Gamma^\mu_{\lambda\nu} - \Gamma^\mu_{\nu\lambda}}{2}, \quad (1.3.49)$$

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。像这样在指标上标注括号，表示其反称部分，而**对称**部分是指

$$\Gamma^\mu_{(\lambda\nu)} = \frac{\Gamma^\mu_{\lambda\nu} + \Gamma^\mu_{\nu\lambda}}{2}. \quad (1.3.50)$$

二者相加即为 $\Gamma^\mu_{\lambda\nu}$ ，因此称为“部分”。反称部分为零，就意味着 $\Gamma^\mu_{\lambda\nu}$ 是（关于指标 λ, ν ）对称的，交换指标的分量一致；反之，则称其是反称的，交换指标的分量会反号。

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⁵⁰。现在研究矢量 U^μ 沿着回路的平行移动。根据

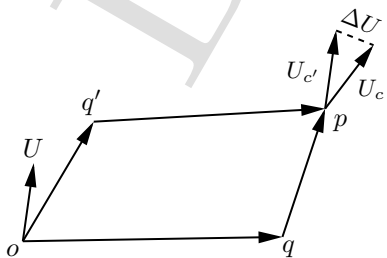


图 1.13: 曲率导致的差值

定义，先移动至 q 处有

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

⁵⁰如果不闭合仍可以构造五边形回路，在第三章中对此详解。

再移动至 p 点有

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

同理沿 o, q', p 有

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

相减得到

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

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。

1.3.6 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。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。

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

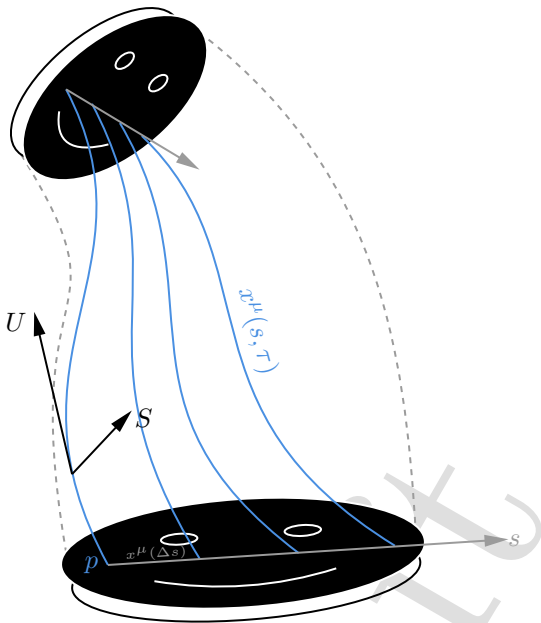


图 1.14: 测地偏离 (s 轴不一定是直的)

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. 它一方面是一张面，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，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的世界线汇，同时也可看作一个立体的世界管. 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⁵¹.

世界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矢或者说速度

$$U^\mu = \frac{\partial x^\mu}{\partial \tau}, \tag{1.3.51}$$

⁵¹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.

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

$$S^\mu = \frac{\partial x^\mu}{\partial s} \quad (1.3.52)$$

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 称作偏离矢量。

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

$$\frac{D^2 S^\mu}{d\tau^2} = U^\rho \nabla_\rho U^\sigma \nabla_\sigma S^\mu, \quad (1.3.53)$$

答案是

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

这称作测地偏离方程。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：

$$U^\rho \nabla_\rho S^\mu = S^\rho \nabla_\rho U^\mu. \quad (1.3.55)$$

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,

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

而另一方面根据定义,

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

从而

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

此即

$$R^\lambda_{\mu\nu\sigma} U^\sigma = -[\nabla_\mu, \nabla_\nu] U^\lambda, \quad (1.3.56)$$

则继续有

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

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

$$\eta^\mu = S^\mu + g_{\sigma\nu} U^\sigma S^\nu U^\mu, \quad (1.3.57)$$

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

$$\frac{D^2 \eta^\mu}{d\tau^2} = R^\mu_{\rho\nu\sigma} U^\rho U^\nu \eta^\sigma, \quad (1.3.58)$$

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

$$\frac{d^2 x^i}{dt^2} = -\eta^{ij} \frac{\partial \phi}{\partial x^j}, \quad (1.3.59)$$

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

$$\frac{d^2 \eta^i}{dt^2} = -\eta^{ij} \frac{\partial^2 \phi}{\partial x^j \partial x^k} \eta^k, \quad (1.3.60)$$

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：

$$\eta^{ij} \frac{\partial^2 \phi}{\partial x^i \partial x^j} = 4\pi G\rho, \quad (1.3.61)$$

而这个操作对应着右边缩并为

$$R^\mu_{\rho\mu\nu} U^\rho U^\nu = R_{\rho\nu} U^\rho U^\nu, \quad (1.3.62)$$

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：

$$R_{\mu\nu} U^\mu U^\nu \sim 4\pi G T_{\mu\nu} U^\mu U^\nu, \quad (1.3.63)$$

或者说

$$T_{\mu\nu}U^\mu U^\nu = \rho. \quad (1.3.64)$$

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,而无论是物体还是流体,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导致了我們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、速度场等概念就更为有用了⁵².

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 \mathbf{v} 推广到了 4 维中的 U , 关系是

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

⁵²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.

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，即动流密度

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

这里 $U_\mu = g_{\mu\nu} U^\nu$ ，也可以看作矢量，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。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⁵³，这称作理想流体。这样根据此前讨论，在某点的相对静止系 \tilde{x}^β 中 T 就成为了对角矩阵：

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

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 \mathbf{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 \mathbf{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：

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

⁵³ 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。

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。虽然目前是第一次正经地看到张量这个词，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，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⁵⁴。比如说 T_{00} 就是 ρ ，描述多少能量经过这一点向未来流动，能量的流动成了能流； 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

⁵⁴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]。

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⁵⁵，一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。因此我们只需考察四个坐标即可：不论所涉及物质场⁵⁶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。上式体现了在微分水平上能动张量的守恒，而将其在两个同时面⁵⁷上积分，Gauss 定理就会给出整体的平衡关系：在 t_1, t_2 之间积分后就得出

$$\int_{t=t_1} T_{0\alpha} d^3x = \int_{t=t_2} T_{0\alpha} d^3x, \quad (1.4.11)$$

对于任意选定的 Lorentz 参考系，这个式子在 $\alpha = 0$ 时就是能量守恒，在其它情况下就是动量守恒。

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

$$T: \mathbf{R}_q^4 \times \mathbf{R}_q^4 \rightarrow \mathbf{R} \quad (1.4.12)$$

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

$$T(U, \tilde{U}) = T_{\mu\nu} U^\mu \tilde{U}^\nu \quad (1.4.13)$$

⁵⁵ 这样一些条件在连续介质力学中由一系列状态方程表示。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其间的相互关系。由于考虑的是理想流体，我们在此不去为这些事操心。

⁵⁶ 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。

⁵⁷ 时空是 4 维的，扔掉时间就得到了我们的空间，这称作时空的余 1 维子空间，可以形象地叫做一张超曲面。所谓的同时面就是这种同时事件所构成的 3 维空间。

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”. 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 有一样多的分量, 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

$$T_{\mu\nu} = \rho g_{\mu\nu} + (\rho + p)U_\mu U_\nu, \quad (1.4.14)$$

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

$$\nabla_{[\alpha} R^\mu{}_{\beta\gamma]\delta} = \nabla_\alpha R^\mu{}_{\beta\gamma\delta} + \nabla_\beta R^\mu{}_{\gamma\alpha\delta} + \nabla_\gamma R^\mu{}_{\alpha\beta\delta} = 0, \quad (1.4.15)$$

令 $\mu = \alpha$ 得

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

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

$$\nabla^\mu R_{\mu\nu} = \frac{1}{2} \nabla_\nu R, \quad (1.4.16)$$

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 $R_{\mu\nu} = 4\pi G T_{\mu\nu}$ 就是

$$R = 4\pi G T,$$

因此

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

即 T 在整个时空里是个常数。但另一方面

$$T = \rho U_\mu U^\mu + p(g^{\mu\nu} g_{\mu\nu} + U_\mu U^\mu) = -\rho + p(\delta_\mu^\mu - 1) = 3p - \rho,$$

取低速情况使得 p 相对于 ρ 而言忽略不计，则 $T \approx -\rho$ ，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

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

它也可以对应于一个变分原理，其泛函是

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

这里我们以后会知道 $\sqrt{-g}d^4x$ 表示时空体积微元， g 是度规的行列式：

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

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:

$$S(g) = \int (R + L_M) \sqrt{-g} d^4x, \quad (1.4.24)$$

于是我们发现可从另一角度定义能动张量:

$$\frac{\partial}{\partial g^{\mu\nu}} (L_M \sqrt{-g}) = -\frac{\sqrt{-g}}{2} T_{\mu\nu} = -\kappa T_{\mu\nu}, \quad (1.4.25)$$

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⁵⁸, 这个泛函也就称为 *Einstein-Hilbert* 作用量.

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 (M, g) , 其中 M 是微分流形——为了把 g 用局部坐标写出来, 需要在最底层的拓扑对象上处处 (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 \mathbf{R}^4 ; 而这若干个坐标系相容地粘贴在一起, “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.

⁵⁸但后来的文献表示其实 Hilbert 并没有先于 Einstein.

水星近日点进动

通过求解在所谓的线性近似下的场方程，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，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，真空方程另一著名的解是 Schwarzschild⁵⁹ 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

$$-A dt^2 + A^{-1} dr^2 + r^2 d\Omega, \quad (1.4.26)$$

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}^+ 为黑洞的视界。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” \mathcal{I}^+ 。不能把信号传送到远处的点之集合称为黑洞区域，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的定量解释。

这些问题花了很长时间才理清，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⁶⁰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。进入黑洞的测地线会在有限固有时内到达奇点，不可无限地延伸下去，称作不可延测地线。接近奇点的宏观观者会被差距强烈的潮汐力撕碎。

⁵⁹Karl 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。

⁶⁰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假设. 第一个是弱宇宙监督假设: 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. 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的 *Cauchy* 曲面了, 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, 宇宙是均匀且各向同性的. 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

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}, \quad (1.4.27)$$

这里 Λ 称为宇宙学常数, 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. 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. 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

2.1 时空的几何学

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，因此我们有必要用严格的术语来陈述其数学框架。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。

2.1.1 闵氏时空

之前已经说过这些矢量的坐标都是实数。矢量空间 \mathbf{R}^4 理解为所有矢量构成的集合，更严谨的表述将稍后再说。我们知道 \mathbf{R}^4 有自然坐标，这个天然的坐标系当然符合我们对“特殊”的需求，因此它的基矢就单独取名为坐标基。在熟悉这些概念之后，我们可以尝试给闵氏时空一个数学的表达：

定义 2.1 (闵氏度规). \mathbf{R}^4 中一点 q 处的闵氏度规是该处矢量空间 \mathbf{R}_q^4 上的双线性函数 $\eta: \mathbf{R}_q^4 \times \mathbf{R}_q^4 \rightarrow \mathbf{R}$ ，如果 $\{e_\mu\}$ 是矢量空间的坐标基，则

$$\eta(e_\mu, e_\nu) = \eta_{\mu\nu}.$$

函数的像 $\eta(v, w)$ 称为二者的内积 $\langle v, w \rangle$ ，或写作 $v \cdot w$ 。

定义 2.2 (闵氏时空). 闵氏时空 \mathbf{R}^{3+1} 是 \mathbf{R}^4 再处处配上一个闵氏度规，或者说直接配上了一个闵氏度规场。有时我们写作 (\mathbf{R}^4, η) 。

我们可以看到，坐标基在 η 的衡量下是正交的（彼此内积为零）、归一的（模长为 1 或 -1 ），那么就说明坐标基（在其衡量下）是正交（归一）基。

狭相讨论的是闵氏时空中的物理学. 在这种情形下, 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“都一样”, 而且还可以和时空本身等同.

定义 2.3 (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、类时或类光矢量. 所有类光矢量之集 C_N 称为矢量空间 \mathbf{R}^4 的光锥. 号差 -2 不外乎正负定义颠倒之.

设 $v = v^\mu e_\mu, w = w^\nu e_\nu$, 根据度规的线性就有 $v \cdot w = \eta_{\mu\nu} v^\mu w^\nu$. 因此类时矢量位于光锥“内部”, 类空矢量则位于光锥“外部”. 从 \mathbf{R}^3 的 Schwartz 不等

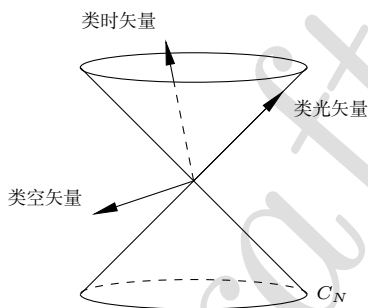


图 2.1: 光锥

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 $v \cdot w < 0$, 并指定 v, w 等价, 那就可在光锥内部定义一种等价关系, 称 v, w 具有相同的时间取向. 存在两种等价方式, 一种是选择并指定未来指向的矢量, 那另一种方式的类时矢量就称为过去指向的. 我们可以看到, 这种分类可以推广到非零类光矢量, 但不能推广到类空矢量. 有确定指向的正交基称为时间指向的, 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 上的一个具有确定指向的、时间指向的正交基称为惯性基矢或者惯性观者, 由于矢量空间和时空本身的等同, 我们也可以把这些矢量都“搬”到时空里来.

定义 2.4. 设 $\{e_\mu\}$ 和 $\{\hat{e}_\nu\}$ 是这样两个惯性基矢, 且之间满足 $e_\mu = \Lambda_\mu^\nu \hat{e}_\nu$, 可见, 矩阵 Λ 将满足:

1. (正交) $\Lambda^T \eta \Lambda = \eta$;
2. (可定向) $\det \Lambda = 1$;
3. (时间可定向) $\Lambda^0_0 \geq 1$.

把任意满足上述三个条件的四阶方阵 Λ 称为 *Lorentz* 变换¹.

这种矩阵的全体集 L 结合矩阵乘法运算, 构成了闵氏几何学的对称群, 正是 *Lorentz* 群. 在后面我们会知道, 群相比于集合稍微多了一种元素的乘法运算结构, 用以更方便地描述对称性. 由正交条件很容易证明, 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. 这种形式的全部矩阵集 R 是 *Lorentz* 群 L 的一个子群, 称为转动群². 虽然此概念对本节所述无用, 但应指出, 在诸多领域中 (如粒子物理) 须考虑由 *Lorentz* 群和时空平移 ($x^\mu \rightarrow x^\mu + \Lambda^\mu$, 其中 Λ^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.

2.2 时空的物理学

为从物理上想像闵氏时空和 *Lorentz* 群, 最好在直观和术语上把 \mathbf{R}^4 中的“矢量”和“点” (矢量一端) 区分开来. \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”.

狭相所感兴趣的, 是由一类特殊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, 光子相对于这种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 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

$$c = \frac{2\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}}{t_2 - t_1}.$$

现在, 在坐标系里每个空间点上, 都放置一只全同的时钟, 且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 (因此每个时钟走

¹有时为了强调条件 2、3, 会分别叫做“固有的”和“正交的”.

²这种旋转变换将保持对象的原样, 也称作 $SO(3)$ 群.

时一致), 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 $\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}/c$. 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. 这种坐标系连同处处存在的所有时钟称为一个参考系. 狭相所研究的观者 (无论是否有相对运动) 是 Newton 力学意义下的惯性观者 (即在所引进的参考系中, 不受“力”粒子之轨迹或是一点, 或是一条以不变速度运动的直线). 一个实验事实是, 所有惯性观者均承认相同的真空光速数值³ c . 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, 取几何单位制, 即光速等于无量纲的 1, 则时间将和距离有“完全”一致的量纲. 后文我们将默认这一点.

在我们的事件集的数学模型 M 中, “惯性观者”这种极为微妙而复杂的概念, 与惯性基矢 $\{e_\mu\}$ 的简单概念是等同的. 若 $x \in M$ 是事件且写成 $x = x^\mu e_\mu$, 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矢或观者, 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 $(\Lambda^\nu_\mu) \in R$, 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 的一个元素:

$$L(\theta) = \begin{pmatrix} \cosh \theta & -\sinh \theta & 0 & 0 \\ -\sinh \theta & \cosh \theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

且若两组惯性基矢由该 Lorentz 变换联系, 则坐标变换具体写出来就是

$$\begin{aligned} \hat{x}^0 &= (\cosh \theta)x^0 - (\sinh \theta)x^1, \\ \hat{x}^1 &= -(\sinh \theta)x^0 + (\cosh \theta)x^1, \\ \hat{x}^2 &= x^2, \\ \hat{x}^3 &= x^3. \end{aligned}$$

³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.

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. 而且上面的公式, 就成为狭相中通常所说的“Lorentz 变换”或者 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, 只要考虑所谓的特殊 Lorentz 变换 $L(\theta)$ 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 $R_1, R_2 \in R$, 使得 $\Lambda = R_1 L(\theta) R_2$. 因为 R_1, R_2 不包含相对运动, 所以全部的运动学都包含在 $L(\theta)$ 之中.

现在假定, x, x_0 是 M 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 间的空间间隔等于光在事件间的时间间隔中传播的距离. 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 $x_0, x \in M$ 之间的 $x - x_0$ 类光, 则在 M 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 $x_0, x \in M$ 之间的 $x-x_0$ 类时, 则在 M 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 2.2.1. 设 x_0, x 和 y 是 M 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 $x_0, x \in M$, $g(x-x_0, x-x_0)$ 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 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 $L(\theta_1)$ 和 $L(\theta_2)$. 根据双曲函数的性质, 有 $L(\theta_1)L(\theta_2) = L(\theta_1 + \theta_2)$. 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 Δx^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.

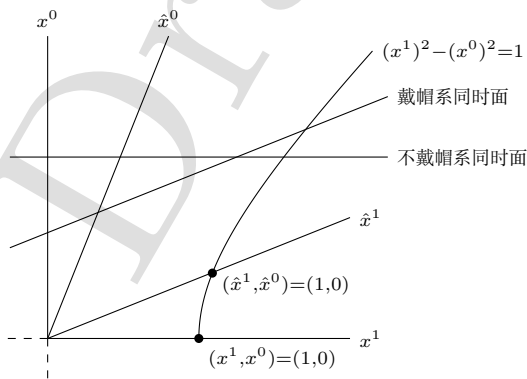


图 2.2: 同时的相对性

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.

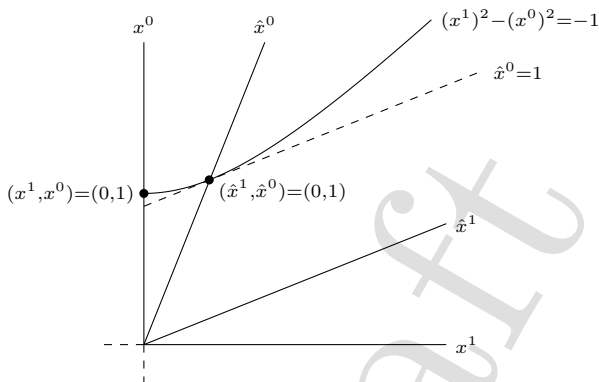


图 2.3: 钟慢效应

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

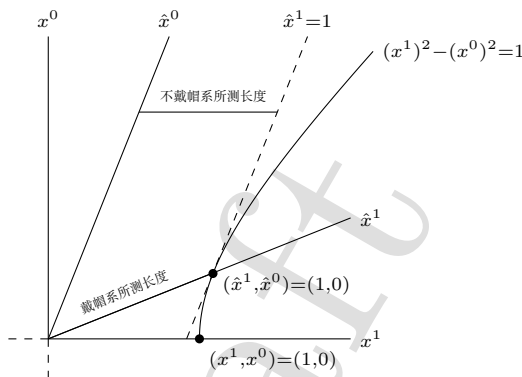


图 2.4: 尺缩效应

一把沿戴帽坐标系 \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时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, 会更容易发

现同时的相对性. 在通过 M 的一点的所有超平面中, 有一种对某参考系的 t 为常数的特殊超平面. 在一定条件下, 一定有保度规的 Lorentz 变换把这个超平面变成过此点的其他超平面, 所以在这个度规下找不到什么东西能把这个特定的超平面区别出来. 这与之前复杂的计算相一致.

必须清楚, 这并不是准确的数学陈述. 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的指导, 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, 但人类将来必须面对.

2.4 力学

前面我们看到, 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. 下面我们将给出进一步的定义.

若 $I \subseteq R$ 是一种间隔, 则映射 $\alpha: I \rightarrow M$ 是 M 中的一条曲线. 在任意惯性基矢下, 对任意 $\xi \in I$, 我们可以写下 $\alpha(\xi) = x^\mu(\xi)e_\mu$. 假设 α 光滑, 即 $x^\mu(\xi)$ 在 I 上是 C^∞ 的, 且速度矢量

$$\alpha'(\xi) = \frac{dx^\mu}{d\xi} e_\mu$$

对任意 $\xi \in I$ 不为零 (几何学习惯把一点的切空间与该矢量本身等同). 这种光滑性的定义显然与 M 中惯性基矢的选取无关. 若对任意 $\xi \in I$,

$$\alpha'(\xi) \cdot \alpha'(\xi) = \eta_{\mu\nu} \frac{dx^\mu}{d\xi} \frac{dx^\nu}{d\xi}$$

是正、负或者零, 则曲线 α 分别称为类空、类时或类光曲线. 若一条类时曲线 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, 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. 相对论限制现在可以严格表述为: 有质粒子的世界线是类时的. 至于类光曲线, 就对应于电磁理论的几何光学极限了.

设 $I = [\xi_0, \xi_1]$, 且 $\alpha: [\xi_0, \xi_1] \rightarrow M$ 是一条类时世界线, 则这种 α 所经历

的这一段线长为

$$\begin{aligned} L(\alpha) &= \int_{\xi_0}^{\xi_1} \sqrt{|g(\alpha'(\xi), \alpha'(\xi))|} \, d\xi \\ &= \int_{\xi_0}^{\xi_1} \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\xi} \frac{dx^\nu}{d\xi}} \, d\xi, \end{aligned}$$

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 2.4.1. 设 x_0 和 x 是 M 中的事件, $x - x_0$ 类时且指向未来, 当且仅当在 M 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. 这一明显的事实称为 “双生子佯谬”.

正如欧氏几何中的曲线具有弧长参数那样, 一条类时世界线具有如下定义的固有时参数或仿射参数⁴.

定义 2.5 (固有时). 对 $[\xi_0, \xi_1]$ 中的任意 ξ , 设

$$\tau(\xi) = \int_{\xi_0}^{\xi} \sqrt{|g(\alpha'(\zeta), \alpha'(\zeta))|} \, d\zeta,$$

此即 α 从 $\alpha(\xi_0)$ 到 $\alpha(\xi)$ 的固有时⁵.

$\tau = \tau(\xi)$ 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

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

⁴只能用仿射参数描述光.

⁵光速不变意味着固有时均匀流逝.

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⁶

$$U(\tau) \cdot U(\tau) = -1.$$

这也说明, 时空中通常不存在静止粒子——即使在空间中静止, 它也会不断向未来迁移, 这种情况下 4-速分量全部给予了时间. 那种“断掉”的世界线在广相中称为不可延世界线. 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

$$\|V\| = \sqrt{\left(\frac{dx^1}{dx^0}\right)^2 + \left(\frac{dx^2}{dx^0}\right)^2 + \left(\frac{dx^3}{dx^0}\right)^2}$$

是粒子的相对速度

$$V = V(x^0) = \frac{dx^i}{dx^0} e_i = V^i e_i$$

在给定的惯性坐标系中所测得的通常量值. 则我们发现

$$U = (1 - \|V\|^2)^{-1/2} (V + e_0).$$

把 M 中一个质点用 (α, m) 来代表, 其中 α 是一条类时世界线, m 是这个粒子的固有质量或静质量, 它是一个正的常数, 不随参考系变换而改变. 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. (α, m) 的 4-动量就可以自然地定义为 $P = mU$. 因此

$$P \cdot P = -m^2.$$

在任意惯性基矢下, 有

$$\begin{aligned} P &= P^\mu e_\mu = mU^\mu e_\mu = m \frac{dx^\mu}{d\tau} e_\mu \\ &= m(1 - \|V\|^2)^{1/2} (V + e_0), \end{aligned}$$

即惯性坐标中 P 的“空间”部分是

$$P = \frac{m}{\sqrt{1 - \|V\|^2}} V,$$

⁶光速不变保证我们可以有这种归一化.

如果 $\|V\| \ll 1$, 那么这个量近似于 mV . 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

$$P^i = \frac{m}{\sqrt{1-\|V\|^2}} V^i = mV^i + \frac{1}{2}mV^i\|V\|^2 + \cdots,$$

这给出了经典动量的分量及其“相对论性修正”. 有时为了保持在形式上与 Newton 力学类似, 会把 $m/\sqrt{1-\|V\|^2}$ 称为“相对论质量”或动质量, 但此属历史遗留问题, 现代不怎么这个概念. 再来看看 P 的时间分量, 即

$$P^0 = -P \cdot e_0 = \frac{m}{\sqrt{1-\|v\|^2}} = m + \frac{1}{2}m\|v\|^2 + \cdots,$$

相应于 Newton 动能项 $(1/2)m\|V\|^2$ 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 $\|V\| = 0$ 的情况 (即粒子相对于观者静止), 则有质能方程

$$E = m,$$

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.

下面会看到, 不仅仅有质粒子有“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, 光子的特性不是固有质量而是频率 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 (Doppler 效应). 而且, 我们没有相应类比出“固有质量”的“固有频率”, 因为不存在光子在其中静止的惯性观者. 为了尝试刻画这种特性, 考虑一点 $x_0 \in M$ 、一条指向未来的类光矢量 N 和一个间隔 $I \subseteq \mathbb{R}$. 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. 对于类光世界线, NL 可以在任意惯性基矢下写成

$$N = (-N \cdot e_0)(d + e_0),$$

其中

$$d = \frac{(N \cdot e_1)e_1 + (N \cdot e_2)e_2 + (N \cdot e_3)e_3}{\sqrt{(N \cdot e_1)^2 + (N \cdot e_2)^2 + (N \cdot e_3)^2}}$$

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 M 中的光子定义成 M 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

$$N = \varepsilon(d + e_0).$$

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

$$\frac{\hat{\varepsilon}}{\varepsilon} = \frac{\hat{\nu}}{\nu} \frac{1 - v \cos \theta}{\sqrt{1 - v^2}} = (1 - v \cos \theta) + \frac{1}{2}v^2(1 - v \cos \theta) + \cdots,$$

其中 v 是两个空间坐标系的相对速度, 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 Doppler 效应公式, 其中展开级数第一项是经典 Doppler 效应公式.

下面我们来考察上面的定义的粒子之间几个简单的相互作用, 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). \mathcal{A} 中的一个接触性相互作用是一个三元组 $(\mathcal{A}, x, \tilde{\mathcal{A}})$, 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 $x \in M$ 是这样的事件, 使得

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 $\alpha(\xi_1) = x$);

x 是 $\tilde{\mathcal{A}}$ 中所有粒子的起点;

\mathcal{A} 的总 4-动量等于 $\tilde{\mathcal{A}}$ 的总 4-动量.

严格地说, 第三个条件称为 4-动量守恒. 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性相互作用. 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⁷.

⁷这个例子在 [5] 第 6 章, §14 中有详细分析.

Draft

Chapter 3

弯曲时空

3.1 引力同惯性的等效原理

3.2 广义协变性

3.3 弯曲几何

3.4 参考系

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¹, 使得 Einstein 和 Grossmann 大为困惑, 他们是用坐标下的方程“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.

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²双曲方程组. 于是方程组的波动性质就是显然的了³.

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.

¹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/>

²实际上是拟线性.

³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 微分流形

4.2 切空间

4.3 微分算子、Lie 导数

4.4 微分形式和上同调论

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, 24
- [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. 1.3.5, 2.2
- [5] Synge J L (1972) *Relativity: The Special Theory*. Amsterdam: North-Holland. 7
- [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.
- [9] Einstein A. *Zur Elektrodynamik bewegter Körper*[J]. Annalen der Physik, 1905, 17(891): 50. 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. 33
- [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) 47
- [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》) 54

Draft

索引

P

Poisson 方程, 7

Y

引力场强, 6

Draft