从零学相对论

梁灿彬　曹周键　著

高等教育出版社
HIGHER EDUCATION PRESS

Congling Xue Xiangduilun

从零学相对论

梁灿彬　曹周键　著

中国科学技术大学
图书馆藏

高等教育出版社·北京
HIGHER EDUCATION PRESS　BEIJING

内容提要

本书由我国相对论领域著名专家北京师范大学梁灿彬教授主笔撰写。本书用浅显而直观的几何语言，较少的数学公式，对狭义相对论与广义相对论的物理基础及前沿领域作了通俗、清晰而富有趣味的介绍，对广大理工科学生和教师理解相对论大有帮助。 除了相对论基础知识，本书还详细讲述了光速为什么是极限速率、高速物体的视觉形象、相对论在全球定位系统(GPS)的重要作用、虫洞和时间机器(以及弒母悖论)等饶有兴味的专题。

本书可作为我国高等学校物理学类、天文学类各专业本科生及研究生的广义相对论课程教材，亦可供有关的科学研究人员和教师参考。 各行各业的相对论爱好者(及优秀高中生)只要稍有微积分初步知识都可将本书选作从零起学习相对论的入门读物。

图书在版编目(CIP)数据

从零学相对论 / 梁灿彬,曹周键著. -- 北京:高等教育出版社,2013.10

ISBN 978-7-04-038121-4

Ⅰ. ①从… Ⅱ. ①梁… ②曹… Ⅲ. ①相对论 Ⅳ. ①O412.1

中国版本图书馆 CIP 数据核字(2013)第 169255 号

策划编辑	忻 蓓	责任编辑	忻 蓓	特约编辑	张竹琪	封面设计	于文燕
版式设计	于 婕	插图绘制	尹 莉	责任校对	杨雪莲	责任印制	尤 静

出版发行	高等教育出版社	咨询电话	400-810-0598
社 址	北京市西城区德外大街 4 号	网 址	http://www.hep.edu.cn
邮政编码	100120		http://www.hep.com.cn
印 刷	化学工业出版社印刷厂	网上订购	http://www.landraco.com
开 本	787mm×960mm 1/16		http://www.landraco.com.cn
印 张	16.25	版 次	2013 年 10 月第 1 版
字 数	290 千字	印 次	2013 年 10 月第 1 次印刷
购书热线	010-58581118	定 价	25.70 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换

版权所有 侵权必究

物 料 号 38121-00

前　言

我从1981年起，在美国芝加哥大学相对论组任访问学者两年，在该组“言必称几何”的浓郁学术气氛的耳濡目染之下学到了用整体微分几何研究广义相对论的先进方法，回国后立即开课推广。我在2000年、2001年出版了《微分几何入门与广义相对论》上、下册，2006年及2009年又与周彬博士合作修订为第二版，并扩充内容成为上、中、下三册，在国内相对论界乃至理论物理学界产生了一定的影响。我从1983年起至今从不间断地开设这方面的硕士、博士研究生系列课程，同时吸引了（并且正在吸引着）无数本科生选修和旁听，取得了良好的效果。我欣慰地看到在听课的研究生和本科生中都涌现出一大批“心领神会”的尖子生，而且这些知识对他们日后的研究工作发挥着重要的作用。然而，要用这种方法学习相对论，物理工作者必须过好微分几何这一关，具体来说就是至少大致读懂该书上册的前5章。我每次讲这5章时，总有许多物理听众由于感到格格不入或者拿不出足够的复习时间而先后掉队（撤退）。然而他们对第6章（用几何语言剖析狭义相对论）甚感兴趣，不少撤退了的听众又纷纷来听第6章，无奈由于缺乏前5章的基础而跟不上。前5章早已成为学习《微分几何与广义相对论》的拦路虎。

一年多前，我曾教过的一个学生张华在采访我之前在网上征集问题，征来的一个问题对我触动巨大。这个问题是：“有许多工科学生对相对论也感兴趣，他们希望了解这个领域，但往往不能花费大量时间在这上面，毕竟，三本书对于非物理专业的同学来说还是太耗费精力了，不管是物理专业还是非物理专业的学生，都是国家的栋梁，手心手背都是肉，梁老师是否会考虑为这些同学做些什么呢？”我清醒地知道，即使是物理专业的学生，也只有少数佼佼者能较好地读完这三册书。我的确应该设法为相对论爱好者（连“手心”带“手背”的“肉”）做些什么。经过大约一周的思考，我毅然决定把《电磁学拓展篇》的写作再次暂停（尽管此书早该出版，属于一拖再拖），立即着手写一本当时暂名为《趣味相对论》的书，基本宗旨是：① 尽量用几何语言；② 尽量通俗易懂。这两点宗旨表面上互相抵触，但经验使我自信有办法做到两者兼顾。1997年9月，我在系主任的支持下大胆地为北京师范大学物理系一年级新生开设了一门名为“趣味近代物理”的选修课（限于课时，后来成为“趣味相对论”），先用一种很容易接受的讲法带领学生“学一点点几何”，在此基础上就可以不用前5章而把第6章的主要

内容娓娓道来,特别是用几何语言对尺缩、钟慢、双子、车库佯谬进行剖析。这是一次“空前绝后”的成功尝试(“空前”是指从未有过这种课程,“绝后”是指一直没有时间再讲一次),大半学生基本听懂,约有10%的学生达到心领神会、言必称几何的美好境界。我相信,以此为基础,我可以写出一本既用几何语言又比较好懂的书。

本书虽然力求易懂,却并非科普著作,因为书中出现许多公式,要求读者有起码的一点微积分知识。书中对许多问题的深入细致、追根问底式的讨论也完全不是科普书的味道。书名《趣味相对论》容易被误解为科普书,所以后来放弃。本书共有9章,前5章讲狭义相对论,后4章讲广义相对论,都是从零开始,所以书名改为《从零学相对论》。我一直认为学习相对论有一条捷径(可惜有太多的人不走),就是尽早学习和使用几何语言。“从零学”还意味着要尽早学会最起码的几何语言。本书从第2章开始带领读者“学一点点几何”,以后各章就尽量用几何语言讲解。人们总以为几何语言难学,所以只顾砍柴,不愿磨刀。但是如果按本书的讲法去学,只要很少的磨刀时间就能砍大量的柴,何乐而不为?为了通俗易懂,全书不讲张量,以好懂得多的“线元”代替度规张量,强调“线元决定几何”。“线元”一词贯穿全书,几乎可说“一个线元吃到底”,居然连GPS(全球定位系统)原理、史瓦西时空的最大延拓、嵌入图、虫洞和时间机器等复杂问题也能依靠线元讲清楚。本书虽然起点为零,但却不惜篇幅、不厌其烦地力图带领读者由浅入深地达到比较深刻的理解境界,以下四句可以作为本书的副标题:

四维眼光,几何语言,从零学起,渐步登高。

虽然本书比《微分几何入门与广义相对论》的三册书在篇幅上小很多,但却包含了该书不曾涉及的许多方面的内容,而且多数都有详细讨论,例如:① 时序和因果关系(§3.6);② 高速运动物体的视觉形象(§4.4);③ 质点必须亚光速的论证(§5.9);④ 全球定位系统(GPS)的相对论修正(§8.6);⑤ 广义相对论的近代实验验证(§8.2,§8.5);⑥ 虫洞和时间机器(第9章)。此外,本书还多处添加了相对论史(重要历史事件和著名人物的“八卦”故事)的介绍。

为了适应不同读者,本书内容分为必读和选读两大部分,分别用宋体和楷体排印。必读部分自成体系,不会由于略去选读内容而影响后续必读内容的学习。各页的脚注(如果有的话)与选读内容类似,初次学习的读者不妨略去全部或部分选读和脚注内容。第1至第5章还配有一定数量的习题,对学好这5章颇有帮助,其中第4章的习题尤为重要,建议读者至少要做其中的多数题目。

相对论的几何语言包括两大方面:① 用时空图讨论问题;② 用4维矢量和张量讨论问题。为了大大削减入门难度,本书的几何语言只限于第①方面(在

这方面做了很大程度的发挥),这也许是本书的最大不足。一个明显的后果就是无法讲电磁场,因为一提电磁场就要涉及电磁场张量$F_{\mu\nu}$。我们已经想好了补救办法:再写一本风格类似但一开始就介绍4维矢量和张量的书(不妨戏称为本书的“妹妹篇”)。用映射定义张量是最优雅的讲法,已经详述于《微分几何入门与广义相对论》上册,其唯一缺点是比较抽象,成为许多物理人的拦路虎。我们初步打算在待写的“妹妹篇”中用坐标变换的常用讲法介绍张量,在此基础上就可以充分自由地用张量语言讨论各种相对论问题。

本书第二作者、中国科学院应用数学研究所副研究员曹周键是前面提到的“趣味近代物理”课选修生中鹤立鸡群的第一名,在本科期间又听全了我的硕士生、博士生系列课程,是我在毕生教学中遇到的最优异的学生之一。我们在这次合作写书中有过无数次讨论,我自己百思不解的若干问题跟他讨论后都受到重要启发,经过讨论往往迎刃而解。书中对不少问题的独特理解都包含着他的重要贡献,他的加盟使得本书质量得以明显提高。

笔者曾邀请北京师范大学物理系裴寿镛教授和中国科学院数学研究所吴小宁副研究员审阅过本书第7章的初稿,他们提出的意见和建议对该章的修改起到了重要作用。此外,北京师范大学相对论组博士生韩愈和广州大学本科生杨锦波也曾应邀阅读过本书的部分重要章节,他们也都提出了很好的修改意见。笔者在此要对他们深表谢意。

梁灿彬

2012年10月11日于北京

郑重声明

高等教育出版社依法对本书享有专有出版权。任何未经许可的复制、销售行为均违反《中华人民共和国著作权法》，其行为人将承担相应的民事责任和行政责任；构成犯罪的，将被依法追究刑事责任。为了维护市场秩序，保护读者的合法权益，避免读者误用盗版书造成不良后果，我社将配合行政执法部门和司法机关对违法犯罪的单位和个人进行严厉打击。社会各界人士如发现上述侵权行为，希望及时举报，本社将奖励举报有功人员。

反盗版举报电话　(010)58581897　58582371　58581879

反盗版举报传真　(010)82086060

反盗版举报邮箱　dd@ hep. com. cn

通信地址　北京市西城区德外大街 4 号　高等教育出版社法务部

邮政编码　100120

目　　录

第 1 章　狭义相对论的引入 …… 1

§1.1　相对性原理 …… 1

§1.2　麦克斯韦电磁理论的挑战 …… 3

§1.3　洛伦兹变换 …… 7

第 1 章习题 …… 12

第 2 章　学一点点几何 …… 13

§2.1　线元与几何 …… 13

§2.2　闵氏几何 …… 15

§2.3　球面几何 …… 17

§2.4　曲率 …… 18

§2.5　测地线 …… 18

第 2 章习题 …… 21

第 3 章　相对论的几何表述 …… 22

§3.1　世界线和时空图 …… 22

§3.2　惯性观者、参考系和坐标系 …… 27

§3.3　光速不变的 4 维表述 …… 30

§3.4　固有时与坐标时的区别和联系 …… 33

§3.5　时空图中的坐标轴 …… 34

§3.6　时序和因果关系 …… 36

§3.7　何谓几何语言？何以要用几何语言？ …… 41

第 3 章习题 …… 44

第 4 章　趣味运动学效应 …… 45

§4.1　尺缩效应和车库佯谬 …… 45

4.1.1　尺缩效应 …… 45

4.1.2　车库佯谬 …… 47

§4.2　钟慢效应 …… 48

4.2.1　通常的钟慢效应 …… 48

4.2.2　用洛伦兹变换讨论尺缩钟慢[选读] …… 50

4.2.3　不同比钟方式有不同结果 …… 54

4.2.4 多普勒频移 …… 57
4.2.5 钟慢尺缩效应的实验验证 …… 57
§4.3 双子效应(佯谬) …… 60
4.3.1 双子效应的戏剧性描述 …… 60
4.3.2 双子效应的几何剖析 …… 61
4.3.3 双子“悖论”的长期论战 …… 62
4.3.4 双子效应的实验验证 …… 68
§4.4 高速物体的视觉形象 …… 72
§4.5 用几何语言导出洛伦兹变换[选读] …… 84
第4章习题 …… 86
第5章 相对论质点力学 …… 88
§5.1 动量和动量守恒 …… 88
§5.2 力的定义 …… 90
§5.3 动能 …… 93
§5.4 能量和能量守恒 …… 95
§5.5 对“质能相当性”的述评 …… 98
§5.6 结合能和质量亏损 …… 101
§5.7 质能动关系式 …… 103
§5.8 光子及其能量和动量 …… 104
§5.9 质点为何必须亚光速? …… 105
第5章习题 …… 106
第6章 广义相对论初步 …… 108
§6.1 引力的实质是时空的弯曲 …… 108
§6.2 弯曲时空的物理定律 …… 113
§6.3 等效原理 …… 114
§6.4 潮汐力与测地偏离 …… 118
§6.5 爱因斯坦场方程 …… 122
第7章 恒星演化和黑洞 …… 126
§7.1 引力是天体物理学的主角 …… 126
§7.2 恒星演化 …… 128
7.2.1 恒星的形成 红巨星 …… 128
7.2.2 白矮星 …… 130
7.2.3 中子星 …… 133
7.2.4 超新星 …… 135

7.2.5　脉冲星 …… 136
§7.3　奇点、视界和黑洞 …… 138
§7.4　黑洞无毛定理 …… 143
§7.5　黑洞的搜寻 …… 149
第8章　史瓦西时空 …… 156
§8.1　史瓦西真空解 …… 156
§8.2　广义相对论的实验验证 …… 160
§8.3　引力钟慢和引力红移 …… 161
8.3.1　引力钟慢效应 …… 161
8.3.2　引力红移 …… 167
§8.4　水星近日点进动 …… 168
§8.5　星光偏转 …… 172
8.5.1　理论预言值和日全食观测 …… 172
8.5.2　用星光偏转检验引力理论[选读] …… 175
8.5.3　牛顿引力论对星光偏角的推导[选读] …… 177
8.5.4　雷达回波的时间延缓[选读] …… 178
§8.6　相对论在 GPS(全球定位系统)中的应用 …… 181
§8.7　史瓦西时空的最大延拓 …… 195
8.7.1　奇点和奇性 …… 195
8.7.2　克鲁斯科尔坐标系 …… 196
8.7.3　克鲁斯科尔最大延拓 …… 199
8.7.4　黑洞、白洞和视界 …… 204
§8.8　伯克霍夫定理 …… 205
§8.9　恒星的引力坍缩和史瓦西黑洞 …… 206
第9章　虫洞和时间机器 …… 212
§9.1　嵌入图 …… 212
§9.2　虫洞 …… 217
9.2.1　史瓦西虫洞 …… 217
9.2.2　可穿越虫洞 …… 219
§9.3　时间机器 …… 221
9.3.1　把虫洞转化为时间机器 …… 221
9.3.2　弒母悖论 …… 224
9.3.3　物理学允许时间机器吗? …… 227
附录A　单位制、量纲和公式转换 …… 229

A.1 单位制基本知识 …… 229
A.1.1 基本单位和导出单位 …… 229
A.1.2 量纲式和量纲 …… 232
A.2 几何单位制 …… 234
A.2.1 几何单位制 …… 234
A.2.2 几何单位制公式转换为国际单位制公式 …… 235
习题答案 …… 238
参考文献 …… 240
索引 …… 243

第1章　狭义相对论的引入

相对性原理是牛顿力学的一个重要出发点。然而，麦克斯韦（以下简称麦氏）的电磁理论问世之后，相对性原理曾一度遭遇过严峻的挑战，最终导致狭义相对论的诞生。因此，要介绍狭义相对论的引入动机，有必要讲清以下三个问题：

1. 什么是相对性原理？
2. 麦氏电磁理论为何曾一度对相对性原理提出过严峻挑战？
3. 上述挑战如何诱发爱因斯坦创立狭义相对论？

§1.1　相对性原理

“温故而知新”。要介绍相对论，首先复习一点牛顿力学。牛顿力学的一个出发点是（伽利略的）相对性原理，它要求所有惯性系平权，就是说，要求同一物理定律在所有惯性系中有相同的数学表达式。这是一个很强的原理，实质上是一个“管定律的定律”——想要成为定律就必须服从这个原理的要求。为了解释“物理定律在所有惯性系中有相同数学表达式”一语的准确含义，我们从介绍“事件”概念开始。“事件”概念非常直观，一声咳嗽、两车相撞以及炸弹爆炸都是实际发生的事件。每个实际事件都要占据空间的一个小区域，而且要持续一小段时间。物理学中的事件概念则是实际事件的模型化，即认为每一事件发生在空间的一点和时间的一瞬（一个时刻）。甚至，不论有没有什么（有意义的）事情发生，空间的一点和时间的一瞬的结合都叫做一个**事件**（event）。如果两件事情在同时同地发生（例如某人在咳嗽的同时眨了一下眼），则认为是同一事件。全体事件的集合称为**时空**（spacetime），因此事件也称为**时空点**（spacetime point）。

借助于惯性系能够对事件进行定量描述。在很多情况下可认为地球在做惯性运动，所以地球可近似地看作惯性参考系，称为**地球参考系**或**地面参考系**。设A是相对于地球静止的一点（称为地面系的一个**空间点**），如果A点在某一时刻t发生了一个事件，记作p，则p可用如下方式表为

$$p=(t,\mathrm{A})。\tag{1-1-1}$$

上式表明事件p的发生时刻是t，发生地点是A点。请注意p与A的区别。p是

事件(时空点),A 是空间点,在其上可以发生一连串不同的事件,各事件的区别在于发生时刻不同,p 只是其中之一。鉴于初学者(甚至不少非初学者)存在着混淆时空点和空间点的“常见病”,本书特意用小写斜体字母(如 p)代表时空点(事件),用大写正体字母(如 A)代表空间点。

为了定量地刻画 A,还有必要引入坐标系。设 O 是地面参考系的一个空间点,以它为原点建立一个相对于地球为静止的直角坐标系,记作 $\{x,y,z\}$ 系,便可给空间点 A 赋予 3 个空间坐标 x_A,y_A,z_A,可表为 $A=(x_A,y_A,z_A)$,从而把式(1-1-1)具体化为

$$p=(t,x_A,y_A,z_A)。\tag{1-1-1'}$$

去掉下标 A,可以一般地把任一事件 p 表为

$$p=(t,x,y,z)。\tag{1-1-2}$$

把 t 称为**时间坐标**,就可以说每一事件 p 有 4 个时空坐标。于是我们就有了一个 4 维坐标系,称为地面参考系内的一个**惯性坐标系**,记作 K,也记作 $\{t,x,y,z\}$,即 K 与 $\{t,x,y,z\}$ 同义。

设有一列火车在地面上匀速直线平动,则火车系也是惯性参考系,也可以类似地建立火车参考系内的一个惯性坐标系,记作 K′或 $\{t',x',y',z'\}$,此系对同一事件 p 又赋予另外 4 个时空坐标:

$$p=(t',x',y',z')。\tag{1-1-3}$$

在 t 上加撇似乎画蛇添足,因为在牛顿力学(以及人们的生活经验)中,同一事件的发生时刻与坐标系无关,总有 $t'=t$。不过还是写成 t' 为好,因为不久将会看到在相对论中 t' 一般不等于 t。

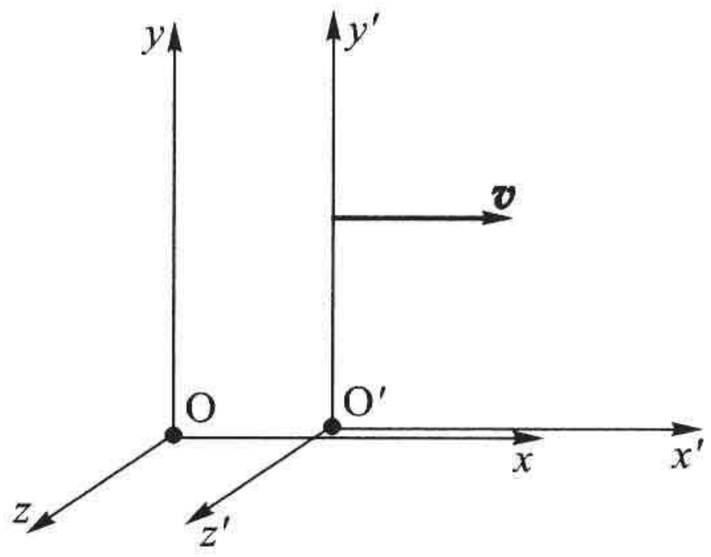


图 1-1　有最简关联的两个惯性坐标系

同一事件的两组坐标 t,x,y,z 和 t',x',y',z' 当然应有某种关系,这一关系称为**坐标变换式**,取决于两个惯性坐标系之间的关联方式。图 1-1 示出一种最简单的关联方式(本书称之为**最简关联**),满足如下条件:① 两系空间坐标轴对应同向;② K′系的空间坐标轴相对于 K 系以速率 $v>0$ 沿 x 轴正向匀速平动;③ 空间坐标原点 O 和 O′在 $t'=t=0$ 时重合(相遇)。在这种情况下不难证明两系坐标的变换关系为(证明见力学教科书)

$$t'=t,\quad x'=x-vt,\quad y'=y,\quad z'=z。\tag{1-1-4}$$

上式就是著名的**伽利略变换式**。由此又不难推出质点的速度变换式。由于本书把字母 $\boldsymbol{v}$ 专用来代表两系之间的相对速度,质点的速度只好改用字母 $\boldsymbol{u}$ 代表。质点的运动对应于一连串(相继发生的)事件,质点在每一时刻 t 位于某空间点

(x,y,z)就是一个事件,所以质点的空间坐标 x,y,z 都是 t 的函数,可记作 $x(t)$, $y(t),z(t)$,于是质点的速度 $\boldsymbol{u}$ 的分量依次为

$$u_x=\frac{\mathrm{d}x(t)}{\mathrm{d}t},\quad u_y=\frac{\mathrm{d}y(t)}{\mathrm{d}t},\quad u_z=\frac{\mathrm{d}z(t)}{\mathrm{d}t},\tag{1-1-5}$$

同理,质点相对于 K′系的速度 $\boldsymbol{u}'$的分量为

$$u'_x=\frac{\mathrm{d}x'(t')}{\mathrm{d}t'}=\frac{\mathrm{d}x'(t)}{\mathrm{d}t},\quad u'_y=\frac{\mathrm{d}y'(t)}{\mathrm{d}t},u'_z=\frac{\mathrm{d}z'(t)}{\mathrm{d}t}。$$

利用式(1-1-4),注意到 v 为常数($\mathrm{d}v/\mathrm{d}t=0$),便得

$$u'_x=u_x-v,\quad u'_y=u_y,\quad u'_z=u_z。\tag{1-1-6}$$

上式就是牛顿力学中的速度变换式,是伽利略变换式的必然结果。

现在就可以说明“物理定律在所有惯性系中有相同数学表达式”的准确含义。以牛顿第二定律为例。谁都知道此定律在 K 系的表达式为 $\boldsymbol{f}=m\boldsymbol{a}$,问题是:若以 m', $\boldsymbol{f}'$, $\boldsymbol{a}'$依次代表 K′系测得的质量、力和加速度,是否也有 $\boldsymbol{f}'=m'\boldsymbol{a}'$? 如果此式成立,就说牛顿第二定律在两系有相同的数学表达式。由于牛顿力学默认 $m'=m$, $\boldsymbol{f}'=\boldsymbol{f}$,所以只需验证 $\boldsymbol{a}'=\boldsymbol{a}$,为此只需验证两者的各分量相等。而这是很容易的,因为

$$a'_x=\frac{\mathrm{d}u'_x}{\mathrm{d}t'}=\frac{\mathrm{d}}{\mathrm{d}t}(u_x-v)=\frac{\mathrm{d}u_x}{\mathrm{d}t}=a_x,\quad [\text{其中第二步用到式(1-1-6)}]$$

再由式(1-1-6)一望而知 $a'_y=a_y$, $a'_z=a_z$,所以 $\boldsymbol{a}'=\boldsymbol{a}$。可见牛顿第二定律服从相对性原理。

事实上,牛顿力学的每一个定律都服从相对性原理,因此都是定律。

然而“好景不长”,因为麦氏在 1873 年发表的电磁理论曾一度对相对性原理提出过严峻的挑战。详见下节。

§1.2　麦克斯韦电磁理论的挑战

麦氏在前人工作的基础上于 1873 年发表了一套电磁理论,其中的麦氏方程可被看作电磁场的演化方程。麦氏理论的一个伟大成果是预言了电磁波的存在(麦氏方程容许波动解),并证明了电磁波在真空中的传播速率与真空中的光速 c 相同(波动解的波速为 c),从而揭示了光的电磁本性。既然相对性原理是“管定律的定律”,麦氏方程(如果想要成为定律)就必须服从,亦即在任一惯性系中都应取相同的数学表达式,从而导致“电磁波相对于任一惯性系的速率都为 c”的结论。然而,设电磁波沿某惯性系的 x 方向以速率 c 传播,根据速度变换式(1-1-6),该电磁波相对于另一惯性系的速率就不为 c,因为两系之间有相对

速度。于是出现如下局面：

① 麦氏方程 + ② 相对性原理 ⇒ 光速对各系皆为 c。　(1 -2 -1)

[与伽利略变换推出的速度变换式(1 -1 -6)矛盾!]

上式似乎表明,如果①和②都成立,就会推出矛盾[与速度变换式(1 -1 -6)矛盾]。于是麦氏理论与相对性原理似乎互不相容。物理学似乎只能在以下两种选择中做一抉择。

选择 A　相对性原理只适用于力学而不适用于电动力学。存在一个特殊的惯性系,麦氏方程在该系成立,真空中的光波相对于该系的速率为 c;在其他惯性系中麦氏方程不成立,真空光速不为 c。这个特殊的惯性系在历史上称为“以太(ether)系”。洛伦兹是坚持这一选择的代表人物。

选择 B　相对性原理对力学、电动力学等一切物理范畴都适用(能“管”所有物理定律),但麦氏电磁理论不正确。应该存在一个不同于麦氏理论的电磁理论,它在所有惯性系中的数学表达式都有相同形式(惯性系之间的坐标变换仍默认为伽利略变换)。

以上讨论给读者这样的印象:相对性原理与麦氏理论水火不容,必须放弃其中一个。这是一个“硬碰硬”的问题:相对性原理和麦氏理论都非同一般地“硬”,放弃哪一个都非常困难。历史上的确有过许多尝试,但都陷入理论上无法自圆其说或者与某些实验事实不符的困境①。于是物理学家遇到了一个两难问题。不过,如果更仔细地思索,就会发现相对性原理与麦氏理论并非真的势不两立,因为在导致矛盾的讨论中还涉及第三者——由伽利略变换推出的速度变换式(1 -1 -6)。人们之所以认为相对性原理与麦氏理论水火不容,是因为他们心中默认速度变换式(1 -1 -6)。假若允许对此式做适当修改,矛盾仍有希望消除。(打个生活中的比喻:去掉第三者仍有希望恢复夫妻关系的和谐。)可是速度变换式(1 -1 -6)似乎是更“硬”的东西,因为它是伽利略变换的推论,而伽利略变换是既合情合理又久经考验的公式,还能有错!? 然而,独具慧眼的爱因斯坦在反复思考后竟然发现这一公式的确有一个错误。问题恰恰出在伽利略变换式(1 -1 -4)的第一式 $t' = t$ 身上。此式的含义是:如果一个惯性系认为某事件在时刻 t 发生,则另一惯性系也认为该事件在时刻 t 发生。这同每个人的生活经验完全吻合,可以称为“时间的绝对性(与坐标系之类的人为选择因素无

---

① 关于这方面的历史(特别是史上有名的实验)请参阅其他教材,例如 Resnick(1968),张元仲(1994)。历史上最有名的实验是迈克耳孙(Michelson)和莫雷(Morley)在 1887 年的干涉仪实验,旨在测定地球相对于以太的速度,但得到的是“零结果”。由于发明这种干涉仪以及所做的多次光学实验,迈克耳孙荣获 1907 年的诺贝尔奖,是美国获此殊荣的第一人。

关)”,由此又可推出“同时性的绝对性”,含义如下。设 p_1,p_2 是两个事件,在惯性系 K 的时空坐标分别为

$$p_1=(t_1,x_1,y_1,z_1),\quad p_2=(t_2,x_2,y_2,z_2)。$$

如果 $t_1=t_2$,我们就说 p_1 和 p_2 对 K 系而言是同时事件。再设 K′系给 p_1,p_2 赋予的时空坐标为

$$p_1=(t_1',x_1',y_1',z_1'),\quad p_2=(t_2',x_2',y_2',z_2'),$$

则由“时间的绝对性”[式(1-1-4)的第一式]自然有 $t_1'=t_1$,$t_2'=t_2$,于是由 $t_1=t_2$ 得到 $t_1'=t_2'$,可见对 K 系为同时的两事件对 K′系也是同时的。这就是“同时性的绝对性”的含义。牛顿力学坚持(默认)伽利略变换,因此在牛顿力学中同时性是绝对的,而且人们从生活经验出发也对此深信不疑。然而爱因斯坦在反复思考后认为“同时性的绝对性”并非无懈可击,他构想出如下反例。

一列火车 A′B′在地面上匀速直线地飞驰而过。假定地面系发现车头 A′和车尾 B′同时遭受雷击(图 1-2)。爱因斯坦认为,对火车系而言,这两个事件并非同时事件,理由如下。

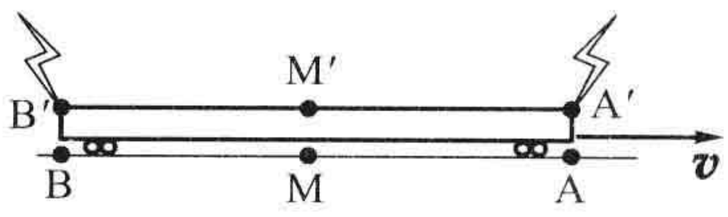


图 1-2　爱因斯坦列车的雷击同时性问题

以 A 和 B 分别代表地面系发现雷击车头时 A′和 B′在地面上留下的印记。(忽略火车高度,认为此时 A′与 A 重合,B′与 B 重合。)以 M 代表 AB 段的中点。所谓“地面系认为两个雷击事件同时发生”,是指车头和车尾遭雷击时所发的光同时传到 M 点①。再以 M′代表火车的中点,随着火车向前行进,M′点迎着车头而背离车尾运动,理应先收到车头来光后收到车尾来光,就是说,车头和车尾发出的光并不同时到达 M′,所以火车系认为“雷击车头”和“雷击车尾”这两个事件并不同时。爱因斯坦就是这样出人意料地提出了对“同时性的绝对性”的质疑,从而也就是对伽利略变换式(1-1-4)中的 $t'=t$ 做出了否定,并以此为突破口创立了狭义相对论。相对论的第一篇(奠基性)论文发表于 1905 年,题为《论运动物体的电动力学》[见爱因斯坦等(1980)]。在这篇论文中,爱因斯坦首次提出了狭义相对论的两个基本假设,介绍如下。

假设 1(相对性原理)　所有惯性系平权,任何物理定律(而不只是力学定律)在所有惯性系的数学表达式都有相同的形式。

假设 2(光速不变原理)　真空中的光速沿任何方向、对任何惯性系都是 c,

① 设某人坐在 M 点,则他“收到车头来光”与“收到车尾来光”这两个事件都发生在他身上,同时与否有明确意义。爱因斯坦正是巧妙地利用“发生在同一个人身上的两事件的同时性”这一已知概念来给“发生在不同地点的两事件的同时性”下定义。

与光源的运动情况无关。

注记1　伽利略的相对性原理只适用于牛顿力学定律，爱因斯坦的相对性原理则适用于所有物理定律。由前面的讨论可知，为此必须找出一组新的坐标变换来代替伽利略变换，这组新的坐标变换叫做洛伦兹变换，见下节。

注记2　假设2的关键内容是“光速对任何惯性系都是c”，而关键的关键是“惯性系”一词。用非惯性系测量光速的结果完全可以不是c。

[选读1-1]

用今天的眼光看问题(“事后诸葛亮”)，麦氏方程组当然是物理定律，所以假设1保证它在任何惯性系都有相同表达式，而由此自然导出光速(电磁波速)沿任何方向、对任何惯性系都是c，所以假设2应看作假设1的推论而不是独立的假设(其实历史上早就有人提出过类似问题并引起过讨论)。至于假设2的最后一句话，即“与光源的运动情况无关”，今天看来也是多余的，因为，既然光速沿任何方向、对任何惯性系都是c，它当然与光源的运动情况无关。但是，从历史上来讲，选择B(见第4页)的支持者们构造了各种各样的、不同于麦氏理论的电磁理论，它们的共同点是：光相对于光源的速率为c，相对于惯性系的速率则取决于光源相对于该系的速度。所以，这些理论认为光速与光源的运动情况有关。后来的双星实验(1913年)及其他实验否定了这一结论。但是爱因斯坦提出上述两个假设是在1905年，那时人们对于光速是否与光源有关还极为重视，也许正是这个原因使爱因斯坦在假设2中添加了最后一句关于光速与光源的运动情况无关的话。论文中的原话是这样的：“下面我们……引入进一步的假设(笔者注：指假设2)——一个乍看起来与前一假设势不两立的假设(笔者注：从伽利略变换看来当然势不两立)——光在真空中以速度c传播，这一速度不依赖于发射体的运动性质。”

[选读1-1完]

[选读1-2]

关于火车遭受雷击的讨论，笔者曾不止一次地被问及一个有趣的问题。本选读用对话方式介绍这一问题及其答案，其中乙代表笔者。

甲　如果站在火车系上看问题，则M′并无运动，正如地面系认为M无运动那样。把在地面系上的思辨过程照样搬到火车系上，岂非M′也同时收到车头、车尾来光？

乙　不能照搬。地面系认为车头发光和车尾发光是同时事件，而我们不敢说火车系也认为这两个事件同时(事实上有待证明的正是它们并不同时)。可

见火车系的已知条件并不充分，难以站在火车系上讨论此问题。而地面系的已知条件——车头、车尾同时发光——是从一开始就给定的。为帮助你获得更清晰的理解，我用稍微不同的方式把问题的“已知”、“求证”和“证明”再次明确地写在下面。

已知：① 车头、车尾都受雷击，雷击时都向外发光；② M 点同时收到车头、车尾来光。求证：M′不同时收到车头、车尾来光。证明：站在地面系上想，假若 M′不随车动，根据已知条件② ，M′自然会同时收光。但实际上 M′随车向前运动，所以他迎着车头来光而背离车尾来光，故不同时收光。说明：如果站在火车系上想，虽然 M′不动，却没有类似于已知条件② 的条件可用，所以得不出 M′同时收光的结论。

甲　我明白了。由此我还体会到爱因斯坦的一个高明之处，就是他只站在地面系讨论这个问题。

乙　很好。　**[选读 1 -2 完]**

§1.3　洛伦兹变换

设两个惯性系之间有最简关联（见图 1 -1 及其所在段），则任一事件在该系的时空坐标 t,x,y,z 和 t',x',y',z'有如下关系：

$$\text{(a)}\ t'=\gamma\left(t-\frac{v}{c^2}x\right),\quad \text{(b)}\ x'=\gamma(x-vt),\quad \text{(c)}\ y'=y,\quad \text{(d)}\ z'=z,\tag{1-3-1}$$

其中

$$\gamma\equiv\frac{1}{\sqrt{1-(v/c)^2}}\ 。\tag{1-3-2}$$

上式称为**洛伦兹变换**。利用狭义相对论的两个基本假设（即相对性原理和光速不变原理），加上对空间、时间均匀性以及空间各向同性性的默认，爱因斯坦在 1905 年的开创性论文中推出了这一变换。在后来的无数教材中都能找到这一变换的证明版本，虽然证法五花八门，但都离不开两个假设以及对均匀性和各向同性性的默认。然而某些数学家（如华罗庚）对于物理教材的这些证法不太满意，于是又发展出“数学家式”的证明思路。本书不拟过多涉及这一问题，只想说明，笔者曾杜撰出一种用“几何语言”导出洛伦兹变换的方法，详见 §4.5。

洛伦兹在相对论诞生前的 1904 年就已发表过式（1 -3 -1）（这正是该式被称为洛伦兹变换的缘由），不过并未给出推导过程。他知道用式（1 -3 -1）代替伽利略变换就意味着修改牛顿力学，而他不认为牛顿力学需要修改，所以只把式

(1-3-1)视为便于计算而引入的辅助工具,其中的t'不是什么物理时间(他称之为局域时间),物理上的坐标变换仍然要用伽利略变换。正是对伽利略变换和牛顿力学的过分坚持使洛伦兹与相对论的创立失之交臂。

相对论物理学与非相对论物理学都把相对性原理作为基本出发点。仅就这一层面而言,相对论与非相对论并无二致。然而,把物理定律的数学表达式从一个惯性系变到另一惯性系时必然涉及两系之间的坐标变换,而这在非相对论和相对论中存在实质性的区别,所以相对性原理在两种理论中有不同的内涵。伽利略提出相对性原理时,心目中只有力学现象,而且坐标变换只有伽利略变换,因此伽利略的(即非相对论的)相对性原理可以表为"力学定律的数学表达式在伽利略变换下形式不变",简言之就是"力学定律具有伽利略协变性"①,爱因斯坦的相对性原理则要求每一物理定律的表达式在各惯性系中形式相同,而且两系之间要用洛伦兹变换,因此爱因斯坦的(即相对论的)相对性原理也可表为"物理定律的数学表达式在洛伦兹变换下形式不变",简言之就是"物理定律具有洛伦兹协变性"。

洛伦兹变换有一系列重要的性质和推论,下面指出5点。

1. 由$\gamma\equiv 1/\sqrt{1-(v/c)^2}$易见$v\ll c$时$\gamma\approx 1$,故式(1-3-1b)近似回到$x'=x-vt$,与伽利略变换式(1-1-4)的第二式一致。此外,通过讨论还能证明式(1-3-1a)在$v\ll c$时也近似回到伽利略变换的$t'=t$(详见选读1-3)。可见伽利略变换是洛伦兹变换的低速近似。事实上,整个牛顿力学都可看作狭义相对论力学的低速近似(见后续内容)。日常生活和工程(包括汽车、飞机甚至火箭和航天领域)中的绝大多数问题都可用牛顿力学解决(杀鸡焉用牛刀)。然而,高能物理学领域经常涉及速率与光速同量级的高能粒子(能量越高速率越大),甚至光子本身,这时牛顿力学就不再适用。例如,当粒子的速率$v=0.99\cdots995c$(共有$2n$个9)时,γ近似为10^n。可见,高能粒子物理学的大部分问题非用狭义相对论不可。

2. 由$\gamma\equiv 1/\sqrt{1-(v/c)^2}$可知当$v>c$时$\gamma$为虚数,但时空坐标$t,x,y,z$和$t',x',y',z'$都是实数,故式(1-3-1)中的$\gamma$必须为实数,因而$v$不允许大于$c$。此外,当$v=c$时$\gamma$无意义(分母为零的分数无意义),可见洛伦兹变换对惯性系之间的相对速率v给出了上限,只允许$v<c$。我们在此第一次看到真空光速在相对论中扮演着极限速率的角色。后面(§5.9)还将讨论超光速导致因果关系破坏的问题。

---

① "协变(covariance)"一词有"协同改变"之意,就是说,虽然公式中的若干物理量在坐标变换下要变,但它们之间能够如此这般地协同改变,致使整个公式的形状不变。

3. 式(1-3-1)表明变量 t',x',y',z' 对 t,x,y,z 有线性的依从关系,不难通过反解求得(留作习题)逆变换式

$$\text{(a)}\ t=\gamma\left(t'+\frac{v}{c^2}x'\right),\quad \text{(b)}\ x=\gamma(x'+vt'),\quad \text{(c)}\ y'=y',\quad \text{(d)}\ z=z'。\tag{1-3-1'}$$

从物理角度看,上式无非是如下简单事实的反映:"火车相对于地面以速度 $\boldsymbol{v}$ 运动"相当于"地面相对于火车以速度 $-\boldsymbol{v}$ 运动"。

4. 仿照牛顿力学速度变换式(1-1-6)的推导方法,不难从洛伦兹变换式(1-3-1)出发推出如下的相对论速度变换式:

$$u'_x=\frac{u_x-v}{1-\frac{v}{c^2}u_x},\quad u'_y=\frac{u_y}{\gamma\left(1-\frac{v}{c^2}u_x\right)},\quad u'_z=\frac{u_z}{\gamma\left(1-\frac{v}{c^2}u_x\right)}。\tag{1-3-3}$$

推导留作习题。对上式要做几点说明。

(1) 式中的 v 是两惯性系的相对速率,是常数。但式中的 $\boldsymbol{u}$ 和 $\boldsymbol{u}'$ 却代表粒子在两个惯性系中的速度,不要求是常矢量,即式(1-3-3)适用于作任意变速运动的粒子。

(2) 当 $v\ll c$ 时有 $vu_x/c^2\ll 1$,式(1-3-3)简化为牛顿力学的速度变换式(1-1-6)。

(3) 根据牛顿力学的速度变换式,当 $v>c/2$,$|\boldsymbol{u}|>c/2$ 时,只要适当选择 $\boldsymbol{u}$ 的方向,就有 $|\boldsymbol{u}'|>c$,即允许超光速。但由相对论速度变换式(1-3-3)可以证明此种情况不会出现。只要粒子相对于某惯性系的速率 $u<c$,它相对于任何惯性系的速率 u' 就都小于 c。下面是最极端的情况:设粒子在惯性系 K 中的速度满足 $u_x=c$,$u_y=u_z=0$,则由式(1-3-3)易得它在惯性系 K′的速度满足

$$u'_x=\frac{c-v}{1-v/c}=\frac{c(c-v)}{c-v}=c,\quad u'_y=u'_z=0。$$

所以 §1.2 讲到的"麦氏理论对相对性原理的挑战"现在就彻底不存在:只要用洛伦兹变换代替伽利略变换,麦氏理论与相对性原理完全可以和谐共存,并行不悖! 现在可以明确指出,麦氏方程天生就没有伽利略协变性:如果从 K 系变到 K′系时采用伽利略变换,那么麦氏方程在 K′系的形式将与在 K 系的形式不同。但是麦氏方程天生就有洛伦兹协变性,它生来就是狭义相对论的天之骄子。

由式(1-3-3)还可求得逆变换式:

$$u_x=\frac{u'_x+v}{1+\frac{v}{c^2}u'_x},\quad u_y=\frac{u'_y}{\gamma\left(1+\frac{v}{c^2}u'_x\right)},\quad u_z=\frac{u'_z}{\gamma\left(1+\frac{v}{c^2}u'_x\right)}。\tag{1-3-3'}$$

5. 狭义相对论中存在着若干饶有趣味的运动学效应,最常见的是:① 尺缩

效应;② 车库佯谬;③ 钟慢效应;④ 双生子佯谬。虽然这些效应都可用洛伦兹变换直接证明(可参阅几乎任一本狭义相对论教材),但存在着一种简单、清晰、深刻得多的、借助于时空几何语言的证明。本书将在第4章详细讲述。

[选读1-3]

上节曾指出,当 $v<<c$ 时洛伦兹变换近似回到伽利略变换,具体来说就是

(A)　$x'=\gamma(x-vt)$　近似回到　$x'=x-vt$,

(B)　$t'=\gamma\left(t-\frac{v}{c^2}x\right)$　近似回到　$t'=t$。

由于 $v<<c$ 时 $\gamma\approx1$,自然就有 $x'=\gamma(x-vt)\approx x-vt$,即(A)自然成立。然而要说明(B)也成立就要略费笔墨。为便于陈述,下面以对话方式讲解,其中乙代表笔者。

甲　由 $\gamma\approx1$ 直接得到的只是

$$t'\approx t-vx/c^2, \tag{1-3-4}$$

由此又怎能得到 $t'\approx t$?

乙　关键是要说明式(1-3-4)右边第二项可被第一项所忽略,即说明 $t-(v/c^2)x\approx t$。

甲　t 和 x 是被观测事件的坐标,原则上可取任意值,凭什么就有 $|t|>>(v/c^2)|x|$?

乙　先看两个特例。① 设事件发生在K系的空间坐标原点O,则 $x=0$,自然有 $t-(v/c^2)x=t$;② 设事件发生在K′系的空间坐标原点O′,则 $x'=0$,由 $x'=\gamma(x-vt)$ 得 $x=vt$,代入式(1-3-4)得

$$t'\approx t-(v^2/c^2)t=t(1-v^2/c^2)\approx t。$$

甲　但这两个只是特例。作为被观测事件的坐标,x 和 t 原则上可各取任何值。如果 x 很大而 t 很小,式(1-3-4)右边第二项还能被第一项忽略吗?

乙　你涉及问题的要害了。我的答案是:“洛伦兹变换在 $v<<c$ 时近似回到伽利略变换”这一提法总是对的,其准确含义如下:

“在约定的允许误差范围内,对任一被观测事件而言,只要 v 与 c 相较足够小,洛伦兹变换都能近似回到伽利略变换。”

理由很简单:这一提法要求你首先给定被观测事件,然后才告诉你 v 要比 c 小多少(这个先后顺序极其重要)。设此事件的 x 与 t 的关系为 $x=Mct$,其中 M 是个很大的(无量纲)数,例如10^2,10^3,…。(这就是你说的“x 很大而 t 很小”的准确含义,须知只有同量纲的量方可比较大小。)代入式(1-3-4)便得

$$t' \approx t - \frac{v}{c^2}Mct = t\left(1 - \frac{v}{c}M\right), \tag{1-3-5}$$

故

$$t - t' = t - t\left(1 - \frac{v}{c}M\right) = t\frac{v}{c}M, \tag{1-3-6}$$

因而 t' 与 t 的相对差别为

$$\frac{t - t'}{t} = \frac{v}{c}M。 \tag{1-3-7}$$

假定事先约好的允许相对误差为 δ（例如 $\delta = 10^{-2}$ 或 10^{-3}），则为使 $(t - t')/t < \delta$，只须

$$\frac{v}{c} < \frac{\delta}{M}。 \tag{1-3-8}$$

例如，设 $M = 10^3$，$\delta = 10^{-2}$，则 $v/c < 10^{-5}$。就是说，只要 $v < 10^{-5}c$，对该事件而言就可在允许误差范围内使用伽利略变换。

甲　您的分析很有帮助。美中不足的是，您写式(1-3-6)时略去了式(1-3-5)中 t' 后面的近似号。严格说来，不写近似号就应补上 γ，即式(1-3-5)应改为

$$t' = \gamma\left(t - \frac{v}{c^2}Mct\right),$$

这对结果的影响如何？

乙　你的意见很对，但按此修改会使推导变得很繁，不过不影响定性结论。

[选读 1-3 完]

本书至今一直采用国际单位制。相对论公式由于经常出现 v/c 而显得冗长，而且运算繁杂。如果改用一种称为“几何单位制”的单位制，公式便可大为简化。涉及单位制时应该分清“量”和“数”。一个物理量与一个被选作单位的同类量做比较就给出一个数（称为用该单位测该量所得的数）。任何物理书中的绝大多数等式都可以看作数的等式。几何单位制把真空光速这个量选做速度这一类量的单位，所以，若以 c 代表真空光速在几何制的数值，便有 $c = 1$（附录 A 对几何单位制有较详细的讨论）。本书从现在开始改用几何单位制，这时洛伦兹变换式简化为

$$\text{(a)}\ t' = \gamma(t - vx),\quad \text{(b)}\ x' = \gamma(x - vt),\quad \text{(c)}\ y' = y,\quad \text{(d)}\ z' = z, \tag{1-3-9}$$

其中

$$\gamma \equiv \frac{1}{\sqrt{1 - v^2}}。 \tag{1-3-10}$$

第1章习题

注记　习题中涉及的洛伦兹变换都是有最简关联的两个惯性系(见图1-1)之间的坐标变换。

1. 通过反解式(1-3-1)证明式(1-3-1′)(洛伦兹变换的逆变换)。

2. 从洛伦兹变换式(1-3-1)出发推出相对论速度变换式(1-3-3)。

3. 设事件 p,q 在惯性系 $\mathrm{K}\equiv\{t,x,y,z\}$ 的坐标(用几何单位制)为 $p=(0,0,0,0)$,$q=(1,2,0,0)$。(a)找出惯性系 $\mathrm{K}'\equiv\{t',x',y',z'\}$,它认为 p 和 q 同时发生。(b)找出惯性系 $\mathrm{K}''\equiv\{t'',x'',y'',z''\}$,它认为 q 先于 p 发生,提前量 $t''_q-t''_p=-1$。

4. 证明以下有用公式:

(a) $\gamma v=\sqrt{\gamma^2-1}$,　(b) $(\gamma-1)/v^2=\gamma^2/(1+\gamma)$,　(c) $\mathrm{d}\gamma=\gamma^3 v\mathrm{d}v$,　(d) $\mathrm{d}(\gamma v)=\gamma^3\mathrm{d}v$。对国际单位制则为

(a) $\gamma v=c\sqrt{\gamma^2-1}$,　(b) $c^2(\gamma-1)/v^2=\gamma^2/(1+\gamma)$,　(c) $c^2\mathrm{d}\gamma=\gamma^3 v\mathrm{d}v$,　(d) $\mathrm{d}(\gamma v)=\gamma^3\mathrm{d}v$。

5. 设惯性系 K′相对于 K 的速率为 $v=c/3$,质点相对于 K 系的速度分量为 $u_x=c/2$,$u_y=u_z=0$,求质点相对于 K′系的速度分量 u'_x、u'_y、u'_z。

6. 飞船相对于惯性系 $\mathrm{K}\equiv\{t,x,y,z\}$ 以常速率 v 沿 x 轴正向飞行,粒子相对于飞船以常速率 nv 沿 x 轴负向飞行。(a)求粒子相对于 K 系的速度分量 u_x 的表达式(分别就牛顿力学和狭义相对论两种情况求解);(b)若 $n=1/2$,$v=c/2$,求 u_x 的数值。

7. 试证两个洛伦兹变换的相继作用仍是洛伦兹变换,而且,设两变换的参数分别为 v_1 和 v_2,则合变换的参数 $v=(v_1+v_2)/(1+v_1v_2)$。

8. 把洛伦兹变换的参数 v 的取值范围从 $0<v<1$ 放宽为 $-1<v<1$,引入新参数 λ,与 v 的关系为 $v=\mathrm{th}\lambda$(th 代表双曲正切),试证:(a)λ 的取值范围是全体实数,即 $-\infty<\lambda<\infty$;(b)以λ 为参数的洛伦兹变换(在几何制中)取如下形式:$t'=t\mathrm{ch}\lambda-x\mathrm{sh}\lambda$,$x'=-t\mathrm{sh}\lambda+x\mathrm{ch}\lambda$。

9. (a)试证以 λ 为参数的洛伦兹变换(见上题)的逆变换是以 $-\lambda$ 为参数的洛伦兹变换。(b)借用参数 λ 再证两个洛伦兹变换的相继作用仍是洛伦兹变换,而且,设两变换的参数分别为 λ_1 和 λ_2,则合变换的参数 $\lambda=\lambda_1+\lambda_2$。

第2章　学一点点几何

§2.1　线元与几何

狭义相对论经常涉及各惯性系之间的坐标变换，面对这种变化纷繁的局面，物理学家使出了从数学家学来的高招——“以不变应万变”。为此就要抓住“万变不离其宗”的那个“宗”，即是要学会抓住在坐标变换下不变的量，称为**不变量**(invariant)，亦称**绝对量**。欧氏空间的常识为此提供重要借鉴。设 L 是 2 维欧氏空间的曲线，以 dL 代表它的一个元段(见图 2 - 1)，dl 代表 dL 的长度(称为**元段长**)，dl^2 代表 dl 的平方[简称为**线元**(line element)]。元段 dL 一经给定，其线元 dl^2 当然与坐标系无关，所以是不变量(绝对量)。下面再关心线元 dl^2 的表达式。设 $\{x,y\}$ 是直角坐标系，由图 2 - 1 显见

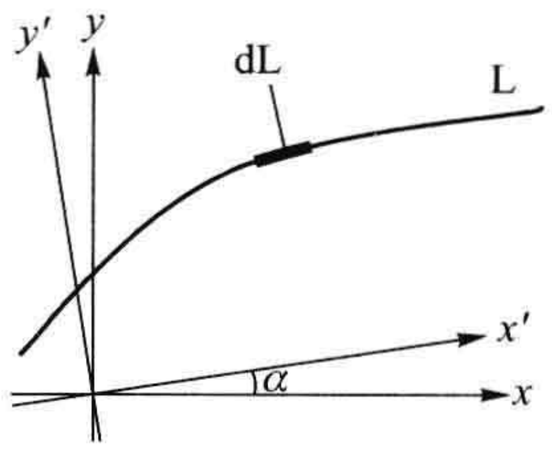


图 2 - 1　欧氏线元表达式在直角坐标变换下不变

$$dl^2 = dy^2 + dx^2。\tag{2-1-1}$$

如果改用极坐标系 $\{r,\varphi\}$，则同一线元 dl^2 的表达式变为

$$dl^2 = dr^2 + r^2 d\varphi^2。\tag{2-1-2}$$

上两式是同一个线元 dl^2 的表达式，坐标变换所改变的只是线元表达式而不是线元本身。要使线元表达式不变，新坐标系必须也是直角坐标系，例如，设 $\{x',y'\}$ 是由 $\{x,y\}$ 系转动 α 角而得的新系(仍见图 2 - 1)，则容易验证两者之间的坐标变换为

$$x' = x\cos\alpha + y\sin\alpha,\ y' = -x\sin\alpha + y\cos\alpha,\tag{2-1-3}$$

故

$$dy^2 + dx^2 = dy'^2 + dx'^2,\tag{2-1-4}$$

因而线元 dl^2 在新系的表达式为

$$dl^2 = dy'^2 + dx'^2。\tag{2-1-5}$$

可见线元表达式在直角坐标系的转动变换下形式不变。其实还有更强的结论：当且仅当采用直角坐标变换(转动加原点平移)时，欧氏线元表达式不变。

在相对论中有非常类似的情况。先讨论 2 维时空。设 p_1,p_2 是两个相邻事

件,在惯性系 K 和 K′的坐标分别为

$$p_1=(t_1,x_1)=(t_1',x_1'),\quad p_2=(t_2,x_2)=(t_2',x_2')。$$

令 $\mathrm{d}t\equiv t_2-t_1$, $\mathrm{d}x\equiv x_2-x_1$, $\mathrm{d}t'\equiv t_2'-t_1'$, $\mathrm{d}x'\equiv x_2'-x_1'$,则由洛伦兹变换式(1-3-9)得

$$\mathrm{d}t'=\gamma(\mathrm{d}t-v\mathrm{d}x),\quad \mathrm{d}x'=\gamma(\mathrm{d}x-v\mathrm{d}t),\qquad (2-1-6)$$

由此不难验证(习题)

$$-\mathrm{d}t^2+\mathrm{d}x^2=-\mathrm{d}t'^2+\mathrm{d}x'^2,\qquad (2-1-7)$$

定义**时空(元)间隔**(spacetime interval)$\mathrm{d}s^2$ 为

$$\mathrm{d}s^2\equiv-\mathrm{d}t^2+\mathrm{d}x^2,(\text{对国际单位制为 }\mathrm{d}s^2\equiv-c^2\mathrm{d}t^2+\mathrm{d}x^2)\quad (2-1-8)$$

则式(2-1-7)表明时空间隔 $\mathrm{d}s^2$ 在洛伦兹变换下形式不变。可见,至少形式地说,时空间隔 $\mathrm{d}s^2$ 很像线元 $\mathrm{d}l^2$,惯性坐标 t,x 很像直角坐标 y,x。不妨建立如下对应:

时空间隔 $\mathrm{d}s^2\leftrightarrow$欧氏线元 $\mathrm{d}l^2$;　惯性坐标 $t,x\leftrightarrow$直角坐标 y,x。

但"很像"并非全同,时空间隔 $\mathrm{d}s^2$ 与欧氏线元 $\mathrm{d}l^2$ 表达式的唯一差别就是在 $\mathrm{d}t^2$ 前有一负号。同欧氏线元 $\mathrm{d}l^2$ 一样,时空间隔 $\mathrm{d}s^2$ 一经选定[由式(2-1-8)定义]就是不变量(绝对量),坐标变换有可能改变的只是它的表达式。例如,仿照欧氏空间引入极坐标系的做法,现在可用下式引入新坐标系$\{\psi,\eta\}$:

$$x=\psi\mathrm{ch}\eta,\quad t=\psi\mathrm{sh}\eta,(\text{其中 ch 和 sh 分别代表双曲余弦和正弦})\qquad (2-1-9)$$

同一间隔 $\mathrm{d}s^2$ 在新系的表达式变为

$$\mathrm{d}s^2\equiv\mathrm{d}\psi^2-\psi^2\mathrm{d}\eta^2。\qquad (2-1-10)$$

小结　存在两个意义的不变性:(1) 欧氏线元 $\mathrm{d}l^2$ 和相对论间隔 $\mathrm{d}s^2$ 都是不变量(绝对量);(2) 欧氏线元(相对论间隔)的表达式在直角坐标(惯性坐标)变换下形式不变。

对于欧氏几何,人们关心曲线的长度、两点之间的距离、两直线之间的夹角等等。这些几何性质都由欧氏线元决定:曲线的长度定义为线元 $\mathrm{d}l^2$ 的开方根 $\mathrm{d}l$ 的积分;两点之间的距离定义为两点之间的最短曲线的长度(因而也取决于线元);两直线之间的夹角定义为弧长与半径长之比(归根结底还是由线元决定)。事实上,欧氏几何的所有内容都由欧氏线元决定,简言之就是**线元决定几何**。这一结论对非欧几何也成立。指定一个与欧氏线元不同的线元就定义了一种非欧几何。线元的改变起着"牵一发而动全身"的作用。现在,只要把时空间隔 $\mathrm{d}s^2\equiv-\mathrm{d}t^2+\mathrm{d}x^2$ 也看作一种线元[称为**闵氏线元**,闵氏是指闵可夫斯基(Minkowski)],它也决定一种几何,与欧氏几何不同,称为(2 维)**闵氏几何**(Minkowski geometry)。定义了闵氏几何的空间称为**闵氏空间**(Minkowski

space)。今后统一用符号 ds^2 代表线元(但元段长则仍记作 dl),所以欧氏线元 $dl^2 = dy^2 + dx^2$ 将改记作 $ds^2 = dy^2 + dx^2$。

以上只是为陈述方便而讨论 2 维情况,不难将讨论推广到维数更高的空间。读者都熟悉 3 维欧氏空间,其线元在直角坐标系的表达式为

$$ds^2 = dx^2 + dy^2 + dz^2, \tag{2-1-11}$$

如果改用球坐标系 $\{r,\theta,\varphi\}$,利用该系与直角系的熟知关系:

$$x = r\sin\theta\cos\varphi, \quad y = r\sin\theta\sin\varphi, \quad z = r\cos\theta, \tag{2-1-12}$$

由微分运算容易求得欧氏线元 $ds^2 = dx^2 + dy^2 + dz^2$ 在球坐标系的表达式:

$$ds^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2)。 \tag{2-1-13}$$

我们再次强调,坐标变换改变的只是线元的表达式而不是线元本身。就是说,式(2-1-13)与(2-1-11)虽然形式不同,但实质一样,都代表 3 维欧氏线元,可表为连等式:

$$ds^2 = dx^2 + dy^2 + dz^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2)。$$

如果改用第三种坐标系(例如柱坐标系或者更复杂的坐标系),欧氏线元 ds^2 的表达式还会取其他更为复杂的形式。欧氏线元只在直角坐标系才有最简单的表达式 $dx^2 + dy^2 + dz^2$。想象地再补上第 4 维,把第 4 个直角坐标记作 w,便可写出 4 维欧氏线元在直角系的表达式:

$$ds^2 = dw^2 + dx^2 + dy^2 + dz^2, \tag{2-1-14}$$

与 4 维闵氏线元

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 \tag{2-1-15}$$

相对应(但请特别注意右边第一项的负号)。

类似地,如果在闵氏时空中选取惯性坐标系以外的坐标系,闵氏线元的表达式也要改变。例如,若选 4 维球坐标系 $\{t,r,\theta,\varphi\}$,其中 r,θ,φ 仍由式(2-1-12)定义,则闵氏线元的表达式改为

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2)。 \tag{2-1-16}$$

若选更复杂的坐标系,闵氏线元的表达式将更为复杂。当且仅当选用惯性坐标系时,闵氏线元才有最简单的表达式 $-dt^2 + dx^2 + dy^2 + dz^2$。

§2.2 闵 氏 几 何

为方便起见,仍先讨论 2 维情况。把 $ds^2 \equiv -dt^2 + dx^2$ 看作线元似乎意味着曲线的长度应定义为 ds 的积分,但 dt^2 前的负号却带来一个微妙之处,因为它导致 ds^2 正负不定。图 2-2 示出了三类曲线。曲线 L_1 的任一元段都有 $ds^2 > 0$,故元段长 dl 可自然定义为 $dl \equiv \sqrt{ds^2}$,曲线 L_1 介于 a_1 和 b_1 点之间的一段的线长

(记作 $l_{a_1b_1}$)自然是

$$l_{a_1b_1} = \int_{a_1}^{b_1} \sqrt{\mathrm{d}s^2}。 \qquad (2-2-1)$$

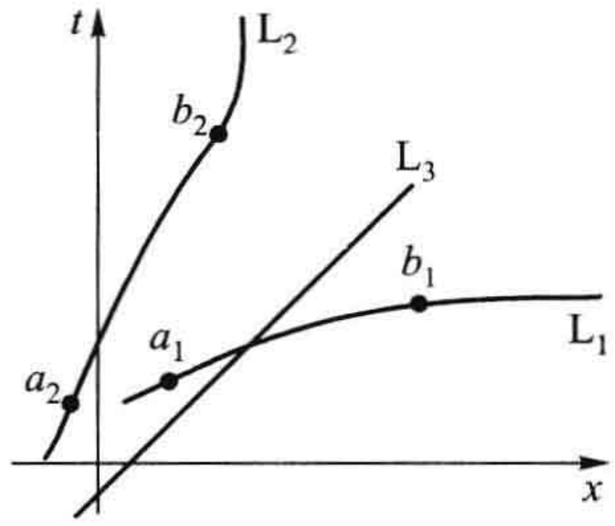


图 2-2　闵氏空间的三类曲线
(对国际制,t 轴改为 ct 轴)

但曲线 L_2 的任一元段有 $\mathrm{d}s^2<0$,如果把元段长 $\mathrm{d}l$ 仍定义为 $\mathrm{d}l\equiv\sqrt{\mathrm{d}s^2}$,则 $\mathrm{d}l$ 为虚数,而虚数线长没有意义。现在应该明确指出,$\mathrm{d}s^2$ 其实只是 $-\mathrm{d}t^2+\mathrm{d}x^2$ 的代表符号,并不代表某个实数 $\mathrm{d}s$ 的平方(个别书用符号 Φ 代替 $\mathrm{d}s^2$,以防止把它看作某实数的平方这一误解)。记号 $\mathrm{d}s$ 并无意义,它总是连同右上角的"2"字一同出现(但不代表平方),有意义的只是整体记号 $\mathrm{d}s^2$。大多数文献之所以用记号 $\mathrm{d}s^2$ 而不用 Φ,目的在于让人感到"$\mathrm{d}s^2$ 像 $\mathrm{d}l^2$",但并非像到可以把 $\mathrm{d}s^2$ 看作 $\mathrm{d}s$ 的平方的程度。明确这点之后,我们约定把元段长 $\mathrm{d}l$ 定义为 $\mathrm{d}l\equiv\sqrt{-\mathrm{d}s^2}$(请注意 $-\mathrm{d}s^2>0$),把曲线 L_2 介于 a_2 和 b_2 点之间的一段的**线长**定义为

$$l_{a_2b_2} = \int_{a_2}^{b_2} \sqrt{-\mathrm{d}s^2}。 \qquad (2-2-2)$$

最后,曲线 L_3 的任一元段有 $\mathrm{d}s^2=0$,我们定义其元段长为零,因而 L_3 的任一线段的线长为零。表面看来似乎很奇怪:这么长长的一条曲线 L_3 的线长竟然为零?其实这正是闵氏几何与欧氏几何的一个重要区别。用欧氏眼光看问题,L_3 的确很长,但用闵氏眼光来看,由定义可知其长度就是为零。

综上所述可知任一类型的曲线段 ab 的线长都可表为

$$l_{ab} = \int_a^b \sqrt{|\,\mathrm{d}s^2\,|}, \qquad (2-2-3)$$

因为第一类曲线有 $\mathrm{d}s^2>0$,上式回到式(2-2-1);第二类曲线有 $\mathrm{d}s^2<0$,上式回到式(2-2-2);第三类曲线有 $\mathrm{d}s^2=0$,上式给出 $l_{ab}=0$。以上讨论不难推广至4维情况,对三类曲线依次命名如下:

(1) 任一元段都有 $\mathrm{d}s^2>0$ 的曲线称为**类空曲线**(spacelike curve);

(2) 任一元段都有 $\mathrm{d}s^2<0$ 的曲线称为**类时曲线**(timelike curve);

(3) 任一元段都有 $\mathrm{d}s^2=0$ 的曲线称为**类光曲线**(lightlike curve,又称 null curve)。

此外也存在这样的曲线,其上某些元段有 $\mathrm{d}s^2>0$ 而另一些元段有 $\mathrm{d}s^2<0$ 或 $\mathrm{d}s^2=0$,我们戏称这些曲线为**不伦不类的曲线**。

图2-3是欧氏空间与闵氏空间的对比图。为画图方便,只对比3维欧氏空间与3维闵氏空间。欧氏空间的任一点 A 称为一个**空间点**,可用直角坐标表为

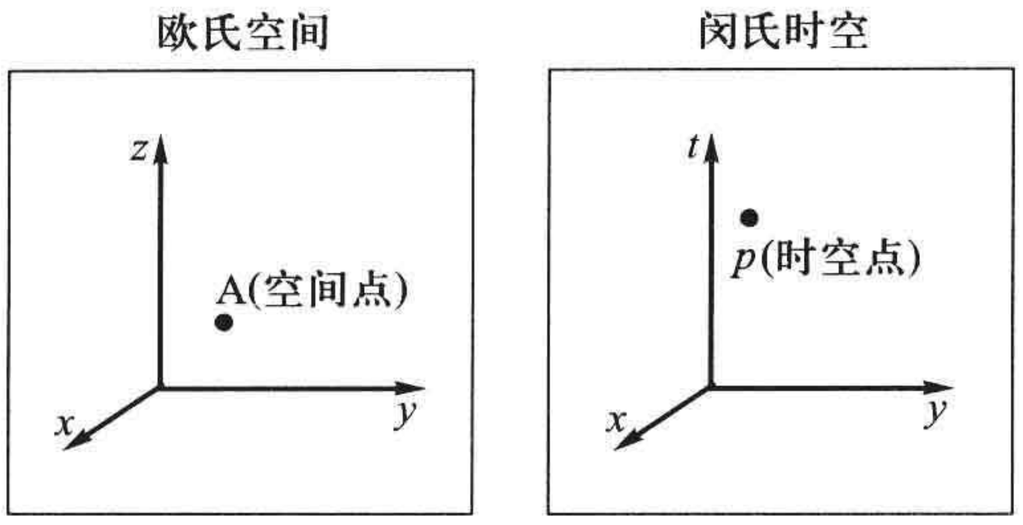


图 2-3 欧氏空间与闵氏时空对比

$$\text{空间点 A} = (z, x, y)。$$

由于闵氏空间必有一维代表时间,所以闵氏空间又称为**闵氏时空**(Minkowski spacetime)。不妨认为闵氏空间是数学家的称谓,而闵氏时空是物理学家的称谓。时空中的任一点 p 称为一个**时空点**(spacetime point),代表一个事件,可用惯性坐标表为

$$\text{时空点 } p = (t, x, y)。$$

你看,从一个纯物理问题出发,经过如此这般的三分析两讨论,“一不留神”地竟然就变成了一个几何问题,就可以借用数学做许多极有帮助的讨论和计算。尽量用几何工具讨论时空问题是国际相对论界的流行趋势(其好处多得不计其数),也是本书的一大特色。

§2.3 球面几何

一切不同于欧氏线元的线元所决定的几何都可称为**非欧几何**。从这一意义上说,闵氏几何也是非欧几何。不过,由于它与欧氏几何很像,线元之间只有一个负号之差,通常也把闵氏几何称为**伪欧几何**(pseudo-Euclidean geometry)。本节再介绍一种很常用的非欧几何,即球面几何,这是一种 2 维非欧几何,其线元可用坐标系 $\{x, y\}$ 表为

$$ds^2 = dx^2 + \sin^2 x\, dy^2, \tag{2-3-1}$$

与欧氏线元的唯一区别是 dy^2 前要补一个因子 $\sin^2 x$。虽然你可能从未见过这样的线元表达式,但下面的一段会让你发现这其实是很常见的线元。

3 维欧氏线元在球坐标系 $\{r, \theta, \varphi\}$ 的表达式为

$$ds^2 = dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\varphi^2)。 \tag{2-3-2}$$

设 S 是以原点为心、以 1 为半径的球面,dL 是躺在 S 上的任一曲线元段,当然也可把它看作 3 维欧氏空间的元段,故式(2-3-2)就是这一特殊元段的线元在

球坐标系的表达式。由于元段躺在球面上,其上每点都有 $r=1$,故式(2-3-2)简化为(改用 $d\hat{s}^2$ 代表):

$$d\hat{s}^2 = d\theta^2 + \sin^2\theta\, d\varphi^2。 \qquad (2-3-3)$$

球面是个2维空间,因为要决定其上一点需要两个坐标(θ 和 φ)。上式实质上是3维空间的欧氏线元在镶嵌于其中的2维球面所诱导而生的线元,称为(标准)**球面线元**,描述球面这个2维空间的几何(**球面几何**)。人们在讨论球面上的问题时,无论自觉与否,都在默认其几何由线元(2-3-3)决定。请注意式(2-3-3)与式(2-3-1)实质一样(差别只在于坐标的记号从 x,y 改为 θ,φ,而这毫无影响),所以式(2-3-1)的确是球面线元。

应该指出,虽然从3维欧氏线元诱导出球面线元的讲法很直观,但球面几何本身完全可以不依托于3维空间而独立存在,它无非就是由式(2-3-1)表示的一种2维几何。

§2.4　曲　　率

利用线元可以定义一个称为**曲率**(curvature)的张量,又称**黎曼张量**,是由线元所决定的诸多几何性质中很重要的一个性质。本书读者不必知道什么是张量,只需记住以下四点。

(1) 各点曲率都为零的空间称为**平直空间**(flat space),否则称为**弯曲空间**(curved space)。如果空间是时空(其中一维代表时间),则平直空间和弯曲空间也分别称为**平直时空**(flat spacetime)和**弯曲时空**(curved spacetime)。

(2) 欧氏空间是平直空间,闵氏时空是平直时空。

(3) 球面是2维弯曲空间的简单例子。球面上任一点的曲率都非零。

(4) 狭义相对论以闵氏时空为背景时空,而广义相对论则以弯曲时空为背景时空(详见第6章)。在弯曲时空中,曲线的类空、类时和类光性的定义与闵氏时空相同,线长仍定义为线元绝对值 $|ds^2|$ 的积分,即式(2-2-3)同样适用于弯曲时空。

§2.5　测　地　线

欧氏空间的曲线中有一类最为简单,这就是直线。**直线**是这样的曲线,它在直角坐标系中的任意两个坐标都有线性关系。以3维欧氏空间为例,直角坐标 x,y,z 满足下式的曲线就是直线:

$$y = ax + b,\quad z = cx + d。\quad (\text{其中 } a,b,c,d \text{ 是常实数}) \qquad (2-5-1)$$

由于任意两个直角坐标系之间的坐标变换都是线性变换，所以，只要式(2-5-1)对一个直角系成立，则它对任一直角系也成立。这就保证直线是绝对的概念，不存在某系认为是直线而另一系认为不是直线的情况。

欧氏空间之所以可以定义直线，是因为它有一类简单的、与众不同的坐标系，这就是直角坐标系。类似地，闵氏时空也有一类简单的、与众不同的坐标系，这就是惯性坐标系，因而也可用类似方法定义直线。以4维闵氏时空为例，其**直线**是这样的曲线，它的惯性坐标 t,x,y,z 之间有如下的线性关系：

$$x=a^1t+b^1,\quad y=a^2t+b^2,\quad z=a^3t+b^3,(\text{其中 } a^1,b^1,a^2,b^2,a^3,b^3 \text{ 是常实数}) \tag{2-5-2}$$

由于任意两个惯性坐标系之间的坐标变换都是线性变换(洛伦兹变换)，所以闵氏时空中的直线概念也是绝对的。

欧氏空间和闵氏空间都是平直空间。对于弯曲空间，不存在像直角坐标系或惯性坐标系那样的与众不同的坐标系，直线难以定义。然而，弯曲空间的曲线中也存在一类最简单的曲线，称为**测地线**(geodesic)。由于数学知识不够，此处不拟给出测地线的定义，只想指出以下三点。

(1) 把测地线定义用于平直空间，结果正是直线。可见测地线概念是直线概念向弯曲空间的推广。

(2) 不难验证球面上的大圆弧满足球面上测地线的定义。事实上，球面上的曲线是测地线当且仅当它是大圆弧。

(3) 鉴于欧氏空间存在“两点之间直线最短”的结论，人们往往把测地线称为短程线。然而这一称谓容易误导。下面仅举两个反例。

第一个反例是关于球面的。下面以对话形式阐述，其中乙代表笔者。

甲 球面上的大圆弧(测地线)一定是其首末两点之间的最短曲线，对吗？

乙 不对。图2-4中的大圆弧SAND就不是点S与D之间最短的线，因为大圆弧SED更短。

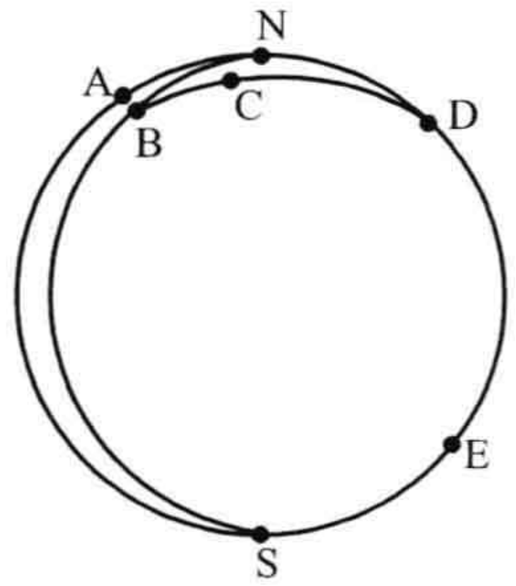


图2-4 测地线未必是极短线

甲 我把“最短”改为“极短”总对了吧？我的“最短”和“极短”是仿照一元微积分中的“最小值”和“极小值”说的，函数在定义域中可以有若干个极小值，其中最小的一个叫最小值。就是说，当我说“SAND是极短线”时，是只与其邻近曲线比较的结果。

乙 问题就在于“大圆弧SAND是极短线”这一结论也不对，因为跟它紧邻的曲线SBCD比它还短(其中BCD是B,D点间的一个大圆弧)，理由如下。

$$\text{SBCD 长} = \text{SB 长} + \text{BCD 长},$$

$$\text{SAND 长} = \text{SA 长} + \text{AN 长} + \text{ND 长},$$

故

$$\begin{aligned}\text{SAND 长} - \text{SBCD 长} &= \text{SA 长} + \text{AN 长} + \text{ND 长} - \text{SB 长} - \text{BCD 长}\\ &= (\text{BN 长} + \text{ND 长}) - \text{BCD 长}\\ &= \text{BND 长} - \text{BCD 长} > 0,\end{aligned} \tag{2-5-3}$$

[其中第二步用到 SA 长 = SB 长(总可选这样的 A,B 点)及 AN 长 = BN 长,末步(即 > 号)是因为 BCD 和 BND 都是连接 B,D 点的线,而前者是测地线后者不是,故 BND 长 > BCD 长。]式(2-5-3)表明测地线段 SAND 并非短程线,连"极短"也不是。

测地线段 SAND 之所以不是"极短",关键在于线上包含南极 S 和北极 N,两者组成线上的一个"共轭点对",即存在从 S 到 N 的、与测地线 SAN 无限邻近的测地线 SBN。可以证明,只要大圆弧上没有共轭点对,它必定是最短线。例如,大圆弧 SA,SB 和 SED 分别都是最短线。

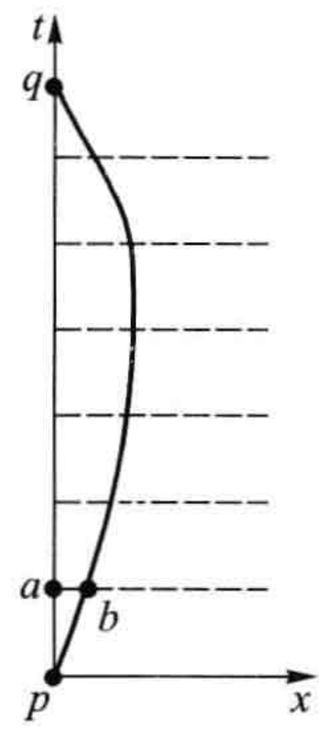


图 2-5　类时测地线 paq 是 p,q 间最长的类时线

上述例子已经表明把测地线称为"短程线"至少不够准确。更有甚者,在闵氏时空及弯曲时空中存在大量的类时测地线,它们不但不是短程线,而且还是长程线。图 2-5 就是闵氏时空的简单例子。图中的 paq 是一条类时测地线(类时直线),我们来证明它是 p,q 点之间所有类时曲线中的最长者,为此只需证明它比任一连接 p,q 点的类时曲线(例如 pbq)都长。先假定 pbq 线躺在 $t-x$ 平面上。用许多与 x 轴平行的直线把测地线 paq 和非测地线 pbq 分成许多小元段,由类时线长定义式(2-2-2)及闵氏线元表达式可知元段 pa 和 pb 的线长依次为

$$(\mathrm{d}l)_{pa} = \sqrt{-\mathrm{d}s^2|_{pa段}} = \sqrt{(\mathrm{d}t^2 - \mathrm{d}x^2)|_{pa段}} = \sqrt{(t_a - t_p)^2 - (x_a - x_p)^2} = t_a - t_p, \tag{2-5-4}$$

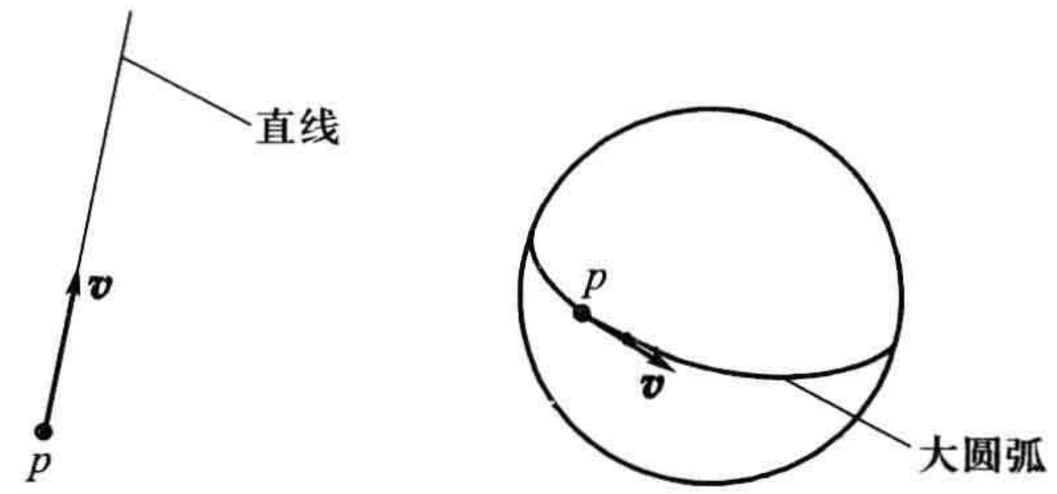
$$(\mathrm{d}l)_{pb} = \sqrt{(\mathrm{d}t^2 - \mathrm{d}x^2)|_{pb段}} = \sqrt{(t_b - t_p)^2 - (x_b - x_p)^2} < t_a - t_p = (\mathrm{d}l)_{pa}。 \tag{2-5-5}$$

[其中倒数第二步是因为 $x_b - x_p \neq 0$ 且 $x_b - x_p < t_b - t_p = t_a - t_p$。]上述结果也适用于其他任一元段,所以 $l_{paq} > l_{pbq}$。不难看出上述结论也适合于 pbq 线不躺在 $t-x$ 平面上的一般情况,可见类时测地线不但不是最短线,而且(与连接首末两点的所有类时线相较)还是最长线!

最后介绍关于测地线的一个重要定理[证明见梁灿彬,周彬(2006)定理 3-3-4],本书 §6.1 要用到。

定理 2 -1 给定一点 p 和 p 点的一个矢量 $\boldsymbol{v}$,存在唯一的测地线,它经过 p 点,而且在 p 点与 $\boldsymbol{v}$ 相切。

下面用简单例子帮助读者接受这一定理。首先,对欧氏空间及闵氏时空,给定一点 p 和 p 点的一个矢量 $\boldsymbol{v}$ 后,显然存在且仅存在一条直线(测地线),它经过 p 点而且在 p 点与 $\boldsymbol{v}$ 相切[图 2 -6(a)]。其次,给定球面的一点 p 和 p 点的一个矢量 $\boldsymbol{v}$ 后,显然存在且仅存在一个大圆弧(测地线),它经过 p 点而且在 p 点与 $\boldsymbol{v}$ 相切[图 2 -6(b)]。



(a) 欧氏或闵氏空间　　(b) 球面

图 2 -6　定理 2 -1 例证

为了便于读者记住这个定理,我们把它归结为如下的“七字真言”:

定理 2 -1(简述)　一点一矢定一测。

第 2 章习题

1. 从洛伦兹变换式(1 -3 -9)出发推出式(2 -1 -7),表明时空间隔在洛伦兹变换下形式不变。

2. 从 $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2$ 出发推出闵氏线元 ds^2 在柱坐标系 $\{t,\rho,\varphi,z\}$ 中的表达式(已知 ρ,φ 与 x,y 的关系为 $x=\rho\cos\varphi$,　$y=\rho\sin\varphi$)。

第3章　相对论的几何表述

§3.1　世界线和时空图

在牛顿力学中，当物体的形状和大小对所关心的问题影响不大时，可以把它看作有质量而无大小的对象，每个时刻只占领空间的一点，这种模型化后的物体称为**质点**。进入相对论后，质点的概念仍然保持，但它（在惯性系中）只能以亚光速运动（理由详见 §5.9）。此外还有一种很重要的点状对象，其质量为零①（因而不叫质点），在惯性系中总以光速运动。这种对象称为**光子**（photon）。质点和光子总称为**粒子**（particle）。这里的粒子是模型语言的粒子，与粒子物理学的具体粒子（质子、中子等）虽有联系但不相同，是完全没有大小的点。

设想你自己是个质点，则你在某一时刻位于某处就是发生在你身上的一个事件，所以你的全部历史由一连串相继发生的事件组成。既然每一事件都是一个时空点，你的全部历史便是时空中的一条曲线（积点成线）。推而广之，任何粒子的全部历史都是时空中的一条曲线，称为该粒子的**世界线**（world line）。

时空和世界线比较抽象，通常用时空图来直观表示。时空图中应画上（至少心中明确）某个惯性系的坐标轴，这个坐标系称为这个时空图的**基准坐标系**②。由于是时空图，时间 t 这一维不可或缺，约定把 t 轴画成竖直向上（代表该系的时间流逝）。然而，即使采用立体画法，平面上最多也只能表现 3 个坐标轴，所以空间坐标轴只能画两个，例如 x 和 y③，都与 t 轴正交。下面举 4 例帮助理解时空图。

例 1　图 3 - 1 中躺在 $t-x$ 面内的 45°斜直线（点划线）代表什么粒子的世界线？

解　由图可知该粒子的 y,z 坐标恒为零，x 坐标随 t 线性增大，所以它位于 x

---

① 在相对论的早期发展中，质量有“静质量”和“运动质量”之分。后来，包括爱因斯坦在内的越来越多的作者都放弃了“运动质量”的概念，谈到质量时一律是指静质量（见 §5.5）。此处的质量也是指静质量。

② 也可用非惯性坐标系为基准坐标系，但应明确说明。

③ 如果运动在某一空间方向（例如 z 向）无分量，只画其他两轴便可。如果运动在 x,y,z 三个方向上都有分量，必要时可以“多侧面”地画时空图，例如“$t-x-y$ 图”，“$t-x-z$ 图”，等等。

轴上而且沿 x 正向匀速运动。由图又知

$$\text{粒子速率}=\frac{x_2-x_1}{t_2-t_1}=\frac{t_2-t_1}{t_2-t_1}=1(=c),\qquad(3-1-1)$$

可见该粒子以光速运动。结论:图中的斜直线可解释为沿 x 轴正向运动的光子的世界线。

注记 1　图 3 -1 是几何单位制中的时空图。若用国际单位制,纵坐标应从 t 改为 ct。

例 2　以图 3 -2 的类时曲线 L_1 为世界线的粒子是质点还是光子?类空曲线 L_2 呢?

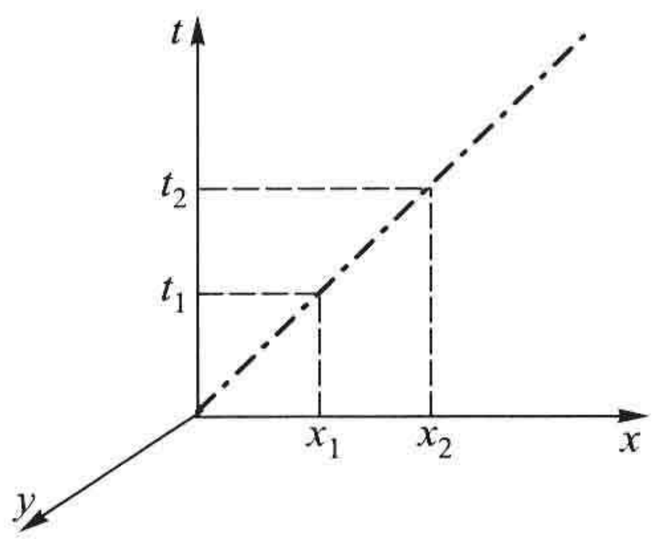


图 3 -1　点划线是光子的世界线

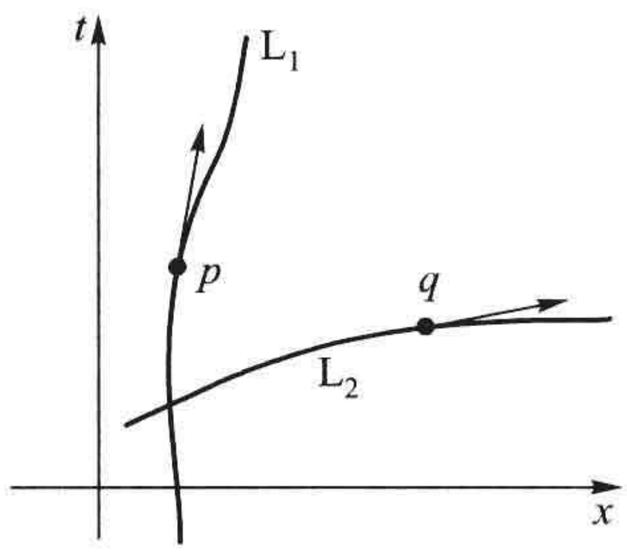


图 3 -2　以惯性系为基准系的时空图。L_1 是质点的世界线,L_2 不是世界线。

解　曲线 L_1 上任一点 p 的、切于该曲线的矢量(图中的箭头)都比 45°更陡(斜率更大),仿照式(3 -1 -1)的讨论可知该粒子在每一时刻的速率 u 都为亚光速(即 $u<1$),所以 L_1 是某个质点(而不是光子)的世界线。反之,从曲线 L_2(类空曲线)在任一点 q 的切矢看出,假定它是某一粒子的世界线,该粒子必定超光速,而超光速粒子在相对论中不存在,可见 L_2 不是任何粒子的世界线。

小结　**质点的世界线是类时曲线,反之亦然;光子的世界线是类光曲线;类空曲线不能充当世界线。**

对于 2 维时空(以 t,x 为时、空坐标),过任一时空点 p 只有两条光子世界线(斜率分别是 +1 和 -1);对于 3 维时空,过一点 p 有无限多条光子世界线,以任一条为母线就扫出一个以 p 为顶点的圆锥面,称为 p 点的**光锥面**(2 维曲面)。对于 4 维时空,每点 p 的光锥面是一个 3 维锥"面"(图 3 -3,图中只能压缩一维)。只要知道一点的光锥面,就能判断

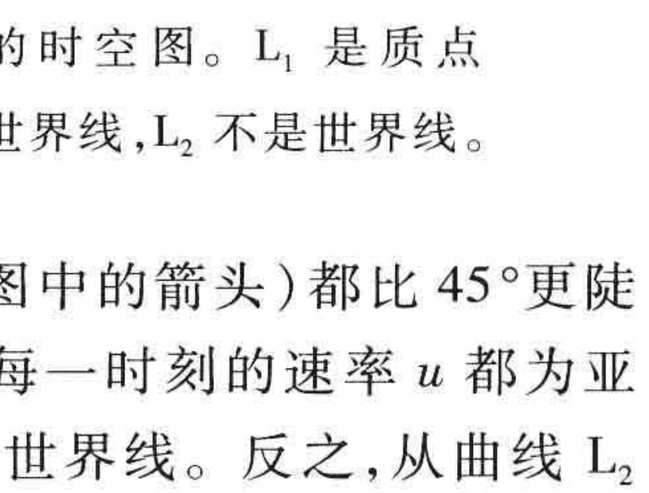
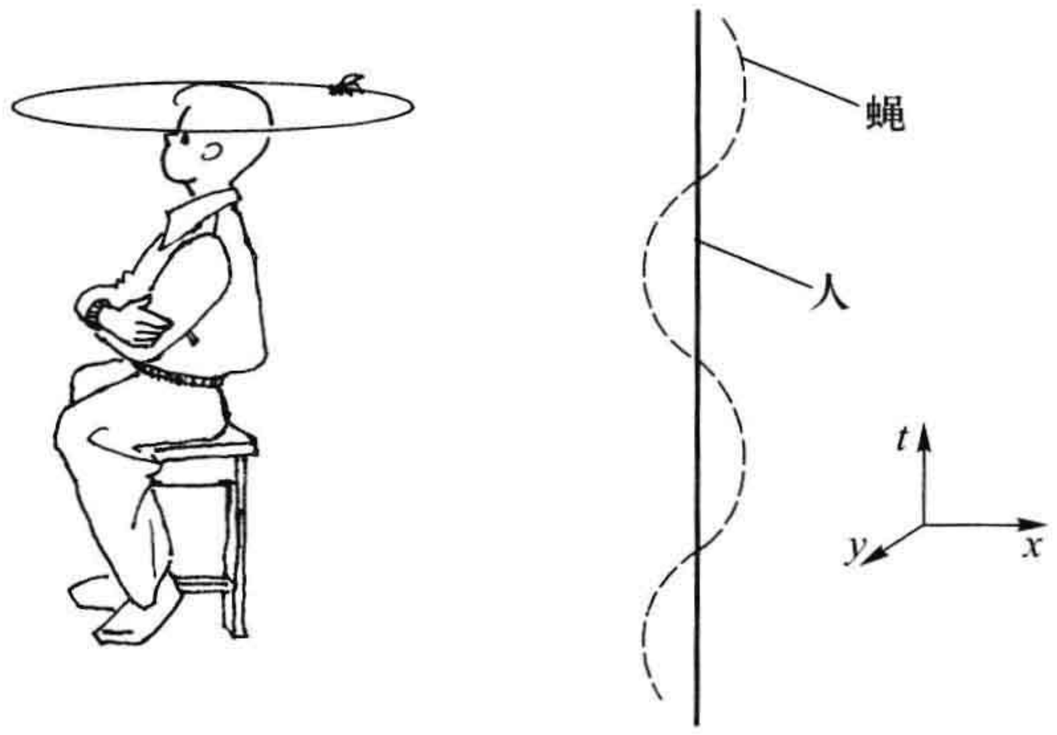


图 3 -3　p 点的光锥面

过该点的曲线在该点的切矢的类时、类空和类光性，就可以避免把在该点为类空的曲线误以为是世界线。

为了帮助读者熟练掌握空间语言（空间图）与时空语言（时空图）之间的“翻译”工作，我们再提一个反方向的“翻译”问题——请读者从给定的空间图出发画出相应的时空图。

例3　某人（看作质点）静止于地面，一只苍蝇绕他做圆周运动[图3-4(a)]，试画出人和蝇在时空图中的世界线（选地面系为基准）。



(a) 空间图　　(b) 时空图

图3-4　人静止于地面，苍蝇绕人做匀速率圆周运动。

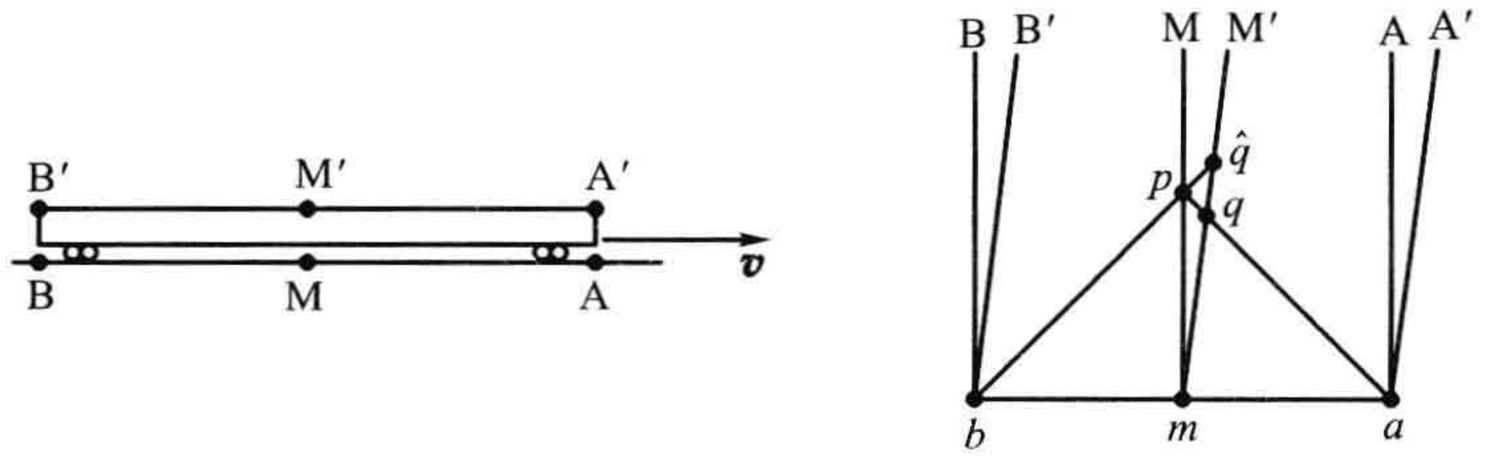
解　因为人相对于地面静止，他的世界线自然是竖直线。至于苍蝇的世界线，读者也都能想到应是一条螺旋线[图3-4(b)]。但是画图时还有若干不确定性。首先，这条螺旋线的螺距是均匀的还是不均匀的？这取决于苍蝇的圆周运动是否匀速率。然而这个问题无法从空间图[图3-4(a)]找到答案（空间图不含此信息）。只有当出题人在空间图旁附一句话“苍蝇做匀速圆周运动”，你才可以把苍蝇的世界线画成等螺距螺旋线。这时又出现第二个问题：螺距该多大？在空间图中同样找不到答案，除非出题人再附上一句话“苍蝇的速率是0.1 m/s”。这一讨论表明，空间图包含的信息量对绘制时空图未必足够。反之，如果先给你时空图，只要把它从上到下压缩为一个平面图，本质上就得到空间图。更有甚者，如果你从时空图底部开始匀速地向上看，你还能看到运动的全过程。建议读者立刻以图3-4(b)为例试试看，你是否看到“人一直不动而苍蝇绕着他做匀速圆周运动”这一过程？以上讨论表明时空图含有比空间图更多的信息。

例4　用时空图重新审视“爱因斯坦列车”的雷击同时性问题（见图1-2及相应文字）。

解　图3-5(b)是以地面系为基准的时空图，现在介绍画图过程。地面的

点 A,M,B[见图 3 -5(a)]相对于地面系皆静止,故世界线都是竖直线。列车的点 A′,M′,B′都以同一速度 v 相对于地面运动,故世界线是三条斜率相同的直线(斜率为 $1/v$),可用下法画出。

(1) 在 A 线上任选一点 a 代表“雷击车头”事件。(再次提醒:务必注意大、小写字母的区别——A 代表空间点,a 代表时空点,是发生在 A 处的一个事件。)由于前已约定 A 是雷击时 A′在地面的印记(见图 1 -2),所以 a 也代表“A′与 A 相遇”的事件,因而 A′线是过 a 点的、斜率为 $1/v$ 的斜直线。



(a) 空间图　　(b) 时空图

图 3 -5　M 同时收到车头、车尾来光,但 M′不同时收到。

(2) 以 b 代表“雷击车尾”事件,它当然应是 B′线与 B 线的交点,但如何确定此点?由于事件 b 和 a 相对于地面系同时发生,而图 3 -5(b)的任一水平线上各点代表(对地面系而言)同时发生的事件(因为它们有相同的 t 值),所以,过 a 点的水平直线与 B 线的交点就是 b。

(3) 以 m 代表 M′与 M 相遇的事件。对地面系而言,这一事件同时于事件 a,因此应在水平线 ab 上,所以 ab 线与 M 线的交点就是 m。由此可画出 M′线。

车头受雷击时(事件 a)向四面八方发射光子,其中一条光子世界线与 M 线交于 p 点,所以

$$p = \text{“M 收到车头来光”事件。}$$

车尾受雷击时(事件 b)也发光,其中一条光子世界线由于与 ap 线关于 M 线对称而交 M 线于 p,可见观者 M 的确同时收到车头和车尾来光。这从另一侧面印证了刚才找出的 b 点是正确的。

现在已经水到渠成。由图可见 M′先收到车头来光(记此事件为 q)后收到车尾来光(记此事件为 $\hat{q}$),可见火车系认为事件 a,b 并不同时。

注记 2　刚才讲过,图 3 -5(b)的任一水平线上各点代表(对地面系)同时发生的事件。通常称这些水平线为(地面系的)**同时线**。至此只涉及 1 维空间,即 2 维时空 $t-x$。推广至 2 维空间,即 3 维时空 $t-x-y$,同时线就发展为**同时面**,见图 3 -6,每一同时面(如 Σ_{t_1})代表某一时刻(如 t_1)发生的所有事件。再推广至最常用的 3 维空间,即 4 维时空 $t-x-y-z$,同时面理应改称为“同时体”,

代表同一时刻发生的所有事件的集合，亦即代表该时刻的全空间。但时空图必须压缩掉一个空间维，故图中仍表现为一张面，而且习惯上也仍称为**同时面**。

一个粒子在3维语言中是个点(0维对象)，在4维语言中就是一条线(1维对象)。考虑一把静止于地面的“无限细”尺子(1维对象)，它的每点的全部历史在(以地面系为基准的)时空图中就是一条竖直线，所以尺子的全部历史就由一个(2维)平面代表(图3-7)，称为尺子的**世界面**(world sheet)，它与任一同时面 Σ_t 的交线就代表地面系认为的、t 时刻的(1维)尺子。再看地球表面(2维球面)，它的每点的历史由一条竖直线代表，地球表面所有点的世界线的总体便是图3-8的圆柱面(地球表面的世界面)。你也许会觉得有什么东西不对，因为任一同时面 Σ_t 与这个圆柱面的交集本应代表 t 时刻的地球表面(2维球面)，但图中的交集却是个(1维)圆周！问题仍然出在时空图至多只能表现两维空间，只要自己在头脑中补上第3个空间维(比如说，z 轴)，这圆周就代表2维球面。如果愿意，你还可以补画人造地球卫星的世界线(图3-9)。

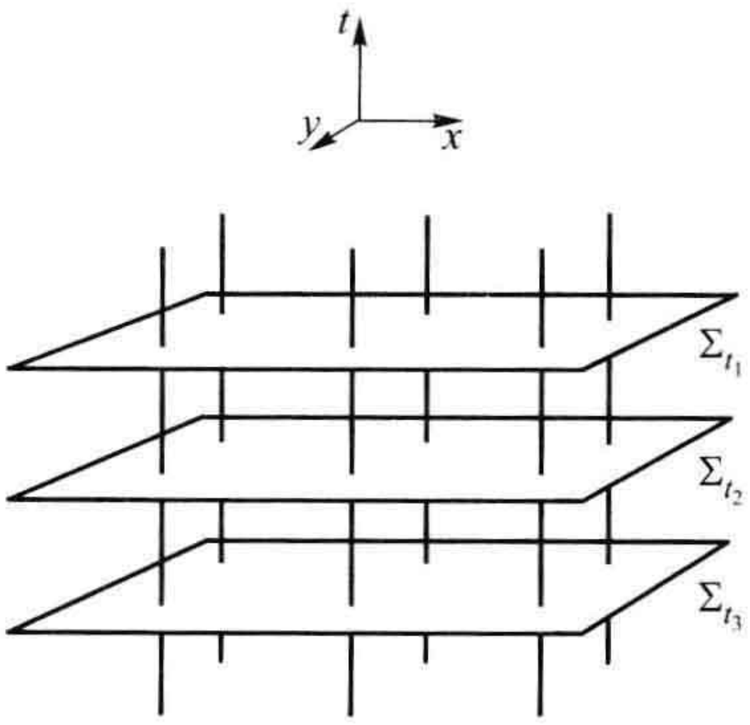


图3-6　每一同时面代表一个时刻的全空间

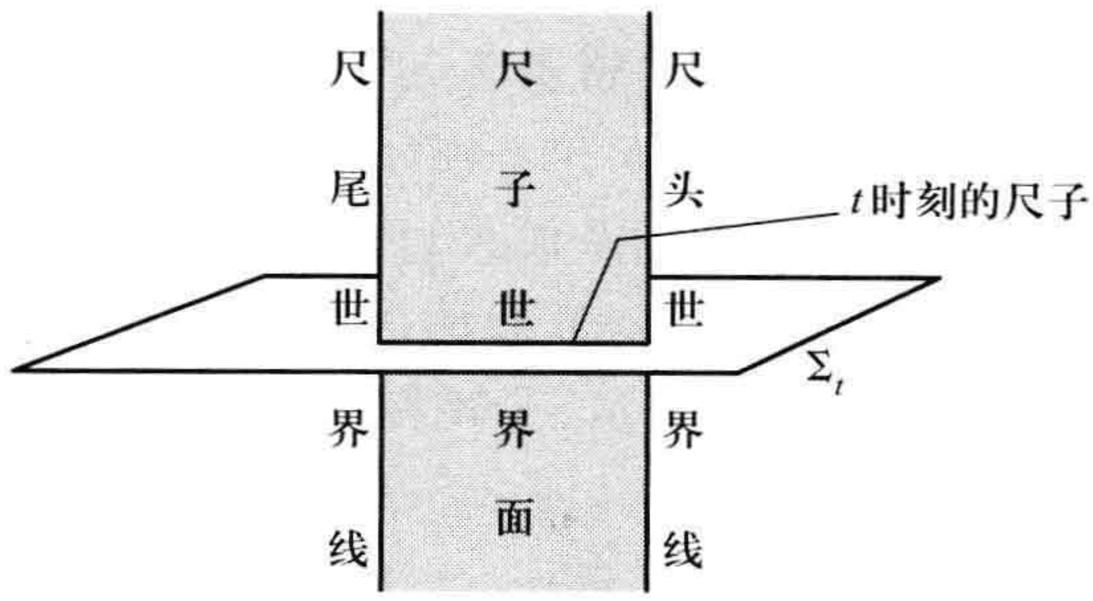


图3-7　尺子的世界面

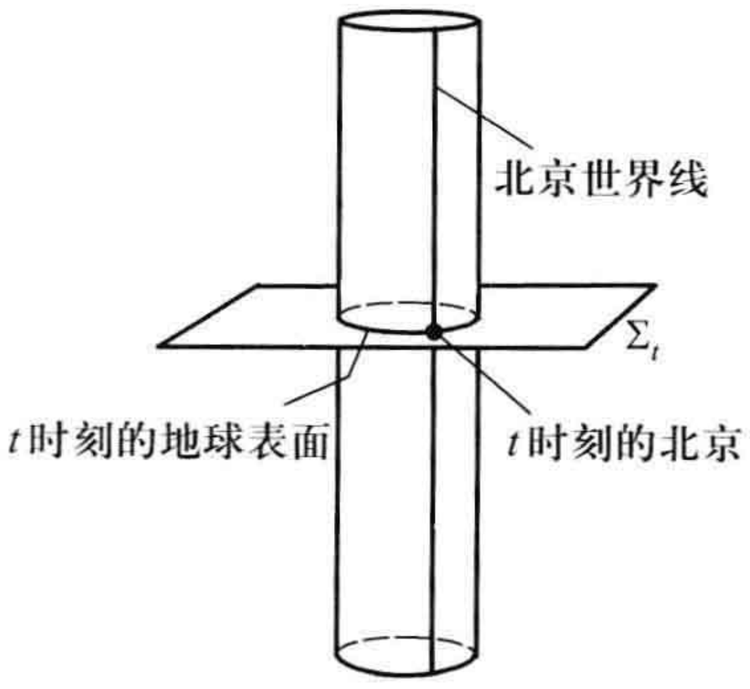


图3-8　圆柱面是地球表面的世界面(压缩掉一维)

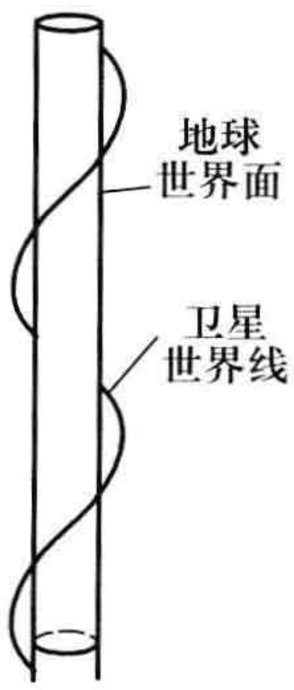


图3-9　人造地球卫星世界线(地球世界面仍压缩掉一维)

§3.2　惯性观者、参考系和坐标系

进行物理观测的人叫观察者。通常把观察者模型化，即看成质点，简称**观者**(observer)。为了观测，观者要携带一个走时准确的钟，叫**标准钟**(standard clock)，该钟的读数称为该观者的**固有时**(proper time)。推而广之，可以认为任一质点(不一定是人)都携带一个标准钟，每个质点都有自己的固有时，都可起到观者的作用。

图 3－10　标准钟定义

然而，什么叫做“走时准确”？下面用几何语言给出明确定义，这也就是标准钟的定量定义。

定义 1　设钟在自己世界线任意两点 p_1 和 p_2 的读数分别是 τ_1 和 τ_2($>\tau_1$)(见图 3－10)。如果

$$\tau_2 - \tau_1 = \text{世界线 } p_1p_2 \text{ 段的线长 } l_{p_1p_2} = \int_{p_1}^{p_2} \sqrt{-\mathrm{d}s^2}, \tag{3-2-1}$$

此钟就称为一个**标准钟**。

注记 3　上式表明世界线上任一元段的固有时间与元段线长的关系为

$$\mathrm{d}\tau = \sqrt{-\mathrm{d}s^2}。 \tag{3-2-2}$$

若光速 c 不取为 1，则上式应改为

$$\mathrm{d}\tau = \sqrt{-\mathrm{d}s^2/c^2}。 \tag{3-2-2'}$$

式(3－2－1)则应改为

$$\tau_2 - \tau_1 = \int_{p_1}^{p_2} \sqrt{-\mathrm{d}s^2/c^2}。 \tag{3-2-1'}$$

注记 4　上述定义涉及线长这一几何性质，在通常教科书中很少见，但可以证明式(3－2－1)中的 τ 与教科书中的固有时完全一致。

注记 5　有必要分清与钟有关的两个概念——**走时率**(rate)和**初始设定**(setting)，前者反映走时的快慢，后者则是对零点的指定。标准钟只对走时率提出要求(世界线上任意两点的读数差等于线长)。我国曾一度实行夏时制，规定某月某日零时“把钟拨快一小时”。这“快”字有可能使人误以为要提高走时率，其实只是改变初始设定。事实上，你为你的手机调整时间的操作也只是在改变初始设定，你是无法改变手机的走时率的。

现在介绍惯性观者的概念。早在伽利略时代，物理学家就已认识到物体的运动有惯性运动与非惯性运动之分。从教学法的角度，不妨把相对性原理做如

下的重新表述。

相对性原理实质上包含以下两部分内容(假设):

(1) 所有观者可以分为两类,即**惯性观者**(inertial observer)和**非惯性观者**。这种区分无须借用其他物体作为参考,所以说是绝对的区分。你可以指着任一观者问:这是否惯性观者?答案非是则否,泾渭分明,绝不模棱两可。

(2) 所有惯性观者互相平权,不存在任何特殊的、优越的、与众不同的惯性观者。特别地,没有一个惯性观者可以被认为处于绝对静止的状态。

然而,什么观者才算是惯性观者?或者说,惯性观者与非惯性观者的根本区别是什么?想象你置身于远离各个星球(因而引力可以忽略)的飞船内。只要你开动发动机(引擎),你就是个非惯性观者,你的身体就会或多或少地有不舒适的感觉。反之,只要发动机完全停止工作,你就是个惯性观者,不舒适感就会消失。请注意你只需根据自己的感觉就能判断自己在做惯性运动与否,完全不必依靠观察窗外的景物(所以说惯性运动与否的区别是绝对的)。当然也可以举日常生活的例子。假定你正在驾车,当汽车启动、刹车或拐弯时,你会或多或少地有不舒适感,这时你是个非惯性观者。反之,当汽车平稳地行驶在笔直又平坦的路面时,你是个惯性观者。读者可能觉得以上描述过于直观、粗略而有失定量性,但这种描述确实能够说明观者的确可以分为两类。以上关于观者的讨论也适用于质点(物体):质点也可以分为两大类:① 做惯性运动的质点(简称**惯性质点**或**自由质点**);② 做非惯性运动的质点(简称**非惯性质点**)。

你不是很想知道惯性观者(质点)的更为准确的、定量的定义吗?下面就是在狭义相对论框架内的惯性观者的数学定义:在闵氏时空中,当且仅当观者的世界线是类时测地线时,这个观者是**惯性观者(自由质点)**。这一定义不但已经通过了无数(实验的以及理论内部自洽性的)考验,而且非常符合美学原则,因为,从物理角度看,惯性观者是既特殊又最简单的一类观者;从数学角度看,类时测地线是既特殊又最简单的一类类时曲线。

观者(无论惯性与否)的任务是观测。观测又可分为两大类。观者对发生在自己身上的事件的观测称为**直接观测**,亦称**当时当地观测**,是最简单、最明确的观测。如果事件发生在观者身外,他当然也可能听到或看见,从而也可进行观测和记录,这称为**间接观测**。在理论方面,狭义相对论中高速运动物体的视觉形象问题就属于间接观测范畴(详见 §4.4);在实用方面,所有天文观测都属于间接观测。间接观测必定涉及从事件到观者的信号传递(例如用声或光),要花费一定时间,讨论起来比较复杂。此处只讨论直接观测。为了能对发生在任何地点的事件进行观测和记录,就需要处处设置观者(每一观者带一标准钟),这些

无处不在的观者的集合(总体)就构成一个**参考系**。参考系内的每个观者都“枕戈待命”地随时准备用自己的钟对发生在自己身上的事件做记录,所以爱因斯坦曾说过:“**在我的相对论里,我的空间每一点都安放了钟;但实际上在我的房间里,甚至连一个钟都没有。**”(那时爱因斯坦经济还很拮据,虽然生活在号称钟表王国的瑞士,却买不起一个钟。)以地面惯性参考系为例,它由无处不在的、相对于地面静止的惯性观者组成(图 3 - 11 画出了其中五个观者的世界线)。设炮仗爆炸事件 a 发生在观者 A 身上,他立即记下自己的标准钟在爆炸时的读数 τ_a,并把它定义为事件 a 在该参考系的时间坐标,记作 t_a,即 $t_a \equiv \tau_a(\mathrm{A})$(括号内的 A 代表 A 钟的读数)。为了进一步定量描述事件 a 的发生点(即 A 的所在地),可以任选一个观者(记作 O)为原点建立直角坐标系(坐标轴相对于地面系无转动),这样就得到一个惯性坐标系 $\{t,x,y,z\}$(O 的世界线就充当 t 坐标轴),称为地面惯性参考系内的一个**惯性坐标系**(图 3 - 11),事件 a 就可以用时空坐标表为

$$a = (t_a, x_a, y_a, z_a)。$$

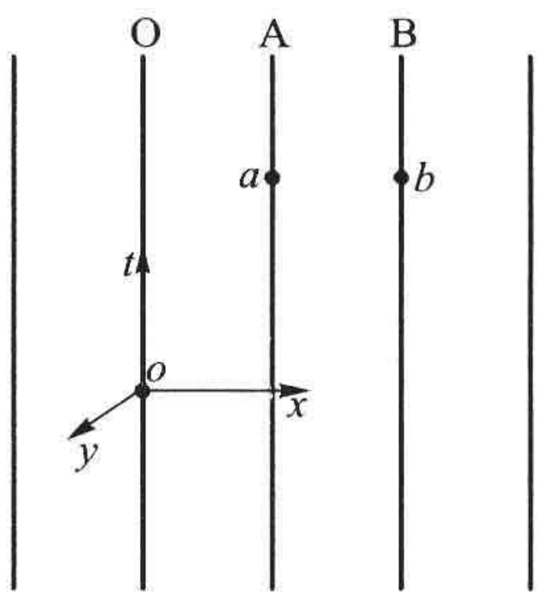


图 3 - 11　地面参考系内的惯性坐标系 $\{t,x,y,z\}$

一个惯性参考系内可以有很多惯性坐标系,因为通过以下三种做法都可获得该参考系内的另一惯性坐标系:① 让 3 个坐标轴(统一地)做一个空间转动;② 改变空间坐标原点(选另一惯性观者为空间坐标原点);③ 把图 3 - 11 的 4 维坐标原点 o(小写)移至观者 O(大写)的世界线的另一点。

在不必细分惯性参考系和惯性坐标系时,可以笼统地称之为**惯性系**。

现在遇到一个重要问题。设事件 a,b 位于图 3 - 11 的某个水平面上。由于每个水平面要代表一张同时面(等 t 面),a,b 点在该坐标系的时间坐标应相等,即应有 $t_a = t_b$,其中 $t_a \equiv \tau_a(\mathrm{A})$,$t_b \equiv \tau_b(\mathrm{B})$。然而,每个标准钟(作为单个的钟)的初始设定完全任意,所以并不保证 $t_a = t_b$。为了解决这一问题,我们约定对同一参考系内的各钟要事先施行**同步**操作,即要求各钟在任一同时面上的读数必须相同。满足这一要求的各钟称为已同步的。由于各钟的世界线介于任意两个同时面之间的线段长度相等,同步操作具有一劳永逸性:只要各钟在某一同时面上读数相等,它们就已同步。乍看起来可用下法简单地实现同步:任选一观者 A 作为“带头羊”,他在把自己的钟设定为零(事件 a)的同时告知另一钟 B“你现在立即把钟调到指零”。但是,“告知”就是发出某种信号,而任何信号的传播都需要时间,若以 b 代表 B 接到通知的事件,则 b 与 a 不会在同一张同时面上。所以,如果 B 执行命令,即在事件 b 把自己的钟设定

为零,必然达不到同步的要求。可见同步操作在相对论中是个并非平凡的过程。下面介绍一种原则上可行的同步操作法[称为雷达法(radar method)]:观者A事先(用任意方式)告知B:"你在身上装一面反射镜,当镜子接收到我发的光时把你的钟调零。"然后A在适当时刻向B发光(事件a_1),它到达B时被反射(事件b,见图3-12)。以a_2代表A回收到反射光的事件,a代表A线上a_1a_2段的中点(以线长衡量),则A在该点把自己的钟调零便可保证B钟与A钟同步。

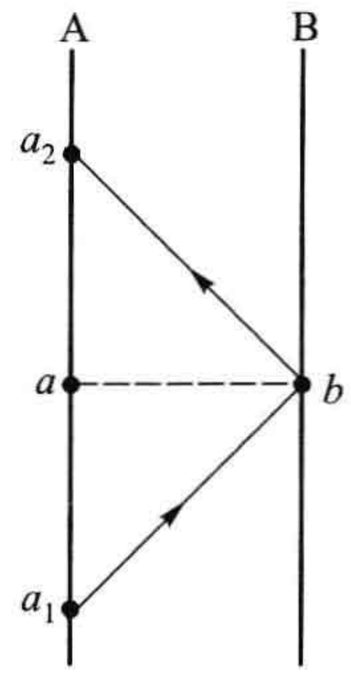


图3-12　钟同步的雷达法

[选读3-1]

为了尽量保证后续内容的清晰性,刚才是先介绍惯性观者,再以此为基础介绍惯性参考系和坐标系(见下节)。这种讲法的好处将会逐渐显现。当然,不用惯性观者概念也能对惯性系下定义(而且这是通常的讲法),这定义如下。

与其他物体距离足够远(因而不受任何物体的作用)的质点称为孤立质点(或自由质点)。如果孤立质点相对于某个参考系做匀速直线运动,这个参考系就称为惯性参考系,简称惯性系。

采用这一定义后,惯性观者的定义就变得非常简单:相对于惯性系做匀速直线运动的观者称为惯性观者。　　**[选读3-1完]**

§3.3　光速不变的4维表述

用惯性坐标系K可以自然地定义任何粒子的速率。设L是粒子的世界线(图3-13),其任一元段dL的线元可表为

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2。\quad (3-3-1)$$

以p代表元段的起点,则粒子在时刻p相对于惯性系K的速率自然定义为

$$u \equiv \frac{\sqrt{dx^2 + dy^2 + dz^2}}{dt}。\quad (3-3-2)$$

于是式(3-3-1)可改写为

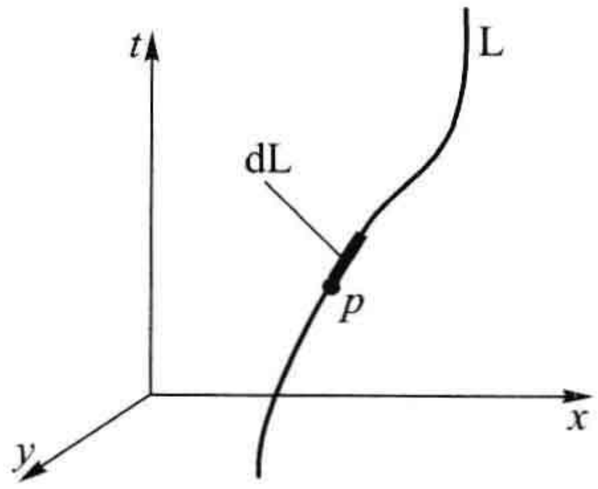


图3-13　粒子世界线上的元段

$$ds^2 = -dt^2\left(1-\frac{dx^2+dy^2+dz^2}{dt^2}\right) = -(1-u^2)dt^2。\qquad (3-3-3)$$

以符号“⇔”代表“等价于”,则上式表明:

$$u=1 \quad \Leftrightarrow \quad ds^2=0 \quad \Leftrightarrow \quad \text{线元为类光},\qquad (3-3-4a)$$

$$u<1 \quad \Leftrightarrow \quad ds^2<0 \quad \Leftrightarrow \quad \text{线元为类时}。\qquad (3-3-4b)$$

利用上两式可把狭义相对论的两个基本“信条”用4维语言重新表述。这两个信条的3维表述是大家熟知的两句话:

(A) 光子相对于任何惯性系的速率 $u=1$(此即光速不变原理的核心内容);

(B) 质点相对于任何惯性系的速率 $u<1$(此即“质点必须亚光速”)。

现在,借助于等价关系式(3-3-4)又可把两个信条用4维语言重新表述如下:

(a) 光子的世界线是(闵氏时空的)类光曲线;

(b) 质点的世界线是(闵氏时空的)类时曲线。

由式(3-3-4)可知4维表述(a)、(b)等价于3维表述(A)、(B)。然而,3维表述总要借助于坐标系,而且还涉及粒子速率 u 的定义,不但不如4维表述简练自足,而且有时还会招致误解。特别是,人们往往把3维表述(A)、(B)简单地记忆为

(A′) 光子的速率 $u=1$;

(B′) 质点的速率 $u<1$。

或者更简单地记忆为“相对论不许超光速”,则更是容易引起误解和不必要的困惑。关键在于,“速率”一词可以有各种不同定义,其中有些在物理上还很合理,很有资格被称为速率,但只要与质点相对于惯性系的速率定义式(3-3-2)不等价,这样的“速率”超过光速 c 就未必违背两个信条,就未必与4维表述(a)、(b)相矛盾(“超光速”质点的世界线仍然可以是类时曲线),因而是完全允许的,不在被禁止之列。然而,如果只记住“相对论不许超光速”,就会误以为连这种“超光速”也有悖于相对论,也必须被禁止。这种误解的例子很多,下面仅举三例。

例1　设 $\{t,x,y,z\}$ 是惯性坐标系,用下式定义新坐标系 $\{T,X,Y,Z\}$:

$$T\equiv t/2,\ X\equiv x,\ Y\equiv y,\ Z\equiv z,\qquad (3-3-5)$$

则

$$U\equiv\frac{\sqrt{dX^2+dY^2+dZ^2}}{dT}\qquad (3-3-6)$$

自然就是质点在新坐标系中的速率。由上式易得

$$U=\frac{\sqrt{\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2}}{\mathrm{d}t/2}=2\frac{\sqrt{\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2}}{\mathrm{d}t}=2u。[末步用到式(3-3-2)]$$

设 $u=2/3$(即 $2c/3$),便有 $U=4/3>1$,即质点超光速。

注记6　本例是一位美籍华人在北京做报告时对狭义相对论提出的一个质疑,本书第一作者当时在听众席。

述评　本例的 $U>1$ 的确违背(B′),但并不违背准确提法(B),因为(B)要求 u 是质点相对于惯性系的速率,而新系 $\{T,X,Y,Z\}$ 不是惯性系。如果听完例1后跟着去大惊小怪,那完全是庸人自扰。

例2　设 $\mathrm{K}\equiv\{t,x,y,z\}$ 是惯性坐标系,质点甲、乙沿 x 轴相背而行,如图3-14所示。以 $x_{甲}$ 和 $x_{乙}$ 代表甲、乙的 x 坐标,则 $R\equiv x_{乙}-x_{甲}$ 是两者的距离。有人喜欢用下式定义"甲乙之间的、由K系测得的相对速率"u:

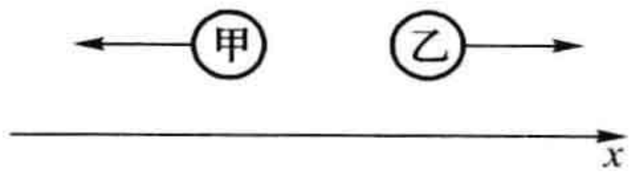


图3-14　例2用图

$$u\equiv\frac{\mathrm{d}R}{\mathrm{d}t},\qquad(3-3-7)$$

于是

$$u=\frac{\mathrm{d}x_{乙}}{\mathrm{d}t}-\frac{\mathrm{d}x_{甲}}{\mathrm{d}t}。$$

而 $\mathrm{d}x_{甲}/\mathrm{d}t$ 和 $\mathrm{d}x_{乙}/\mathrm{d}t$ 分别是甲和乙相对于K系的速率(记作 $u_{甲}$ 和 $u_{乙}$),故

$$u=u_{乙}-u_{甲}。$$

只要取 $u_{乙}=-u_{甲}=2/3$(即 $2c/3$),便有 $u=4/3>1$,即超光速。

述评　由式(3-3-7)定义的 u 并非哪个质点相对于哪个惯性系的速率,它大于光速 c 并不违背信条(B)。

事实上,例1的 $u=2/3$ 和例2的 $u_{乙}=-u_{甲}=2/3$ 都不会改变两例所涉及的质点的世界线是类时曲线这一事实,根本不违背基本信条的4维表述(b),什么问题也没有。

例3　本例是宇宙论中经常被问及的一个有趣而且实际的问题,详见选读3-2。

[选读3-2]

宇宙是个"无所不包"的最大时空。由于存在大量物质,这个时空是弯曲的,不存在什么惯性系。不过,宇宙存在一个特殊的、各向同性的参考系,称为宇宙静系,其中的时间坐标 t 称为宇宙时,它的每一同时面 Σ_t 代表宇宙在时刻 t 的全空间(也就是3维语言中的全宇宙)。在宇宙这个浩瀚的"大海"中,一个星系(例如我们的银河系)就如同"沧海之一粟",可以被处理为宇宙时空中的一条世

界线，它与每一同时面 Σ_t 的交点就代表“t 时刻的该星系”。观测表明宇宙自创生至今一直在膨胀之中，表现为每一星系都觉得其他星系不断离自己退行而去。作为比喻，考虑一个布满蚂蚁的气球。当气球膨胀时，每一蚂蚁都看到其他蚂蚁在退行。没有一只蚂蚁是特殊的。同气球膨胀一样，宇宙的膨胀也没有中心。任意两个星系的距离 D 都是宇宙时 t 的函数，可记作 $D(t)$，两星系之间的退行速率 u 自然定义为 $D(t)$ 的时间导数：

$$u \equiv \mathrm{d}D(t)/\mathrm{d}t。 \tag{3-3-8}$$

宇宙论有一个哈勃定律，断言当今任意两个星系之间的退行速率 u 与它们的距离 D 成正比：

$$u = HD, \tag{3-3-9}$$

其中 H 是比例常数（哈勃常数）。由上式可知，当 D 足够大时 u 可以大于光速。许多人对此感到疑惑：这是否违背相对论的基本信条？答案是并不违背。关键在于，虽然用式(3-3-8)给退行速率下定义非常合理，但它并不等价于式(3-3-2)的速率定义[①]，这样的 u 大于光速未必违背相对论。判断是否违背相对论的最佳办法是用基本信条的 4 维表述(b)，也就是看看每个星系的世界线是否为类时曲线。事实上，用 4 维语言讨论宇宙论时自始至终都默认每一星系的世界线是类时线[详见梁灿彬，周彬(2006)第 10 章]，因此由式(3-3-8)定义的 u 超过光速并不带来任何问题。读者由此应该再次悟出（我们提倡一个“悟”字）4 维语言的优越性。试想：只记住“相对论不许超光速”而不知道对速率的定义有所要求就容易上当，而 4 维表述(b)（即“质点的世界线是类时线”）则简练自足，毫无误导之虞。唯一缺点是不懂世界线和类时线的读者无从理解，如读天书。不过，按照本书第 2、3 章的讲法很快就能学会有关概念（就像用傻瓜照相机前稍微学一学简单用法），又何乐而不为呢？　**[选读 3-2 完]**

§3.4　固有时与坐标时的区别和联系

惯性坐标系 $\{t,x,y,z\}$ 中的坐标 t 也称为**惯性坐标时**，简称**坐标时**，与固有时 τ 既有区别又有联系。坐标时与固有时有如下两点重要区别：

（1）固有时只对世界线上的点而言，脱离世界线就没有固有时可言。如果两条不同的世界线 L_1 和 L_2 交于 p 点，则 p 点作为 L_1 的一点的固有时可以不同于它作为 L_2 的一点的固有时。因此，“固有时”前面应该加一个定语，即明确说

---

① 事实上也无从等价，因为式(3-3-2)涉及惯性坐标 t,x,y,z，而宇宙（作为弯曲时空）并不存在惯性坐标系。

成“某世界线的固有时”。当然，每次都加定语会过于啰嗦，但至少要在心中明确，做到心照不宣。（遗憾的是，有太多的人在谈及固有时的时候心中对此定语若明若暗。你如果问他：“谁的固有时？”他可能答不出来，甚至从来就没有想过。）反之，坐标时与世界线无关，对任一时空点 p 都可谈及它的惯性坐标时。

（2）同一时空点 p 在不同惯性坐标系中可以有不同的惯性坐标时。因此，“坐标时”前面应该加一个定语（至少做到心照不宣），即明确说成“某某惯性坐标系的坐标时”。而固有时则与坐标系毫无关系。应该防止另一极端，那就是乱加定语，例如，如果说“p 点在某坐标系的固有时”或者“p 点作为某世界线的一点的坐标时”，都会贻笑大方。

如果 p 点在某条世界线 L 上，当然可以谈及它作为 L 线的一点的固有时 τ。如果又选定了一个惯性坐标系 $\{t,x,y,z\}$，当然也可谈及 p 点在这个坐标系的坐标时 t。我们想找出 τ 与 t 的关系。设 dL 是 L 的一个元段，起点为 p（仍见图 3－13）。注意到 L 是类时曲线，由式（3－2－1）可知元段 dL 的固有时间

$$\mathrm{d}\tau = \sqrt{-\mathrm{d}s^2}, \tag{3-4-1}$$

故

$$\mathrm{d}\tau^2 = -\mathrm{d}s^2, \tag{3-4-2}$$

设 u 是以 L 为世界线的质点相对于所论坐标系的速率，则上式与式（3－3－3）结合便得

$$\mathrm{d}\tau^2 = (1-u^2)\,\mathrm{d}t^2。 \tag{3-4-3}$$

再利用符号 $\gamma_u \equiv 1/\sqrt{1-u^2}$ 就可把上式改写为

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \gamma_u。 \tag{3-4-4}$$

上式就是惯性坐标时 t 与固有时 τ 的微分关系式。当 L 是所选惯性系内的一个惯性观者时，由式（3－3－2）可知其 $u=0$，故 $\gamma_u=1$，注意到 $\tau=0$ 时 $t=0$，对式（3－4－4）积分便得 $t=\tau$。这是很自然的，因为，根据图 3－11 所涉及的一段，任一时空点 p（当时记作 a）的坐标时 t_p 本来就定义为 $t_p \equiv \tau_p$，其中 τ_p 是该系过 p 点的那条惯性观者世界线在 p 点的固有时。

§3.5　时空图中的坐标轴

前已讲过，画时空图前应选定基准坐标系，并把该系的 x 轴和 t 轴画成横平竖直。仍以地面系 $\{t,x,y,z\}$ 和火车系 $\{t',x',y',z'\}$ 为例，假定两系有最简关联。选定地面系为基准，则其 x 轴和 t 轴如图 3－15 所示。我们希望再画出 t' 轴和 x' 轴。由于 t' 轴上有 $x'=y'=z'=0$，t' 轴就是火车系的空间坐标原点 O′ 的世界线。

O′点显然相对于地面沿 x 正向匀速运动，故其世界线应是图中标有 t' 的斜直线。至于 x' 轴，许多初学者会不假思索地画成图中的虚直线。然而这是错的！理由如下。既然已知 t,x 轴而待求 t',x' 轴，应该从两系坐标之间的关系出发，这关系由洛伦兹变换

$$t'=\gamma(t-vx),\quad x'=\gamma(x-vt)\qquad(3-5-1)$$

给出。先检查 t' 轴。既然 t' 轴上各点有 $x'=0$，代入式(3－5－1)得 $x-vt=0$，即

$$t=\frac{1}{v}x+0。\qquad(3-5-2)$$

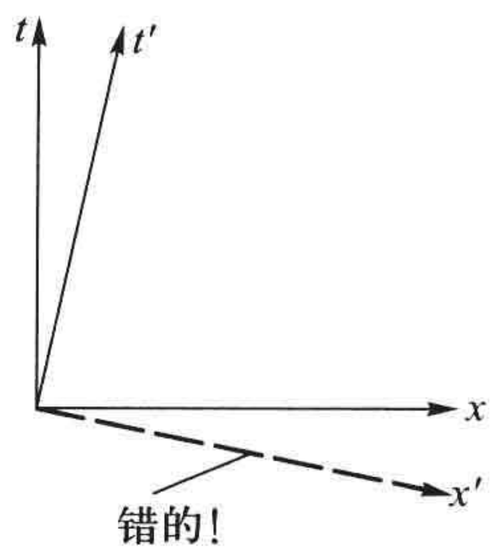


图 3－15　初学者易犯的错误

根据解析几何，在以 x 和 t 为横平竖直的坐标轴的情况下，上式所代表的曲线是一条过原点的直线(“过原点”是因为截距为零)，而且斜率为 $v^{-1}(>1)$。可见图 3－15 的 t' 轴是正确的。另一方面，x' 轴上各点有 $t'=0$，代入式(3－5－1)得

$$t=vx,\qquad(3-5-3)$$

可见 x' 轴是过原点的直线，斜率为 $v(>0)$。然而图 3－15 的“x' 轴”的斜率却为负，所以必定是错的。正确的 x' 轴示于图 3－16 中。由于 t' 轴和 x' 轴的斜率互为倒数，它们对称地分居于图中的点划线(光子世界线)的两侧。常常有人提出这样的问题：t 轴与 x 轴是正交的，而 t' 轴与 x' 轴却不正交，这岂非意味着两个惯性系互不平权？其实这只是时空图的“欺骗”，是人们习惯于用欧氏几何想问题的结果。对时空图应当用闵氏几何想问题。根据闵氏几何，图 3－16 的 t' 轴也正交于 x' 轴，就是说，两轴互相闵氏正交而欧氏非正交。(本书并未给出闵氏正交的定义，读者只能承认此结论。)反之，图 3－15 的 t' 轴与“x' 轴”却是欧氏正交而闵氏非正交的。如果愿意，你也可以选 $\{t',x'\}$ 系为基准，图 3－17 就是所得的时空图，它与图 3－16 貌似不同，实质等价。

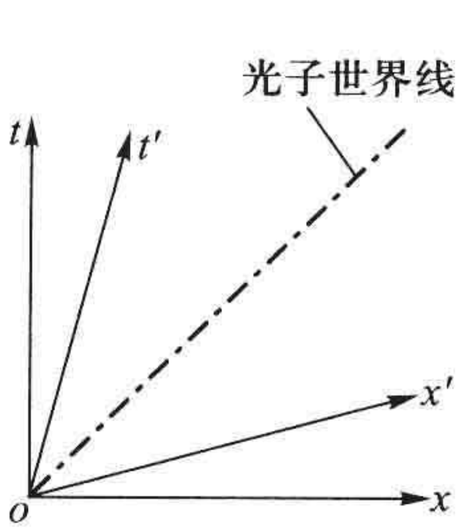


图 3－16　t',x' 轴的正确画法

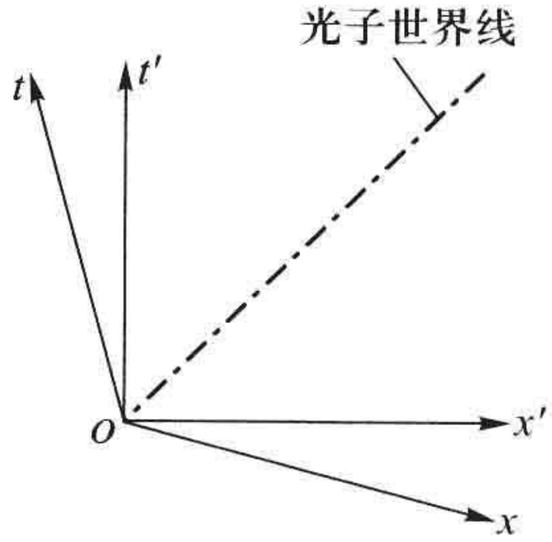


图 3－17　以 $\{t',x'\}$ 系为基准的时空图，等价于图 3－16。

时空图的“欺骗性”不但体现在正交性上，还体现在曲线长度的判断中。设 $p=(t_p,x_p)$ 为任一时空点，op 是连接 o 与 p 的直线段（见图 3－18），其线长按闵氏几何为

$$l_{op}=\sqrt{|-t_p^2+x_p^2|},$$

可见双曲线 $-t^2+x^2=\alpha$（常数）上各点与 o 点所连直线等长，例如 op 与 oq 等长，尽管直观看来（即按欧氏几何看来）不等。图 3－18 中的双曲线称为**校准曲线**，有助于迅速判断两段直线孰长孰短。校准曲线既可以是类时的（当 p 点在第一、三象限），也可以是类空的（当 p 点在第二、四象限）。

图 3－16 表明不同惯性系有不同的同时面族，由此可以清楚地看出同时性的相对性。考虑图 3－19 的事件 p 和 q。由于它们位于 $\{t,x\}$ 系的两张不同的同时面上，所以该系认为事件 p 和 q 不同时。然而，p,q 位于 $\{t',x'\}$ 系的同一张同时面 $\Sigma_{t'_p}$ 上，所以 $\{t',x'\}$ 系认为它们是同时事件。

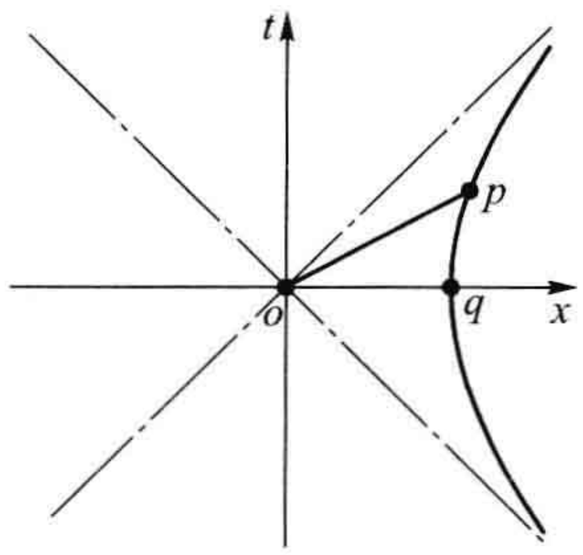


图 3－18　双曲线是校准曲线

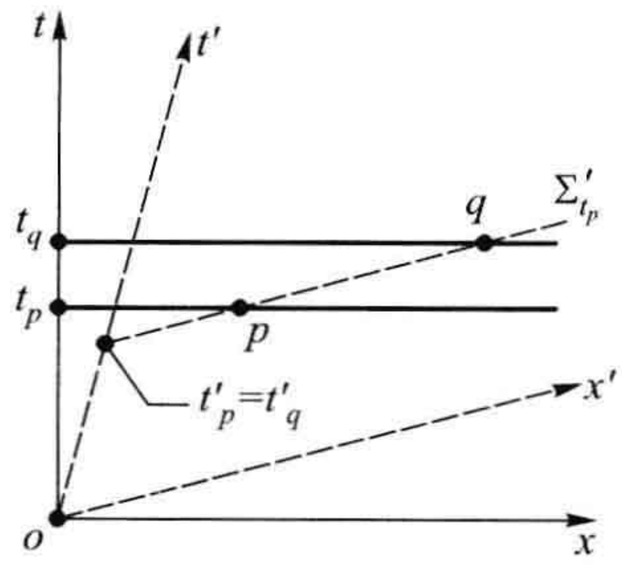


图 3－19　p,q 对 $\{t',x'\}$ 系是同时事件，但对 $\{t,x\}$ 系却不同时。

§3.6　时序和因果关系

以 p_1 和 p_2 分别代表某人的出生和死亡事件，设它们在某惯性系 $\mathrm{K}\equiv\{t,x,y,z\}$ 的时间坐标为 t_1 和 t_2，则显然有 $\Delta t\equiv t_2-t_1>0$（出生先于死亡）。设 p_1 和 p_2 在另一惯性系 $\mathrm{K}'\equiv\{t',x',y',z'\}$ 的时间坐标为 t'_1 和 t'_2。如果牛顿力学适用，由时间的绝对性可知 $t'_1=t_1,t'_2=t_2$，故 $\Delta t'\equiv t'_2-t'_1=\Delta t>0$，仍是出生先于死亡。但是，在狭义相对论中时间是相对的，$\Delta t'$ 与 Δt 一般不等。如果竟然出现 $\Delta t'<0$ 的情况，岂不是 K′系认为死亡先于出生？这种违反因果律的结论是任何人都不能接受的，因此很有必要把这个问题弄个水落石出。

先讨论 2 维闵氏时空这一简单情况，然后再推广到 4 维闵氏时空。设 p_1 和 p_2 是任意两个事件，它们在惯性系 K 和 K′的坐标分别为

$$p_1=(t_1,x_1)=(t_1',x_1'),\quad p_2=(t_2,x_2)=(t_2',x_2'),\qquad(3-6-1)$$

令 $\Delta t\equiv t_2-t_1,\Delta x\equiv x_2-x_1$。对 2 维时空,两惯性系 K 和 K′之间的坐标变换只能是如下最简单的洛伦兹变换:

$$t'=\gamma(t-vx),\quad x'=\gamma(x-vt)。\qquad(3-6-2)$$

过去一直约定 $v>0$,现在放宽为 v 可正可负(只需 $|v|<1$),$v<0$ 代表 K′系沿 K 系 x 轴的负向运动。由式(3-6-2)得

$$\Delta t'=\gamma(\Delta t-v\Delta x)。\qquad(3-6-3)$$

不失一般性,设 $\Delta t>0$,这意味着 K 系认为事件 p_1 先于 p_2 发生。如果 $\Delta t'>0$,就说洛伦兹变换(3-6-2)是**保时序的**;如果 $\Delta t'<0$,就说变换(3-6-2)是**改时序的**。

时序能否保持,关键在于两事件的时空间隔。事件 p_1 和 p_2 的**时空间隔** I_{12} 定义为

$$I_{12}\equiv-\Delta t^2+\Delta x^2。\quad(\text{易证它在洛伦兹变换下形式不变})(3-6-4)$$

间隔分为三类:$I_{12}>0$ 的间隔叫**类空的**;$I_{12}<0$ 的间隔叫**类时的**;$I_{12}=0$ 的间隔叫**类光的**。间隔的类空、类时和类光性都是绝对的,不随惯性系而变。下面是本节的关键性定理。

定理 3-1

(A) 若两事件的间隔为类时或类光($I_{12}\leqslant0$),则任何洛伦兹变换都保时序;

(B) 若两事件的间隔为类空($I_{12}>0$),则必有改时序的洛伦兹变换。

证明　在式(3-6-3)涉及的各量 $\Delta t',\Delta t,\Delta x,v$ 中,除约定 $\Delta t>0$ 外,其他都可取正、负或零值。

(A) 由式(3-6-4)可知,$I_{12}\leqslant0$ 保证

$$|\Delta x|\leqslant\Delta t。\qquad(3-6-5)$$

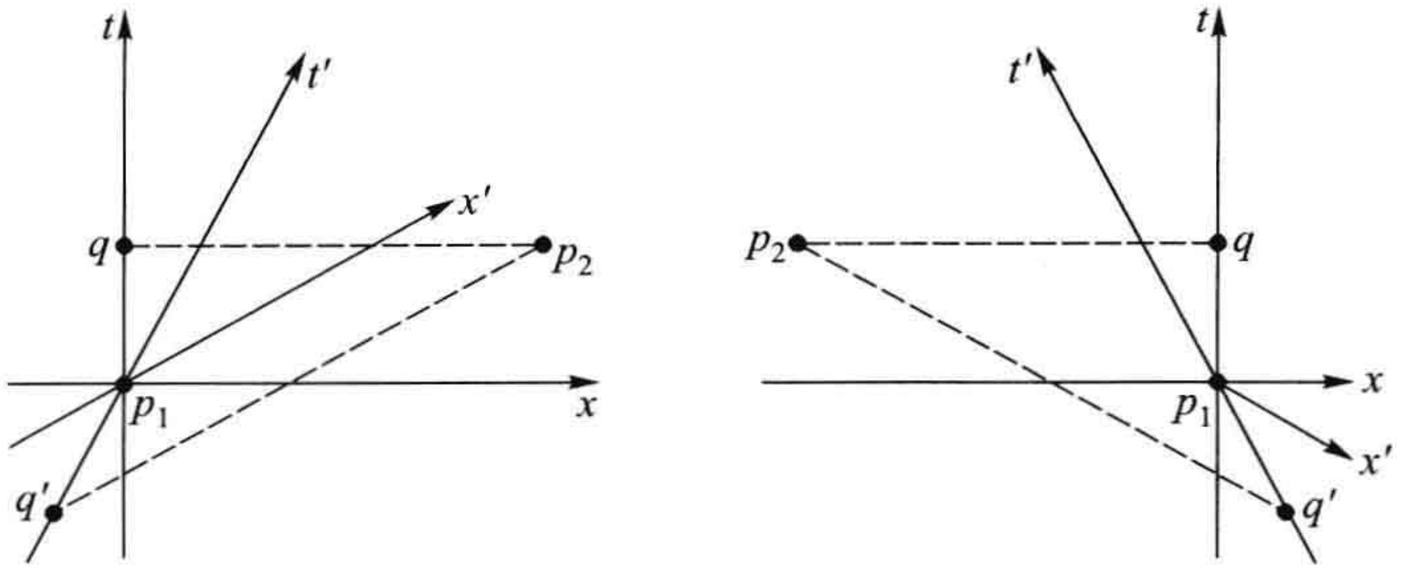
注意到 $|v|<1$,又有 $|\Delta x|>|v||\Delta x|\geqslant v\Delta x$,与式(3-6-5)结合得 $\Delta t>v\Delta x$,代入式(3-6-3)便得 $\Delta t'>0$(保时序)。可见,当 $I_{12}\leqslant0$ 时任何洛伦兹变换都保时序。

(B) 由式(3-6-4)可知

$$I_{12}>0\Leftrightarrow\Delta t<|\Delta x|\Leftrightarrow\Delta t/|\Delta x|<1。\qquad(3-6-6)$$

现在说明,在 $I_{12}>0$ 的情况下,无论 Δx 为正为负,必有改时序的洛伦兹变换。

若 $\Delta x>0$,则 $I_{12}>0$ 保证 $\Delta t/\Delta x<1$[式(3-6-6)]。只要取 $v>\Delta t/\Delta x$,便有 $v\Delta x/\Delta t>1$,由式(3-6-3)便得 $\Delta t'<0$,以这个 v 为参数的洛伦兹变换便改时序。图 3-20(a)是这种变换的时空图,由图可清晰地看到 $\Delta t>0$ 而 $\Delta t'<0$。(不失一般性,图中以 p_1 作为惯性系$\{t,x\}$的坐标原点。)



(a) $v>0$ 的情况　　(b) $v<0$ 的情况

图 3 – 20　改时序的洛伦兹变换。对 $\{t,x\}$ 系，　p_2 与 q 同时，晚于 p_1；对 $\{t',x'\}$ 系，　p_2 与 q'同时，早于 p_1。

若 $\Delta x<0$，则 $I_{12}>0$ 保证 $\Delta t/|\Delta x|<1$[式(3 – 6 – 6)]。只要取 $v<0$ 而且 $|v|>\Delta t/|\Delta x|$，便可使 $\Delta t<|v||\Delta x|=v\Delta x$，再由式(3 – 6 – 3)便得 $\Delta t'<0$，这个洛伦兹变换便改时序。图 3 – 20(b)是这种变换的时空图，由图可见 $\Delta t>0$ 而 $\Delta t'<0$。　[证毕]

下面揭示定理 3 – 1 的物理意义。以 p_1 为原点建惯性系 $\{t,x\}$(图 3 – 21)。当 $I_{12}<0$ 时 $\Delta t/|\Delta x|>1$，由图 3 – 21 可知直线段 p_1p_2 是类时线段(因而 p_2 位于 p_1 的未来光锥面以内)，我们说 p_2 与 p_1 有**类时联系**。既然如此，原则上就可以认为 p_1 和 p_2 分别是某个信号的发出和接收事件。例如，p_1 代表某人的出生，p_2 代表他的死亡；又如，p_1 代表子弹从手枪射出，p_2 代表子弹中靶。总之，有类时联系的两个事件 p_1 和 p_2(其 $I_{12}<0$)原则上可以有因果联系，如果竟然存在一个洛伦兹变换使得新系认为 p_2 先于 p_1 发生(改变时序)，就会因果倒置，这是按任何标准衡量都不允许的。定理 3 – 1(A)的重要性就是保证这种改时序的洛伦兹变换不会存在。此外，定理 3 – 1(A)的 $I_{12}\leqslant 0$ 还包含擦边情况 $I_{12}=0$，这时 p_2 位于 p_1 的未来光锥面上，两者也可有因果联系(例如 p_1 和 p_2 分别代表光子枪的发射和中靶事件)，定理3 – 1(A)也保证不存在改时序的洛伦兹变换。

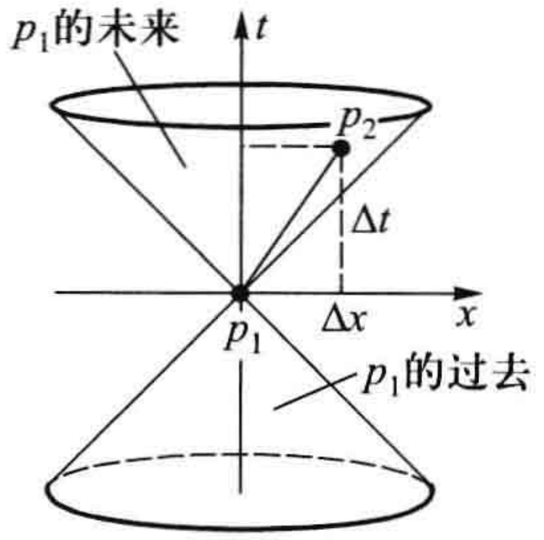


图 3 – 21　事件 p_1，p_2 有类时联系

反之，定理 3 – 1(B)则表明，只要 p_1 和 p_2 有 $I_{12}>0$，就总可找到改时序的洛伦兹变换。假若这两个事件竟然也有因果联系，在新的惯性系中岂非会因果倒置？幸好，任何两个满足 $I_{12}>0$ 的事件一定不可能有因果联系，证明如下。

设事件 p_1 和 p_2 满足 $I_{12}>0$，则 $(\Delta t)^2<(\Delta x)^2$。不失一般性，可设 $x_2>x_1$，即

$\Delta x>0$,因而 $\Delta t<\Delta x$(如图 3 - 22)。以 L 代表连接 p_1 和 p_2 的任一曲线,则线上任一点的 t 便是 x 的函数,记作 $f(x)$,定义域为 $[x_1,x_2]$。根据微积分学的中值定理,在 x_1 和 x_2 之间必定存在一个 $\bar{x}$,满足

$$\left.\frac{\mathrm{d}f(x)}{\mathrm{d}x}\right|_{x=\bar{x}}=\frac{f(x_2)-f(x_1)}{x_2-x_1}=\frac{\Delta t}{\Delta x}。$$

由 $I_{12}>0$ 知 $\Delta t/\Delta x<1$,故 $\left.[\mathrm{d}f(x)/\mathrm{d}x]\right|_{x=\bar{x}}<1$,表明曲线 L 在点 $f(\bar{x})$ 的斜率小于 1,因而 L 不是类时或类光曲线。可见,假若竟有一个信号从事件 p_1 传到 p_2,在传播过程中至少有某些时候会超光速,而这样的信号在相对论中不存在。这就证明了 p_1 和 p_2 不存在任何因果联系。虽然确有洛伦兹变换改变它们的时序,但这种改变无关紧要,不会导致任何因果疑难。

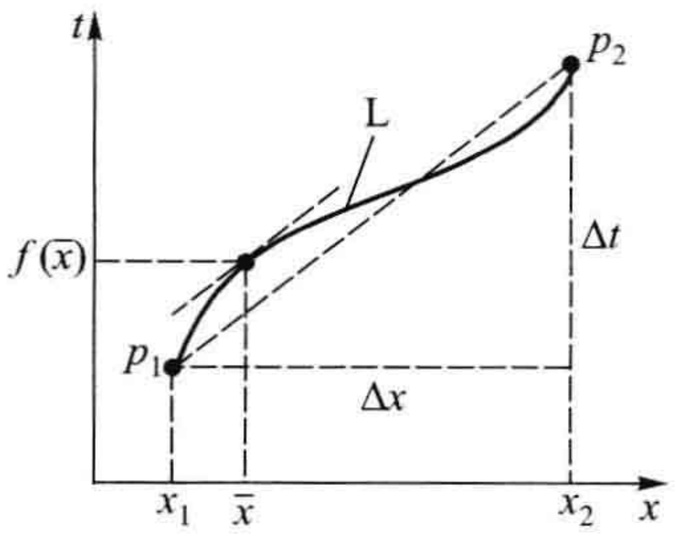


图 3 - 22　用中值定理证明曲线为类空

刚才只就 2 维闵氏时空做了讨论,现在将结论推广至 4 维闵氏时空。设事件 p_1 和 p_2 在某惯性系 $\mathrm{K}\equiv\{t,x,y,z\}$ 的坐标为

$$p_1=(t_1,x_1,y_1,z_1),\quad p_2=(t_2,x_2,y_2,z_2),$$

它们的**时空间隔** I_{12} 定义为

$$I_{12}\equiv-\Delta t^2+\Delta x^2+\Delta y^2+\Delta z^2,\qquad(3-6-7)$$

其中 $\Delta t\equiv t_2-t_1,\Delta x\equiv x_2-x_1,\Delta y\equiv y_2-y_1,\Delta z\equiv z_2-z_1$。间隔类空、类时和类光性的定义与 2 维时空相同。由 2 维的讨论可知,现在的要害问题是要证明如下定理。

定理 3 - 2　若对 p_1 和 p_2 有 $I_{12}\leqslant 0$,而且在某惯性坐标系 $\mathrm{K}\equiv\{t,x,y,z\}$ 中 $\Delta t>0$,则在任何惯性坐标系都有 $\Delta t'>0$(这意味着有类时或类光联系的两事件在任何惯性系中都有相同时序,因而不会导致因果疑难)。

注记 7　表面看来这一定理的证明与定理 3 - 1(A)的证明很像,只需把 2 维洛伦兹变换(3 - 6 - 2)改成 4 维洛伦兹变换:

$$t'=\gamma(t-vx),\quad x'=\gamma(x-vt),\quad y'=y,\quad z'=z。\qquad(3-6-8)$$

然而问题并非如此简单。关键在于上式只是两个有最简关联(见图 1 - 1 及其所在段)的惯性系之间的坐标变换,而我们要证明的是在任何惯性坐标系中都有 $\Delta t'>0$,其中许多系与 K 系之间都不是最简关联。由于式(3 - 6 - 8)在狭义相对论中被频繁使用,久而久之,就会误以为这是一个“包打天下”的变换而忘记有时还会遇到更复杂的洛伦兹变换。现在,为了证明定理 2,恰恰必须涉及这种复杂变换。这是 4 维闵氏时空中任意两个惯性系 K 和 K′之间的坐标变换,两系之间的关系比最简关联要复杂,由以下三点表征:① 两系空间坐标原点在 $t'=$

$t=0$ 时相遇；② $\{x',y',z'\}$ 相对于 $\{x,y,z\}$ 系以任意速度 $\boldsymbol{v}=v_1\boldsymbol{e}_1+v_2\boldsymbol{e}_2+v_3\boldsymbol{e}_3$ 做匀速平动，其中 $\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_3$ 是沿 K 系 x,y,z 轴正向的单位矢；③ 两系空间坐标轴对应同向。这两个系的坐标变换由如下的矩阵等式表出[证明见梁灿彬，周彬(2009)中册小节 G.9.1]：

$$\begin{bmatrix} t' \\ x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \gamma & -\gamma v_1 & -\gamma v_2 & -\gamma v_3 \\ -\gamma v_1 & 1+\dfrac{(\gamma-1)v_1^2}{v^2} & \dfrac{(\gamma-1)v_1v_2}{v^2} & \dfrac{(\gamma-1)v_1v_3}{v^2} \\ -\gamma v_2 & \dfrac{(\gamma-1)v_2v_1}{v^2} & 1+\dfrac{(\gamma-1)v_2^2}{v^2} & \dfrac{(\gamma-1)v_2v_3}{v^2} \\ -\gamma v_3 & \dfrac{(\gamma-1)v_3v_1}{v^2} & \dfrac{(\gamma-1)v_3v_2}{v^2} & 1+\dfrac{(\gamma-1)v_3^2}{v^2} \end{bmatrix} \begin{bmatrix} t \\ x \\ y \\ z \end{bmatrix}, \tag{3-6-9}$$

其中

$$v\equiv\sqrt{v_1^2+v_2^2+v_3^2}<1,\quad \gamma\equiv 1/\sqrt{1-v^2}。 \tag{3-6-10}$$

式(3-6-9)称为**最一般的洛伦兹变换**。为了证明定理2，必须且只需证明式(3-6-9)的变换不会导致 Δt 变号。

定理3-2的证明　由式(3-6-9)得 $t'=\gamma(t-v_1x-v_2y-v_3z)$，故

$$\Delta t'=\gamma(\Delta t-v_1\Delta x-v_2\Delta y-v_3\Delta z)。 \tag{3-6-11}$$

注意到 $\boldsymbol{v}=v_1\boldsymbol{e}_1+v_2\boldsymbol{e}_2+v_3\boldsymbol{e}_3$ 是3维矢量，再定义3维矢量 $\Delta\boldsymbol{X}\equiv\boldsymbol{e}_1\Delta x+\boldsymbol{e}_2\Delta y+\boldsymbol{e}_3\Delta z$，就可把式(3-6-11)表为

$$\Delta t'=\gamma(\Delta t-\boldsymbol{v}\cdot\Delta\boldsymbol{X})。 \tag{3-6-11'}$$

以 α 代表矢量 $\boldsymbol{v}$ 与 $\Delta\boldsymbol{X}$ 的夹角，$|\boldsymbol{v}|$ 和 $|\Delta\boldsymbol{X}|$ 代表两矢量的长度[其中 $|\boldsymbol{v}|$ 即式(3-6-10)的 v，故 $0<v<1$]，则

$$\boldsymbol{v}\cdot\Delta\boldsymbol{X}=|\boldsymbol{v}||\Delta\boldsymbol{X}|\cos\alpha\leqslant|\boldsymbol{v}||\Delta\boldsymbol{X}|<|\Delta\boldsymbol{X}|。 \tag{3-6-12}$$

由式(3-6-7)又知 $I_{12}\leqslant 0$ 保证 $|\Delta\boldsymbol{X}|\leqslant\Delta t$，代入式(3-6-12)给出 $\boldsymbol{v}\cdot\Delta\boldsymbol{X}<\Delta t$，再代入式(3-6-11')便知 $\Delta t'>0$。　[证毕]

[选读3-3]

正文把式(3-6-9)称为“最一般的洛伦兹变换”其实不够准确，本选读对此加以准确化。感到不好理解的读者不读也罢。

洛伦兹变换包含两种类型，仍以地面系和火车系为例说明。设 $\mathrm{K}\equiv\{t,x,y,z\}$ 是地面参考系内的一个惯性坐标系，让3个空间坐标轴(统一地)做一个空间转动就得到一个新坐标系 $\tilde{\mathrm{K}}\equiv\{t,\tilde{x},\tilde{y},\tilde{z}\}$，同一事件 p 既可表为 (t,x,y,z) 又可表

为$(t,\tilde{x},\tilde{y},\tilde{z})$。从K系到$\tilde{\mathrm{K}}$系的变换是地面参考系内的坐标系变换(所以时间坐标$t$不变),称为**转动**,是洛伦兹变换的第一种类型。我们熟悉的洛伦兹变换式(3-6-8)则属于洛伦兹变换的第二种类型,称为**伪转动**,是从地面参考系内的惯性坐标系$\mathrm{K}\equiv\{t,x,y,z\}$到火车参考系内的惯性坐标系$\mathrm{K}'\equiv\{t',x',y',z'\}$的变换。由于约定K和K′系有最简关联,所以式(3-6-8)最为简单,因而又称为**最简单的伪转动**。如果让K′系的3个空间坐标轴(统一地)做一个空间转动,而且这个转动与从K系到$\tilde{\mathrm{K}}$系的转动一样,所得到的新系$\tilde{\mathrm{K}}'\equiv\{t',\tilde{x}',\tilde{y}',\tilde{z}'\}$与$\tilde{\mathrm{K}}$系的关系就比最简关联要复杂,两系的坐标变换就是式(3-6-9),称为**最一般的伪转动**,其特点就是式(3-6-8)后面一段的①、②、③点。不应将它称为最一般的洛伦兹变换,因为转动也是洛伦兹变换。转动与伪转动的关键区别是:转动不改变参考系,伪转动则必然涉及参考系的变换。最后给出洛伦兹变换的明确定义:两个惯性坐标系之间的、保证空间坐标原点在$t'=t=0$时重合的坐标变换称为**洛伦兹变换**,它既可以是转动,又可以是伪转动,还可以是两者的叠加(相继作用)。所有洛伦兹变换的集合构成群(洛伦兹群)。这方面的详细讨论可在梁灿彬、周彬(2009)中册§G.9中找到。　**[选读3-3完]**

§3.7　何谓几何语言？何以要用几何语言？

在变化纷繁的研究对象中抓住不变的东西是相对论学家从数学家借用过来的高招。要学好相对论,最关键的(偏偏又是最容易忽视的)一点就是要善于区分绝对的和相对的对象(包括量、概念及其他),并且要紧抓不放。如果一个对象与某种人为选择的因素有关,该对象就称为**相对的**(relative),否则称为**绝对的**(absolute)。下面举例说明。

例1　事件本身是绝对的,但事件的坐标是相对的,因为对于同一事件,不同坐标系会给出不同的时空坐标,即$p=(t,x,y,z)=(t',x',y',z')$,但通常$t'\neq t,x'\neq x$,等等。

例2　时空,作为所有事件的集合,是绝对的,但"每一时刻的全空间"(同时面)的概念则是相对的,因为不同惯性系有不同的同时面族。

一个惯性系有一个同时面族(不妨比喻为一副扑克牌,见图3-6),每张同时面代表一个时刻的全空间,从时空图底部逐渐向上看,就会看到运动的全过程[仍可用图3-4(b)为例],这个"逐渐向上看"就代表时间的流逝,所以说时空被这个惯性系做了"3+1分解"。如果改用另一惯性系,其同时面族就是"另一副扑克牌"(参见图3-19),所以说这是对同一时空的另一种"3+1分解"。可见,时空是绝对的,但分解而得的空间和时间都是相对的。

例3　给定几何(例如给定闵氏几何)后,曲线元段 dL 的线元 ds^2 本身是绝对的,但线元表达式却是相对的,例如闵氏线元在惯性坐标系 $\{t,x,y,z\}$ 的表达式为

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2, \tag{3-7-1}$$

而在4维球坐标系 $\{t,r,\theta,\varphi\}$ 的表达式则为

$$ds^2 = -dt^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2)。 \tag{3-7-2}$$

如果愿意(有时也很必要),你还可以选择更为奇特的坐标系使闵氏线元 ds^2 的表达式复杂到无人可以一眼看穿的程度。例如,下式虽然不算太复杂:

$$ds^2 = -[1+g(t)x]^2dt^2 + dx^2 + dy^2 + dz^2, \quad \text{其中 } g(t) \text{是 } t \text{ 的任意函数,} \tag{3-7-3}$$

你能看出它其实是闵氏线元吗?为了给出证明,只需找到一个能将上式变为式(3-7-1)的坐标变换。以 T,X,Y,Z 代表新坐标,则这一坐标变换是

$$T = \int[\mathrm{ch}G(t)]dt + x\mathrm{sh}G(t),\ X = \int[\mathrm{sh}G(t)]dt + x\mathrm{ch}G(t),$$

$$Y = y,\ Z = z,\ \text{其中 } G(t) \equiv \int g(t)dt。 \tag{3-7-4}$$

作为习题,请读者自行验证式(3-7-3)的线元在坐标系 $\{T,X,Y,Z\}$ 的确取如下形式:

$$ds^2 = -dT^2 + dX^2 + dY^2 + dZ^2。$$

上式说明 $\{T,X,Y,Z\}$ 是惯性坐标系,而式(3-7-3)无非是闵氏线元在非惯性坐标系 $\{t,x,y,z\}$ 的表达式。一个常见误解是一见式(3-7-3)就以为是弯曲几何。产生误解的根源就是不会透过坐标表象看清几何实质,好比抓住"鸡毛",丢掉"令箭"。正是:"不识几何真面目,只缘心在坐标中。"所以,我们应该提高警惕,在见到复杂的线元表达式时切莫"只重衣冠不认人",其实它可能只是同一个"人"(闵氏线元)穿了一身奇特的"衣服"。

例4　世界线,作为一连串事件的集合,是绝对的。但画在时空图中的"世界线"却是相对的,因为其形状还取决于画图时所选择的基准坐标系。应该学会把绝对的世界线同画在时空图中的"世界线"区分开。仍以图3-5的火车头A′为例。A′的"世界线"在以地面系为基准的时空图中是斜直线;但在以火车系为基准的时空图中则是竖直线。作为有益的练习,读者不妨把图3-5(b)改画成以火车系为基准的时空图。进一步说,如果故意选一个不是惯性系的、更为复杂的坐标系为基准画时空图,则A′的"世界线"甚至可以是一条奇形怪状的曲线。下面的比喻有助于理解:录制电视剧时,对同一场戏通常要用几台摄像机从不同角度拍摄。不妨说那场戏是绝对的(只有这么一场戏),但录像带记录的画

面则是相对的(依赖于拍摄角度)。

以上例子说明,同一个绝对对象在不同参考系(坐标系)中会成为不同的相对对象。因此,在注意这些相对对象的区别的同时,还应特别注意它们所描述的是同一绝对对象("万变不离其宗")。如果只看见对象在坐标变换时变来变去,却抓不住万变所不离的"宗",就有可能"找不着北",自乱阵脚,造成各种迷惑和误解(例子不胜枚举)。

上述讨论说明坐标语言有时会遮盖研究对象的绝对性质,在平面几何学中也有这样的例子。设 x,y 是 2 维欧氏空间的直角坐标,则方程

$$x^2 + y^2 = R^2\,(\text{常数})$$

代表圆。但通过坐标变换可把上式变得使你认不出它是圆。要突出圆的绝对性质可改用不借助坐标系的语言:圆是与某定点距离相等的动点的轨迹。不妨把这种语言称为(纯)几何语言。

相对论中的 3 维看法总是借助于某参考系(或坐标系)看 4 维时空的产物(是借某系对时空做 3 + 1 分解的结果),天生就是相对的,所以直接涉及绝对对象的语言只能是 4 维语言(时空语言),通常就称之为相对论的**几何语言**。用时空图讨论问题是几何语言的重要方面,此外,几何语言也包括用 4 维矢量和张量讨论问题。限于篇幅,本书不拟涉及 4 维矢量和张量,书中的几何语言主要体现在:① 从观念上彻底分清绝对和相对的事物;② 尽量用时空图讨论问题。我们虽然强调几何语言的诸多优越性,但毫不否认坐标语言具有便于表达和计算等等重要好处。

相对论几何语言的创始人是大数学家**闵可夫斯基**(Minkowski),他曾是爱因斯坦在苏黎世理工学院就读时的数学老师。由于爱因斯坦经常逃课(以便有更多时间独自学习和思考),他对爱因斯坦的总体印象不佳,甚至曾称他为"懒骨头"。但后来爱因斯坦的相对论引起了闵可夫斯基的强烈兴趣,他敏锐地觉察到,只有把空间和时间有机结合而成的 4 维时空[他称之为**世界**(world)]才是有实质性意义的东西,并且用 4 维几何语言对相对论做了十分优雅的改写。他在 1908 年的一次"通俗讲话"中介绍了这一革命性的思想①。这一讲话其实非常不通俗,因为它基本上是此前他的一篇专业性很强的文章②的重述。该文一开头就说:"***现在我要向你们提出的时空观是在实验物理学的土壤上产生的,其力***

① 第 28 届德国自然科学家和医生大会于 1908 年在科隆举行,闵可夫斯基在会上发表了这一讲话。

② 该文被收入一本论文集中,其英译本出版于 1923 年,作者为 Lorentz, Einstein, Minkowski 和 Weyl,后来又被译成汉语并出版,书名为《相对论原理》,A. 爱因斯坦等著,赵志田、刘一贯译,孟昭英校,科学出版社出版(1980)。

量就在这里。这些观点是根本性的。从现在起，孤立的空间和孤立的时间注定要消失成为影子，只有两者的统一才能保持独立的存在。”可惜当时的爱因斯坦对这一重大贡献不感兴趣，他还语带嘲讽地说过：“重要的是内容而不是数学，数学对什么都能证明。”[见 Brain(1996)P.76]他甚至把闵可夫斯基的发现称为“多余的学问(superfluous erudition)”[见 Cropper(2001)P.220]。爱因斯坦后来还说过，由于数学家使用如此复杂的语言描述相对论，“我就再也不懂相对论了”[见 Clark(1971)P.159]。然而，此一时彼一时，几年后的爱因斯坦终于认识到闵可夫斯基贡献的巨大意义，并且承认，如果没有4维语言，“相对论仍将处于襁褓之中”[见 Cropper(2001)P.220]。特别是，为了把引力也纳入相对论中(为了建立广义相对论)，爱因斯坦从1912年起就认识到闵可夫斯基的4维绝对时空是极其必要、无法替代的(广义相对论的背景时空是弯曲时空，详见第6章)。他在1916年发表的关于广义相对论的文章中写道：“闵可夫斯基为狭义相对论给出的4维表述大大促进了相对论的推广，他是清晰地认识到空间和时间坐标等价性的第一位数学家；这种4维表述使广义相对论的构建成为可能。”遗憾的是，闵可夫斯基没能活着听到爱因斯坦对于自己的贡献的肯定，他英年早逝，45岁(1909年)时卒于阑尾炎。

第3章习题

1. 用图3-12的雷达法还可测定观者A、B之间的空间距离D。设A钟的读数在事件 a_1 和 a_2 分别为 τ_1 和 τ_2，试用这两个数表出D。

2. 改用火车系为基准系重画“爱因斯坦列车”雷击问题的时空图，并由此得出与用图3-5(b)所得结论相同的结论。

3. 你能认出 $ds^2 = -x^2dt^2 + dx^2 + dy^2 + dz^2$ 是闵氏线元吗？试用坐标变换

$$T = x\mathrm{sh}t,\ X = x\mathrm{ch}t,\ Y = y,\ Z = z,$$

把它变为 $ds^2 = -dT^2 + dX^2 + dY^2 + dZ^2$ 的形式。

4. 试证 $ds^2 = -t^{-4}dt^2 + dx^2 + dy^2 + dz^2$ 是闵氏线元。提示：只需寻找坐标系 $\{T,X,Y,Z\}$ 把它变为 $ds^2 = -dT^2 + dX^2 + dY^2 + dZ^2$ 的形式。

5. 用坐标变换式(3-7-4)把复杂线元式(3-7-3)变成 $ds^2 = -dT^2 + dX^2 + dY^2 + dZ^2$ 的形式。

第4章　趣味运动学效应

§4.1　尺缩效应和车库佯谬

稍微接触过狭义相对论的人都听说过尺缩、钟慢等饶有兴味的效应。当你进一步听到“双生子悖论”和“车库悖论”时，除了觉得更加有趣之外，还会平添许多疑惑和不解。其实，只要思维清晰，利用洛伦兹变换就可把这些效应解释清楚。不过，对于某些问题（特别是“双子悖论”），在听完用洛伦兹变换的分析后往往仍感到若明若暗，甚至产生这样那样的误解。反之，本书推荐的4维几何语言（也可称为时空语言）可以对这些效应给出让你叹为观止的清晰剖析。这一方法至少有以下优点：① 简单易学性（前提是你基本掌握第2、3章的内容）；② 深刻清澈性（透过时空图可以一眼看穿问题实质）；③ 误解免疫性。（不但不易产生误解，还可澄清过去的诸多误解。）下面就用此法逐一剖析四个效应。

4.1.1　尺缩效应

此效应可用一句话描述：“当尺子沿尺长方向相对于你运动时，你会测得它变短。”简言之就是“动尺缩短”甚至“尺缩”①。但是“缩”字容易误导，有人据此猜测：由于运动，尺子内部的弹性机制可能会受到影响，从而导致尺子收缩。由下面的几何剖析就可看出这是严重的误解。

对熟悉4维语言的人来说，尺子本来就不是1维的，它是个2维对象（尺子的世界面）。这是个绝对的对象，与参考系无关。既然是2维对象，“尺子有多长？”就是个意义不明的问题。只有选定参考系后尺长才有意义。设选定惯性系K，就有一系列的同时面，任选一张，例如 Σ_0（即 $t=0$ 的同时面，见图4－1），

---

① 又称为**洛伦兹－斐兹杰惹收缩**。迈克耳孙和莫雷的1887年实验的目的是证明以太存在性，结果却是迹象全无。为了说明这一“零结果”与以太的存在性并无矛盾，**斐兹杰惹**（Fitzgerald）于1892年大胆提出一个假设（后来又被洛伦兹强化）：所有物体在相对于以太系运动的方向上的长度都要缩短，缩短因子为 $\sqrt{1-v^2/c^2}$。但这一假设后来得不到实验支持，只有用爱因斯坦的相对论方可正确解释尺缩效应。

则它代表 K 系认为的、在 $t=0$ 时刻的全空间，故 Σ_0 与尺子世界面的相交线段 oa 就代表 K 系认为的、$t=0$ 时刻的（1 维）尺子，于是线长 l_{oa} 就是 K 系在该时刻测得的尺长（静尺长）。然而，由于同时性的相对性，K′系在 $t'=0$ 时刻的同时面 Σ'_0 与 Σ_0 不重合（仍见图 4－1），它与尺子世界面的相交线段 ob 不同于 oa，而 ob 段代表 K′系认为的、$t'=0$ 时刻的尺子，所以 l_{ob} 才是 K′系测得的尺长。由于 K′系认为尺子在动，所以 l_{ob} 是动尺长。ob 和 oa 本来就是两个不同的直线段，线长不同又何足大惊小怪？但是，不懂几何语言的人总觉得“尺子明明只有一把”，既然同一把尺子有两种长度，想必是“动尺的内部机制起了变化，导致尺缩。”殊不知从 4 维角度看来尺子（世界面）根本就没有变化，更没有收缩可言。尺长不同的本质原因是：虽然尺子只有一把（是绝对的），但不同惯性系有不同的同时面导致不同惯性系测到不同的 1 维尺子（1 维尺是相对的），而不同的 1 维尺有不同尺长当然不足为怪。尺缩效应不过类似于“盲人摸象”而已。

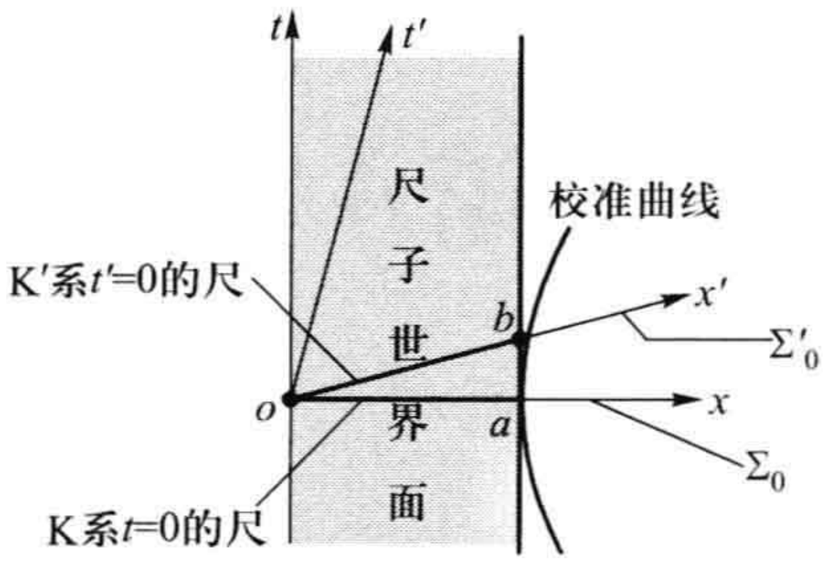


图 4－1　尺子世界面是绝对的。oa 和 ob 分别是 K 系和 K′系在 $t=0$ 和 $t'=0$ 时的 1 维尺。

既然 l_{oa} 和 l_{ob} 分别代表静、动尺长，所余问题无非是比较 l_{oa} 和 l_{ob}。由于直角三角形斜边最长，似乎 $l_{ob}>l_{oa}$，岂非动尺较长？然而这也是时空图的“欺骗”（“斜边最长”只是欧氏几何的结论）。过 a 作校准曲线（见 §3.5）便知 $l_{ob}<l_{oa}$，可见动尺较短。欲求两者之间的定量关系，只需计算两段线长。线长是绝对量，计算结果同所选坐标系无关。为便于比较，我们用同一坐标系（K 系）计算。注意到 o 点在该系的坐标为（0,0,0,0），由线长公式（2－2－1）及闵氏线元在该系的表达式（2－1－8）得

$$l_{oa}=\int_o^a\sqrt{\mathrm{d}s^2}=\int_o^a\sqrt{\mathrm{d}x^2-\mathrm{d}t^2}=\int_o^a\sqrt{\mathrm{d}x^2-0}=\int_o^a\mathrm{d}x=x_a,\qquad(4-1-1)$$

$$l_{ob}=\int_o^b\sqrt{\mathrm{d}s^2}=\int_o^b\sqrt{\mathrm{d}x^2-\mathrm{d}t^2}\ 。$$

由式（3－5－3）又知 x' 轴满足 $t=vx$，故

$$l_{ob}=\int_o^b\sqrt{\mathrm{d}x^2-(v\mathrm{d}t)^2}=\sqrt{1-v^2}\int_o^b\mathrm{d}x=\gamma^{-1}x_b=\gamma^{-1}x_a。\qquad(4-1-2)$$

与式（4－1－1）对比便得动、静尺长的定量关系：

$$l_{ob}=\gamma^{-1}l_{oa},\qquad \text{即}\qquad \text{动尺长}=\text{静尺长}/\gamma。\qquad (4-1-3)$$

4.1.2　车库佯谬

设汽车与车库静长相等。汽车匀速进库时，司机想：“动库变短，车装不下。”司库想：“动车收缩，装下有余。”司机的想法对吗？司库的想法对吗？利用时空图可以获得清晰的认识。不失一般性，设汽车沿地面系的 x 轴正向行驶。为了由简入繁，我们分两种情况讨论。

（情况 1）　车库并无后墙，其“后墙”只是一条画在地上的直线。

以地面系为基准画时空图（图 4－2）。库门和库“墙”自然是两条竖直线。车头的世界线是一条斜率大于 1 的斜直线，与库“墙”世界线的交点记作 b，过 b 的水平直线便是车库系的一条同时线，故 l_{ab} 就是车库静长。车尾世界线应与车头世界线平行，但这样的直线很多，选择时应体现已知条件——车与库静长相等。设 a 是过 b 的水平线与库门世界线的交点，过 a 作如图所示的校准曲线（双曲线），它与汽车系过 b 的同时线交于 d，再过 d 作与车头世界线平行的直线便是车尾的世界线（这样作出的车尾世界线确能保证车与库有相等静长）。时空图一旦画出，所余问题就迎刃而解。司库认为 a,b 点所在的同时线代表车头抵墙这一时刻（t_b 时刻）的全空间，直线段 ab 和 eb 分别代表此时刻的库和车，由图可见 $l_{ab}>l_{eb}$，所以汽车装下有余。司机则认为 b,d 所在的同时线代表车头抵墙这一时刻（t'_b 时刻）的全空间，直线段 db 和 fb 分别代表此时刻的车和库，由图可见 $l_{db}>l_{fb}$，所以库比车短，车装不下。可见两人看法都对，关键是同时性的相对性导致结论的相对性。不许提出这样的问题：“到底装下还是装不下？”结论的相对性使这种绝对式的问题没有意义。

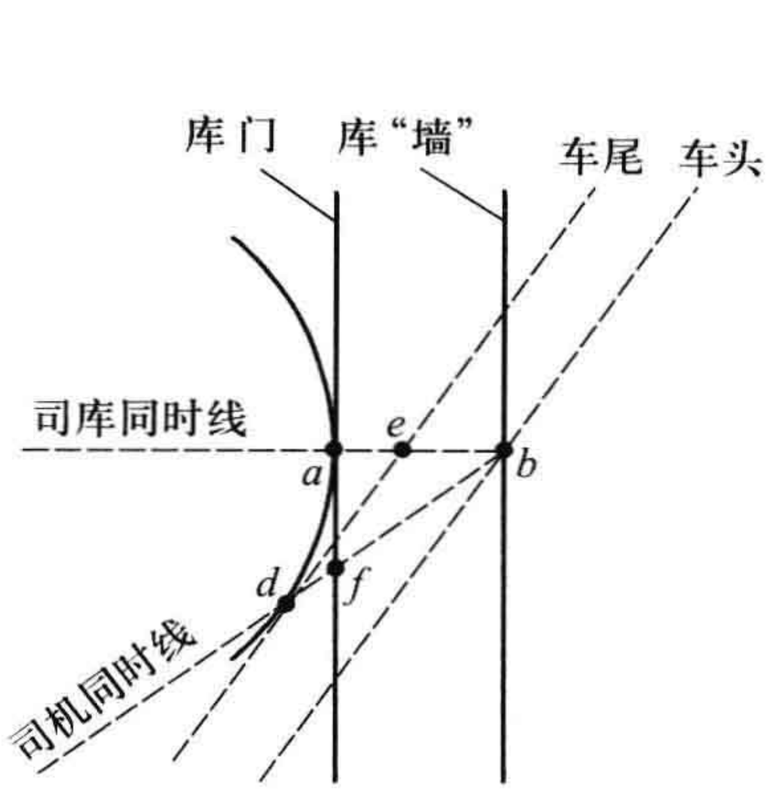


图 4－2　车库佯谬时空图
（无真实后墙的情况）

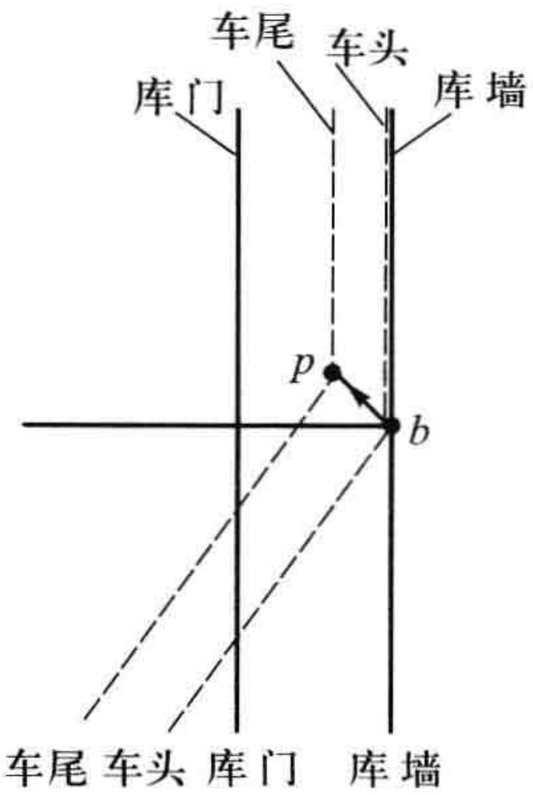


图 4－3　车库佯谬时空图
（有坚硬后墙的情况）

（情况2）　车库有坚硬后墙。

图4-3是这种情况的时空图。车头在撞到后墙时（事件b）被迫停下，世界线当即变成竖直线（与库墙世界线重合）。然而“车头撞墙（因而停止前进）”的信息传到车尾需要时间，车尾在获悉此信息之前仍继续前进。通常认为车头撞墙时会发出冲击波，只当冲击波传到车尾时才被车尾获悉（事件p），车尾才不得不停下。于是汽车将被压缩到在库中的确装下有余的程度（绝对的，谁看都装得下）。

§4.2　钟慢效应

4.2.1　通常的钟慢效应

考虑惯性系K的两个标准钟C_1，C_2和惯性系K′的一个标准钟C′。三钟的世界线示于图4-4。从K系看来，C_1，C_2钟静止而C′钟运动。开始时C′钟与C_1钟重合（事件o），两钟调成指零。一段时间后，C′钟与C_2钟重合（事件b）。由“固有时间等于线长”可知C′钟在b点的读数等于l_{ob}。C_2钟与C_1钟同属K系，x轴是K系的同时线，既然C_1钟在o时读数为零，C_2钟在c时读数也应为零。故C_2钟在b时的读数$l_{cb}=l_{oa}$。过a作校准曲线可知$l_{ob}<l_{oa}=l_{cb}$，故K系认为C′钟（动钟）较慢。然而K′系对此有不同意见，因为该系认为与事件o同时的是事件d（图4-5）而非c。既然C_2钟在c指零，在d就必有读数$\delta>0$。待C_2钟运动到与C′钟重合时（事件b），虽然C′的读数l_{ob}小于C_2的读数l_{cb}（两系都承认这一事实），但不说明C′钟较慢，因为在C′读数为零的同时（按K′的同时线判

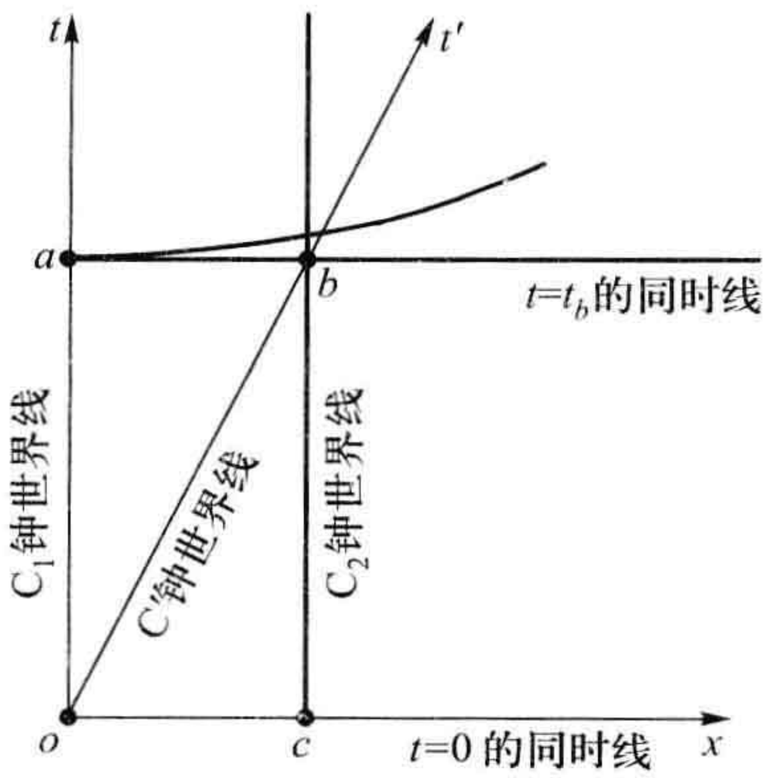


图4-4　K系根据同时线$t=t_b$和$t=0$认为C′钟较慢。

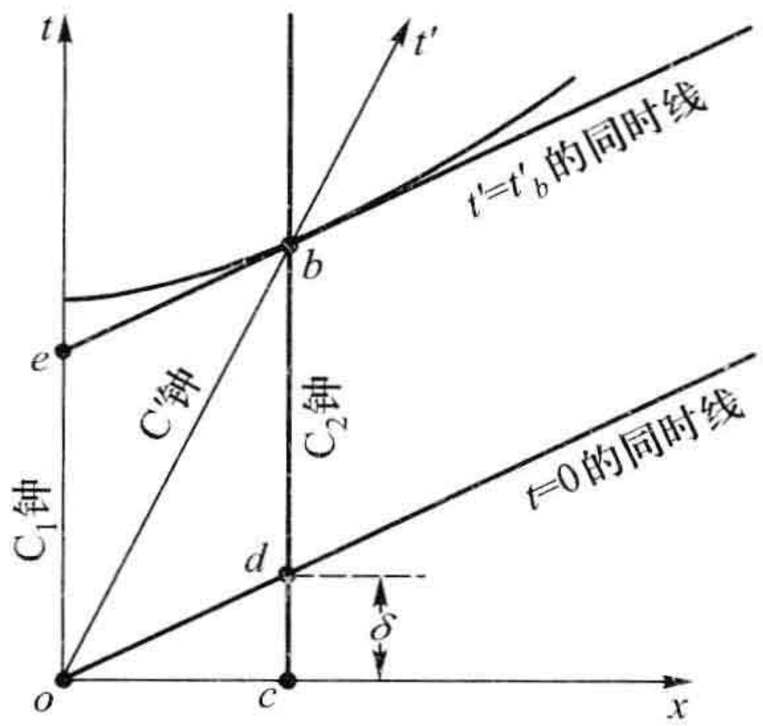
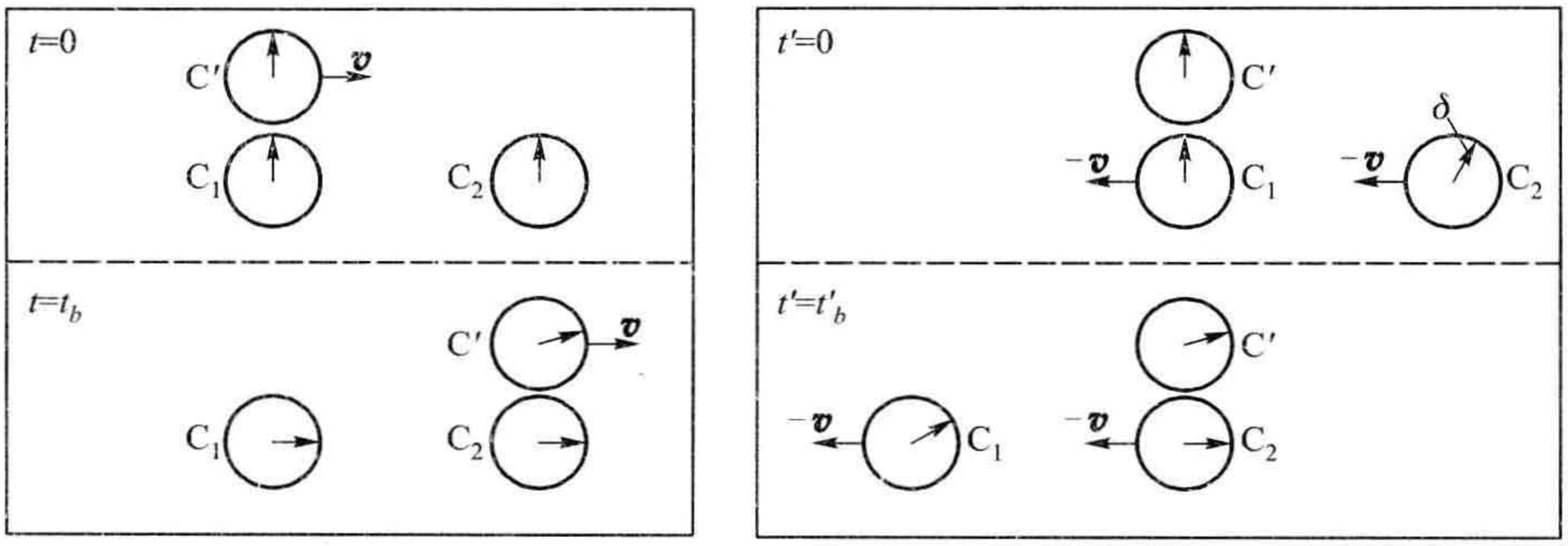


图4-5　K′系根据同时线$t'=t'_b$和$t'=0$认为C_2钟较慢。

断)C_2 读数已是 δ(即 C_2 作了“偷跑”),故应先从 C_2 在 b 的读数 l_{cb} 减去 δ 再与 l_{ob} 比较,即 K′认为应比较 l_{db} 与 l_{ob}。由过 b 的校准曲线可知 $l_{ob} > l_{oe} = l_{db}$,故 K′系认为 C_2 钟较慢,仍是动钟较慢。图 4-6 是以上讨论的 3 维图示,其中(a)和(b)分别为 K 和 K′系的 3 维看法。由此可再次看出 3 维看法依赖于参考系,只有时空图以及用 4 维语言的表述才与参考系无关。



(a) K 系看法　　(b) K′系看法

图 4-6　惯性系 K 和 K′的 3 维看法

下面用图 4-4 计算两钟读数的定量关系,为此只需计算线段 oa 和 ob 的线长。由式(2-2-2)可知

$$l_{oa} = \int_o^a \sqrt{-\mathrm{d}s^2} = \int_o^a \sqrt{\mathrm{d}t^2 - \mathrm{d}x^2} = \int_o^a \sqrt{\mathrm{d}t^2} = t_a - t_o = t_a,$$

$$l_{ob} = \int_o^b \sqrt{-\mathrm{d}s^2} = \int_o^b \sqrt{\mathrm{d}t^2 - \mathrm{d}x^2} = \int_o^b \sqrt{\mathrm{d}t^2 - (v\mathrm{d}t)^2} = \sqrt{1-v^2}\int_o^b \mathrm{d}t$$

$$= \gamma^{-1} t_b = \gamma^{-1} t_a,$$

其中第三步是因为直线 ob 满足 $x = vt$[见式(3-5-2)]。对比两式便得动、静钟所走时间的定量关系:

$$l_{ob} = \gamma^{-1} l_{oa} < l_{oa}。 \tag{4-2-1}$$

可见动钟较慢。通常称此为**钟慢效应**(moving clocks run slow)。应该特别强调,这一效应至少涉及三个标准钟(图 4-4 的 C_1,C_2 和 C′),前两个是惯性系 K 内事先同步好的钟,第三个钟相对于 K 系做匀速直线运动。如果没有 C_2,虽然同样可从图 4-4 知道 $l_{ob} < l_{oa}$,却无从得出“K 系觉得 C′钟慢”的结论,关键在于“觉得”两字无从谈起。事件 b 不在 C_1 钟的世界线上,C_1 钟不可能对它有直接感觉。C_1 钟对 b 进行测量的唯一办法是接收从 b 发来的信号(例如光或声),而这就涉及光的传播时间所带来的问题。(不是不能这样做,而是这样做时必须考虑这一问题。)其实,借用 C_2 钟得出“C′钟较慢”的结论时就已巧妙地发挥了光信号的“使者”作用,因为在把 C_2 钟与 C_1 钟调整同步时已经用到了光信号

(见§3.2末段)。总之,如果只有两个钟 C_1 和 C',就无法用上述方式比钟。这是常遭忽视的一个关键点。

注记1　式(4-2-1)不但说明动钟较慢,还给出了慢的程度。英文文献常用"rate"代表快慢,本书愿意译作"走时率"。但如果说"动钟的走时率较小",就会带来混淆。我们从讨论开始就约定三个钟都是标准钟,因此不但走时率相同,而且都是标准走时率(所走的时间等于线长)。不过英语文献的确存在"动钟的rate较小"的说法,我们建议把这句话中的走时率改为"表观走时率",把钟慢效应说成"动钟的表观走时率较小"。此外,也有不少作者愿意把"rate"译作"速率"[但要防止与钟(作为质点)的运动速率相混淆],这时最好把钟慢效应说成"动钟的表观速率较小"。

注记2　用时空图讨论时经常要计算线长,特别是要用到"横平竖直"地摆放着的直角三角形(例如图4-1的三角形 oab)的边长关系。由于关心的是闵氏线长,欧氏结论"勾股弦公式"不适用,代之而成立的是如下两个公式(把"长直角边"和"短直角边"分别简称为"长边"和"短边"):

$$\textbf{长边长} = \gamma \times \textbf{斜边长};\tag{4-2-2}$$

$$\textbf{短边长} = v \times \textbf{长边长}。\tag{4-2-3}$$

[其中 v 代表两惯性系间的相对速率(对国际单位制 v 要改为 v/c),$\gamma \equiv 1/\sqrt{1-v^2}$。]前面在推导式(4-2-1)和式(4-1-3)时已分别就图4-1和图4-4两种情况证明了式(4-2-2);式(4-2-3)的证明则留作习题。对于长边类时的情况(例如图4-4),此式的物理意义很清楚:静钟认为动钟在长边代表的时间(l_{oa})内走了短边代表的距离(l_{ab}),式(4-2-3)无非是"距离等于速率乘时间"。长边类空的情况虽然没有这样的物理意义,但式(4-2-3)照样成立。以下把式(4-2-2)和式(4-2-3)简称为横竖三角形关系。

4.2.2　用洛伦兹变换讨论尺缩钟慢[选读]

本小节前半部分用对话形式叙述,其中乙代表笔者。

甲　大多数教材都用洛伦兹变换(而不像您的书那样用时空图)证明尺缩钟慢效应,您对此有何述评?

乙　只要思维清晰,用洛伦兹变换当然也能讲清楚。

甲　不少教材在讲钟慢效应时都写出以下公式:

$$t_1 = \gamma(t'_1 + vx'_1),\ t_2 = \gamma(t'_2 + vx'_2),\tag{4-2-4}$$

$$t_2 - t_1 = \gamma[(t'_2 - t'_1) + v(x'_2 - x'_1)] = \gamma(t'_2 - t'_1) > t'_2 - t'_1,\tag{4-2-5}$$

然后说:"$t_2 - t_1$ 和 $t'_2 - t'_1$ 分别是静、动钟经历的时间,所以动钟较慢。"但我觉得式(4-2-4)是站在K′系写出的,它认为自己静止而K系运动,故 $t_2 - t_1$ 和 $t'_2 -$

t_1' 应该分别是动、静钟经历的时间，如此说来，式(4-2-5)岂非表明动钟较快？

乙 不应说“式(4-2-4)是站在K′系写出的”。以其第一式为例，它无非是同一事件在两系的坐标 t_1、x_1 与 t_1'、x_1' 之间的关系，不存在站在哪个系的问题(无论哪个系都承认这一关系)。要弄清 t_2-t_1 是动钟还是静钟经历的时间，必须一步一步地讲清楚。下面是我们用洛伦兹变换证明钟慢效应的讲法。首先，这一效应的比钟方式是K系的两个钟 C_1、C_2 与K′系的一个钟C′做比较。我们约定站在K系看问题，所以也可说是两个静钟与一个动钟比较。其次，在写公式前应该明确定义两个事件：$p_1=(t_1,x_1)=(t_1',x_1')$ 代表C′钟与 C_1 钟相遇的事件[图4-6(a)的上图，亦即时空图4-4的事件 o]；$p_2=(t_2,x_2)=(t_2',x_2')$ 代表C′钟与 C_2 钟相遇的事件[图4-6(a)的下图，亦即时空图4-4的事件 b]，这样就知道：① $x_2'=x_1'$，所以式(4-2-5)的第二个等号成立；② 把C′钟从 C_1 走到 C_2 看作一个过程，t_2-t_1 和 $t_2'-t_1'$ 分别就是K系的钟(静钟)和K′系的钟(动钟)测得的、这一过程经历的时间，所以 $t_2-t_1>t'_2-t'_1$ 说明动钟较慢，而且动钟的表观走时率只有静钟的 $1/\gamma$ 倍。

甲 这样讲是清楚的，但我还有个问题：如果站在K′系看问题，则C′是静钟而 C_1、C_2 成了动钟。$t_2-t_1>t_2'-t_1'$ 岂不是表明动钟较快？

乙 这时 t_2-t_1 不再能代表动钟在上述过程中经历的时间，因为K′系认为 C_1、C_2 两钟并不同步。

甲 不是早就约定K系的各钟都是同步的吗？

乙 是的。同步是指各钟在任一同时面上读数相同，而K′系与K系有不同的同时面族，所以，K系认为同步的各钟，K′系必然认为不同步(认为自己指零时 C_2 钟已有读数 $\delta>0$，即“偷跑”了)，导致 t_2-t_1 不再代表动钟在上述过程中经历的时间，详见围绕图4-5和图4-6(b)所做的讨论。

甲 证明尺缩时是否也要先定义事件？

乙 是的。为此，首先要明确静、动尺长的概念。设尺子沿 x' 轴静止在K′系，该系测得的尺长(静尺长)显然就是尺子两端的坐标差 $x_2'-x_1'$。但K系测得的尺长(动尺长)则要说清楚。由于运动，尺子两端的坐标值不断变化，只有同时测得尺子两端的空间坐标，其差值才是动尺长。具体操作如下：K系观者在某时刻记下尺头和尺尾在地面留下的印记(依次记作 A_1 和 A_2，如图4-7)，等尺子走后再慢慢测量 A_1 和 A_2 的空间坐标。定义两个事件：$p_1=(t_1,x_1)=(t_1',x_1')$ 代表尺头经过 A_1 的事件；$p_2=(t_2,x_2)=(t_2',x_2')$ 代表尺尾经过 A_2 的事件，便有 $t_1=t_2$，$x_2-x_1=$ 动尺长，$x_2'-x_1'=$ 静尺长。由洛伦兹变换得

$$x_1'=\gamma(x_1-vt_1),\quad x_2'=\gamma(x_2-vt_2),\qquad (4-2-6)$$

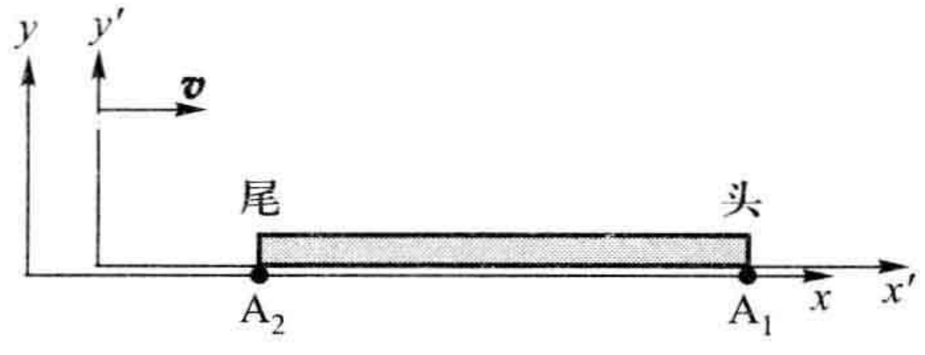


图 4－7　动尺长的测量

$$x_2' - x_1' = \gamma[(x_2 - x_1) - v(t_2 - t_1)] = \gamma(x_2 - x_1), \qquad (4-2-7)$$

可见动尺较短，其表观长度只有静尺长的 $1/\gamma$ 倍。

甲　我明白了。但还有一个问题：p_1 和 p_2 对 K′系并不同时，为什么 $x_2' - x_1'$ 可以看作静尺长？

乙　静尺头尾的坐标不随时间而变（x_1' 和 x_2' 都是常数），无须同时测量。

用时空图解题和用洛伦兹变换解题各有优缺点。但是无论如何，解题前画出时空图总是很有帮助的。下面是三个例子。

例 1　事件 p_1 和 p_2 在 K 系的时间和空间间隔分别为 8×10^{-7} s 和 600m，要使它们在 K′系为同时事件，K′与 K 系的相对速率 v 应为何值？

解

（解法 1）用时空图求解

图 4－8 中的 l 和 n 都代表线长，其国际制单位是 m。已知 $l = 600$ m，$n = c(t_2 - t_1) = (3 \times 10^8$ m/s$) \times (8 \times 10^{-7}$ s$) = 240$ m。将式（4－2－3）用于图 4－8 得

$$\frac{v}{c} = \frac{n}{l} = \frac{240\ \text{m}}{600\ \text{m}} = 0.4,$$

故

$$v = 0.4 \times (3 \times 10^8\ \text{m/s}) = 1.2 \times 10^8\ \text{m/s}。$$

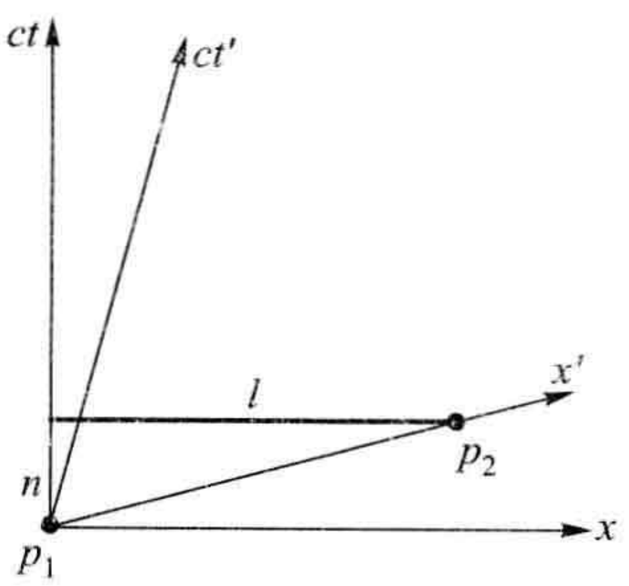


图 4－8　例 1 用图

（解法 2）用洛伦兹变换求解

$$t_1' = \gamma\left(t_1 - \frac{v}{c^2}x_1\right), \quad t_2' = \gamma\left(t_2 - \frac{v}{c^2}x_2\right), \qquad (4-2-8)$$

由已知条件 $t_1' = t_2'$ 得 $t_2 - t_1 = \dfrac{v}{c^2}(x_2 - x_1)$，故

$$v = c^2\,\frac{t_2 - t_1}{x_2 - x_1} = (3 \times 10^8\ \text{m/s})^2 \times \frac{8 \times 10^{-7}\ \text{s}}{600\ \text{m}} = 1.2 \times 10^8\ \text{m/s}。\quad [\text{解毕}]$$

例 2　飞船以 $v = 0.1c$ 的速率相对于地球匀速直线飞行。设地面系测得某运动员以 $\Delta t = 10$ s 的时间跑完 $l = 100$ m，(a)飞船测得跑道有多长？(b)飞船测

得运动员跑了多少米?

解 本题及下题用几何单位制求解,只在最后代入数值时才改用国际单位制。

与 $v=0.1c$ 相应的 $\gamma=1/\sqrt{1-0.1^2}=1/0.995$。

(a) 由尺缩关系可知飞船测得的跑道长度:

$$l'=l/\gamma=0.995\times100\ \text{m}=99.5\ \text{m}。$$

(b) **(解法 1)** 用洛伦兹变换求解

设起跑和冲刺事件分别为 $p_1=(t_1,x_1)=(t_1',x_1')$ 和 $p_2=(t_2,x_2)=(t_2',x_2')$,则 $t_2-t_1=\Delta t=10\ \text{s}$, $x_2-x_1=l=100\ \text{m}$。以 λ 代表飞船测得运动员所跑长度(从时空图 4-9 可清楚看出 $\lambda\neq l'$),则 $\lambda=x_1'-x_2'$(由图 4-9 可知 $x_2'-x_1'<0$),再用洛伦兹变换 $x_1'=\gamma(x_1-vt_1)$, $x_2'=\gamma(x_2-vt_2)$ 便得

$$\lambda=\gamma[v(t_2-t_1)-(x_2-x_1)]。$$

将 $v=0.1c$, $t_2-t_1=10\ \text{s}$, $x_2-x_1=100\ \text{m}$ 代入上式得

$$\lambda\approx3\times10^8\ \text{m}。\qquad\text{[解毕]}$$

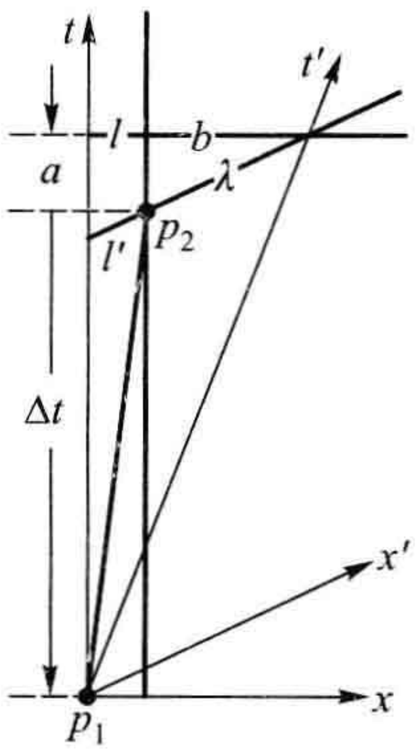


图 4-9 例 2 用图, l,l',a,b 均为线长。

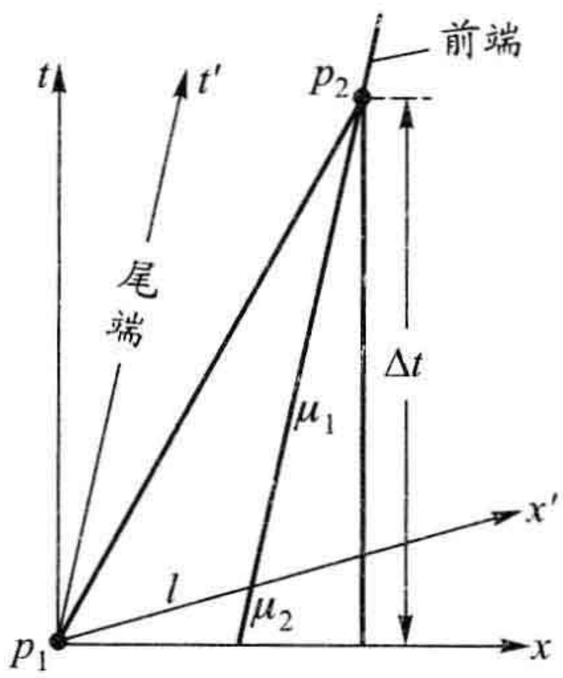


图 4-10 例 3 用图

(解法 2) 用时空图求解

由式(4-2-2)得

$$\lambda=b/\gamma,\ \text{其中}\ \gamma\equiv1/\sqrt{1-v^2}。\qquad(4-2-9)$$

由式(4-2-3)得 $l+b=v(a+\Delta t)$ 及 $a=vb$,联立解得 $b(1-v^2)=v\Delta t-l$,即 $b=\gamma^2(v\Delta t-l)$,注意到 $\Delta t=t_2-t_1$,代入式(4-2-9)便得

$$\lambda=\gamma[v(t_2-t_1)-(x_2-x_1)]\approx3\times10^8\ \text{m}。\qquad\text{[解毕]}$$

例 3 静长为 l 的容器以速率 v 沿 K 系 x 轴正向匀速平动,由容器尾端放出的粒子相对于容器以速率 u' 匀速向前端运动,求 K 系测得的、粒子从尾端到达前端所用的时间 Δt。

解

(解法1) 用洛伦兹变换求解

设粒子“起跑”和“冲刺”事件分别为 $p_1=(t_1,x_1)=(t_1',x_1')$ 和 $p_2=(t_2,x_2)=(t_2',x_2')$，则 $l=x_2'-x_1'=u'(t_2'-t_1')$，　$\Delta t=t_2-t_1$。由洛伦兹变换得

$$\Delta t=t_2-t_1=\gamma[(t_2'-t_1')+v(x_2'-x_1')]=\gamma\left(\frac{l}{u'}+vl\right)=\frac{l}{\sqrt{1-v^2}}\left(\frac{1}{u'}+v\right), \tag{4-2-10}$$

对国际单位制则为

$$\Delta t=\frac{l}{\sqrt{1-(v/c)^2}}\left(\frac{1}{u'}+\frac{v}{c^2}\right)。$$　　[解毕]

(解法2) 用时空图求解

选K系为基准系画时空图4-10，由式(4-2-2)得

$$\Delta t=\gamma(\mu_1+\mu_2)。 \tag{4-2-11}$$

为看出 μ_1、μ_2 与静长 l 的关系，再以 $\{t',x'\}$ 系(容器所在系)画一时空图(读者自画)，将式(4-2-3)用于该图得 $\mu_1=l/u'$ 及 $\mu_2=vl$，代入式(4-2-11)便得式(4-2-10)。　　[解毕]

4.2.3　不同比钟方式有不同结果

狭义相对论和广义相对论都经常谈到钟慢效应，它一定涉及钟与钟的读数比较问题。谈及钟慢效应时，常遭忽视的一个问题就是钟与钟如何比较，也就是**比钟方式**问题。式(4-2-1)给出的是通常比钟方式的结果，这一方式已清楚地约定在图4-4中，实质上是要比较 l_{oa} 和 l_{ob}，结果是 $l_{ob}<l_{oa}$，即 C′钟较慢。应该强调的是，比钟方式是多种多样的，不同方式导致不同结果。“种瓜得瓜，种豆得豆。”因此，在讨论比钟问题时必须事先明确约定比钟方式的每一细节。(可惜许多人对此很不注意，他们往往不问清比钟方式就忙于推出结论。) 上小节的通常比钟方式和结果虽然广为人知(所有狭义相对论教材都要介绍)，却只是众多比钟方式的一种，其特点是必须有三个钟 C_1，C_2 和 C′，前两个是同一惯性系内事先同步好的钟。在只有两个钟 C 和 C′的情况下倒是存在着很有直观意义的比钟方式。例如，图4-11表示携带 C 钟的观者 G 采用如下办法比钟：他在时刻 a 用左右两眼分别看钟 C 和 C′。所谓“用右眼看 C′”是指右眼在时刻 a 收到 C′在某时刻 e 发来的光(光子从 e 沿着指向未来的光子世界线到达 a)。设两钟在相遇时(事件 o)都指零，则 C 钟在 a 点的读数等于 l_{oa}，C′钟在 e 点的读数等于 l_{oe}。由于 $l_{oe}<l_{oa}$，观者 G 将同样得出“动钟较慢”的结果，只是比图4-4的方式慢得更甚。为定量计算慢的程度，可过 e 作水平线交 C 钟世界线于 f(见

图 4 - 12)。令 $\tau \equiv l_{oa}$, $\tau' \equiv l_{oe}$, 则待求量应为 τ'/τ。为简单起见, 记 $m \equiv l_{of}$, $n \equiv l_{fa}$, 则 n 也等于 l_{fe}(理由见注记 3)。由图易见

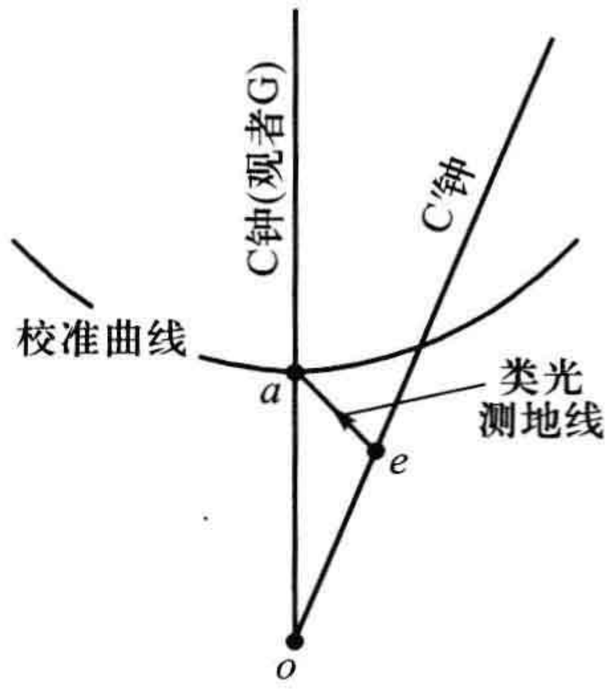


图 4 - 11 G 在 a 时用左右眼分别看 C 和 C′钟, 发现动钟更慢。

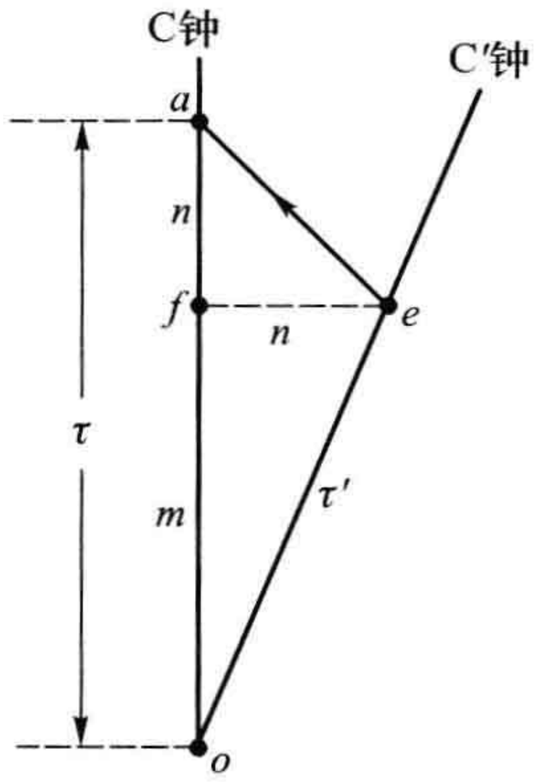


图 4 - 12 求 τ'/τ 的简捷方法

$$\tau = m + n, \tag{4-2-12}$$

以 $\{t, x\}$ 代表 C 钟所在惯性系, 则该系认为 C′钟在时间 m 内走过距离 n, 以 u 代表 C′钟相对于该系的速率, 便有[其实就是式(4 - 2 - 3)]

$$n = um。 \tag{4-2-13}$$

此外, 由式(4 - 2 - 2)又知

$$\tau' = \gamma^{-1} m。 \tag{4-2-14}$$

于是

$$\frac{\tau'}{\tau} = \frac{m}{\gamma(m+n)} = \frac{1}{\gamma(1+n/m)} = \frac{1}{\gamma(1+u)} = \frac{\sqrt{1-u^2}}{1+u} = \sqrt{\frac{1-u}{1+u}}。 \tag{4-2-15}$$

上式表明:

(1) $\tau'/\tau < 1$, 可见动钟较慢;

(2) $\tau'/\tau < 1/\gamma$, 可见动钟慢的程度比用图 4 - 4的比钟方式更甚。

以上两例已经说明不同比钟方式带来不同结果。甚至还可以举出这样的比钟方式(图 4 - 13), 它竟导致动钟较快! 设 C 和 C′钟相遇时(事件 o)均指零, 则观者 G 在时刻 a 用两眼分别看 C 和 C′钟都得负的读数。由图易见 $l_{oa} < l_{oe}$, 故 C′钟的读数比 C 钟的读数负得更甚。于是观者 G 认为动钟较快。这个乍听似

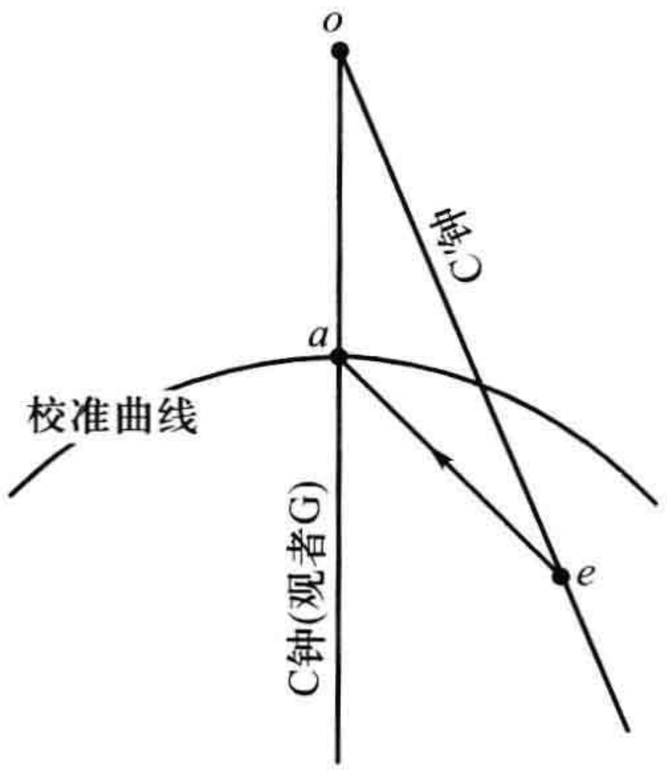


图 4 - 13 G 在 a 时用左右眼分别看 C 和 C′钟, 发现动钟更慢。

乎荒谬的结论其实无可非议，它不过是图4－13的特定比钟方式的结果。因此，讨论比钟问题时必须事先说明比钟方式的每一细节，而为此最好先画时空图。如果只用3维语言构思比钟方式，有时可能会缺少细节（已知条件不全），但一画时空图就能发现。

作为习题，请读者求出图4－13的 τ'/τ 之值，其中 $\tau \equiv -l_{ao}$，$\tau' \equiv -l_{eo}$。答案是

$$\tau'/\tau = \sqrt{(1+u)/(1-u)} > 1。 \qquad (4-2-16)$$

注记3　式(4－2－12)前曾说过，若设 $n \equiv l_{fa}$，则也有 $n \equiv l_{ef}$（图4－12）。这是为什么？你也许会说："三角形 aef 的斜边与两直角边的夹角都是45°，所以等腰。"然而这是欧氏几何的结论，不足信。不过用闵氏几何也不难证明这一结论（在这个问题上欧氏眼光不带来欺骗），为此只需借用基准坐标系 $\{t,x\}$ 分别计算线长 l_{fe} 和 l_{fa}（请注意前者类空后者类时）：

$$l_{fe} = \int_f^e \sqrt{\mathrm{d}s^2} = \int_f^e \sqrt{-\mathrm{d}t^2 + \mathrm{d}x^2} = \int_f^e \sqrt{\mathrm{d}x^2} = x_e - x_f,$$

$$l_{fa} = \int_f^a \sqrt{-\mathrm{d}s^2} = \int_f^a \sqrt{\mathrm{d}x^2 - \mathrm{d}t^2} = \int_f^a \sqrt{\mathrm{d}t^2} = t_a - t_f = x_e - x_f = l_{fe},$$

其中倒数第二步用到光子世界线 ea 是45°斜直线。

[选读4－1]

关于比钟方式的多样性问题，著名相对论学家、芝加哥大学教授**格柔池**（Geroch）在他的小册子"General relativity from A to B"[Geroch(1978)]中有若干精辟的论述，下面是该书第132—133页的部分摘译：

现在考虑"走时过快"和"走时过慢"这两个词汇，它们是我们日常生活中的日常用词。……人们可以用许多不同方式借实验来判断一个钟是否"走时过快"。你可以把它放在一个走时正常的钟旁；你可以从一定距离处看它并与正常钟作比较；你可以借助光脉冲的反弹回收来制造一种"公平比较"的方式……；如此等等。在日常生活中所有这些不同比较方式都用"走时过快"一词来表述，其原因当然是因为在日常生活中它们都给出相同答案，没有必要发明不同的词汇来描述不同的比较方式……反之，在相对论中，不同比较方式一般给出不同答案。那个单一的、无所不包的词汇"走时过快"不再够用。如果你想知道结果，你必须对你打算做的实验的细节给出过细的描述，而不能只问"它是否走时过快"。……据我所知，爱斯基摩人有25个不同词汇来表达雪，……而我们却只有一个。我敢保证，如果光走得比实际情况慢得多，我们在日常生活中就会有类似于25个那样多的不同词汇来表达"走时过快"。　　**[选读4－1完]**

4.2.4 多普勒频移

光波是光源内部机制的某种振动向外发出的电磁波。最简单的振动是简谐振动(有单一的频率),相应的光波是简谐电磁波(因波长单一而称为**单色光**)。设观者 G_2 携带的简谐光源向观者 G_1 发光,光源的两个相邻波峰(正最大值)发生在 p_2 和 p_2'(图 4-14),则线段 p_2p_2' 的线长 $\Delta\tau_2$ 就是 G_2 测得的光波的周期,而图中的 p_1p_1' 段的线长 $\Delta\tau_1$ 则是 G_1 收到的光波的周期。过 p_1 作 G_2 的平行线 G_3,以 $\Delta\tau_3$ 代表 p_1q 段的线长,则 $\Delta\tau_3=\Delta\tau_2$,再用小节 4.2.3 的计算方法不难证明

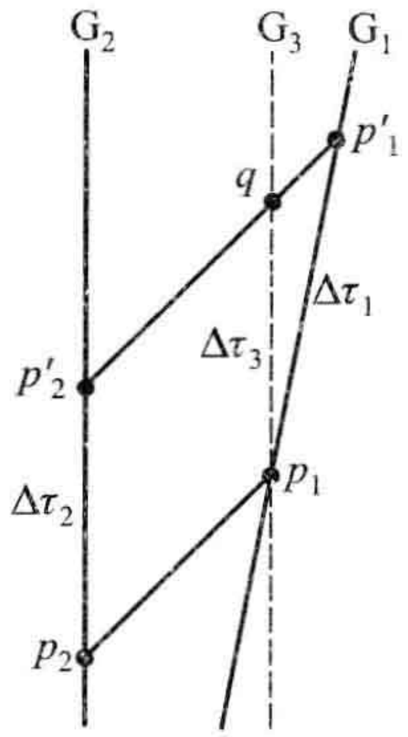


图 4-14 光波的多普勒红移

$$\Delta\tau_3/\Delta\tau_1=\sqrt{(1-u)/(1+u)}, \qquad (4-2-17)$$

[证明留作练习,其实此题与图 4-12 本质一样,上式与式(4-2-15)本质相同。] 由上式得

$$\Delta\tau_2/\Delta\tau_1=\sqrt{(1-u)/(1+u)}。 \qquad (4-2-17')$$

以 λ_2 和 λ_1 分别代表 G_2 和 G_1 测得的光波波长,由波长与周期成正比便得

$$\lambda_1/\lambda_2=\sqrt{(1+u)/(1-u)}>1。 \qquad (4-2-18)$$

可见 G_1 测得的波长较大,即 G_1 看到的颜色比 G_2 看到的颜色向红的方向偏移。这就是所谓的**红移**(redshift)。这种红移起因于相对运动,与声波的多普勒频移类似,故称为光波的**多普勒红移**。图 4-14 代表光源与观者相背(渐行渐远)而行的情况,结果出现红移。反之,如果两者相向运动(渐行渐近),则观者测得的波长比光源测得的要短,故称为**多普勒蓝移**。

4.2.5 钟慢尺缩效应的实验验证

验证钟慢效应的关键是要找到一个高速运动的标准钟(读数差等于世界线长)。物理学界早就知道许多高能粒子以接近光速的高速运动,而它们的衰变固有寿命又使它们可以充当标准钟,自然难能可贵。高能粒子有两大来源,一是高能加速器,二是宇宙射线(宇宙空间中各种高能天体现象产生的高能粒子流)。取自这两种来源的多种高能粒子(如 π 介子、μ 子①和 K 介子)都曾被用

---

① 过去曾把 μ 子称为 μ 介子,后来发现其性质与电子很像,在粒子分类中应该与电子一样属于轻子(lepton)而不属于介子(meson),故改称为 μ 子。但 μ 子与电子也有两个显著区别:① μ 子质量约为电子质量的 207 倍;②电子几乎"长生不老",其平均寿命为 10^{32} 年;μ 子的平均寿命却只有 2μs。(不过,在粒子世界中 2μs 已经是很长的寿命了。)

来成功地验证过钟慢效应,而且精度不断提高[详见张元仲(1994)]。此处只介绍利用宇宙线的 μ 子所做的实验。这方面的首次实验是 Rossi 和 Hall (1941)完成的。后来,Frisch 和 Smith (1963)对此又做了重要的实质性改进,得出了更为精确的结果,并将实验过程及仪器拍成一部很有影响的教学电影。μ 子是在高空大气层中由 π 介子衰变而来的,它们又要衰变为其他粒子。虽然一大堆 μ 子中的每一个可以有非常不同的存活期(寿命),但是可以统计性地谈及它们的平均寿命[定义见式(4-2-19)后]。因此,可以把一大堆 μ 子整体地看作一个标准钟。不过,为了帮助理解,不妨先对单个 μ 子做一理想化讨论。一个静止 μ 子从产生到衰变的时间称为它的静止寿命。考虑一个静止寿命为 2μs(即 2×10^{-6} s)的 μ 子,在 6 km 的高空产生后向地球以 $u=0.995c$ 的高速飞奔而下。如果不考虑相对论效应,就算它以光速飞行,在存活期内也只能走过

$$(3\times10^{8}\ \text{m/s})\times(2\times10^{-6}\ \text{s})=600\ \text{m},$$

根本无望到达地面。然而,由于地球系 K 认为 μ 子在高速运动,动钟较慢,所以测得它的寿命("运动寿命")比静止寿命 2μs 要长。图 4-15 是这一情况的时空图。虽然静止寿命只有 $\Delta t'=2\mu s$,但由钟慢效应可知 $\Delta t=\gamma\Delta t'$,其中

$$\gamma\equiv\frac{1}{\sqrt{1-u^2/c^2}}=\frac{1}{\sqrt{1-(0.995)^2}}\approx10,$$

故地球系测得该 μ 子的运动寿命是静止寿命的 10 倍,即

$$\Delta t=10\Delta t'=10\times2\mu s=20\mu s。$$

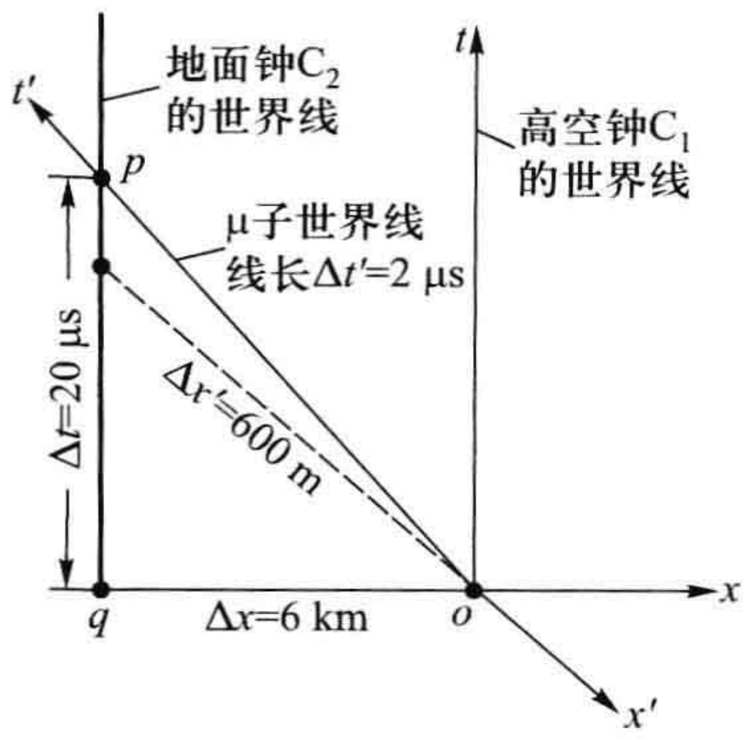


图 4-15　μ 子"从天而降"时空图
事件 o 代表 μ 子产生,事件 p 代表 μ 子死亡。

μ 子在这段时间内走过的距离为

$$\Delta x=u\Delta t=(0.995\times3\times10^{8}\ \text{m/s})\times(2\times10^{-5}\ \text{s})\approx6\ 000\ \text{m}。$$

可见该 μ 子在存活期内正好能走过从产生处(6 km 高空)到衰变处(地面)的距离。请注意这里的钟慢效应同样涉及三个钟,其一是 μ 子(地球系 K 认为的动钟),另外两个则是 K 系的、事先同步好的高空钟 C_1 和地面钟 C_2,它们在事件 o 和 q 都指零。

有人问:站在 μ 子系 K′的立场思考,虽然地球(带着地面钟 C_2 和高空钟 C_1)向自己高速飞来,但自己的寿命只有 2 μs,而且出生时(事件 o)与高空钟 C_1 相遇,地面钟怎能在如此短暂的时间内走完全程 6 km? 答案是:现在应考虑尺缩效应。把 C_1 和 C_2 看作一把很长的尺子的头和尾,则 $\Delta x=6$ km 就是静尺长。

但 K′系认为尺子在高速运动,由尺缩效应可知动尺长只有(仍见图 4 - 15)

$$\Delta x' = \gamma^{-1}\Delta x = 600\ \text{m}。$$

可见,虽然 μ 子认为自己只存活了 2μs(而且出生时遇到尺头 C_1),但死亡(衰变)时恰能遇到尺尾(地面)。就是说,μ 子认为,虽然地面在 2μs 内向着自己只走了 600m,但这正好就是(缩短后的)尺长,所以地面能在自己死亡前及时赶到自己身边。有兴趣的读者不妨补画以 μ 子系 K′为基准系的时空图,由于接近光速,尺子头、尾的世界线将异常靠近。

可见,μ 子实验对钟慢、尺缩效应都给出令人信服的验证,是非常成功的"一箭双雕"的事例。

[选读 4 - 2]

然而以上只是为教学法目的而设的简化讨论。在宇宙线的实验中无法对 μ 子做单个跟踪测量,只能对处于某一速率段内的一批 μ 子进行观测。作为放射性粒子,μ 子的衰变也服从如下的指数衰变律:

$$n = n_0 e^{-t/\tau_0}, \qquad (4-2-19)$$

其中 n 和 n_0 分别代表 t 秒和 0 秒时的 μ 子数,τ_0 是常数。上式表明当 $t=\tau_0$ 时 $n = n_0/e$,可见,经过 τ_0 秒后的 μ 子数只剩下开始时的 1/2.7 倍,人们就把 τ_0 称为 μ 子的**平均寿命**(average lifetime)①。式(4 - 2 - 19)是狭义相对论问世前就有的实验公式,适用于静止的放射性物质。现在自然要问:它对高速 μ 子也适用吗?以 K 和 K′分别代表地球系和 μ 子所在的惯性系。由于 K′认为 μ 子静止,式(4 - 2 - 19)当然适用,只不过要把式中的 t 改为 K′系的时间坐标 t',即

$$n = n_0 e^{-t'/\tau_0},(\text{其中 } n \text{ 和 } n_0 \text{ 应理解为 } t' \text{秒和 0 秒时的 μ 子数})$$

再以 u 代表 μ 子相对于地球的速率,令 $\gamma \equiv 1/\sqrt{1-u^2/c^2}$,由钟慢效应便有 $t' = t/\gamma$,故

$$n = n_0 e^{-t/\gamma\tau_0}, \qquad (4-2-20)$$

现在应该强调,τ_0 代表的只是静止 μ 子的平均寿命,由上式不难看出 $\gamma\tau_0$ 才是运动 μ 子的平均寿命。下面就可简介 Frisch 和 Smith (1963)的实验。

μ 子在高空产生后以大于 $0.99c$ 的速率向地球飞奔而下,某些 μ 子在飞行中会衰变,所以 μ 子总数不断减少。实验者选择美国的华盛顿山②的山顶实验

---

① 莫将"平均寿命"与"半衰期"相混淆,后者是指粒子数减半所经历的时间。

② 即 Mt. Washington,位于美国的 New Hampshire(新罕布什尔州)。

室和哈佛大学①的地面实验室分别作为测量μ子数目的中途站和终点站。在两站各用仪器测出速率介于0.995 0c和0.995 4c之间的μ子数,便可依次充当式(4-2-20)的n_0和n。实验测得值为

$$n_0 = 563, \quad n = 408。$$

在实验室中还可用设备拦住μ子,从而测量其静止时的平均寿命,得值为$\tau_0 = 2.2 \times 10^{-6}$s。此外,已知山顶站与地面站的高度差$L = 1\ 907$m,故式(4-2-20)的$t$应为

$$t = \frac{L}{u} = \frac{1\ 907\ \text{m}}{0.995 \times (3 \times 10^8\ \text{m/s})} = 6.4 \times 10^{-6}\ \text{s}。$$

把式(4-2-20)的γ看作待求量(而不是u的函数,即不管$\gamma \equiv 1/\sqrt{1-u^2/c^2}$),将以上数值代入该式便得

$$\exp\left(-\frac{1}{\gamma}\frac{t}{\tau_0}\right) = \frac{n}{n_0} = \frac{408}{563} = 0.72,$$

故

$$0.72 = \exp\left(-\frac{1}{\gamma}\frac{6.4}{2.2}\right) = \exp\left(-\frac{2.9}{\gamma}\right),$$

进而求得

$$\gamma = \frac{2.9}{\ln(1/0.72)} = 8.8。$$

这一γ值是根据实测的一批数据(如563和408)求得的,实验者称之为γ的**观测值**,并记作$\gamma_{\text{观测}} = 8.8$。另一方面,因子γ又出现在狭义相对论的质能关系式$E = \gamma m_\mu c^2$中。[此式即第5章的式(5-5-2),式中的m_μ代表μ子的静质量。]实验时又测得μ子在飞行中的能量平均值E,代入$E = \gamma m_\mu c^2$中便可求得$\gamma = 8.4$,实验者称之为γ的理论预言值,不妨记作$\gamma_{\text{预言}} = 8.4$,在实验误差范围内可以认为与$\gamma_{\text{观测}} = 8.8$符合得相当理想。　　**[选读4-2完]**

§4.3　双子效应(佯谬)

4.3.1　双子效应的戏剧性描述

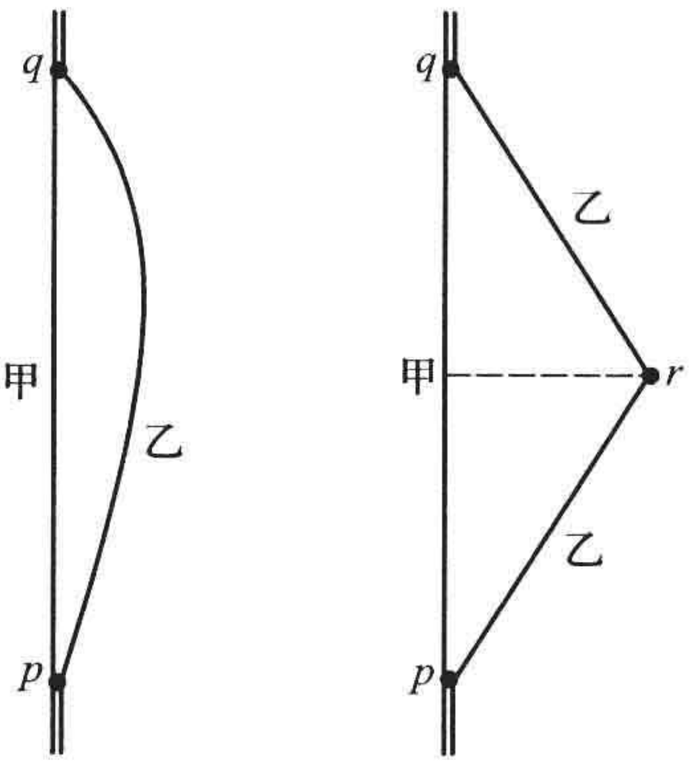
乘坐高速飞船遨游太空的某君归来后竟成了特大新闻人物,因为所有人都

① 即Harvard University,位于美国的Cambridge,Massachusetts(马萨诸塞州的坎布里奇市)。马萨诸塞州是新罕布什尔州的邻州。

发现他比他的孪生兄弟较为年轻。这就是所谓的双子佯谬。把“双子佯谬”再戏剧化为“夫妻佯谬”则更为刺激。一对年龄相同的俊男靓女喜结良缘，欢度蜜月。蜜月后丈夫因公出差，乘坐高速飞船到其他星球处理公务。归来时人们惊讶地发现：夫君依旧是风度翩翩的白马王子，妻子已然是风烛残年的耄耋老妪。此时人们不免想起“山中方七日，世上已千年”的民间传说。这种怪事可能发生吗？根据狭义相对论，从理论上说的确可能，而且早在 1971 年就已被实验所证实，当然不是用人而是用铯原子钟（详见小节 4.3.4）。为了造成前面形容的巨大年龄差异，飞船必须经历（至少在部分时间内经历）非常极端的运动状态，这是人体所无法承受的。但是，从纯理论上讲（抛开人体的生理限制），“飞船方七日，世上已千年”在原则上是可以实现的。理由如下。

4.3.2　双子效应的几何剖析

利用 4 维几何语言可对双子效应给出最为简单、清晰和准确的剖析。首先说明一点。从理论上说，这一效应完全与地球无关，但形象（戏剧性）的陈述难免涉及地球。为了免除毫无必要的复杂性，人们在做纯理论讨论时默默约定：① 近似认为地球做惯性运动；② 忽略地球的引力场，因而只涉及平直时空。图 4 - 16(a) 是在此约定下的时空图（以地面系为基准）。兄弟甲守在家中，是惯性观者，其世界线是竖直线（类时测地线）。兄弟乙外出遨游，其世界线不可能为测地线，否则将一去不复返。p，q 两点分别代表分手和重逢事件。已知分手时甲乙年龄相等，重逢时还等吗？如果不等，孰大孰小？为回答这一问题，只需比较甲乙两人在事件 p,q 之间各自经历的固有时间，也就是比较甲乙两线介于 p,q 之间的线长 $l_{甲}$ 和 $l_{乙}$。因为闵氏时空中两点间的类时测地线是该两点间的类时曲线的最长者（见 §2.5 末），所以 $l_{甲} > l_{乙}$，可见重逢时乙比甲年轻。若要定量地求得 $l_{甲}$ 与 $l_{乙}$ 的关系，就要把乙的世界线具体化，因为两点之间的不同曲线一般有不同线长。最简单的情况如图 4 - 16(b)，这时乙的世界线是由两段直线组成的折线，借助于横竖三角形关系（§4.2 注记 2）易得 $l_{甲}=\gamma l_{乙}$，其中 $\gamma\equiv 1/\sqrt{1-u^2}$，$u$ 是两段斜直线相应的速率。利用此式不难算出，①设分手时两人皆为 20 岁，欲使重逢时甲为 60 岁而乙为 30 岁，只需 $u\approx 0.968c$；② 为了实现“飞船方七日，世上已千年”，只需



(a) 一般情况　　(b) 最简特殊情况

图 4 - 16　双子效应时空图

$u \approx 0.999\ 999c$。

你看,只要你学了“一点点几何”,双子效应就是如此简单明了,几句话便可讲得一清二楚。与那些3维语言的、长篇大论的讨论相比(姑且不论其中还有许多似是而非甚至根本错误的讨论),你不觉得4维语言在这个问题上具有“四两拨千斤”的功效吗?但是,如果你连这一点点几何都不懂,就只好用3维语言思考,就要碰到一系列需要很高智商方可解决的问题。谈到智商,不妨打个比喻。小学算术的鸡兔同笼问题属于难题,它考验着小学生(及其家长)的智商。进入初中后,学了代数方程就有法可循,只要列出方程求解,鸡兔问题便易如反掌。本书笔者一直认为,在某种意义上说,数学是高智商的数学家为一般智商的大众发明的,就像“傻瓜式照相机”是为芸芸众生发明的那样。这种例子不胜枚举。4维语言颇像傻瓜相机,它使大量相对论问题变得简单容易。双子问题就是典型实例。在4维语言普及之前,双子问题难倒了不知多少物理学家,在历史上多次掀起论战,详见下一小节。

[选读4-3]

本选读采用对话方式,其中乙代表笔者。

甲　请您说明“钟慢效应”、“时间膨胀”和“双子效应”的联系和区别。

乙　“钟慢效应”和“时间膨胀”是对同一现象的两种不同称谓。这一现象已示于图4-4中。图中o和b是发生在惯性系K′内的C′钟上的两个事件,该图前后的讨论表明,不同惯性系测量这两个事件的坐标时间间隔得值不同,其中最小者就是K′系自己测得的时间间隔(即$l_{ob} < l_{oa}$),其值就是C′钟的固有时间(钟的固有时间等于钟所在惯性系的坐标时间)。因为K系认为C′钟在运动,所以$l_{ob} < l_{oa}$表明动钟较慢,此即“钟慢效应”,又称“时钟延缓效应”。英文文献更多地称此效应为“时间膨胀(time dilation)”,据说是因为它反映动钟的时间尺度被表观地抻长了(the scale of time is apparently stretched out)。

甲　“双子效应”与“钟慢效应”都是关于钟的读数的效应,两者是否也是同一效应?

乙　不,两者非常不同,根本区别就是:钟慢效应涉及的都是做惯性运动的钟,因此结果必然是相对的——你觉得我慢,我觉得你慢;而双子效应的两钟之中必定有一个不做惯性运动,否则一去不返,一旦分手就不再重逢。因此结果是绝对的——守在家中的兄弟年龄较大。　**[选读4-3完]**

4.3.3　双子“悖论”的长期论战

爱因斯坦在1905年的第一篇相对论论文中就讨论过时钟问题,并对后来被

称为双子效应的问题给出过正确的定量预言[他在该文§4给出了相当于式(4-3-8)(第一个等号)的公式]。然而当时的物理学家和哲学家对此众说纷纭,莫衷一是。反对者的主要看法是(改用双子方式陈述):家中的兄弟甲之所以觉得归来的兄弟乙较年轻(钟走得较慢),是因为他觉得自己静止而乙曾经先离开后返回(走了一圈),因而是个动钟。但是,站在乙的立场,他会觉得自己静止而甲曾经先离开后返回(走了一圈),因而甲成为动钟。按照同样推理,重逢时必定会觉得甲较年轻。于是,兄弟重逢时都将觉得对方年轻,这当然不合理。于是成为悖论,这就是著名的**时钟悖论**(clock paradox)或**双子悖论**(twin paradox)。这一"悖论"曾经引起许多物理学家的多次论战,成为20世纪持续时间最长、争论最激烈的物理辩论之一。仅以1956—1958年的一轮辩论高潮为例,辩论双方以物理学家**麦克利**(McCrea)和物理兼哲学家**丁苟**(Dingle)为代表人物,文章发表在《自然(Nature)》、《科学(Science)》和《发现(Discovery)》等重要杂志上。引发这场辩论的种子是麦克利在1951年针对时钟悖论发表的短文[McCrea(1951)],他在该文称:"据我所知,人们至今尚不清楚这里其实并无悖论。"然后对一种最简单的情况[实质上就是图4-16(b)]进行计算并得出 $l_{乙}=\gamma^{-1}l_{甲}$ 这一正确结论。1956年(当时认为太空飞行有望实现,双子悖论再度备受关注),丁苟发表长文[Dingle(1956)]批驳麦克利,他认为,根据相对性原理,一切都是相对的,双子甲乙运动情况完全一样(是对称的),因此重逢时应有相同年龄。他甚至针对图4-16(b)的情况做了长篇讨论和计算(但都是错的),竟然"证明"了两人重逢时年龄相等。在稍后的交锋中①,麦克利针锋相对地、正确地指出,相对论并不认为一切都是相对的,乙有加速度而甲没有,正是这一绝对的差别(非对称性)导致重逢时年龄不同。从物理角度看,乙必须几次开启和关闭引擎,这是甲乙双方的实质性区别,使用相对性原理无助于抹杀这一区别;从数学角度说,甲的世界线是测地线(因而最长),而乙的世界线是由几段测地线组成的非测地线(因而比甲线短)。丁苟则反驳说,既然麦克利也承认引擎每次工作时间极短,由引擎导致的甲乙两人的"绝对差别"也就消失。至于测地线,丁苟甚至说"他关于测地线的谈论完全是对所论问题的荒唐古怪的、毫无用处的推广。"鉴于丁苟固执地坚持甲乙两人毫无差别,完全对称(这是导致错误结论的总根源),麦克利在稍后的一篇短文中甚至说(括号内容是本书笔者所加):"如果空间旅行者(返航时)的运动足够剧烈,正是这个旅行者,而不是丁苟(他待在地球上),将会摔死。丁苟与这样一个最终变成尸体的旅行者是不对称

① 丁苟与麦克利交锋的论文很多,其中一部分被集结为一篇题为 Relativity and space travel 的文章,发表在1956年(April 28 和 September 29)的 Nature 杂志上。

的。”两人接着还有若干回合的交锋,然后以双方坚持己见的方式暂时休战。其他学者后来还掀起过大小不等的论战高潮,直至 1971 年的原子钟环球飞行实验(见下一小节)证实了两钟历时不等才算大体平息。在 4 维几何语言已被国际相对论界广为应用的今天,双子问题在理论上是如此简单,“悖论”不悖,至多不过是个佯谬(假的谬误)而已。

应该特别强调的是,许多人顾名思义地以为“在相对论中一切都是相对的”,这是一种极其有害的误解。除了双子问题之外,笔者还曾多次看到这一误解在不同问题上带来的灾难性后果。其实,要想学好相对论,更重要的是要学会在遇到绝对的对象时善于紧抓不放。不错,钟慢效应(见小节 4.2.1)是相对的——甲觉得乙慢,乙觉得甲慢,而且慢的程度相同。但那是因为前提本身就是相对的——两钟都做惯性运动,因而,根据相对性原理,两者必然平权。然而双子问题中的两者却有绝对的差异,因为甲做惯性运动而乙做非惯性运动(否则不可能与甲重逢①)。前提的这种绝对差别必然带来后果的绝对不同——兄弟俩重逢时年龄不等,根本不存在什么悖论! 至于两人中谁较年轻,前面已看到用世界线长几乎一望而知,易如反掌。然而,在相对论发展的早期,多数物理学家尚未掌握几何语言(不知“世界线长”为何物,更不知它等于固有时间),而用物理语言把“乙比甲年轻”的道理彻底讲清又绝非易事,曾经难住过不知多少物理学家(爱因斯坦是少数例外)。由于百思不解,他们纷纷以为只能求助于广义相对论。于是“要讲清双子问题非用广义相对论不可”的说法就应运而生并且泛滥成灾,至今仍有大量物理工作者持有这种观点。导致这一误解的一个重要原因是,他们以为,① 为计算乙所经历的时间必须使用与乙固连的坐标系(非惯性系);② 只要涉及非惯性系就属于广义相对论范畴。我们的评述是:① 乙经历的时间就是其世界线长,线长是几何量(是绝对的),与坐标系无关,根本没有必要舍近求远、自找麻烦地用与乙固连的坐标系(非惯性系)计算;② 退一万步说,就算你用非惯性系计算,这也同广义相对论无关。广义相对论是爱因斯坦的引力理论(当引力足够弱时近似回到牛顿的引力论),它认为引力实质上是 4 维时空的弯曲在 3 维空间的表现(详见 §6.1)。因此广义相对论与狭义相对论的根本区别在于时空背景——前者是弯曲时空;后者是平直(闵氏)时空。由于早已约定忽略地球引力,双子问题的默认时空背景当然是闵氏时空,因而与广义相对

---

① 你也许会以为图 4 - 16(b)中的乙也做惯性运动,其实不然。该图把乙的世界线近似画成折线,在转折点 p,r,q 上画成方向突变式的硬拐弯。物理过程通常是渐变的,严格说应该对这三处适当进行小范围的“磨光”。磨光后的整条世界线当然不是测地线(虽然其中有两大段测地线),所以乙仍做非惯性运动。但这三处磨光对线长的影响很小(而且“要多小有多小”),因此 $l_{甲}=\gamma l_{乙}$ 仍近似成立。

论毫无关系。诚然，在相对论的早期，在几何意识尚未深入人心的情况下，人们（包括爱因斯坦）曾经以坐标系为标准对广义相对论和狭义相对论划界，认为只要涉及非惯性系就算是广义相对论范畴。然而这种划界方式有太多的人为性。仍以双子问题为例，本书的计算只用到惯性系（用于计算线长），而有些人则还选了非惯性系，如果坚持以坐标系划界，就只好说“本书的双子问题属于狭义相对论范畴”，而“那些人的双子问题属于广义相对论范畴”。你不觉得这种讲法很丑陋吗？你不觉得改用绝对的时空几何来划界要漂亮得多吗？事实上，现代国际广义相对论界的统一划界标准是：凡以闵氏时空为背景的物理学都属于狭义相对论物理学，而广义相对论物理学则必然涉及弯曲时空（详见 §6.1）。讨论相对论问题时，一个十分重要而又常遭忽视的步骤就是事先明确约定时空背景，即约定所讨论的物理现象在什么时空中发生。双子问题的前提约定是整个现象发生在闵氏时空中，因此自然属于狭义相对论范畴。不幸的是有人甚至走得更远，误以为加速度会造成时空弯曲①，于是就认为双子问题非用广义相对论解决不可。这时就已经不是划界标准问题，而是大错。你看，明明约定在平直时空中讨论问题，他三转两转竟然转出个弯曲时空来，与“时空的绝对性”大相径庭，焉能不错！？

与双子佯谬类似的另一热门话题是爱因斯坦转盘问题，也常被误以为涉及广义相对论，其实前提约定也是整个现象（转盘及其上观者的运动）发生在闵氏时空，因此也属于狭义相对论范畴。分析这一问题的最清晰的工具仍然是 4 维几何语言，只是它比双子问题更复杂，详见梁灿彬，周彬（2009）中册 §14.2。

非惯性运动的一个更为常用的同义语就是加速运动，所以你更常听说的结论是“有加速度的兄弟较年轻”。这一提法固然不错，但对加速度一词应当小心。谁都知道加速度是相对于参考系而言的，对地面系来说，甲为静止而乙有加速度。但是，相对于与乙固连的参考系，则乙为静止而甲有加速度。根据“有加速度的兄弟较年轻”的提法，现在岂非又得出甲较年轻的结论？悖论之说岂非又死灰复燃？我们的回答是：物理学家早已自觉或不自觉地形成一种习惯，凡提到加速度而又不说明所相对的参考系时，都默认相对于惯性系而言。只有在这种默契下，加速运动才是非惯性运动的同义语，“有加速度的兄弟较年轻”的提

① 很多人听说过“等效原理”的一个常见提法，即“加速度与引力等效”。如果他又知道“引力就是时空弯曲”，便会推出“兄弟乙（加速观者）认为时空是弯曲的”这一错误结论。这一推理犯了“偷换概念”的逻辑错误。“加速度与引力等效”一语中的“引力”只是一种假引力，它不由物质产生，不对应于时空弯曲，只因加速观者感到“好像置身于引力场中”而得名。而“引力就是时空弯曲”一语中的“引力”却是由物质产生的真引力。真假引力不分是讲解等效原理时的常见病。详见 §6.3。

法才正确。我们在§3.3已经讲过4维表述优于3维表述,不妨认为现在又添加了一个例证。非惯性运动的提法是指世界线为非测地线,属于4维表述;加速运动的提法则由于加速度依赖于参考系而属于3维表述,只有补上"相对于惯性系"的默契才是明确且正确的。但是加上这一补充自然就有拖泥带水之感;而非惯性运动的提法则简练自足,无任何误解之虞。与此类似的是电动力学中关于"电荷只当做加速运动时才有辐射"的结论,在记住这一结论的同时也必须记住"加速度是指相对于惯性系"的默契。然而,"电荷只当做非惯性运动才有辐射"的提法则简练自足。

虽然用世界线长(图4-16)能够如此简单地把双子效应讲得一清二楚,但是,时至今日,几何语言仍然远未真正普及到每个相对论工作者和学习者,于是关于这一佯谬的文章在各种杂志上依然屡见不鲜,层出不穷,数不胜数。英国相对论学家**伦德勒**(Rindler)在他的书中有一段话[Rindler(1977)P.45]对这种情况给出了有趣的挖苦性描述:

"……(双子)悖论很易解决,但它那种令人激动的特殊魅力使争论一代接一代地经历着这样的循环:起先感到困惑不解,继而由于有所领悟(有时是误解)而洋洋自得,并且迫不及待地写成文字发表,仿佛此前无人懂过。在这方面发表的文章简直不可胜数,然而它们共同的有用部分至多不过填满几页纸而已。"

[选读4-4]

自1905年以来,有关双子问题的文献浩如烟海。在结论正确的文献中,绝大多数都只讨论图4-16(b)那种最简单情况,并通过具体计算乙所经历的固有时间来证实它比甲经历的时间短。(计算方法五花八门,有的还非常复杂。)这在原则上是对的,不过既然上面对图4-16(a)的简单分析已就最一般情况证明了$l_{乙}<l_{甲}$,是不是就不必再对具体曲线做计算了?退一步说,若要对图4-16(b)做计算,只要记住固有时间等于线长,用水平线(图中虚线)把甲、乙线都等分为两部分,由横竖三角形关系式(4-2-2)便知$l_{甲}=\gamma l_{乙}$。这与各种其他计算方法相较不是简单得多吗?有一种对图4-16(b)的算法也许值得一提,因为它在新、旧文献中颇为多见[例如Muller(1972);钱尚武(1982)],而且其中有一问题需要澄清。这一算法的4维"译文"如下(意思都是原作者的,只不过他们都用难懂的3维表述,此处改成4维表述):设K′和K″分别是直线段pr和rq相应的惯性系,Σ′和Σ″分别是K′和K″系过r点的同时线(图4-17)。当乙处在pr段时,他认为事件m与r同时,故在自己经历pr段的过程中甲经历的时间只有l_{pm}(短于l_{pr})。另一方面,乙在rq段中又认为n与r同时,故在自己经历rq段的

过程中甲经历的时间只有 l_{nq}（短于 l_{rq}）。原文指出，最重要的是，乙在 r 点前后因同时线的突变而觉得甲的年龄有 l_{mn} 的突变量。因此，根据乙的观察，

甲在 p,q 之间的年龄增量 $=l_{pm}+l_{mn}+l_{nq}$，

乙在 p,q 之间的年龄增量 $=l_{pr}+l_{rq}$。

由于用 3 维语言（当然也没画时空图），原文只好利用钟慢效应等公式逐段计算，再把结果代入上两式，最后得到

甲在 p,q 之间的年龄增量 $=\gamma\times$ 乙在 p,q 之间的年龄增量。

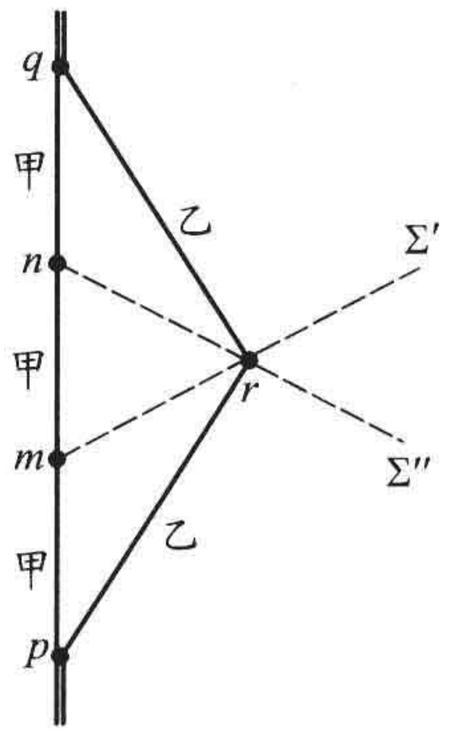


图 4－17　把甲线分段相加

这一结果无疑正确，因为，从 4 维角度看，它无非就是把甲线分为三段并把线长逐一相加而已。然而，除去麻烦之外，这种算法还存在一个概念性问题，关键在于乙对甲的“观察”和“认为”在物理上无法实现。正如小节 4.2.3 所强调的，讨论钟慢效应时除两个钟 C_1 和 C' 外还必须有第三个钟 C_2，而现在偏偏缺少第三个钟。作为局外人，我们可以指着时空图说乙在 pr 段时认为事件 m 与 r 同时，因此应该比较甲钟在 m 和乙钟在 r 的读数；但是，作为局中人的乙却无从知道甲钟在 m 的读数，因为他无法观察（感知）到甲线上的信息。如果你还坚持说可以，那么有一篇文章[孟广达等(1997)]会对你发难。该文指出，接受上述讲法将会导致“返老还童”的离奇推论，其 4 维“译文”如下。设乙的世界线如图 4－18所示，则乙在接近 r 时将根据同时线 Σ' 认为甲的年龄为 l_{pm}（约定甲在 p 时年龄为零），刚离开 r 后根据同时线 Σ'' 又认为甲的年龄为 l_{pn}（$<l_{pm}$），于是乙在 r（掉头）的瞬间观察到甲返老还童。对此返老还童之说你有何评论？其实关键问题仍在于乙对甲的“观察”和“认为”在物理上无法实现。乙从甲获得任何信息的必要条件是甲向乙发出光波或其他信号，而由图 4－19 可知甲线上任意两个事件（如 e_1 和 e_2）的先后顺序在经过光子世界线传到乙线后（为 f_1 和 f_2）都被保持，乙决不会看到返老还童现象。不过，如果甲先在 n 点向乙发出声信号再在 m 点向乙发出光信号，则由图 4－20 可知乙先看见甲在 m 点的形象（一个老人）后听见甲在 n 点发来的声音（甲刚出生时的啼哭声）。在某种意义上可以把这说成“乙感到甲有返老还童现象”，但这在原则上并无不可，它不过是“先看到闪电后听见雷声”这一人所共知现象的一个合理发展而已。再退一步说，如果坚持把乙在每一时刻 $t_乙$ 的同时线与甲线的交点定义为“乙认为的、甲在 $t_乙$ 时刻的年龄”，那么这个“年龄”已没有通常年龄的意义，用它陈述的结论可以稀奇古怪（包括返老还童），但无伤任何物理实质。　**[选读 4－4 完]**

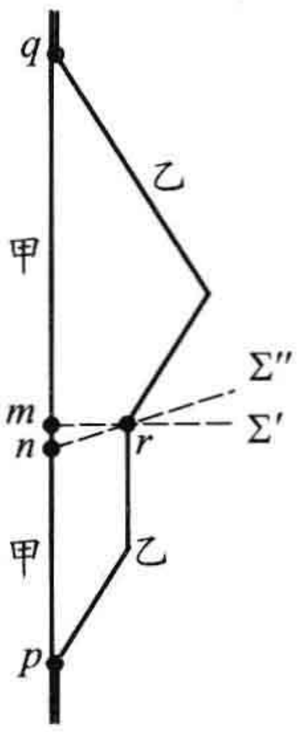


图 4 – 18　返老还童?

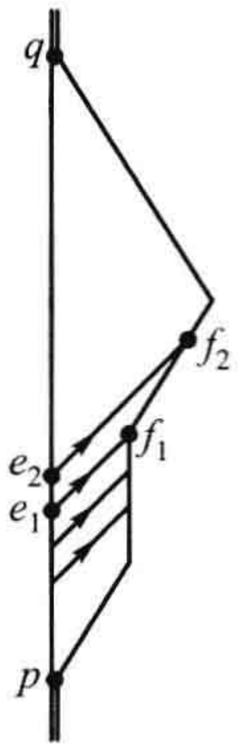


图 4 – 19　事件顺序不因用光传播而变。

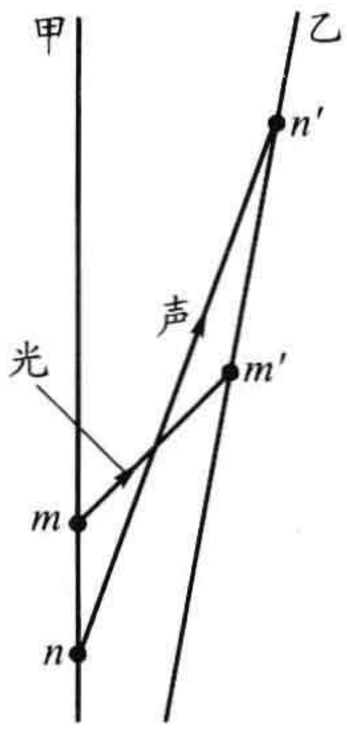


图 4 – 20　乙先见到老人后听到他出生时的啼哭声。

4.3.4　双子效应的实验验证

事实胜于雄辩。双子问题的长期理论争辩呼唤着实验事实的发言。然而载人星际旅行谈何容易！单是造成可观测的年龄差异所需要的高速目前就无望解决,更何况还有人的生理限制等各种实际问题。注意到原子钟测量精度的不断提高,自然想到以钟代人,以飞机代飞船。普通飞机环球一周造成的时间差虽然只有10^{-7}s 的量级,却已可被当时的原子钟精确地测量。于是 Hafele 和 Keating 就在 1971 年借助于商用飞机进行了带钟环球飞行实验,并得出了与理论符合的结果,特别是证实了飞行钟与地面钟经历的时间的确有所不同[见 Hafele 和 Keating(1972)]。然而,实验与理论研究毕竟存在很大差别:理论研究可以约定若干简化前提,而实验则必须真刀真枪地操作。例如,对双子问题做理论讨论时可以忽略地球引力场并认为地球是惯性系,而实验面对的却是真实的地球,应该充分考虑地球的各种实际情况带来的影响,再通过量级估算决定哪些影响非修正不可。估算表明必须计及的是两大因素,即地球的自转(因而不是惯性系)和地球的引力(因而要用广义相对论)。为便于理解,暂时忽略地球引力,讨论如下的半理想化环球飞行问题。赤道上某实验室有两个已同步的铯原子钟 $\hat{C}$ 和 C_0。钟 $\hat{C}$ 随着飞机以 1 600 km/h 的速率向西飞行,环球一周后返回实验室与钟 C_0 重逢,试判断两钟读数孰大孰小,并求出读数差。也许你会不假思索地说:“$\hat{C}$ 钟读数较小,因为它外出做非惯性运动。”但这个答案不对。实验涉及的是个有自转的地球,想象一个惯性观者从北极上空俯视地球(图 4 – 21),则实验室的 C_0 钟随着地球的自转而(从西向东地)做匀速圆周运动,因而不是惯性运动。地球自转使赤道各点都以 1 600 km/h 的线速率向东运动,以同样速率向西飞行的

运动恰好与地球自转相抵消，所以飞机（因而 $\hat{\mathrm{C}}$ 钟）做惯性运动。两钟的世界线如图 4－22 所示。下面计算两钟的读数差。不失一般性，令两钟在分手时（事件 p）都指零，重逢时（事件 q）读数分别为 $\hat{\tau}$ 和 τ_0。由于忽略地球引力，地球附近有平直时空。为便于定量计算，以下的公式都采用国际单位制。

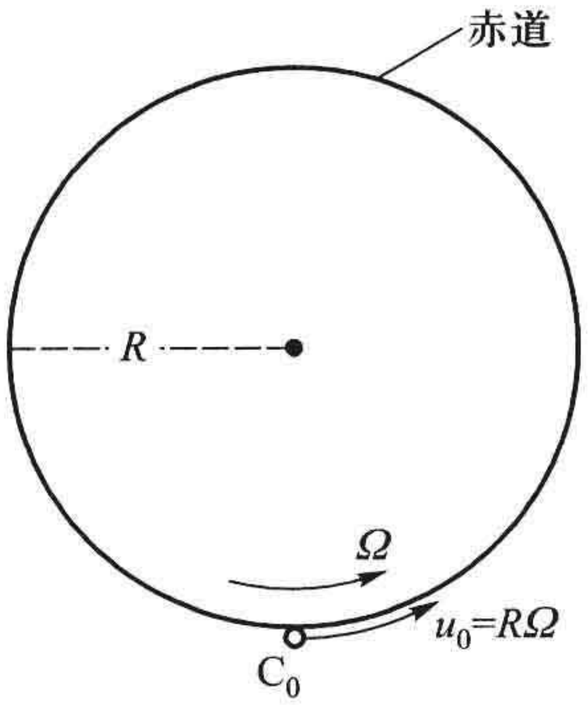


图 4－21　从北极上空俯视地球。C_0 不是惯性钟。

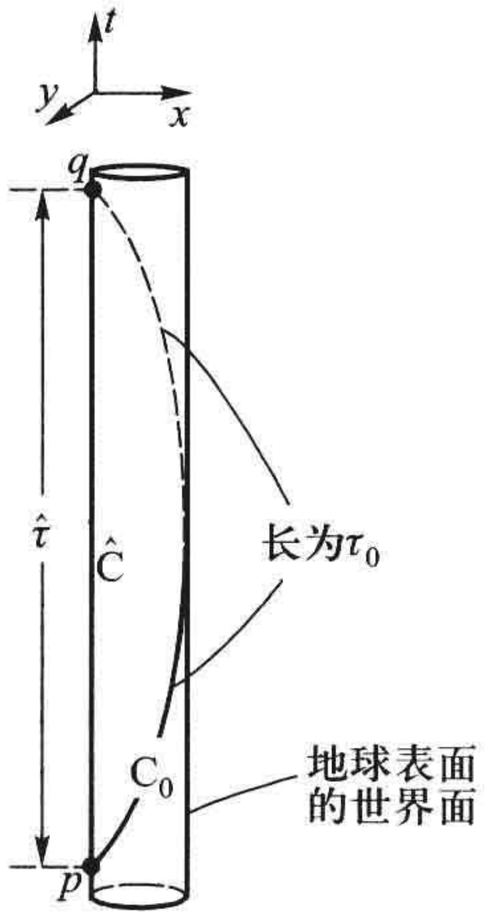


图 4－22　飞机上的钟 $\hat{\mathrm{C}}$ 才是惯性钟。

把地球看作惯性质点，以 $\mathrm{K}\equiv\{t,x,y,z\}$ 代表地球所在的惯性系（其 x,y,z 轴不随地球自转而转动），引入与之相应的球坐标系 $\{t,r,\theta,\varphi\}$，则闵氏线元在国际单位制中取如下形式：

$$\mathrm{d}s^2=-c^2\mathrm{d}t^2+\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2=-c^2\mathrm{d}t^2+\mathrm{d}r^2+r^2(\mathrm{d}\theta^2+\sin^2\theta\mathrm{d}\varphi^2)。\tag{4-3-1}$$

以 Ω 和 R 分别代表地球的自转角速率和赤道半径，把上式用于 C_0 线的任一元段，注意到赤道上有 $r=R,\mathrm{d}r=0,\theta=\pi/2,\mathrm{d}\theta=0$ 及 $\mathrm{d}\varphi=\Omega\mathrm{d}t$，得

$$\mathrm{d}s^2=-c^2\mathrm{d}t^2+R^2\mathrm{d}\varphi^2=-c^2\mathrm{d}t^2+R^2\Omega^2\mathrm{d}t^2=-[1-(R^2\Omega^2/c^2)]c^2\mathrm{d}t^2。\tag{4-3-2}$$

再以 u_0 代表赤道任一点（相对于惯性系 K）的线速率，则 $u_0=R\Omega$，故

$$\mathrm{d}s^2=-[1-(u_0^2/c^2)]c^2\mathrm{d}t^2。\tag{4-3-3}$$

以前熟悉的几何单位制公式 $\mathrm{d}\tau=\sqrt{-\mathrm{d}s^2}$ 在国际制中的形式为 $\mathrm{d}\tau=\sqrt{-\mathrm{d}s^2/c^2}$，故由图 4－22 可知

$$\tau_0=\int_p^q\sqrt{-\frac{\mathrm{d}s^2}{c^2}}=\int_0^{\hat{\tau}}\sqrt{1-\left(\frac{u_0}{c}\right)^2}\,\mathrm{d}t=\hat{\tau}\sqrt{1-\left(\frac{u_0}{c}\right)^2}=\gamma_0^{-1}\hat{\tau},\tag{4-3-4}$$

$$\gamma_0 \equiv \frac{1}{\sqrt{1-(u_0/c)^2}}。$$

（其中第二步是因为我们取 p 点的坐标时 $t_p=0$，从而 $t_q=\hat{\tau}$。）上式中的

$$u_0 = 赤道线速率 = 1\ 600\ \text{km/h} = \frac{1\ 600\times10^3\ \text{m}}{3\ 600\ \text{s}} \approx 444\ \text{m/s},$$

或者，用《中国大百科全书》P. 59 的更准确数据 $u_0=465$ m/s 求得

$$(u_0/c)^2 = [465/(3\times10^8)]^2 = 2.4\times10^{-12}, \qquad (4-3-5)$$

既然 $(u_0/c)^2$ 如此小，就可把 $\gamma_0^{-1}=[1-(u_0/c)^2]^{1/2}$ 展开为泰勒级数并保留到二阶小项：

$$\gamma_0^{-1} = [1-(u_0/c)^2]^{1/2} \approx 1-(u_0/c)^2/2, \qquad (4-3-6)$$

代入式(4-3-4)给出

$$\tau_0 = [1-(u_0/c)^2/2]\ \hat{\tau}。 \qquad (4-3-7)$$

由图 4-22 又知 $\hat{\tau}$ 等于地球自转一周的时间(24 小时)，故

$$\hat{\tau} = 24\ h = 24\times3\ 600\ \text{s} = 8.6\times10^4\ \text{s},$$

于是环球飞行一周所造成的时间差为

$$\hat{\tau}-\tau_0 = [(u_0/c)^2/2]\ \hat{\tau} = (2.4\times10^{-12})\times(8.6\times10^4\ \text{s})/2 \approx 10^{-7}\ \text{s}。 \qquad (4-3-8)$$

以上只是半理想化环球飞行的理论结果。在实际操作中，实验者携带原子钟搭乘商用飞机先后进行了东飞和西飞实验，情况比半理想化飞行更为复杂。此外，飞行钟与地面钟位于地球表面附近的不同高度，根据广义相对论（详见小节 8.3.1），地球引力场还会给两钟的读数差提供一项不可忽略的影响。实验者先从理论上考虑了所有不可忽略的影响并求得两钟读数差的理论预言值，然后再做实验，发现实验结果与理论预言值符合得很好。详见选读 4-5。

[选读 4-5]

实验借助于商用飞机携带 4 只铯原子钟环球飞行了两次，第一次向东，第二次向西。我们先就东飞和西飞问题做一讨论。

设飞机 C 在赤道上空以速度 v 向东飞行（$v>0$ 代表真向东，$v<0$ 代表向西），则它相对于惯性系的速度（记作 u）为（如此低速下当然可用牛顿速度变换式）

$$u = R\Omega + v。 \qquad (4-3-9)$$

当 $v=0$ 和 $v=-R\Omega$ 时 $u=R\Omega=u_0$ 和 $u=0$，C 的世界线分别就是图 4-22 的 C_0 线和 $\hat{C}$ 线。无论 v 取何值，只要飞机 C 在赤道上空飞行，由式(4-3-1)可知对 C 线的任一元段有

$$ds^2 = -c^2dt^2 + R^2d\varphi^2。 \qquad (4-3-10)$$

因 $d\varphi/dt$ 是飞机的角速率，故飞机的线速率 $u = Rd\varphi/dt$，代入上式得

$$ds^2 = -c^2dt^2 + u^2dt^2 = -\left(1 - \frac{u^2}{c^2}\right)c^2dt^2。 \qquad (4-3-11)$$

上式无非是用 u 代替式(4-3-3)中的 u_0，所以，把 C_0 线的环球时间表达式(4-3-7)的 u_0 改为 u 便得 C 线的环球时间(记作 τ)：

$$\tau = [1 - (u/c)^2/2]\,\hat{\tau} = [1 - (R\Omega + v)^2/2c^2]\,\hat{\tau}。 \qquad (4-3-12)$$

当 $v=0$ 时还原为

$$\tau_0 = [1 - (R^2\Omega^2/2c^2)]\,\hat{\tau} = [1 - (u_0/c)^2/2]\,\hat{\tau}, \qquad (4-3-7')$$

两式相减给出

$$\tau - \tau_0 = -[(2R\Omega v + v^2)/2c^2]/\hat{\tau}。 \qquad (4-3-13)$$

上式除以式(4-3-7′)得

$$\frac{\tau - \tau_0}{\tau_0} = \frac{-(2R\Omega v + v^2)/2c^2}{1 - (R^2\Omega^2/2c^2)}。 \qquad (4-3-14)$$

上式右边分母

$$1 - (R^2\Omega^2/2c^2) = 1 - (u_0/c)^2/2 \approx 1 - 10^{-12} \approx 1,$$

[其中第二步用到式(4-3-5)。] 所以式(4-3-14)可简化为

$$(\tau - \tau_0)/\tau_0 \approx -(2R\Omega v + v^2)/2c^2,$$

于是

$$\tau - \tau_0 = -(2R\Omega v + v^2)\tau_0/2c^2。 \qquad (4-3-15)$$

可见，当 $v>0$(东飞)时 $\tau<\tau_0$，即环球时间比地面惯性参考钟经历的时间短("丢失时间")；当 $v<0$(西飞)而且 $|v| \approx R\Omega$ 时 $\tau>\tau_0$，即环球时间比地面钟的时间长("赢得时间")。

除了地球有自转外，地球引力场也会带来可观测的影响。飞机与地面的高度差造成引力势差，根据广义相对论，其影响体现为一个附加项，式(4-3-15)要改为

$$\tau - \tau_0 = [2gh - (2R\Omega v + v^2)]\tau_0/2c^2。 \qquad (4-3-16)$$

其中 $h \ll R$ 是飞行高度，g 是地面重力加速度(详细推证见第 8 章选读 8-1)。上式就是从狭义相对论和广义相对论出发求得的时间差。在与实验数据对比之前还要考虑一系列实际情况，例如，出于经济考虑而采用商用飞机，而商用飞机并不在赤道正上空飞行，其纬度、高度和速度在飞行期间也不保持常数。考虑所有这些因素后得到一个修正公式，由此可求得飞行钟与地面惯性参考钟的时间差的理论预言值：东飞钟为 $(40 \pm 23) \times 10^{-9}$ s，西飞钟为 $(275 \pm 21) \times 10^{-9}$ s。

实际飞行是在 1971 年 10 月进行的。第一次为东飞，持续时间为 65.4 小时，其中飞行时间是 41.2 小时；第二次(约一周后)为西飞，持续时间为 80.3 小

时,其中飞行时间是48.6小时。与地面惯性参考钟的时间差的实验值如下:东飞钟为$(59 \pm 10) \times 10^{-9}$ s,西飞钟为$(273 \pm 7) \times 10^{-9}$ s。Hafele and Keating (1972)的结尾说:

"无论如何,对于'经历闭合旅程后两钟读数是否相等'这一问题,看来已经没有进一步争论的余地,因为我们已经查明两钟读数的确不等。"　[选读4-5完]

§4.4　高速物体的视觉形象

稍懂相对论的人都知道"动尺收缩",于是容易产生如下误解:"如果列车在我的视野内高速飞驰,我看到的车长将比它的静长要短。"事实上,这一误解与狭义相对论同时诞生,因为爱因斯坦在他的相对论奠基性论文《论运动物体的电动力学》(1905)一文§4开头中就有这样的陈述[见爱因斯坦等(1980)]:

"因此,在静止状态下测量为球状的刚体,当处于运动状态时,在静系看来就具有旋转椭球的形状,其三个轴(笔者注:似应为半轴)为$R\sqrt{1-v^2/c^2}$,R,R。"①

从此开始,类似说法就不断出现在浩如烟海的作者的文字中。著名科学家兼科普作家**伽莫夫**②在他的著名科普著作《物理世界奇遇记》(科学出版社,1978年版第5页)中更是淋漓尽致地把这一误解推到了极致。书中描述主人公汤普金斯先生梦游一座奇特城市,市内的光速比c小很多。于是他看到骑车者连人带车都是扁的;当他自己骑车时,又看到被压扁了的街道、橱窗和行人。总之,上述误解统治着物理学界长达50余年,直至**特列尔**(Terrell)和**彭罗斯**(Penrose)于1959年分别载文澄清为止。特列尔首次明确指出,这一误解起因于人们混淆了两个含义不同的词汇——"测量"和"看见"。**看见**(seeing)是指观者用眼睛去看(或用相机拍照),其结果称为**视觉形象**,它由发自物体各点并同时到达眼睛的所有光子组成。由于各点与眼睛的距离未必相同,这些光子未必从各点同时发出。反之,**测量**(measurement)是指特意安排的一种操作,它要"测出"物体各点在同一时刻的位置,这些点组成的图像称为**测量形象**。所谓同一时刻,是指观者所在参考系的一张同时面。小节4.1.1讲过,尺缩效应是指尺子世界面与同时面(图4-1的Σ_0')的交线(动尺)沿运动方向缩短。推而广之,任何运动物体在观者所在系的任一同时面上的图像都沿运动方向缩短。可见,运动物体的测量形象必有尺缩,但

---

① 原文接着说:"于是球(因而无论任何形状的刚体也一样)在Y,Z方向上的尺寸,看来不因运动而有所改变,但X的尺寸看来则按1:$\sqrt{1-v^2/c^2}$的比率缩短了,即v值愈大,收缩愈厉害。当$v=c$时,从'静'系统看来,一切运动物体都收缩成扁平形状了。"

② 伽莫夫是"大爆炸宇宙论"的主要创始人之一。

其视觉形象则有待慎重讨论。有趣的是,特列尔在慎重考虑后竟证明了如下的一般结论[见 Terrell(1959)]:只要物体离观者足够远(物体对观者的张角足够小),高速物体的视觉形象就毫无尺缩,它无非是物体静止时的形象绕某轴转某角的结果。后人称此为**特列尔转动**。限于本书的既定范围,此处不能介绍这一有趣结论的证明,但将讨论两个特例,并讲清两例中的特列尔转动的原因、转轴和转角。

鉴于视觉形象与测量形象存在差别,讨论相对论时就要特别注意用词。通常的习惯是:"测得"或"认为"是指测量形象(或测量结果);"看到"或"看见"是指视觉形象。有时也使用"看来"一词,虽然有个"看"字,但与"测得"或"认为"同义。

例 1　高速运动立方体的视觉形象

边长为 L 的立方体相对于惯性系 K 高速匀速运动,速度矢量 $\boldsymbol{v}$ 平行于立方体的一条边,在图 4 -23 中画成向右。该系的某观者从正下方很远处观看,这个"很远"保证从立方体发出并到达观者的光线为平行光。图 4 -23 是立方体在 K 系的某时刻 t_2 的测量形象。如果 $\boldsymbol{v}=0$,观者只能看到立方体的正面(对观者是正面,对图 4 -23 是底面),即正方形 ABCD。当 $\boldsymbol{v}\neq 0$ 时,由于尺缩,水平边长 BC 只有竖直边长的 γ^{-1} 倍,故正面 ABCD 的测量形象是压扁了的正方形。然而,$\boldsymbol{v}\neq 0$ 还带来另一产物:由于 E,F 点比 B,A 点离观者较远,它们在 t_2 之前的时刻 $t_1 \equiv t_2 - L/c$(那时它们位于 E′,F′)发出的光子将在 t_2 时刻到达正面,并与正面各点在 t_2 时刻所发光子同时到达观者,形成视觉形象。类似的思考也适用于侧面 ABEF 的任一点(例如侧面中点 M 在时刻 $t_2 - L/2c$ 所发光子对视觉形象也有贡献),于是视觉形象如图 4 -24 的下图,正面 ABCD 虽因尺缩而被压扁,但由于侧面 ABEF 的贡献,整个视觉形象在横向还有所"拉长"。其实这正是特列尔转动的后果——远方观者观看立方体所得的视觉形象

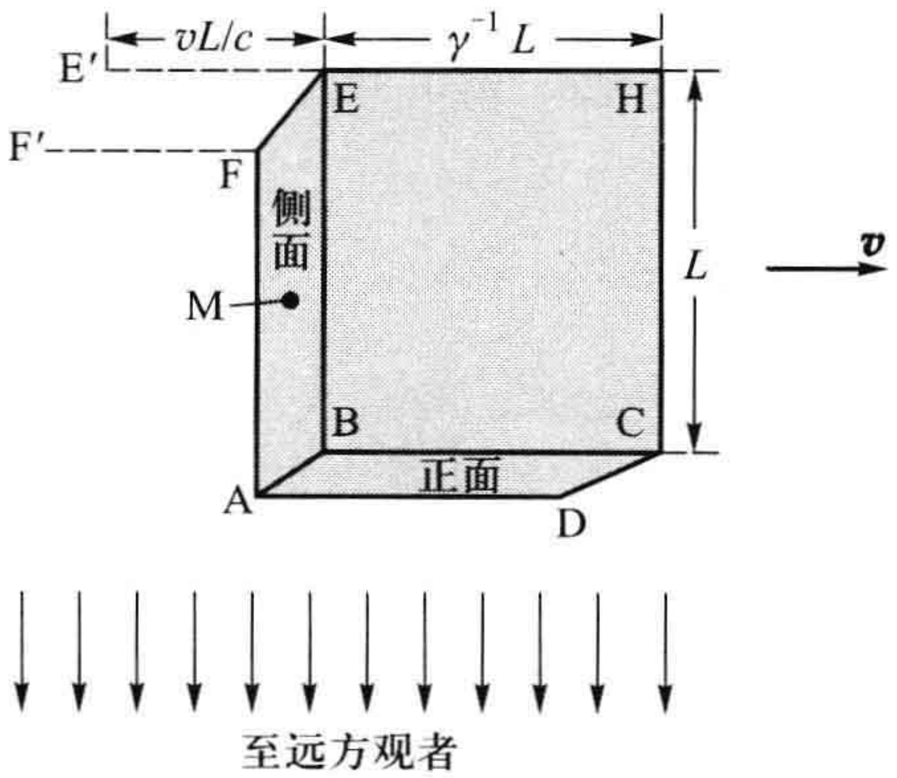


图 4 -23　(测量形象)立方体向右运动,观者在正下方很远处观看。

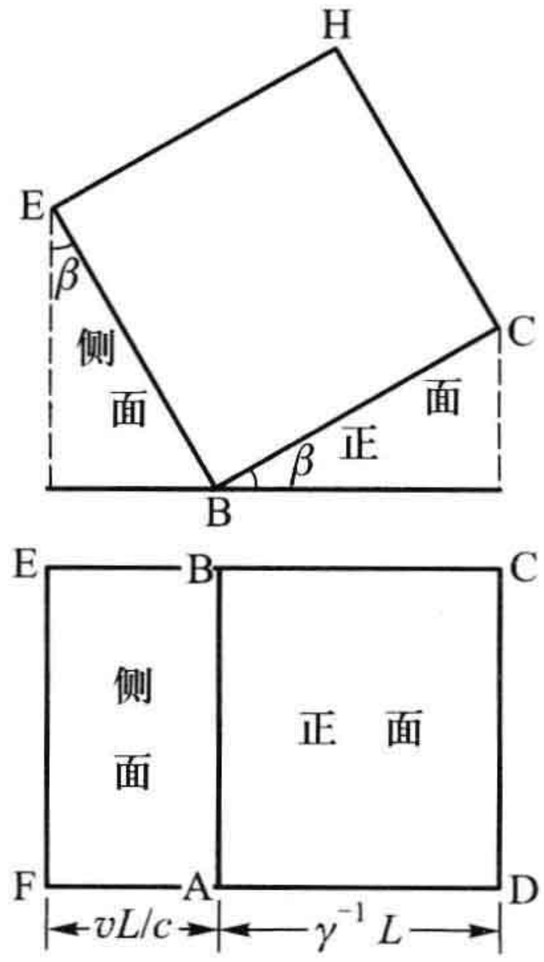


图 4 -24　下图:运动立方体的视觉形象
上图:静止立方体绕 AB 轴转 β 角

相似于静止立方体转了角度 β(图 4 -24 上图)。由图可知 β 满足

$$\sin\beta = (vL/c)/L = v/c。 \qquad (4-4-1)$$

就是说,远方观者看到的图像相当于把静止立方体绕 AB 轴转角度 β 所得的图像。

注记 4　绘制图 4 -23、图 4 -24 时取 $v = 0.5c$,因而 $\gamma^{-1} \equiv \sqrt{1-(v/c)^2} = \sqrt{3}/2 \approx 0.866$。

例 2　高速运动球体的视觉形象

把例 1 的立方体换成半径为 R 的球体,其他条件不变。想象地将球体切成许多薄片,各片平行于 $\boldsymbol{v}$ 和视线决定的平面(图 4 -25 的纸面)。把每片看作一个半径为 $r \leqslant R$ 的圆盘,由于观者的视线平行于这个圆盘("从盘边看盘"),看到的是一段直线。如果 $\boldsymbol{v}=0$(球体静止),观者能看到下半个圆周(图 4 -25 的 A 和 B 是两个"擦边可见点"),表现为一段直线 $\alpha\beta$,长为 $2r$。因为球体由许多圆盘组成(最大半径为 R),所以观者看 $\boldsymbol{v}=0$ 的球体所得的轮廓是半径为 R 的圆周。当 $\boldsymbol{v}\neq 0$ 时,圆盘的测量形象应有尺缩,其周边成为椭圆,长半轴长为 r,短半轴长为 $\gamma^{-1}r$(图 4 -26),椭圆方程为

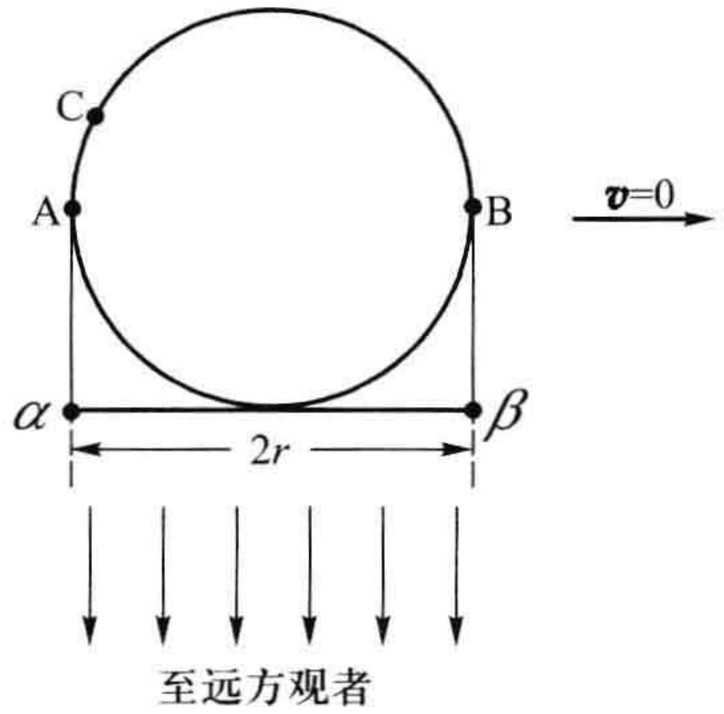


图 4 -25　$\boldsymbol{v}=0$ 时薄片的测量形象为圆形,半径为 $r(\leqslant R)$。A,B 为擦边可见点。

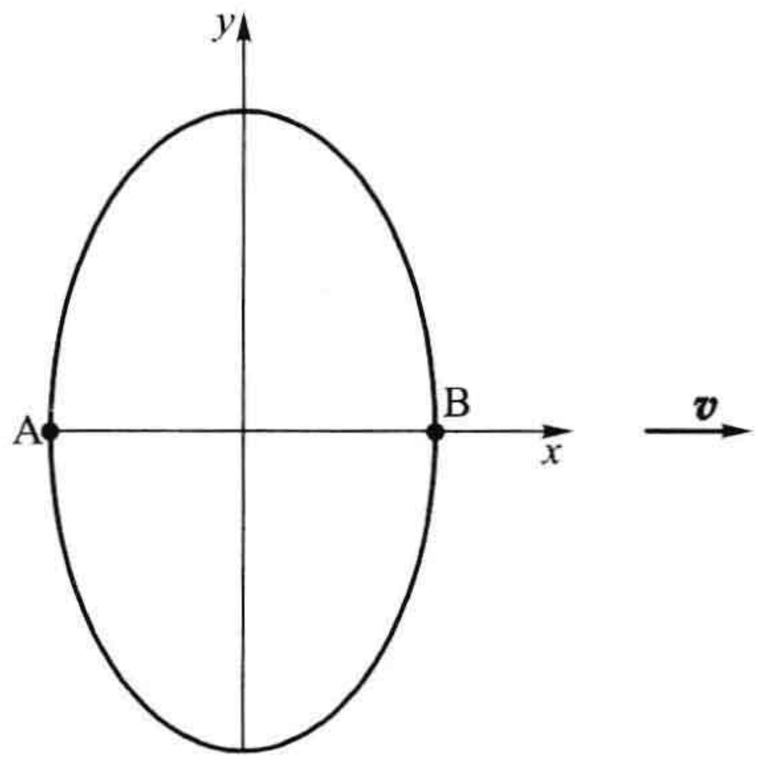


图 4 -26　$\boldsymbol{v}\neq 0$ 时的测量形象

$$\frac{x^2}{\gamma^{-2}r^2} + \frac{y^2}{r^2} = 1, \quad \text{即 } \gamma^2 x^2 + y^2 = r^2。 \qquad (4-4-2)$$

现在要证明观者看运动圆盘所得的视觉形象仍是长为 $2r$ 的直线段(没有尺缩)。从测量形象(图 4 -26)如何求得视觉形象?关键是运动圆盘的擦边可见点与静止圆盘(图 4 -25)不同。图 4 -25 的 A 点之所以是左擦边点,是因为上半圆周的任一点 C 向下所发光子都被圆盘自身遮挡住。然而,当 $\boldsymbol{v}\neq 0$ 时圆盘向右高速平动,只要 C 点比 A 点高得不多,圆盘的右移就会给从 C 向下发出的光子让路,所以左擦边点将沿椭圆适当上移至某点 A′(图 4 -27)。另一方面,图 4 -26的 B 点向下所发光子在 $\boldsymbol{v}\neq 0$ 时将被右移过来的圆盘遮挡,所以右擦边

点将沿椭圆下移至某点 B′。

设图 4 -27 的 A′在某时刻 t_1 所发光子在某一瞬刻到达观者 G,发光事件记作 $p_1=(t_1,\mathrm{A}')$。令 $\Delta t\equiv \mathrm{A}'\mathrm{D}/c$,设 O 是椭圆中心,在过 O 的水平线上取 $\hat{\mathrm{O}}$ 点使$\mathrm{O}\hat{\mathrm{O}}=v\Delta t$,则圆盘在时刻 $t_2=t_1+\Delta t$ 到达以 $\hat{\mathrm{O}}$ 为心的新椭圆(图 4 -28)。以 B′代表新椭圆的右擦边可见点,考虑另一发光事件$p_2=(t_2,\mathrm{B}')$,则事件 p_1,p_2 发出的光子将同时到达图 4 -28 的水平直线段 $\alpha'\beta'$。由于此直线段上各点与观者距离相等,观者看到的就是一段直线$\alpha'\beta'$,下面证明其长度恰为 $2r$。关键是要确定擦边点 A′和 B′。为此,设椭圆盘在无限小时间 $\mathrm{d}t$ 内从图 4 -29 的实线移到虚线。过 A′作竖直线交虚椭圆于 $\tilde{\mathrm{Q}}$,过 $\tilde{\mathrm{Q}}$ 作水平线交实椭圆于 Q。作为擦边点,A′满足如下要求:A′向下所发光子恰好在 $\mathrm{d}t$ 时间内到达虚椭圆($\tilde{\mathrm{Q}}$ 点),$\tilde{\mathrm{Q}}$ 点因向右运动而给它让路。问:该光子从 $\tilde{\mathrm{Q}}$ 继续往下走时会不会被虚椭圆所挡?答:不会,因为“A′是擦边可见点”意味着实椭圆弧$\overset{\frown}{\mathrm{AA}'}$的每点都是可见点,而 $\tilde{\mathrm{Q}}$ 点无非是 Q 点在虚椭圆的体现,自然是虚椭圆的可见点。不妨把从 $\tilde{\mathrm{Q}}$ 继续往下走的光子看作 $\tilde{\mathrm{Q}}$ 所发的光子,它当然不会被虚椭圆所挡。可见擦边点 A′满足下式($\mathrm{d}x$ 和 $\mathrm{d}y$ 含义见图 4 -29):

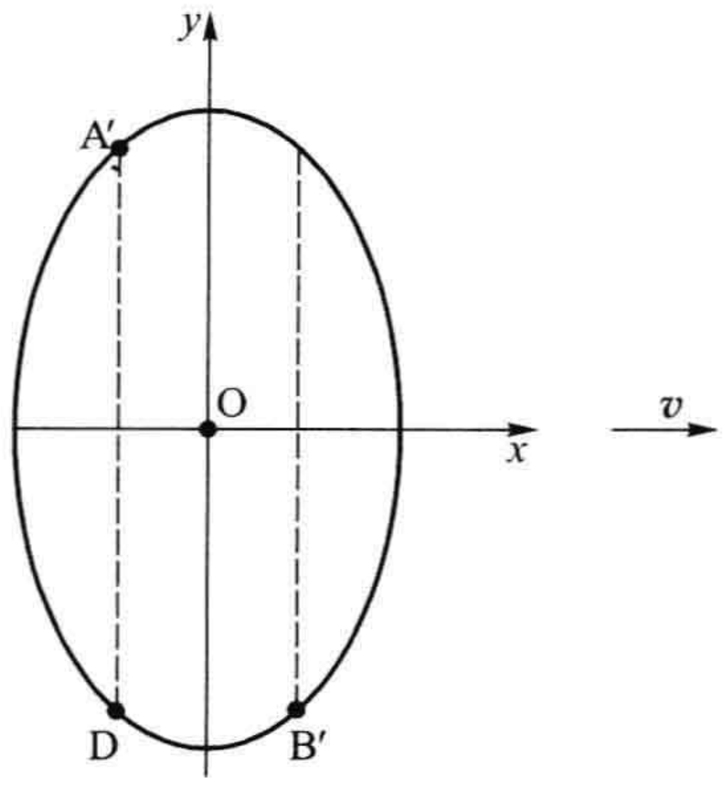


图 4 -27　A′,B′为擦边可见点

(取 $v=0.8c$,因而 $\gamma^{-1}\approx 0.6$, $\gamma\approx 1.67$)

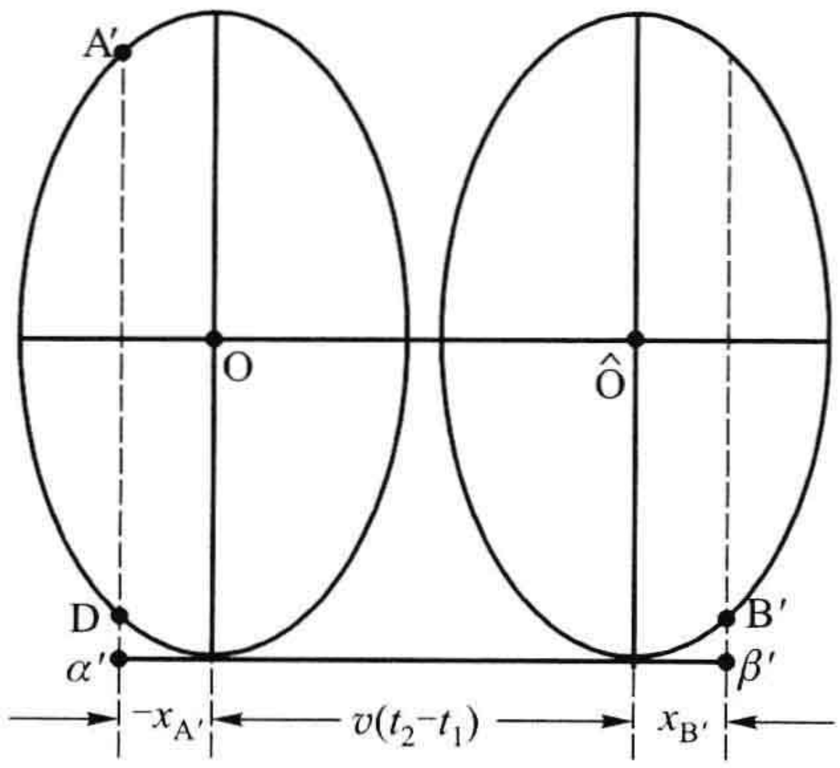


图 4 -28　A′在 t_1 时刻所发光子(记作 γ_1)在 $t_2=t_1+\Delta t$ 时到 D,此时椭圆心恰好移至 $\hat{\mathrm{O}}$,故 B′在 t_2 时所发光子与 γ_1 同时到达 $\alpha'\beta'$线。

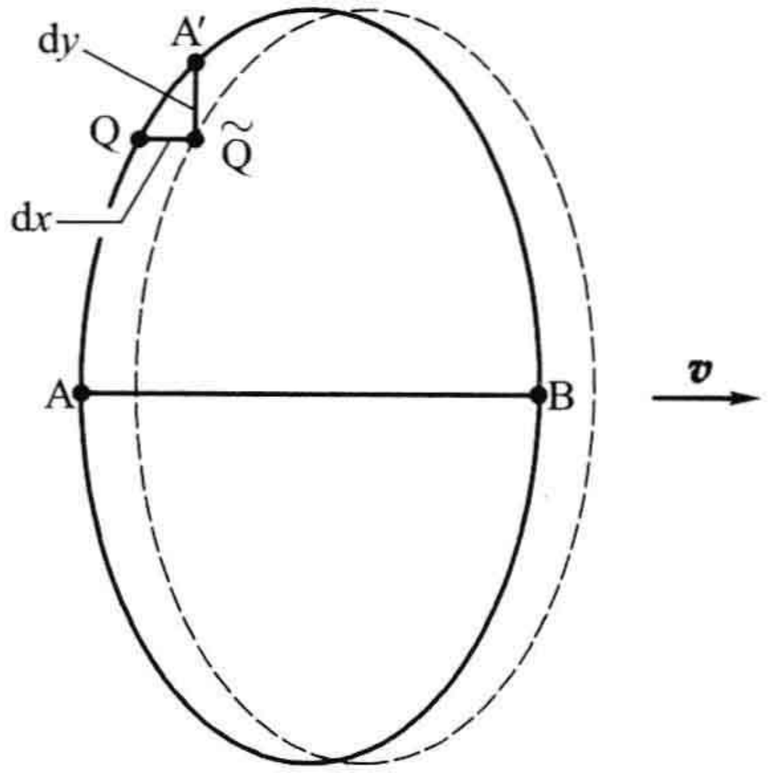


图 4 -29　左擦边可见点 A′的确定

$$dx = v\,dt,\quad dy = c\,dt,\quad 即\ dx/dy = v/c。\tag{4-4-3}$$

再注意到 ① A′和 Q 都是实椭圆的点；② dt 代表无限小时间，就可通过对椭圆方程(4-4-2)求微分得到

$$y/x = -\gamma^2\,dx/dy = -\gamma^2 v/c。\tag{4-4-4}$$

作为左擦边点，A′的坐标 $x_{A'}$ 和 $y_{A'}$ 应满足上式：

$$y_{A'}/x_{A'} = -\gamma^2 v/c。\tag{4-4-5}$$

作为椭圆的点，A′的坐标又应满足椭圆方程(4-4-2)，故

$$r^2 = \gamma^2 x_{A'}^2 + y_{A'}^2。\tag{4-4-6}$$

以式(4-4-5)代入上式，略加运算得

$$r^2 = \gamma^4 x_{A'}^2,$$

故

$$x_{A'} = -\frac{r}{\gamma^2},\quad y_{A'} = -\gamma^2\,\frac{v}{c}\left(-\frac{r}{\gamma^2}\right) = \frac{v}{c}r。\tag{4-4-7}$$

同理还有

$$x_{B'} = r/\gamma^2,\quad y_{B'} = -(v/c)r。\tag{4-4-8}$$

现在就可计算 $\alpha'\beta'$ 的长度。由图 4-28 知

$$\alpha'\beta' = -x_{A'} + v(t_2 - t_1) + x_{B'}。\tag{4-4-9}$$

因为 $t_2 - t_1$ 是光子从 A′到 D 的时间，所以

$$t_2 - t_1 = (y_{A'} - y_{B'})/c = 2vr/c^2。\tag{4-4-10}$$

将式(4-4-10)、(4-4-7)、(4-4-8)代入式(4-4-9)，经简单运算得

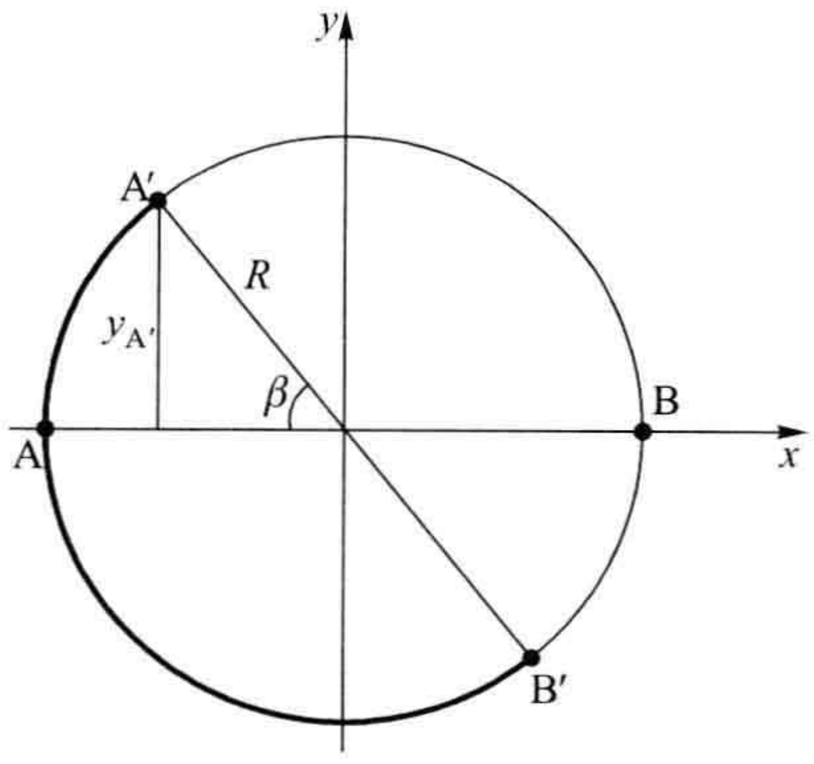


图 4-30　观者看到的是转了 β 角的半圆周 $\overset{\frown}{A'AB'}$(粗线)

$$\alpha'\beta' = r/\gamma^2 + 2r\,(v/c)^2 + r/\gamma^2 = 2r\ 。$$

这就是所要证明的。因为被观测球体由许多圆盘组成(最大半径为 R)，所以在

$v\neq0$ 时观者看到的轮廓与 $v=0$ 时的轮廓一样，都是半径为 R 的圆周。然而两者仍有一个区别：由图 4－28 可知 α' 和 β' 分别对应于 A′和 B′，所以在 $v\neq0$ 时看到的不是下半圆周 $\overset{\frown}{\mathrm{AB}}$ 而是图 4－30 的左下半圆周 $\overset{\frown}{\mathrm{A'AB'}}$（粗线）。就是说，当球体运动时，其视觉形象仍是球体（看不到尺缩），但却是转过 β 角后的球体，由图 4－30 可知 β 满足

$$\sin\beta = y_{A'}/R = v/c, \tag{4-4-11}$$

其中末步用到式(4－4－7)（改 r 为 R）。这正是球体情况下的特列尔转动。

上述两例都只讨论了最简单的情况——观者的视线垂直于物体的运动方向。更一般的讨论将涉及视线与运动方向成任意角的情形。虽然讨论更为复杂，但特列尔转动的结论仍然成立：只要张角足够小，高速物体的视觉形象就相似于静止物体转过一个角度。然而，如果张角不很小，可以预期，物体的不同部分会有不同的转角，因此物体的整体视觉形象将有畸变。不过也有例外：对运动球体而言，无论张角有多大（无论在多么近的地方观看），其视觉形象总是球形。这是彭罗斯[Penrose(1959)]证明的，我们将在选读 4－6 中介绍用时空图对此的证明。

高速物体的视觉形象是一个非常复杂的课题。上面讨论的轮廓问题只是它的一个主要方面，此外还有两个效应：① 物体亮度的角分布会因运动而改变（例如，从各向同性改为像汽车头灯那样发出聚集在一个小立体角内的光线）；② 由多普勒频移带来的颜色改变。这些方面的文献很多，我们只推荐一篇入门读物——Weisskopf(1960)。

电脑模拟是研究高速物体视觉形象的有力工具。这一工作始于 20 世纪 80 年代末期，现在国际上已有多个小组进行研究。以接近光速的高速开车在日常生活中是不可能的，因此，伽莫夫只能让他的主人公梦游光速很低的奇特城市。但是电脑使得“接近光速开车拍摄”成为可能，摄得的彩色录像并非科幻作品，唯一带有“科幻味道”的只是把光速降得很低（例如 $c=5$ m/s）。改变光速不过是对时空尺度的某种重新标度化，并不改变物理，所以拍得的录像在物理上是正确的。从录像和照片不但可以看出物体的形状，而且可以看到由多普勒频移导致的颜色改变。图 4－31 示出其中的一张照片。现在有多个国外网站介绍这方面的成果，除了照片和录像之外，你还可以亲自设定有关参数进行实时演练。有兴趣的读者可以查看下列网站：

① http://www.anu.edu au/Physics/Searle,

② http://www.anu.edu au/Physics/Savage/TEE/,

③ http://www.vis.uni－stuttgart.de/relativity/,

④ http://www.spacetimetravel.org/,

⑤ http://www.adamauton.com/warp/。

(a)“照相机”与街道相对静止　　(b)“照相机”以接近光速的速率相对于街道运动

图4-31　以接近光速的速率开车？这种在日常生活中不可能的事情可以用电脑模拟来做。本图是德国图宾根(Tubingen)市中心的照片。由于不满足张角很小的条件,右图有很大畸变。

[选读4-6]

本选读借用时空图证明如下结论:无论视角大小,高速运动球体的视觉形象总是球形(轮廓为圆周)。用时空图讨论高速物体视觉形象的文献极为罕见,我们能找到的仅有一篇[Penrose(1959)]中的一小段。以下内容主要是笔者钻研的结果。讨论中采用几何单位制,其中光速 $c=1$,所以光锥面是45°圆锥面。为便于讲解,我们讨论两个问题:

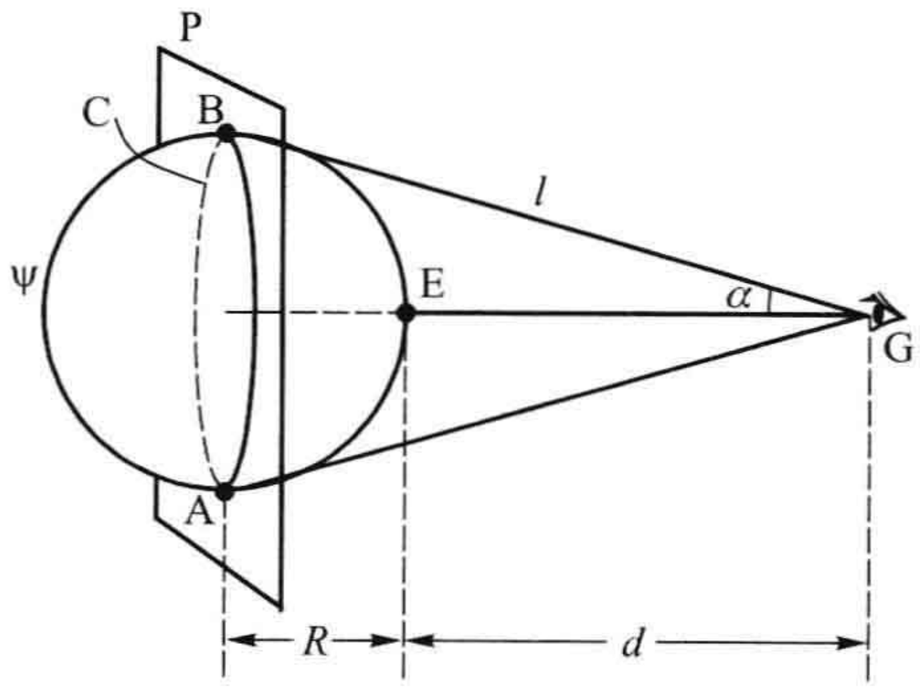


图4-32(空间图)　观者G与球面ψ相对静止,他看到的轮廓是圆周C。

问题1　与球体相对静止的观者G看到什么？

问题2　与球体相对运动的观者G′看到什么？

问题1很简单,答案自然是球形(轮廓是圆周)。但是为了用时空语言找出问题2的答案,最好先用时空图讲清问题1中G看到圆周的道理。这一讲解虽然稍费笔墨,但一经掌握,看似复杂的问题2马上迎刃而解。

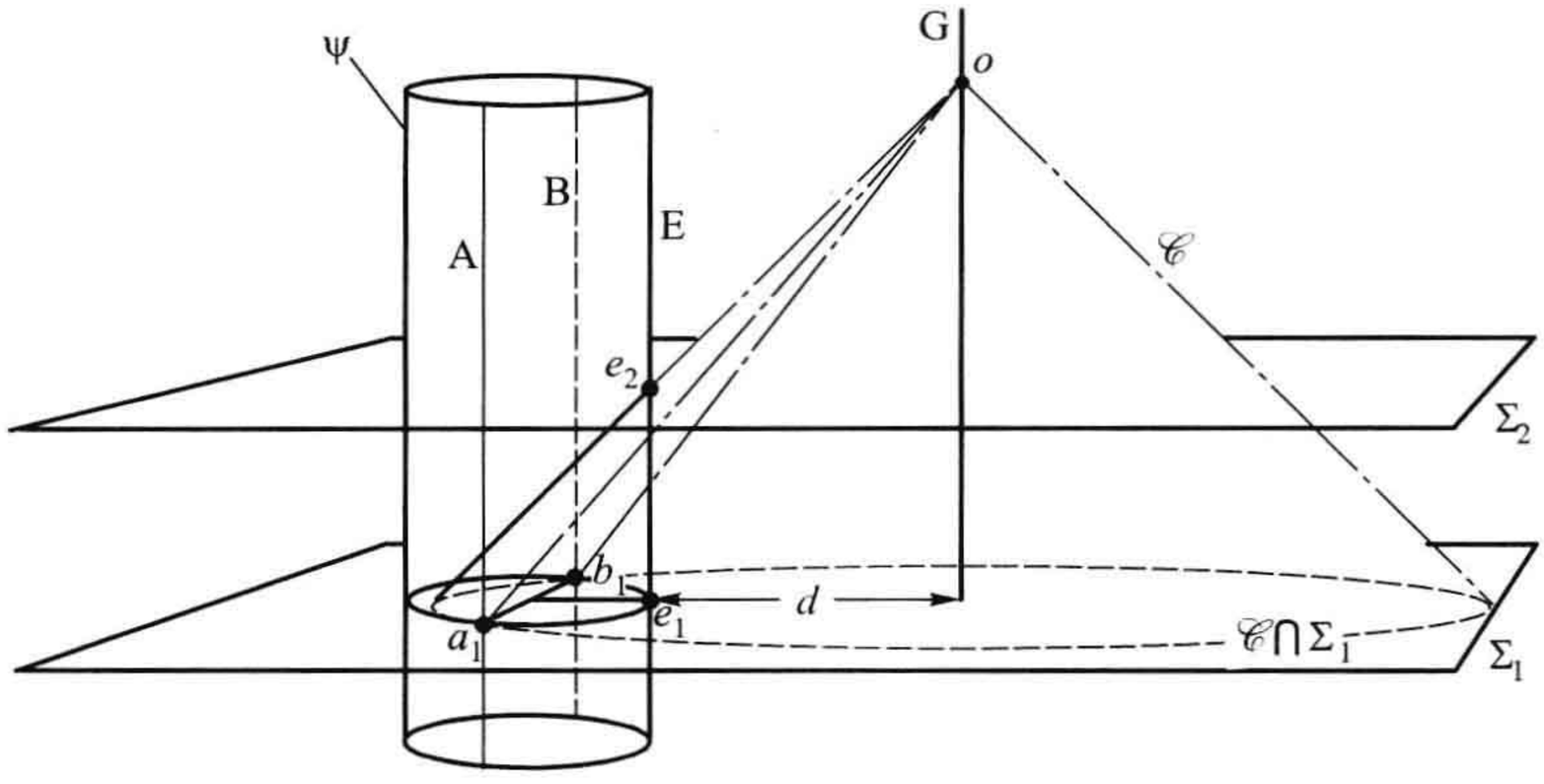


图4－33　观者G与球面ψ相对静止的时空图。a_1o,b_1o和e_2o是三条有用光线。

以G为心作任意半径的球面,并称之为G的一个"天球面",简称**天球**。任何光子在到达G前必定穿过天球,因而留下印记。由于天球各点与G距离相等,所有光子在天球上的印记便构成视觉形象。

以ψ代表被看球体的表面(2维球面)。为便于同稍后的时空图对比,约定G在球的右侧(而不再是下侧)观看,如图4－32。由于球体不透明,球面ψ的左半部分被右半部分挡住,对视觉形象有贡献的只有右半个球面,记作$\psi_{右}$①。选球体所在的惯性系K为基准系画时空图(图4－33),则球面ψ的世界面(仍记作ψ)是竖直圆柱面(这是压缩一维的结果),观者G的世界线是竖直线。在G线上任取一点o作为观测瞬刻,则在此瞬刻到达G的光子构成视觉形象,这些光子的世界线必定躺在o点的过去光锥面(记作$\mathscr{C}$)上。圆柱面ψ上每点都代表一个发光事件,所发光子的世界线都是指向未来的类光测地线(45°斜直线),但只有进入o点(因而躺在$\mathscr{C}$上)的那些线才对视觉形象有贡献。有贡献的类光测地线称为**有用光线**。类光测地线是有用光线当且仅当它发自$\psi_{右}$与$\mathscr{C}$的交集(记作$\psi_{右}\cap\mathscr{C}$)。交集$\psi_{右}\cap\mathscr{C}$的边界点发出的有用光线称为**边缘光线**,它们组成**边缘光线集**,视觉形象的轮廓就由边缘光线集决定。图4－33示出三条有用光线(点划线)——a_1o,b_1o和e_2o,分别从图4－32的A,B和E点发出,a_1o和b_1o还是边缘光线。以t_0代表事件o在K系的时间坐标,则A,B点在时刻

① 当观者与球的距离不够远时,$\psi_{右}$应理解为右边小半个球面。

$t_1 \equiv t_0 - l/c$ 发出的光子恰在 t_0 时刻到达 G，发光事件为 $a_1 = (t_1, \mathrm{A})$ 及 $b_1 = (t_1, \mathrm{B})$，收光事件为 $o = (t_0, \mathrm{G})$。事件 a_1, b_1 对 K 系是同时事件，所以在时空图中位于 t_1 时刻的同时面 Σ_1 上。E 点在较晚时刻 $t_2 \equiv t_0 - d/c > t_1$ 发出的光子也在 t_0 时刻到达 G，所以发光事件 $e_2 = (t_2, \mathrm{E})$ 位于较晚的同时面 Σ_2 上。以 w 代表圆周 C 所发光子的速度的水平分量，则 $w = c\cos\alpha$（图 4－32）。以 P 代表圆周 C 所在平面（仍见图 4－32），想象 P 从时刻 t_1 开始以速率 w 向观者 G 匀速平移，则圆周 C 在时刻 t_1 所发光子都与平面 P“齐步走”，就是说，无论 P 走到哪里（例如走到图 4－34 的虚线处），光子总在 P 上。因此，若以 $\mathscr{P}$ 代表平面 P 的世界面（是个 3 维平面，但在图 4－35 中只能画成 2 维平面），则边缘光子的世界线必定躺在 $\mathscr{P}$ 上。另一方面，这些世界线是类光测地线，当然也躺在光锥面 $\mathscr{C}$ 上，所以

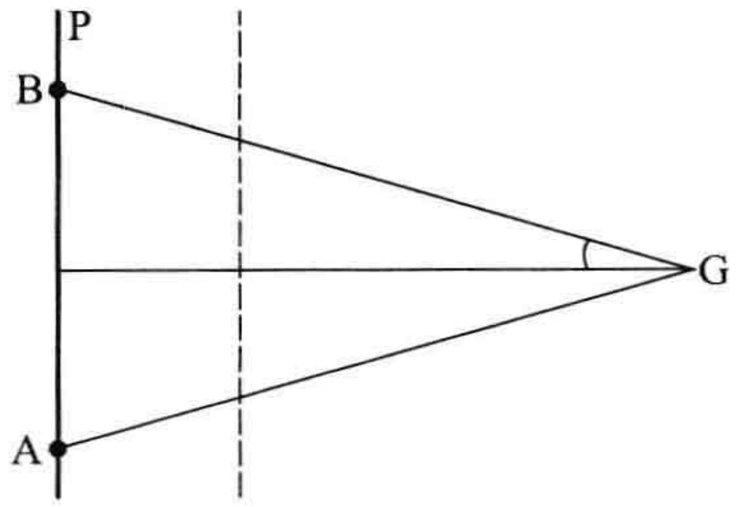


图 4－34　圆周 C 所发光子与平面 P“齐步走”。

$$\text{边缘光线集} = \mathscr{P} \cap \mathscr{C}。\text{（见图 4－35）} \qquad (4-4-12)$$

光锥面 $\mathscr{C}$ 与同时面 Σ_1 的交集 $\mathscr{C} \cap \Sigma_1$ 是个 2 维球面（但在图 4－35 中画成 1 维圆周），不妨取作 G 在 t_1 时刻的天球。由于所有边缘光线在天球上的印记构成视觉形象的轮廓，所以

$$\text{G 看到的视觉形象（轮廓）} = \text{G 的天球} \cap \text{边缘光线集}$$
$$= (\mathscr{C} \cap \Sigma_1) \cap (\mathscr{P} \cap \mathscr{C}) = (\mathscr{C} \cap \Sigma_1) \cap \mathscr{P}。 \qquad (4-4-13)$$

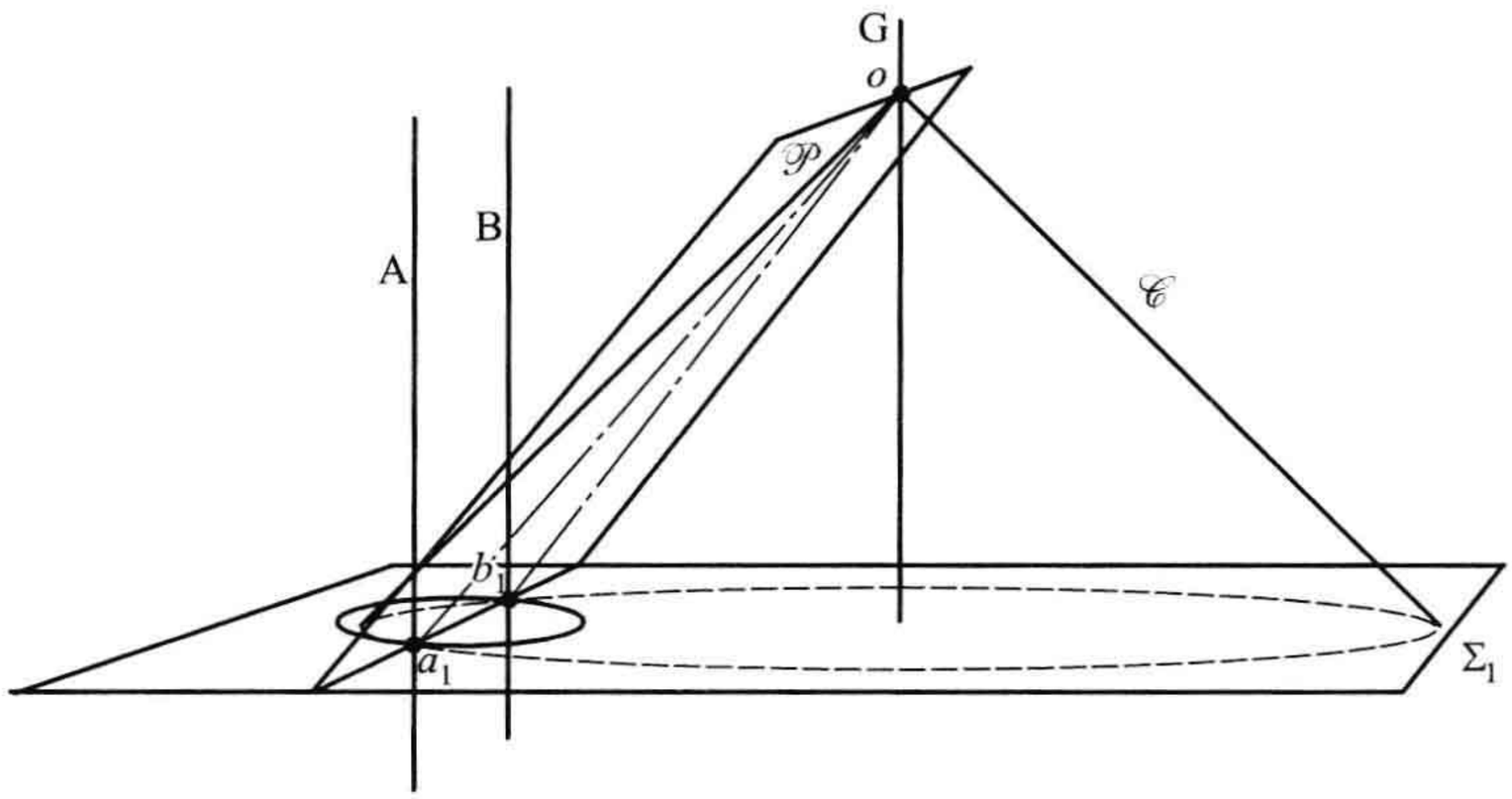


图4－35　边缘光线集 $= \mathscr{P} \cap \mathscr{C}$。由于压缩一维，图中只能显示两条边缘光线，即 a_1o 和 b_1o。

因为 $\mathscr{C} \cap \Sigma_1$ 是 2 维球面（图 4－35 中画成 1 维圆周），而 $\mathscr{P}$ 是 3 维平面（图中画成 2 维平面），所以 $\mathscr{C} \cap \Sigma_1$ 与 $\mathscr{P}$ 的交集 $(\mathscr{C} \cap \Sigma_1) \cap \mathscr{P}$ 是 1 维圆周（图中画

成两个点 a_1, b_1），可见问题 1 的答案是

G 看到的视觉形象（轮廓）=1 维圆周。

这个答案是不证自明的，之所以要花费如此篇幅，是因为它给问题 2 的回答做了极好的铺垫。问题 2 是：与球相对运动的观者 G′看到什么？选读前（例 2）的讨论是站在观者的立场上，认为球在运动。现在用时空图讨论，不如站在球的立场上，认为球静止而观者 G′运动（可沿任何方向以任何速率 $v<1$ 运动）。就是说，仍以球体所在惯性系 K 为基准画时空图，只是用运动观者 G′代替静止观者 G。在这一时空图中，球面 ψ 的世界面仍是竖直圆柱面（仍是静止球面），但运动观者 G′的世界线要画成斜直线。设 G′与 G 的世界线交于 o 点（只需在图 4－33 中补画一条过 o 点的斜直线），这意味着 G′与 G 相遇于事件 o，两人都在事件 o 看到球面 ψ 的视觉形象，只是 G′认为这是运动球面的形象。无论是 G 还是 G′，由于共用事件 o，对视觉形象轮廓造成贡献的是同一个（绝对的）边缘光线集 $\mathscr{P}\cap\mathscr{C}$ ①，集合里的光线由圆周 C 上各点在时刻 $t_1\equiv t_0-l/c$（K 系的时间坐标）同时发出（不过对 G′所在系 K′而言一般就不同时）。于是仿照前面的讨论可知

G′看到的视觉形象（轮廓）=G′的天球∩边缘光线集 =G′的天球∩（$\mathscr{P}\cap\mathscr{C}$）。

这无非是将式（4－4－13）中 G 的天球换成 G′的天球。观者 G 可取同时面 Σ_1 与 $\mathscr{C}$ 的交集为天球，观者 G′则应取他所在系 K′的某张同时面 Σ_1' 与 $\mathscr{C}$ 的交集为天球。既然 G′线是斜直线，同时面 Σ_1' 就是斜置平面（是 3 维平面压缩一维的结果）。于是

$$\text{G′看到的视觉形象（轮廓）}=(\mathscr{C}\cap\Sigma_1')\cap(\mathscr{P}\cap\mathscr{C})=(\mathscr{C}\cap\Sigma_1')\cap\mathscr{P}。\tag{4-4-14}$$

这会带来什么不同？先降一维讨论，并暂时按 3 维欧氏几何考虑。这时 $\mathscr{C}\cap\Sigma_1$ 是 2 维圆锥面 $\mathscr{C}$ 与 2 维平面 Σ_1 的交集，而且 Σ_1 垂直于锥面 $\mathscr{C}$ 的轴，所以交集 $\mathscr{C}\cap\Sigma_1$ 是圆周。然而平面 Σ_1' 不垂直于锥面 $\mathscr{C}$ 的轴，故 $\mathscr{C}\cap\Sigma_1'$ 是椭圆而不是圆周。这似乎就出问题。然而时空图涉及的是闵氏（而不是欧氏）几何（这是关键！），其特点是光锥面在惯性坐标变换下形状不变（光锥面是纯几何对象），就是说，在以 K′为基准系的时空图中，光锥面与图 4－33 的完全一样，但 G′的世界线是竖直线，同时面 Σ_1' 是水平面，由此便知 $\mathscr{C}\cap\Sigma_1'$ 同样是圆周（补上一维就是 2 维球面），于是（$\mathscr{C}\cap\Sigma_1'$）∩$\mathscr{P}$ 同样是两个点（a_1' 和 b_1'），补上一维就是圆周。所以可得结论：

---

① 虽然在引入 $\mathscr{P}$ 时借用过观者 G（图 4－32），但一经引入，$\mathscr{P}$ 就是 4 维时空中的一个 3 维平面，其关键好处就是用它与光锥面 $\mathscr{C}$ 的交集来确定边缘光线集（边缘光线同时到达 o 点，G 和 G′看到的边缘光线都是这些光线）。

G′看到的视觉形象(轮廓)=1维圆周。

最后用对话形式讲解一个重要问题,其中乙代表笔者。

甲　本选读自身是很清楚的,但似乎与前面例2的讨论有矛盾。本选读认为,无论对观者G还是G′,边缘光线的发光事件是相同的(例如a_1,b_1),但在例2中,擦边可见点对观者G是A,B,对观者G′则是A′,B′(见图4-27)。这不是矛盾吗?

乙　其实两者是一致的,只不过由于例2使用3维语言而不容易抓住绝对的东西,因而不容易看出两者的一致性。你先回答一个问题:例2中的事件$p_1=(t_1,\mathrm{A}')$对应于图4-33的哪个时空点?

甲　这正是我的问题所在。我在图4-33中找不到一条世界线,使得p_1在此线上。

乙　竖直线A就是你要找的世界线,线上的时空点a_1就是事件p_1。类似地,例2的$p_2=(t_2,\mathrm{B}')$就是图4-33的时空点b_1。

甲　但a_1,b_1是同时事件(时刻都是t_1)而p_1,p_2却不同时($t_1\neq t_2$),a_1,b_1怎能是p_1,p_2?

乙　这是同时性的相对性使然。事件a_1和b_1相对于球体(因而G)所在系K为同时,相对于G′所在系K′就不同时。$p_1=(t_1,\mathrm{A}')$和$p_2=(t_2,\mathrm{B}')$中的t_1和t_2都是例2中观者所在系的时间坐标,而该观者认为球体在动,所以相当于本选读的K′系,就是说,t_1和t_2是K′系的坐标(没加撇的原因是当时认为球在动而观者静止),所以$t_1\neq t_2$。结论:a_1,b_1对K系为同时事件与p_1,p_2对K′系为不同时事件没有矛盾。

甲　图4-30的A′点是A点做特列尔转动的结果,用时空图的讨论怎么见不到特列尔转动?

乙　这就涉及特列尔转动的起源——相对论光行差现象。假定你在有雨无风的室外撑伞站立。由于无风,你看见雨点竖直下落,所以你的伞柄应该竖直。但是,如果你撑伞快速骑车,你会觉得雨点倾斜而下,你的伞柄就应与地面斜交。把雨点改为光子也有类似的定性结果:设观者G和G′有相对运动,则他们观看同一光线时会测得来光方向有所不同。这种现象称为光行差(aberration),在非相对论物理学中早已存在,但在相对论中有不同表达式。在每个观者的天球面上取角坐标θ,φ(对应于纬度和经度),设观者G′沿$\theta=0=\theta'$的方向相对于G以速率v运动,与G相遇于事件o。外来光子到达G和G′前在每个天球面上留下一点作为印记,两点的角坐标θ,φ和θ',φ'服从如下关系(称为相对论光行差公式):

$$\sin\theta' = \frac{(1-v^2)^{1/2}\sin\theta}{1+v\cos\theta}, \quad \varphi' = \varphi。 \tag{4-4-15}$$

(若用非几何单位制,则 v 改为 v/c。) 利用上式可以证明[见 Terrell(1959)],只要运动物体对观者的张角足够小,物体发出的光线在两个天球面上的印记组成的图像就是相似图形,不过两者之间转过一个角度 $\theta'-\theta$,这就是特列尔转动。请注意这两个图像是同一组光线在不同天球面上的表现,天球面是相对的(依赖于观者),但光线是绝对的(对两个观者都是这组光线,即前面说的有用光线集)。总之,特列尔转动来自光行差,而光行差是3维语言(空间语言)的产物;用时空图讨论时重点是抓住绝对的东西,所以没有突出特列尔转动。

甲　光行差公式(4-4-15)如何推导?

乙　最简单和最清晰的推导是用4维语言,但限于本书的既定范围(不出现4维公式),只好从略。

甲　例1和例2的特列尔转动角 β 可用光行差公式证明吗?

乙　当然可以。对例1和例2,$\theta=0=\theta'$ 的方向就是物体的运动方向,即朝右,而光子的方向朝下,所以光子在天球面上的印记(点)的坐标为 $\theta=\pi/2$,代入式(4-4-15)得

$$\sin\theta' = (1-v^2)^{1/2},$$

因而特列尔转动角 $\beta\equiv\theta'-\theta$ 的正弦为

$$\sin(\theta'-\theta) = \sin(\theta'-\pi/2) = \cos\theta' = \sqrt{1-\sin^2\theta'} = v = v/c,$$

这正是式(4-4-1)和式(4-4-11)。

甲　在讨论视觉形象时,用时空语言比用3维语言(空间语言)有什么主要优点?

乙　至少可列出以下几点。

(1) 例1和例2只讨论了张角很小的情况,而本选读适用于任意张角。

(2) 例2只讨论了观者视线与球体速度正交的情况,而本选读适用于任意情况,即G′在图4-33中的世界线可以是过 o 点的任一斜直线(只需保证类时性)。

(3) 用时空图可以抓住所有绝对的、与观者无关的东西,例如 o 点的过去光锥面以及有用光线集和边缘光线集(因而能跟踪物体每点所发的光子),而3维语言(空间语言)的讨论则容易陷入同时面的图像不能自拔,看不清若干问题的实质,如坠五里雾中,例如一见到特列尔转动就误以为G和G′看到的光线不同,而时空图4-33则清晰地告诉你两者看到的光线一样(有用光线集是绝对的,不因观者而异)。

[选读4-6完]

§4.5　用几何语言导出洛伦兹变换[选读]

爱因斯坦在第一篇相对论论文(1905)中从以下两点出发推出了洛伦兹变换:① 相对性原理和光速不变原理(狭义相对论的两个基本假设);② 空间和时间的均匀性(关于空间和时间性质的假设)。现在,在学过了用闵氏几何表述的狭义相对论后,我们愿意用几何语言重新推出洛伦兹变换,这在某种意义上也可看作是一种"自洽性"验证。推导的关键出发点是:时空的背景几何是闵氏几何,其线元在任一惯性坐标系$\{t,x,y,z\}$的表达式为

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 。$$

在几何语言中,惯性坐标系$\{t,x,y,z\}$的空间坐标原点可看作一个惯性观者(记作G),其世界线就是时空图中的t坐标轴。我们要推导的洛伦兹变换涉及两个有最简关联的惯性系,满足如下条件:①$\{t',x',y',z'\}$系相对于$\{t,x,y,z\}$系以亚光速$v>0$沿x轴正向匀速平动;②空间坐标原点G和G′在$t'=t=0$时相遇;③ 两系空间坐标轴对应同向。有待证明的是

$$t'=\gamma(t-vx),\ x'=\gamma(x-vt),\ y'=y,\ z'=z,\ \gamma\equiv 1/\sqrt{1-v^2} 。 \tag{4-5-1}$$

上式实质上只涉及t,x与t',x'之间的变换,所以只需就2维闵氏时空给出证明。设p是任一时空点,只需证明

$$t'_p=\gamma(t_p-vx_p),\quad x'_p=\gamma(x_p-vt_p) 。 \tag{4-5-2}$$

图4-36是以$\{t,x\}$为基准系的时空图,其中G′线满足$t=v^{-1}x$。因为约定两系空间坐标原点G和G′在$t'=t=0$时相遇,故图中q点满足$t'_q=t_q=0$。将闵氏线元用于光子世界线有

$$0=\mathrm{d}s^2=-\mathrm{d}t^2+\mathrm{d}x^2 ,$$

所以$\mathrm{d}t/\mathrm{d}x=\pm 1$,表明过任一时空点$p$的光子世界线有两条,分别是斜率为$\pm 1$的直线。把斜率为$+1$的那一条记作A线,另一条记作B线。以$a$代表A线与G线的交点,由直线方程的"两点式"得$t_p-t_a=x_p-x_a$。因$a$在G线($t$轴)上,故$x_a=0$,所以

$$t_a=t_p-x_p 。 \tag{4-5-3}$$

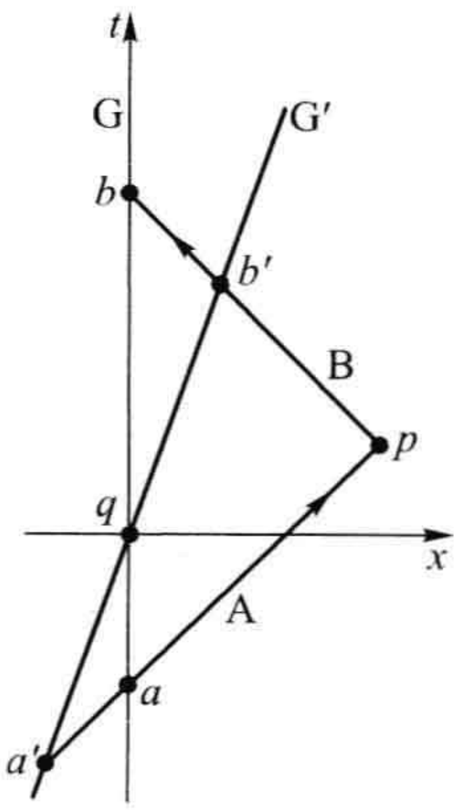


图4-36　洛伦兹变换推导用图

以$\mathrm{d}\tau$代表G线任一元段的线长(固有时间),由类时曲线的线长定义可知

$$\mathrm{d}\tau = \sqrt{-\mathrm{d}s^2} = \sqrt{\mathrm{d}t^2 - \mathrm{d}x^2} = \mathrm{d}t \text{ 。}$$

选 q 点为 τ 的零点(选 $\tau_q = 0$),则 G 线有 $\tau = t$,故式(4-5-3)可改写为

$$\tau_a = t_p - x_p \text{。} \qquad (4-5-4\mathrm{a})$$

类似地,以 b 代表 B 线与 G 线的交点,注意到 B 线的斜率为 -1,又有

$$\tau_b = t_p + x_p \text{。} \qquad (4-5-4\mathrm{b})$$

由上两式易得

$$t_p = (\tau_b + \tau_a)/2, \quad x_p = (\tau_b - \tau_a)/2 \text{。} \qquad (4-5-5\mathrm{a})$$

若改用 $\{t', x'\}$ 系为基准画时空图,则光子世界线 A 和 B 仍是斜率为 ±1 的直线。以 a' 和 b' 分别代表 A 线和 B 线与 G′线的交点,以 τ' 代表 G′线的固有时(选 q 为零点,即 $\tau'_q = 0$),仿照前面的推导又得

$$t'_p = (\tau'_{b'} + \tau'_{a'})/2, \quad x'_p = (\tau'_{b'} - \tau'_{a'})/2 \text{。} \qquad (4-5-5\mathrm{b})$$

如果能再证明(稍后将补证)

$$(\mathrm{a})\ \tau'_{a'} = \tau_a/\gamma(1-v), \quad (\mathrm{b})\ \tau'_{b'} = \tau_b/\gamma(1+v), \qquad (4-5-6)$$

代入式(4-5-5b)便得待证等式(4-5-2):

$$t'_p = \frac{1}{2}\left[\frac{\tau_b}{\gamma(1+v)} + \frac{\tau_a}{\gamma(1-v)}\right] = \frac{1}{\gamma}\,\frac{1}{1-v^2}\left[\frac{1}{2}(\tau_b + \tau_a) - \frac{1}{2}v(\tau_b - \tau_a)\right] = \gamma(t_p - vx_p), \qquad (4-5-7\mathrm{a})$$

$$x'_p = \frac{1}{2}\left[\frac{\tau_b}{\gamma(1+v)} - \frac{\tau_a}{\gamma(1-v)}\right] = \frac{1}{\gamma}\,\frac{1}{1-v^2}\left[\frac{1}{2}(\tau_b - \tau_a) - \frac{1}{2}v(\tau_b + \tau_a)\right] = \gamma(x_p - vt_p), \qquad (4-5-7\mathrm{b})$$

其中末步用到式(4-5-5a)以及 $\gamma \equiv 1/\sqrt{1-v^2}$。

下面补证式(4-5-6)。G′线的 $a'q$ 段的线长等于

$$\tau'_q - \tau'_{a'} = \int_{a'}^{q} \sqrt{-\mathrm{d}s^2} \text{。}$$

线长是几何量(绝对量),可借任一坐标系计算。借 $\{t, x\}$ 系的计算给出

$$\tau'_q - \tau'_{a'} = \int_{a'}^{q} \sqrt{\mathrm{d}t^2 - \mathrm{d}x^2} = \int_{a'}^{q} \sqrt{\mathrm{d}t^2 - (v\mathrm{d}t)^2}$$

$$= \sqrt{1-v^2}\,(t_q - t_{a'}),$$

其中第二步是因为 G′线满足 $t = v^{-1}x$。利用 $t_q = 0$ 及 $\tau'_q = 0$ 便得

$$\tau'_{a'} = \sqrt{1-v^2}\,t_{a'} = \gamma^{-1}t_{a'} \text{。} \qquad (4-5-8)$$

a 及 a' 都在 A 线上,由 A 线的“两点式”得

$$t_a = t_{a'} - (x_{a'} - x_a) = t_{a'} - vt_{a'} = (1-v)t_{a'},$$

其中第二步用到 $x_{a'} = vt_{a'}$ 及 $x_a = 0$。再代回式(4-5-8),利用 $t_a = \tau_a$ 便得待证

式(4-5-6a)。同理可证式(4-5-6b)。整个证明至此结束。

在推得式(4-5-5a)后,细心的读者会问:由图4-36可知 $\tau_b > \tau_a$,从式(4-5-5a)看出 x_p 恒为正。然而 p 点的 x 坐标当然可正可负(还可为零),岂非矛盾?其实,图4-36只示出 $x_p > 0$ 的情况,因为 p 点位于G线右边。$x_p < 0$ 对应于 p 点位于G线左边的情况,只要坚持"把斜率为+1的光子世界线记作A"的约定,便有 $\tau_a > \tau_b$(见图4-37),式(4-5-5a)就给出 $x_p < 0$(当 p 点位于G线上时还给出 $x_p = 0$)。可见式(4-5-5a)总是对的。往往遇到这种情况:解析推导是普适的,但画图只能表现部分可能情况。当然应以推导为准。

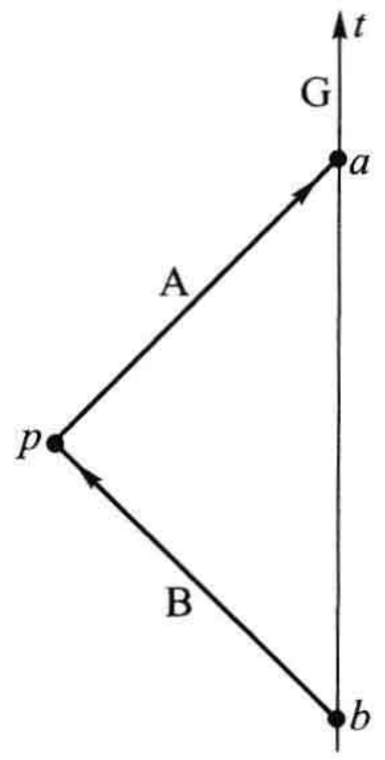


图4-37　p 点在G线左边时 $\tau_a > \tau_b$

第4章习题

1. 惯性系K认为做共线运动的粒子1和2的速率分别为 $u_1 = 0.8\,c$ 和 $u_2 = 0.6\,c$,在 $t = 0$ 时两者相距10 m,在 $t = \hat{t}$ 时1追上2。画出时空图并由图计算(a)$\hat{t}$ 值;(b)粒子1和2(用自己的标准钟)测得的、与 $\hat{t}$ 相应的时间 $\hat{\tau}_1$ 和 $\hat{\tau}_2$。

2. 在实验室中以0.8 c 的速率做直线运动的粒子飞行3 m后衰变,与该粒子一起运动的观者测得它在衰变前存活了多长时间?

3. 地球人发现一艘以0.8 c 运动的飞船将在10 s内撞上迎面飞来的彗星。飞船人认为还有多少时间可以改变航向?

4. 从惯性系K看来(认为,测得),位于某地A的两标准钟甲、乙指零时开始以速率 $v = 0.6\,c$ 一同做匀速直线运动,两钟指1秒时到达某地B。甲钟在到达B时立即以速率 v 向A地匀速返回并停在A地,乙钟在B地停留1秒后以速率 v 向A地匀速返回。另有丙钟一直待在A地,且当甲、乙离A地时也指零。(a)画出甲、乙、丙的世界线;(b)求乙钟返回A地时三钟的读数 $\tau_{甲}$、$\tau_{乙}$ 和 $\tau_{丙}$。

5. 标准钟A、B静止于某惯性系的同一空间点。A钟从某时刻开始以速率 $u = 0.6\,c$ 匀速直线飞出,2 s(根据A钟)后以 u 匀速直线返航。已知分手时两钟皆指零。(a)求重逢时两钟的读数;(b)当A钟指3 s时A看见B钟指多少?

6. 惯性钟C和C′相对速率 $u = 0.6\,c$,相遇时读数皆为零.(a)携带C钟的观者G在两钟相遇后10 s时用眼睛看到C′钟的读数是多少?(b)G在两钟相遇前10 s时用眼睛看到C′钟的读数是多少?

7.(单选题)双生子A、B静止于某惯性系K中的同一空间点上。A从某时刻(此时A、B年龄相等)开始向东以速率 u 相对于K系做惯性运动,一段时间后B以速率 $v > u$ 向东追上

A,则相遇时 A 的年龄

(a) 比 B 大,　　(b) 比 B 小,　　(c) 与 B 等。

(提示:画出时空图,利用测地线可以一望而知答案,无需计算。)

8. 远方星体以 0.8 c 的速率(匀速直线地)离开我们,我们测得它辐射来的闪光按 5 昼夜的周期变化。用时空图求星上观者测得的闪光周期。

9. 改以 K′系为基准系重画尺缩效应的时空图,并借此图证明静尺长等于动尺长的 γ 倍。

10. 静长 $l=5$ m 的汽车以 $u=0.6\,c$ 的速率匀速进库,库有坚硬后墙。为简化问题,假定车头撞墙的信息以光速传播,车身任一点接到信息立即停下。(a)设司库测得在车头撞墙的同时车尾的钟 $C_{尾}$ 指零,求车尾"获悉"车头撞墙这一信息时 $C_{尾}$ 的读数;(b)求车完全停下后的静长 $\hat{l}$;(c)用 u 表出新旧静长比 $\hat{l}/l$。

11. 惯性观者 G 和 G′相对速率为 $u=0.6\,c$,相遇时把钟读数都调为零。用时空图讨论:(a)在 G 所属的惯性参考系看来(以其同时观判断),当 G 钟读数为 5 μs 时,G′钟的读数是多少?(b)当 G 钟读数为 5 μs 时,他看见 G′钟的读数是多少?

12. (a)求图 4 - 13 的 τ'/τ 值,即证明式(4 - 2 - 16)。(b)在 $u=0.6\,c$ 和 $u=0.8\,c$ 两种情况下求出 τ'/τ 的数值。

13. 惯性质点 A、B、C 排成一直线并沿此线相对运动(图 4 - 38),相对速率 $u_{BA}=0.6\,c$,$u_{CA}=0.8\,c$,A、B 所在惯性系各为 K_A 和 K_B。设 K_B 系认为(测得)C 走了 60 m,画出时空图并求 K_A 认为(测得)这一过程的时间 Δt。

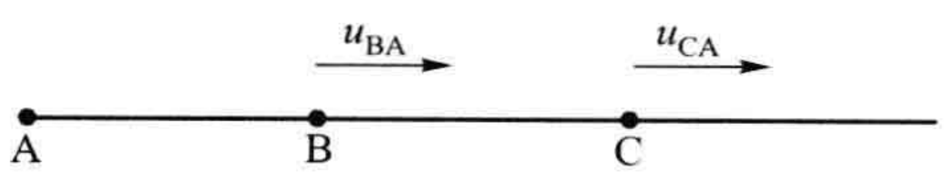


图 4 - 38　习题 13 用图

14. A、B 是同一惯性系的两个惯性观者,他们互相发射中子,每一中子以相对速率 0.6 c 离开中子枪。设 B 测得 B 枪的中子发射率为 $10^4\ s^{-1}$(即每秒发 10^4 个),求 A 所发中子(根据自己的标准钟)测得的 B 枪的中子发射率(要求画时空图求解)。

15. 惯性系 K 和 K′相对速率为 v。p_1 和 p_2 对 K 系而言是同地事件,它们的时间间隔对 K 系和 K′系分别为 4 s 和 5 s。(a)求 v;(b)求 K′系测得的、p_1 与 p_2 的空间间隔(距离)。

16. 补证式(4 - 2 - 3)。

第5章　相对论质点力学

§5.1　动量和动量守恒

在牛顿力学中，质点的动量 $\boldsymbol{p}$ 定义为质量 m 与速度 $\boldsymbol{u}$ 的乘积：

$$\boldsymbol{p}\equiv m\boldsymbol{u}, \tag{5-1-1}$$

质点所受的力定义为它的动量的时间变化率，即

$$\boldsymbol{f}\equiv\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}。 \tag{5-1-2}$$

两式结合得

$$\boldsymbol{f}=\frac{\mathrm{d}(m\boldsymbol{u})}{\mathrm{d}t}=m\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t}, \tag{5-1-3}$$

而质点的加速度 $\boldsymbol{a}\equiv\mathrm{d}\boldsymbol{u}/\mathrm{d}t$，所以上式又可表为

$$\boldsymbol{f}=m\boldsymbol{a}。 \tag{5-1-4}$$

反之，由 $\boldsymbol{f}=m\boldsymbol{a}$ 也容易推出 $\boldsymbol{f}=\mathrm{d}\boldsymbol{p}/\mathrm{d}t$，因此也可把 $\boldsymbol{f}=m\boldsymbol{a}$ 看作 $\boldsymbol{f}$ 的定义，就是说，质点所受的力定义为它的质量与加速度的乘积。读者会问：$\boldsymbol{f}=m\boldsymbol{a}$ 不是人所共知的牛顿第二定律吗？怎么说成是力的定义？选读 5－1 对这个问题将有比较详细的讨论。

现在从相对性原理的角度考察两个小球的碰撞。先看牛顿力学。以 $\boldsymbol{p}_1$ 及 $\boldsymbol{f}_1$ 分别代表球 1 的动量及所受的力（来自球 2），则 $\boldsymbol{f}_1=\mathrm{d}\boldsymbol{p}_1/\mathrm{d}t$。同理还有 $\boldsymbol{f}_2=\mathrm{d}\boldsymbol{p}_2/\mathrm{d}t$。牛顿第三定律要求 $\boldsymbol{f}_1=-\boldsymbol{f}_2$，于是

$$0=\boldsymbol{f}_1+\boldsymbol{f}_2=\frac{\mathrm{d}\boldsymbol{p}_1}{\mathrm{d}t}+\frac{\mathrm{d}\boldsymbol{p}_2}{\mathrm{d}t}=\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t},$$

其中 $\boldsymbol{p}\equiv\boldsymbol{p}_1+\boldsymbol{p}_2$ 代表由两球组成的系统的总动量。上式表明系统的总动量不随时间而变，所以动量守恒。可见动量守恒是力的定义同牛顿第三定律相结合的产物。如果从另一惯性系观测同一碰撞过程，利用由伽利略变换导出的速度变换式（1－1－6），不难看出动量仍然守恒。所以动量守恒的确有伽利略协变性，因而的确是（牛顿力学中的）定律。然而，如果在狭义相

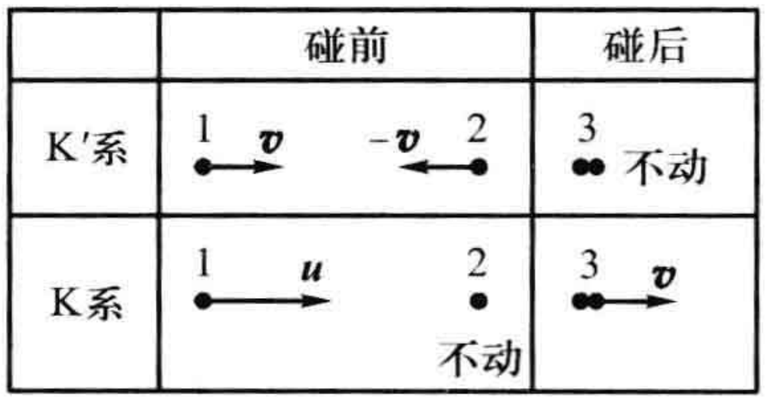


图 5－1　全同小球的完全非弹性碰撞

对论中仍然采用动量的牛顿定义，即 $\boldsymbol{p}\equiv m\boldsymbol{u}$（其中 m 为常量），则下面的例子足以说明动量守恒不具备洛伦兹协变性。考虑两个小球的完全非弹性碰撞。设 K′系测得两球碰前的速度等值反向（因而总动量为零），则由对称性可知碰后速度皆为零，表明碰撞过程对 K′系而言动量守恒。设 K 是另一惯性系，球 2 相对于 K 系静止，我们再从 K 系考察同一碰撞过程（见图 5－1）。首先求出碰后所得的复合体 3 对 K 系的速度（简记作“3 对 K”）：

$$3\text{ 对 K}=3\text{ 对 }2=\text{ K′对 }2=-(2\text{ 对 K′})=\boldsymbol{v},$$

再求球 1 对 K 系的速度（记作 $\boldsymbol{u}$）。以 $\boldsymbol{u}'$代表 1 对 K′系的速度，则由图知 $\boldsymbol{u}'=\boldsymbol{v}$。以 u 简记 u_x，利用相对论速度变换式（1－3－3）得

$$u\equiv u_x=\frac{u'_x+v}{1+vu'_x/c^2}=\frac{v+v}{1+vv/c^2}=\frac{2v}{1+v^2/c^2}。\tag{5-1-5}$$

注记 1　本章许多公式都含有 c 和 c^2（甚至 c^4），为了显示这些系数，本章一律采用国际单位制。

设两球的质量皆为 m，则 K 系测得的两球总动量在碰撞前后各为

$$\text{碰前总动量（大小）}=mu+0=\frac{2mv}{1+v^2/c^2},$$

$$\text{碰后总动量（大小）}=2mv。\text{（用到牛顿质量守恒律）}$$

上两式表明碰撞前后总动量不等，所以动量守恒对 K 系不成立。然而由图可知对 K′系而言动量显然是守恒的，可见动量守恒不具备洛伦兹协变性，因而不是定律。这时有两种选择，或者放弃动量守恒律，或者通过修改有关物理量的定义给动量守恒赋予洛伦兹协变性。鉴于守恒律对物理学的非常重要性，当然选择后者。为了找到修改思路，先做如下考虑。设质点被恒力加速，按照牛顿第二定律，只要时间足够长，其速率必将超过光速，与狭义相对论相悖。为摆脱矛盾，不妨猜测质点的质量在相对论中随速率增大而增大。（因为如果这样，质点在恒力下的加速度将越来越小，速率就有望永不达到光速。）于是想到这样的修改方案：动量仍定义为质量乘速度，但质量不再是常量而与速率 u 有关，记作 m_u（恒为正），称为**运动质量**（亦称**相对论质量**）。现在沿着这一思路重新审查图 5－1 中 K 系的动量守恒问题。既然碰前球 2 静止而球 1 以速率 u 运动，两者的运动质量应分别为 m_0（称为**静质量**）和 m_u，故

$$\text{碰前总动量（大小）}=m_uu+0=\frac{2m_uv}{1+v^2/c^2},\tag{5-1-6}$$

$$\text{碰后总动量（大小）}=M_vv,\tag{5-1-7}$$

其中 M_v 代表两球复合体 3 的运动质量。默认碰撞前后总的运动质量不变，即 $m_u+m_0=M_v$［这是很自然的默认，其深刻含义将在式（5－4－5）后关于能量守

恒的一段论述中阐明],则式(5-1-7)成为

$$\text{碰后总动量(大小)} = (m_u + m_0)v, \tag{5-1-8}$$

对比式(5-1-6)和式(5-1-8)可知,为使动量守恒对K系也成立,必须且只需

$$m_u = m_0 \frac{1 + v^2/c^2}{1 - v^2/c^2}, \tag{5-1-9}$$

而由式(5-1-5)出发的简单计算表明

$$\sqrt{1 - \frac{u^2}{c^2}} = \frac{1 - v^2/c^2}{1 + v^2/c^2}, \tag{5-1-10}$$

与式(5-1-9)对比便得

$$m_u = \frac{m_0}{\sqrt{1 - u^2/c^2}}。 \tag{5-1-11}$$

可见只有承认 m_u 随速率 u 按上式变化方可保证图5-1的碰撞过程对K系有动量守恒。所以狭义相对论中质点的动量应定义为

$$\boldsymbol{p} \equiv m_u \boldsymbol{u}, \text{ 其中 } m_u \text{ 由式(5-1-11)给出。} \tag{5-1-12}$$

以上虽然只就图5-1的过程做了讨论,但后来发现这样定义的动量在任何过程中都是守恒量。

令

$$\gamma_u \equiv 1/\sqrt{1 - u^2/c^2}, \tag{5-1-13}$$

则相对论动量亦可表为

$$\boldsymbol{p} = \gamma_u m_0 \boldsymbol{u} = \frac{m_0 \boldsymbol{u}}{\sqrt{1 - u^2/c^2}}。 \tag{5-1-14}$$

相对论动量(矢量)的大小则为

$$p = \gamma_u m_0 u = \frac{m_0 u}{\sqrt{1 - u^2/c^2}}。 \tag{5-1-15}$$

若质点的速率 $u = c$,则上式右边(以零为分母的分数)无意义;又若质点的速率 $u > c$,则上式右边为虚数,但物理上要求质点的动量有意义且其大小为实数,所以应有 $u < c$,可见质点必须亚光速。

§5.2　力的定义

有了动量定义就可对力下定义。在狭义相对论中仍把质点所受的力 $\boldsymbol{f}$ 定义为质点动量的时间变化率:

$$\boldsymbol{f}\equiv \mathrm{d}\boldsymbol{p}/\mathrm{d}t,\ \text{其中}\ \boldsymbol{p}\ \text{由式(5-1-14)定义。} \qquad (5-2-1)$$

相对性原理要求上式有洛伦兹协变性,就是说,对另一惯性系 K′应有

$$\boldsymbol{f}'\equiv \mathrm{d}\boldsymbol{p}'/\mathrm{d}t'。 \qquad (5-2-1')$$

从以上两式出发,由 $\boldsymbol{p}'$ 与 $\boldsymbol{p}$ 以及 t' 与 t 的变换关系就可求得 $\boldsymbol{f}$ 与 $\boldsymbol{f}'$ 的变换关系,略。

[选读 5-1]

牛顿第二定律是众所周知的。为了证明它是物理定律,中学老师会用演示仪器分别测定物体的质量 m、所受的力 f 以及获得的加速度 a,列成表格,从中看出 f 值总等于 m 与 a 值的乘积。然而你可曾注意到,在做这类实验之前,老师并未对力和质量(作为两个物理量)下过定量的定义?老师和课本的确说过“力是物体对物体的作用”、“质量是物体所含物质的多少”一类的话,但这都不是(定量的)定义。要对一个物理量下定义,至少应该给出一个在原则上是可行的测量方案(虽然在当时的实验条件下可能难以操作),然而上面关于力和质量的说法完全没有给出任何测量方案。试想,公式 $f=ma$ 涉及的三个量中竟然有两个尚未定义,何谈用实验(测量)来验证这个公式?事实上,演示实验中的测力计和天平(质量测量仪)分别是根据力和质量的定义设计的,在力和质量尚未定义的情况下,怎能知道这些仪器的读数分别是力和质量?这样一来,$f=ma$ 还能算是实验定律吗?实事求是地说,牛顿在发表他的划时代巨著《自然哲学的数学原理》时,对某些微妙的内在逻辑关系并未完全理顺。作为在科研战线上冲锋陷阵的勇士,出现这种纰漏是不难理解的——搜集整理战利品的细致工作不妨留待打扫战场的士兵们完成。事实的确如此。与牛顿运动定律有关的逻辑问题的澄清是后人的贡献(特别是牛顿逝世后约 200 年的马赫)。下面简介关于这一问题的现代认识。

(1) 对质量下定义:以 m_1, m_2 分别代表物体 1 和 2 的、有待定义的质量,设它们碰撞时获得的加速度分别是 a_1 和 a_2(这是早有定义的量),则它们的质量之比 m_1/m_2 定义为

$$m_1/m_2\equiv a_2/a_1。 \qquad (5-2-2)$$

再任选某一物体的质量(记作 m_1)作为质量的单位(令 $m_1=1$),则任一物体的质量(记作 m_2)便可通过与物体 1 的碰撞而测得。这就是关于质量的一个原则可

行的测量方案①。

(2) 对力下定义:物体所受的力 $\boldsymbol{f}$ 定义为它的质量 m 与它的加速度 $\boldsymbol{a}$ 的乘积,即

$$\boldsymbol{f}=m\boldsymbol{a}。$$

照此说来,$\boldsymbol{f}=m\boldsymbol{a}$ 只是力的定义而不是什么定律!许多物理工作者在初次听到这一看法时都十分不满。他们说:"定律是客观存在的,而定义则是人为下的,你把定律降格为定义,岂不是把全部物理'干货'都丢光了吗!?"其实这是只知其一,不知其二。$\boldsymbol{f}=m\boldsymbol{a}$ 本身只是力 $\boldsymbol{f}$ 的定义,与它相联系的所有物理"干货"都体现在不同场合下力的具体表达式中。例如,系在弹簧上的木块所受的弹性力 f 与木块的位移 x 满足胡克定律:

$$f=-kx,\quad (k\text{ 为常数}) \tag{5-2-3}$$

其中的 f 无非是木块的质量 m 与加速度 $a\equiv \mathrm{d}^2x/\mathrm{d}t^2$ 的乘积($f=ma$),与式(5-2-3)结合乃得

$$m\mathrm{d}^2x/\mathrm{d}t^2=-kx, \tag{5-2-4}$$

而这正是木块的动力学方程,包含了这一问题的全部"干货"。又如,电磁场($\boldsymbol{E}$, $\boldsymbol{B}$)中的带电质点所受的电磁力由洛伦兹力公式给出:

$$\boldsymbol{f}=q(\boldsymbol{E}+\boldsymbol{u}\times\boldsymbol{B}),(q\text{ 和 }\boldsymbol{u}\text{ 是质点的电荷和速度}) \tag{5-2-5}$$

其中的 $\boldsymbol{f}$ 无非是质点的质量 m 与加速度 $\boldsymbol{a}\equiv \mathrm{d}\boldsymbol{u}/\mathrm{d}t$ 的乘积,代入上式乃得

$$m\mathrm{d}\boldsymbol{u}/\mathrm{d}t=q(\boldsymbol{E}+\boldsymbol{u}\times\boldsymbol{B})。 \tag{5-2-6}$$

上式可充当电磁场中带电质点的动力学方程,包含了这一问题的全部"干货"。

可见,极端地说(当然我们并不提倡),即使把"力"这个词汇从物理学名词中删除,物理学(包括力学!)照样可以存活,而且不受任何实质性的影响。事实上,越是靠近现代的物理学(例如量子力学),"力"这个词就越是不常出现。

英国相对论学家**伦德勒**在他的书[Rindler (1982)]的 101 页中写下这么一段话(下面是笔者的中译文):

"让我们回顾一下牛顿第二定律的两个熟悉的形式:$\boldsymbol{f}=m\boldsymbol{a}$, $\boldsymbol{f}=\mathrm{d}\boldsymbol{p}/\mathrm{d}t$。严格说来,这只是'半个定律',因为它只是力的定义,直到它与另外的'半定律'结合之前都不具备任何物理内容。所谓另外的'半定律'可以是引力 $\boldsymbol{f}=-Gm_1m_2r^{-3}\boldsymbol{r}$,或者洛伦兹力定律 $\boldsymbol{f}=q(\boldsymbol{E}+\boldsymbol{u}\times\boldsymbol{B})$。把这两个'半定律'结合起来方可给出质点的运动。"

---

① 无独有偶。在库仑定律发表时,物理界对于电荷尚未下过定量的定义。电荷的严格定义是后来的科学家[特别是高斯(Gauss)]给出的,基本做法也是先定义两个点电荷的比值,再指定一个点电荷为电荷单位,详见梁灿彬(2012)第 5,6 页。

当然,伦德勒关于$\boldsymbol{f}=m\boldsymbol{a}$是"半定律"的称谓只是为了形象而采用的戏称,究其实质,它就是一个定义。

再回头讨论式(5-2-2),它是牛顿力学中质量m的定义。无独有偶:严格说来相对论的静质量m_0也应补定义(给出原则可行的测量方案)。考虑两个小球的简单碰撞过程:碰前球2静止而球1向右碰球2,碰后两球都有速度。以m_1和m_2依次代表两球的质量,u_1和u'_1依次代表球1碰撞前后的速度,u'_2代表球2碰后的速度(均指沿右向单位矢的分量),则由牛顿力学的动量守恒律知

$$m_1u_1=m_1u'_1+m_2u'_2, \tag{5-2-7}$$

故

$$m_2=m_1(u_1-u'_1)/u'_2, \tag{5-2-7'}$$

因为u_1、u'_1和u'_2原则上都可测量,选m_1为单位便得m_2。所以上式也可看作牛顿力学中的质量定义。受此启发就不难找到相对论中静质量m_0的定义。利用相对论动量表达式(5-1-14)可以写出上述碰撞过程的动量守恒式:

$$\gamma_{u_1}m_{01}u_1=\gamma_{u'_1}m_{01}u'_1+\gamma_{u'_2}m_{02}u'_2, \tag{5-2-8}$$

因而

$$m_{02}=m_{01}(\gamma_{u_1}u_1-\gamma_{u'_1}u'_1)/\gamma_{u'_2}u'_2, \tag{5-2-8'}$$

测得u_1、u'_1和u'_2后自然有γ_{u_1}、$\gamma_{u'_1}$和$\gamma_{u'_2}$,选m_{01}为单位,上式便可充当球2静质量m_{02}的定义。　　[选读5-1完]

§5.3 动　　能

在牛顿力学中,质点的动能E_k是质点速率u的函数,可记作$E_k(u)$,它是由以下两个要求定义的:①质点静止时动能为零,即$E_k(0)=0$;②E_k的时间变化率等于质点所受的力的功率$\boldsymbol{f}\cdot\boldsymbol{u}$。由此得

$$\frac{\mathrm{d}E_k(u)}{\mathrm{d}t}=\boldsymbol{f}\cdot\boldsymbol{u}=\frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}\cdot\boldsymbol{u}=\boldsymbol{u}\cdot\frac{\mathrm{d}(m\boldsymbol{u})}{\mathrm{d}t}=\frac{1}{2}m\frac{\mathrm{d}(\boldsymbol{u}\cdot\boldsymbol{u})}{\mathrm{d}t}=\frac{1}{2}m\frac{\mathrm{d}u^2}{\mathrm{d}t}=mu\frac{\mathrm{d}u}{\mathrm{d}t}。$$

为了在求定积分时不与积分限的符号相混淆,先把上式的自变量u改记作u',即

$$\frac{\mathrm{d}E_k(u')}{\mathrm{d}t}=mu'\frac{\mathrm{d}u'}{\mathrm{d}t},$$

再从$u'=0$到$u'=u$做定积分得

$$E_k(u)-E_k(0)=m\int_0^u u'\mathrm{d}u'=\frac{1}{2}m(u^2-0)=\frac{1}{2}mu^2。$$

注意到$E_k(0)=0$,便得到动能作为速率的函数的表达式:

$$E_k(u) = mu^2/2。 \tag{5-3-1}$$

在相对论中，力所做的功和功率仍用牛顿力学的定义，质点的动能仍按上述两点要求定义，结果的不同来自动量 $\boldsymbol{p}$ 定义的不同（从 $\boldsymbol{p}\equiv m\boldsymbol{u}$ 改为 $\boldsymbol{p}\equiv m_u\boldsymbol{u}$）：

$$\frac{\mathrm{d}E_k(u)}{\mathrm{d}t} = \boldsymbol{f}\cdot\boldsymbol{u} = \frac{\mathrm{d}\boldsymbol{p}}{\mathrm{d}t}\cdot\boldsymbol{u} = \boldsymbol{u}\cdot\frac{\mathrm{d}(m_u\boldsymbol{u})}{\mathrm{d}t} = m_u\boldsymbol{u}\cdot\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} + \boldsymbol{u}\cdot\boldsymbol{u}\frac{\mathrm{d}m_u}{\mathrm{d}t} = m_u u\frac{\mathrm{d}u}{\mathrm{d}t} + u^2\frac{\mathrm{d}m_u}{\mathrm{d}t}, \tag{5-3-2}$$

其中 $\mathrm{d}m_u/\mathrm{d}t$ 可借用式(5-1-11)表示为

$$\frac{\mathrm{d}m_u}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{cm_0}{\sqrt{c^2-u^2}}\right) = \frac{m_u u}{c^2-u^2}\frac{\mathrm{d}u}{\mathrm{d}t}, \tag{5-3-3}$$

代入式(5-3-2)得

$$\frac{\mathrm{d}E_k(u)}{\mathrm{d}t} = (c^2-u^2)\frac{\mathrm{d}m_u}{\mathrm{d}t} + u^2\frac{\mathrm{d}m_u}{\mathrm{d}t} = c^2\frac{\mathrm{d}m_u}{\mathrm{d}t}。 \tag{5-3-4}$$

先将自变量 u 改记作 u'，即

$$\frac{\mathrm{d}E_k(u')}{\mathrm{d}t} = c^2\frac{\mathrm{d}m_{u'}}{\mathrm{d}t},$$

再从 $u'=0$ 到 $u'=u$ 做定积分得

$$E_k(u) - E_k(0) = c^2\int_{m_0}^{m_u}\mathrm{d}m_{u'} = m_u c^2 - m_0 c^2。$$

注意到 $E_k(0)=0$，便得到动能的相对论表达式：

$$E_k(u) = m_u c^2 - m_0 c^2。 \tag{5-3-5}$$

将上式改写为

$$E_k(u) = \frac{m_0 c^2}{\sqrt{1-u^2/c^2}} - m_0 c^2 = m_0 c^2[(1-u^2/c^2)^{-1/2} - 1], \tag{5-3-6}$$

再把 $(1-u^2/c^2)^{-1/2}$ 展开为泰勒级数：

$$(1-u^2/c^2)^{-1/2} = 1 + \frac{1}{2}(u/c)^2 + \frac{3}{8}(u/c)^4 + \cdots,$$

当 $u \ll c$ 时只保留上式右边前两项（只保留到二阶小项），便得

$$E_k(u) \approx m_0 c^2\left[1 + \frac{1}{2}\left(\frac{u}{c}\right)^2 - 1\right] = \frac{1}{2}m_0 u^2,$$

可见相对论动能在 $u \ll c$ 时近似回到牛顿动能，这再次从一个侧面表明牛顿力学是狭义相对论的低速近似。

[选读 5 -2]

前已指出，只要用运动质量 m_u 代替牛顿动量 $\boldsymbol{p}=m\boldsymbol{u}$ 中的 m，所得结果便是相对论动量 $\boldsymbol{p}=m_u\boldsymbol{u}$。有鉴于此，某些初学者以为只要用 m_u 代替牛顿动能 $E_k=mu^2/2$ 的 m 便得相对论动能，然而这是误解。因为

$$E_k=m_uc^2-m_0c^2=m_uc^2(1-m_0/m_u)=m_uc^2[1-(1-u^2/c^2)^{1/2}],\tag{5-3-7}$$

把 $(1-u^2/c^2)^{1/2}$ 展开为泰勒级数：

$$(1-u^2/c^2)^{1/2}=1-\frac{1}{2}(u/c)^2-\frac{1}{8}(u/c)^4-\frac{1}{16}(u/c)^6-\cdots,$$

再代回式(5 -3 -7)给出

$$E_k=m_uc^2\left[\frac{1}{2}\left(\frac{u}{c}\right)^2+\frac{1}{8}\left(\frac{u}{c}\right)^4+\frac{1}{16}\left(\frac{u}{c}\right)^6+\cdots\right]。\tag{5-3-8}$$

在非低速情况下高阶项不能忽略，所以 $E_k\neq m_uu^2/2$。可见，非低速情况下的相对论动能既不能表为 $m_0u^2/2$，也不能表为 $m_uu^2/2$。　**[选读 5 -2 完]**

§5.4　能量和能量守恒

质点动能的相对论表达式(5 -3 -5)可改写为

$$m_uc^2=E_k(u)+m_0c^2。\tag{5-4-1}$$

我们至今只知道 $E_k(u)$ 是质点的动能，因而 m_uc^2 和 m_0c^2 都有能量的量纲。它们有什么物理意义吗？爱因斯坦大胆地把 m_uc^2 解释为质点在速率为 u 时的能量，记作 E，即

$$E=m_uc^2,\tag{5-4-2}$$

于是 m_0c^2 自然就是速率为零时的能量，称为**静能**(rest energy)，记作 E_0，即

$$E_0=m_0c^2,\tag{5-4-3}$$

所以就有

$$E=E_k+E_0,\tag{5-4-4}$$

即质点的能量 E 等于动能加静能(所以也把 E 称为质点的**总能**)。由式(5 -4 -2)及式(5 -1 -11)可得质点能量(总能)的表达式：

$$E=\frac{m_0c^2}{\sqrt{1-u^2/c^2}}。\tag{5-4-5}$$

同式(5 -1 -15)后面的讨论类似，上式表明质点的速率必须满足 $u<c$，否

则质点的能量 E 失去意义。从本章开始至今的讨论对象都是质点。不过，我们也不妨试着把式(5-4-5)用于光子。光子的 $u=c$ 导致式(5-4-5)的分母为零，要使分数有意义，除非分子也为零(这时分数为不定值)。所以，如果默认式(5-4-5)对光子也成立，就能推出光子的静质量为零。近代物理中有不止一种途径能导致光子静质量为零的结论。

能量概念在物理学中总是同能量守恒捆在一起的，因为引入(定义)一个不守恒的能量将不会有什么用处。当爱因斯坦首次大胆地把 $m_u c^2$ 解释为质点的总能量时，他也默认(假定)这是一个守恒量。这一守恒性业已取得迄今所有实验(特别是大量的高能粒子实验)的支持。而且，非常有趣的是，相对论中的能量守恒律与动量守恒律的关系非常密切：只要动量守恒律成立，便可纯理论地证明能量守恒律成立[证明见梁灿彬，周彬(2006)上册选读6-3-1]。其实，也许你已经注意到，§4.1 在证明动量在 K 系中守恒时就已暗暗地用到了能量的守恒性，因为在写出式(5-1-8)时默认了一个公式 $m_u + m_0 = M_v$，此式乘以 c^2 后两边分别就是碰撞前后的总能量。可见，假若能量不守恒，动量也不会守恒。

下面再来讨论静能。“静止物体也有能量”的提法本身并不惊人，人们在相对论诞生前就早已知晓。例如，子弹从枪口射出时枪身后退，子弹和枪身都获得动能，这份能量必定是由开枪前静止枪身内部的能量转化而来的——谁都知道枪身内的火药存在着化学能。然而，“物体的静能等于其静质量的 c^2 倍，即 $E_0 = m_0 c^2$”则是爱因斯坦的划时代重大贡献。非相对论中与静能最类似的能量品种就是物体的内能(火药的化学能也是内能的一部分)，然而谈及内能时总要选定一个状态作为参考态(指定它的内能为零)①。所以内能是某种“相对量”，即相对于某参考态而言的量。由于参考状态原则上可以任选，内能可以灵活到一个任意常数。然而相对论却并非如此。在相对论中，物体的静能 $m_0 c^2$ 也可称作它的内能，但其值是绝对的(完全确定的)，不允许有任何灵活余地②。作为后果，我们不免好奇地想知道一个普通物体(例如一袋方便面)的内能大约有多大(而在非相对论中这是个没有意义的问题)。答案是：大得惊人。一个 $m_0 = 1$ g 的物体(一袋方便面的静质量的1%)的静能竟然达到

$$m_0 c^2 = 10^{-3}\ \text{kg} \times (3 \times 10^8\ \text{m/s})^2 = 9 \times 10^{13}\ \text{J},$$

---

① 例如，当计算涉及饱和蒸汽在某一状态的内能值时可查表解决，而该表[见 Sonntag 的《Introduction to Thermodynamics, Classical and Statistical》书末的附表 A.1.1 和 A.1.2] 是以温度为 0.01℃，压强为 0.611 3 kPa 这一状态为参考态的，即规定饱和液体在该态的内能为零。

② 静质量的绝对性有其更为深刻的原因。根据广义相对论的爱因斯坦方程(见 §6.5)，任何静质量都提供引力，而引力就是时空的弯曲(见 §6.1)。时空曲率的绝对性保证了静质量的绝对性(不存在零点选择的灵活性)。

大约相当于在日本广岛爆炸的原子弹所释放的能量(亦即 2 万吨 TNT 炸药爆炸时所释放的能量)！此等巨大的静能可以被开发(转化为其他能量形式)和利用吗？如果我说送你一亿元巨款,并已存入银行,但声明任何人(包括你)都无权支取,这种空头支票又有什么用处？不过从开枪的例子我们早已知道至少火药的一部分静能(化学能)是可以被开发和利用的。可惜的是火药的化学能只是全枪的静能的很小一部分,所占比例连“九牛一毛”都不到。这是因为化学能只涉及分子之间的化学力。如果一步步地深入到越来越微观的领域,可被利用的静能比例就会越来越大。以原子核的裂变为例。设静质量为 M_0 的静止铀核(母核)裂变为两块(两个子核),其静、动能依次为 $m_{01}c^2$, E_{k1} 和 $m_{02}c^2$, E_{k2},则由能量守恒定律可知

$$M_0c^2 = (m_{01}c^2 + E_{k1}) + (m_{02}c^2 + E_{k2}),$$

故

$$M_0c^2 - (m_{01}c^2 + m_{02}c^2) = E_{k1} + E_{k2} > 0。 \qquad (5-4-6)$$

上式表明系统总静能在裂变前后的差值恒为正,可见静能在裂变过程中并不守恒。其实我们是从能量(总能)守恒出发推出了静能不守恒,这并不奇怪,因为过程中还涉及动能 $E_{k1} + E_{k2}$,正是静能的差值 $M_0c^2 - (m_{01}c^2 + m_{02}c^2)$ 转化成了子核的动能。由 M_0, m_{01} 和 m_{02} 的值就可求得子核动能的值。上述讨论就是原子弹的理论基础(虽然只是大为简化的讨论)。请注意这一讨论是以相对论的能量公式为基础的,所以人们说原子弹是相对论的产物。

虽然原子弹的放能远比开枪时的放能要大,但也不过相当于 1 克物质的静能而已,试想爆炸前的原子弹有多少克物质啊！可见原子弹爆炸时释放的能量仍然只是原子弹总静能的“冰山一角”①。这是因为原子弹涉及的也只是原子核的范畴,事实上,内能的绝大部分(约 99%)存贮在构成物体的基本粒子中。每个基本粒子都有一份静能,而且在适当条件下可以转化为其他形式的能量。例如,电子(electron)的反粒子叫做**正电子**(positron),其静质量与电子相同,其电荷与电子电荷等值异号(即为正,故称正电子),当电子与正电子相遇时会湮灭(消失)并产生两个光子,电子和正电子原来具有的能量(主要是静能)全部转化为两个光子的动能(请注意光子的静能为零)。综上所述可知,物体含有的巨大静能并非空头支票而是实实在在的能量。

---

① “九牛一毛”和“冰山一角”都是生活用语,虽然形象,却欠准确。定量地说,一克物质所含静能为 9×10^{13} J,而燃烧一克煤所放出的能量约为 3×10^4 J,与静能之比仅为 3×10^{-10},故言化学能只占“九牛一毛”。而原子核核心部分的静质量约有 10^{-4} 倍转化为释放的能量,虽比 3×10^{-10} 大很多,仍属很小,故言“冰山一角”。

让我们再回到守恒律。在牛顿力学中有动量、能量和质量守恒律。在相对论中,已经讲过动量和能量是守恒的。至于质量是否守恒,这首先取决于你谈的是运动质量 m_u 还是静质量 m_0。由于 $E \equiv m_u c^2$[式(5-4-2)],能量 E 守恒也就是运动质量 m_u 守恒,两者说的是一回事。余下的问题是:静质量是否守恒?答案是否定的,因为由 $E_0 = m_0 c^2$ 可知静质量 m_0 正比于静能 E_0,而式(5-4-6)分明表明静能并不守恒。"静能不守恒"是应该强调的结论。为防止概念混淆,此处有必要指出"守恒量"和"不变量"是两个完全不同的概念。**守恒量**(conserved quantity)是指在物理过程中保持不变(不随时间而变)的量,强调物理过程;**不变量**(invariant)则是指不随坐标系、参考系或观察者之类的人为因素的改变而改变的量,强调坐标系等的改变,并不涉及时间。举例来说,在相对论中,能量是守恒量而非不变量,静质量是不变量而非守恒量,带电粒子的电量则既是守恒量又是不变量(讨论从略)。

§5.5　对"质能相当性"的述评

读者大概早已知道 $E = mc^2$ 是爱因斯坦相对论的最成功的标志性公式,它表明质量 m 与能量 E 相当(或等价),因为两者在数值上只差一个常值倍数 c^2。通常称此式为质能关系式,它反映**质能相当性**。可惜这一说法颇嫌粗略,因为在相对论中质量有运动质量 m_u 与静止质量 m_0 之分,能量也有总能量 E 与静止能量 E_0 之别。不免要问:什么"质"与什么"能"相当?答案似乎不难给出,因为你会把 $E = mc^2$ 写成两个准确公式:

$$\text{(a)}\ E = m_u c^2; \qquad \text{(b)}\ E_0 = m_0 c^2, \tag{5-5-1}$$

并且说:(a)运动质量相当于总能量;(b)静止质量相当于静能量。两者都是质能相当性的反映。许多相对论教材也都若明若暗地这样讲过。然而,爱因斯坦后来建议把运动质量 m_u 的概念从相对论中删除,从而导致 $E \equiv m_u c^2$ 不再有意义,于是式(5-5-1)的解说就成了问题。不错,狭义相对论的原始表述中存在着静质量 m_0、静能量 E_0、运动质量(又称相对论质量)m_u 和总能量 E 等四个概念。然而,$E = m_u c^2$ 和 $E_0 = m_0 c^2$ 表明这四个量中只有两个是独立的。经历了狭义相对论发展初期的一些曲折后,爱因斯坦在1948年的一次私人通信中写道:"给物体引入质量 $M = m/\sqrt{1 - v^2/c^2}$ 的概念并无益处,我们无法对它下一个清晰的定义。除了'静质量'm 外最好不引入其他质量概念。"随着4维语言在相对论中的使用,越来越多的相对论学者认识到运动质量的概念利少弊多,纷纷放弃。例如,朗道和栗弗席兹的《场论》1948年版中的"质量"一词一律是指静质

量,书中绝无“运动质量”或“相对论质量”的提法。事实上,大量近代文献都只保留质量和能量两个概念,“质量”m 一律是指静质量,是不变量而非守恒量(既然只保留一个质量,就无须再冠以“静”字,也不必对 m 再加下标“0”);而能量则是指质点的总能,是守恒量而非不变量,与 m 的关系为[见式(5-4-5)]

$$E=\frac{mc^2}{\sqrt{1-u^2/c^2}}=\gamma_u mc^2。(m\text{ 现在代表静质量})\qquad(5-5-2)$$

虽然4个原始词汇只保留了能量和质量两个,但真正放弃的只有一个,就是那个本来就应放弃的“运动质量”m_u。静能一词虽未列入保留之列,但它可看作“总能量”一词的特例:质点的静能无非是质点相对于某一特殊惯性系的总能,该系认为质点在所关心的时刻(瞬时)是静止的。

既然 m_u 概念已被删除,公式 $E=m_uc^2$ 自然丧失意义,于是真正的“质能相当性”就只体现为静能相当于静质量,即 $E_0=mc^2$(其中的 m 不再写成 m_0)。可见,相对论的标志性公式虽然依据历史习惯写成 $E=mc^2$,但必须将式中的 E 和 m 分别理解为静能和静质量。这种理解丝毫无损于“质能相当性”的光辉,事实上,上节关于原子弹放能机制的特例正是基于 $E_0=m_0c^2$ 以及能量守恒律做出的。[当时指出 M_0c^2 和 $m_{01}c^2+m_{02}c^2$ 分别是裂变前后的静能,而两者之差就是裂变放出的能量。]

“质能相当性”导致“质即是能,能即是质”,于是人们索性创造一个新名词——质能(mass-energy)——来称呼质量和能量。在不少教科书中,质能既可以指静质量和静能量,也可以指运动质量和总能量。然而,放弃了运动质量概念之后,“质”字的唯一解释就是静质量,质能一词就只保留一种意义,即静质量和静能。这是应该注意的。

美国相对论学者 Adler 曾发表题为“爸,质量真与速度有关吗?”的文章[Adler(1987)],详细分析了引入相对论质量的利弊得失。文章一开头就说:“我的儿子在听了中学物理第一天的课后向我提出了本文题目的问题。我的回答是‘不!’,‘不过,是的,……’,‘实际上,不是的,但是别告诉你的老师。’结果他第二天就退选了物理课。”文章描写了物理教学界在围绕着相对论质量这个问题上的混乱局面,指出了引入相对论质量的种种弊端,不妨一阅。此后还有许多这方面的讨论文章。

[选读5-3]

本选读以对话方式介绍笔者在删除运动质量概念这一问题上的意见,其中乙代表笔者。

甲　为什么运动质量 m_u 的概念应该删除？

乙　首先是因为它根本没有存在的必要，其次是它的存在容易导致某些混淆。

甲　那么，为什么许多书（包括您这本书）都要引进这一概念？

乙　让我们先回顾本书引入 m_u 的过程。第一步，如果在相对论中仍用牛顿的动量定义 $\boldsymbol{p}\equiv m\boldsymbol{u}$（其中 m 为常量），则动量守恒定律失效。第二步，由牛顿的 $\boldsymbol{f}=m\boldsymbol{a}$ 可知恒力作用下的质点必将超光速。为防止超光速可以假定质量随速率 u 而变（成正变），于是把动量定义为 $\boldsymbol{p}\equiv m_u\boldsymbol{u}$。第三步，为保证动量守恒，$m_u$ 只能取 $m_u=m_0/\sqrt{1-u^2/c^2}$ 的形式。这种“三部曲”讲法的好处是便于讲解和接受，有其教学法上的优越性；缺点是人为地引进了多余的运动质量概念，而且往往还认为它反映惯性。

甲　为什么说 m_u 是多余的概念？

乙　首先是因为从后来的 $E=m_uc^2$ 可知 m_u 与 E 只差一个常数因子 c^2，两者保留一个便已足够，而 E（作为总能）有非常明确的含义和非常重要的地位，自然只能删除 m_u 的概念。其次，在爱因斯坦创立狭义相对论（1905 年）后不久，闵氏在 1908 年率先向同行们阐述了 4 维时空的概念及其重要性，随后，有越来越多的学者逐渐认识到用 4 维语言讨论相对论比以前的 3 维语言具有无与伦比的优越性，而且顺便也发现 m_u 在 4 维语言中毫无地位。[其唯一“地位”就是：m_u 等于总能 E 除以 c^2。反之，m_0（作为不变量）在 4 维语言中却不可或缺。] 所以说 m_u 是多余的概念。

甲　您刚才说“而且往往还认为它反映惯性”。难道 m_u 不反映质点（物体）的惯性吗？

乙　Adler（1987）详述了他认为 m_u 不反映惯性的理由（详见该文），指出只有静质量才是物体惯性的量度（反映与物体相对静止的观者所感到的该物体的惯性）。

甲　您在对话开头时说过“其次是它的存在容易导致某些混淆”。m_u 的存在会导致什么混淆？

乙　首先，只要保留 m_u，谈及质量时就总要加上定语（“运动”或“静止”），如果为陈述简单而不加定语，就容易造成混乱。其次，正文中已经指出，对“质能相当性”的正确理解只能是“静质量相当于静能，静质量为 m_0 的静止物体的能量为 m_0c^2”，但在使用 m_u 的教材中，不少作者也常把“质能相当性”理解为“运动质量相当于总能量，运动质量为 m_u 的运动物体的能量为 m_uc^2”，然而这其实是同义反复，因为在用 4 维语言的讲述中从来不定义 m_u，如果硬要引入 m_u，

就只好把它定义为 E/c^2。何必如此画蛇添足呢? [选读5-3完]

§5.6 结合能和质量亏损

我们一直在讨论质点。按照定义,当物体的形状和大小对所关心的问题影响不大时,该物体就可被看作质点。因此,任何客体(小至微观粒子、大到一个星系)在适当条件下都可当作质点处理。常见的质点是由若干粒子构成的复合体(如原子核),质点的静能自然是指复合体在质心静止时的能量,它等于以下三部分之和:①复合体内各粒子的静能之和(恒为正);②各粒子相对于质心系的动能之和(为正或为零);③各粒子之间的相互作用能。以氘的原子核为例。氘核由两个核子(一个质子和一个中子)组成,两者在核力的吸引下结合成一个稳定系统,只有从外界提供适当能量方可把两者分开,所需的最小能量就叫做氘核的结合能。原则上可以根据核力的大小来计算结合能,但核力过于复杂,通常都走捷径,就是利用质能关系式来计算结合能。

在原子物理学中质量和能量的常用单位分别是 a. m. u.(原子单位,简记为 u)和 MeV(兆电子伏),与相应的国际单位制单位的关系为

$$1\ \mathrm{u}\approx 1.66\times 10^{-27}\ \mathrm{kg}\ ,\quad 1\ \mathrm{MeV}\approx 1.6\times 10^{-13}\mathrm{J}\ 。\qquad (5-6-1)$$

由质能关系式又知

$$1\ \mathrm{kg}\ 质量相应的静能 = 1\times(3\times 10^{8})^{2}\ \mathrm{J} = \frac{9}{1.6}\times 10^{29}\mathrm{MeV},\qquad (5-6-2)$$

故

$$1\ \mathrm{u}\ 质量相应的静能 = 1.66\times 10^{-27}\times\left(\frac{9}{1.6}\times 10^{29}\mathrm{MeV}\right) = 931.5\ \mathrm{MeV}^{①}。$$

由查表可得质子、中子和氘核的静质量依次为

$$m_{\mathrm{p}} = 1.007\ 28\ \mathrm{u},\quad m_{\mathrm{n}} = 1.008\ 67\ \mathrm{u},\quad m_{\mathrm{d}} = 2.013\ 55\ \mathrm{u},$$

故

$$m_{\mathrm{p}} + m_{\mathrm{n}} = 2.015\ 94\ \mathrm{u} > m_{\mathrm{d}},$$

这正是静质量不守恒的具体事例。$m_{\mathrm{d}} < m_{\mathrm{p}} + m_{\mathrm{n}}$ 表明只有从外界吸取能量方可使氘核分裂为自由质子和中子,而这份能量的最低数值就是氘核的结合能,故

氘核的结合能 $= [(m_{\mathrm{p}} + m_{\mathrm{n}}) - m_{\mathrm{d}}]c^2$ = 质量$[(m_{\mathrm{p}} + m_{\mathrm{n}}) - m_{\mathrm{d}}]$相应的静能。

注意到 $(m_{\mathrm{p}} + m_{\mathrm{n}}) - m_{\mathrm{d}} = 0.002\ 389\ \mathrm{u}$,便得

$$氘核的结合能 = 0.002\ 389\times 931.5\ \mathrm{MeV} = 2.225\ \mathrm{MeV}。$$

---

① 直接乘除得值为 933.8 MeV,考虑到所涉及的公式都是近似等式,更准确的结果为 931.5 MeV。

反之，一个质子和一个中子结合为一个氘核时必定放出能量，其值等于氘核的结合能。一般而言，质子和中子结合成任一原子核时放出的能量称为该原子核的**结合能**(binding energy)。以 E_0 代表原子核的静能，E_{0i} 代表第 i 个核子的静能，则原子核的结合能 E_b 也可等价地定义为①

$$E_b \equiv \sum_{i=1}^{A} E_{0i} - E_0，\text{其中 } A \text{ 是核内的核子总数。} \qquad (5-6-3)$$

与结合能相应的那部分静质量 Δm 则称为**质量亏损**(mass defect)，即

$$\Delta m \equiv \sum_{i=1}^{A} m_i - m, \qquad (5-6-4)$$

其中 m 及 m_i 分别是原子核及第 i 个核子的静质量。质量亏损的存在是静质量不守恒的表现。

有时我们关心的不是原子核而是由若干子体系结合而成的任一体系，或者，虽然讨论的是原子核，但关心的是它分裂为若干个子体系(但不是核子)的问题。这时有必要把结合能的定义推广。设体系由 N 个子系结合而成，则该体系"关于这些子系(而不是关于所有核子)"的结合能定义为 $\left(\sum_{i=1}^{A} m_i - m\right) c^2$，其中 m 和 m_i 分别是体系和第 i 个子系的静质量。请注意在谈及体系的结合能时必须明确(至少心中有数)这是关于什么子系的结合能。例如，设静质量为 m 的原子核分裂为两个子核(静质量分别为 m_1 和 m_2)，则该原子核关于这两个子核的结合能就等于 $(m_1 + m_2 - m)c^2$。如果笼统地谈到"原子核的结合能"而对子系不加声明时，就是默认每个核子都是一个子系，这时结合能就是 $\left(\sum_{i=1}^{A} m_i - m\right)c^2$ [即式(5-6-3)]。

设静质量为 m 的系统自动分裂为两个部分(子系)，静质量和速率依次为 m_1，u_1 和 m_2，u_2，则由能量守恒及式(5-4-5)可知

$$mc^2 = \frac{m_1 c^2}{\sqrt{1 - u_1^2/c^2}} + \frac{m_2 c^2}{\sqrt{1 - u_2^2/c^2}} > m_1 c^2 + m_2 c^2, \qquad (5-6-5)$$

故

$$(m_1 c^2 + m_2 c^2) - mc^2 < 0。 \qquad (5-6-6)$$

可见该体系(关于那两个子系)的结合能为负。于是得出结论：

---

① 若干教材中的结合能定义与本书差一负号，即 $E_b \equiv E_0 - \sum_{i=1}^{A} E_{0i}$，文字提法相应地也要把"放出"与"吸收"互换。这类教材举例：朗道，栗费席兹(1948)38 页；赵凯华，罗蔚茵(1995)410 页；Ugarov(1979)160 页；虞福春，郑春开(1992)297 页。本书则与多数教材(特别是英美教材)一致。

结论 1　能够自动分裂的体系必有负的结合能。

再由“逆否命题”便有如下推论：

推论 1　结合能为正的体系不会自动分裂(因而是稳定的)。

§5.7　质能动关系式

把式(5－1－14)及式(5－5－2)重新写成(其中 m 和 γ 是 m_0 和 γ_u 的简写)

$$(\mathrm{a})\ \boldsymbol{p}=\gamma m\boldsymbol{u},\quad (\mathrm{b})\ E=\gamma mc^2。\tag{5-7-1}$$

以 p 和 u 分别代表 $\boldsymbol{p}$ 和 $\boldsymbol{u}$ 的大小,则上两式结合给出

$$p=\gamma mu=\frac{1}{c^2}Eu。\tag{5-7-2}$$

由式(5－7－1b)得

$$E^2=\gamma^2m^2c^4=\frac{m^2c^4}{1-u^2/c^2},$$

故

$$m^2c^4=E^2-\frac{1}{c^2}E^2u^2。\tag{5-7-3}$$

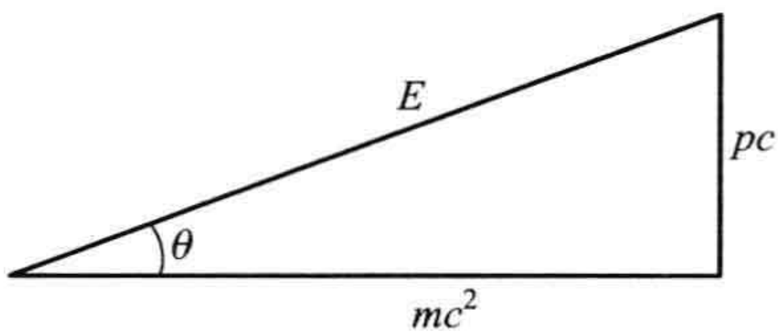


图 5－2　质能动三角形。对瞬时静止参考系有 $\theta=0$,三角形退化为水平线段。

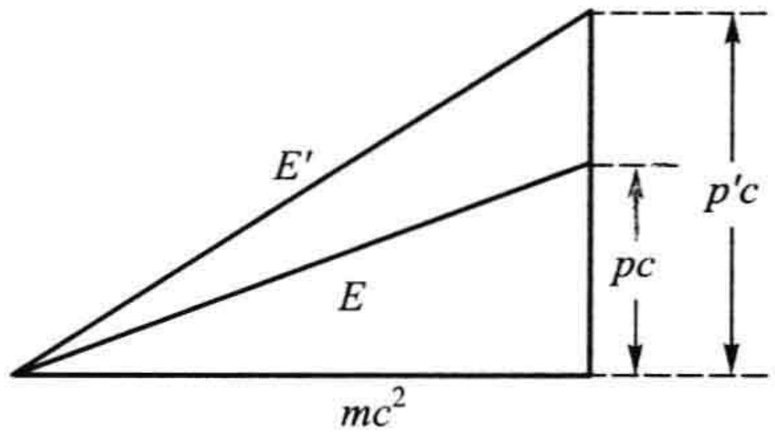


图 5－3　不同参考系中的动量和能量

再利用式(5－7－2)便得

$$E^2=m^2c^4+p^2c^2,\ \text{或(注意到 } E>0)\quad E=\sqrt{m^2c^4+p^2c^2}。\tag{5-7-4}$$

上式称为质点的**质能动关系式**。以 mc^2, pc 为直角边长作直角三角形,则由式(5－7－4)可知斜边之长就是能量 E(图 5－2)。请注意质点的静质量 m 是不变量,而 pc 及 E 都与参考系有关。由图 5－2 不难证明(习题)$\sin\theta=u/c$,所以 θ 依赖于参考系。对瞬时静止参考系有 $u=0$,故 $\sin\theta=0$,因而 $p=0$, $E=mc^2$(静能),三角形退化为一条水平线段。选择非瞬时静止参考系就可使 $u\neq0$。图 5－3

示出两个不同参考系测得的两个三角形，我们看到，在 $p' > p$ 的同时也有 $E' > E$。

§5.8　光子及其能量和动量

关于光的本性，物理学家一直存在着微粒说和波动说两种对立观点，前者的代表人物是牛顿，后者的代表人物是惠更斯。由于牛顿的成就和威望，微粒说在17世纪一直占上风。从19世纪开始，杨氏和菲涅耳的卓越工作使人们改而普遍相信波动说。1873年问世的麦氏电磁理论进一步揭示了光的电磁本性，使人们认识到光不但是一种波动，而且就是电磁波。光的微粒说至此几乎销声匿迹。19世纪末叶至20世纪初，普朗克为了克服黑体辐射研究中的难题而首先提出能量的“量子化”思想，把最小能量单位的携带者称为“量子”。爱因斯坦在他发表相对论第一篇文章的同一年(1905)、同一卷刊物上发表了他关于光电效应的文章(后来为此获诺贝尔奖)①，认为光波的能量集中在一个个以光速运动的微粒上，每个微粒都有一份相同的能量 E，与光的频率 ν 有正比关系：$E = h\nu$，其中 h 为普朗克常数。他把这些微粒称为**光子**②。根据这一思想，光子就是光波的最小微粒，一个频率为 ν 的单色光束可以看作由大量光子组成的整体，设束中的光子个数为 N，则该光束的能量就是 $Nh\nu$。

光的微粒说和波动说过去一直是两种对立的观点——光要么就是微粒，要么就是波动。然而，从1905年爱因斯坦的光电效应文章开始，人们逐渐认识到光在某些方面更像微粒，在某些方面则更像波动，这就是“波粒二象性”思想的萌芽，后来发展起来的量子力学和量子场论进一步认为波粒二象性是一切基本粒子的共同属性。

在相对论的理论讨论中，光子和质点合称为**粒子**。光子与质点的重要区别是：①光子总以光速运动；②光子的静质量为零[见式(5-4-5)后的讨论]。由于光子是一种非常重要的粒子，特别是它与其他粒子(质点)经常发生相互作用(例如康普顿散射就是静止电子对光子的散射)，为了确保能量和动量在相互作用中的守恒性，必须认为光子有一定的能量和动量。上面说过，光子的能量 E 最早由爱因斯坦定义，即 $E = h\nu$。光子的动量(大小)p 则可以借它的能量下定义。为此，我们来观察对质点成立的式(5-7-1)，它包含的两个等式 $\boldsymbol{p} = \gamma m\boldsymbol{u}$

---

① 爱因斯坦在1905年《物理年鉴》第17卷上发表了三篇重要文章，除了狭义相对论的那篇《论运动物体的电动力学》和关于光电效应的文章外，还有一篇关于布朗运动的文章。是年他才26岁，时任瑞士一个专利局的默默无闻的职员。

② 这当然只是对光子的早期认识。对光子的准确认识和描述必然要用到后来发展起来的量子场论，本书无意涉及。

和 $E=\gamma mc^2$ 对光子而言失去意义(因为 $m=0,\gamma=\infty$),但两式结合给出的 $p=Eu/c^2$[式(5-7-2)]对光子($u=c$)来说有意义,即 $p=E/c$。人们就把此式用作光子动量 p 的定义。总之,光子的能量 E 和动量 p 的公认定义是

$$E\equiv h\nu,\quad p\equiv E/c=h\nu/c,\quad \text{其中 } \nu \text{ 是光子的频率,} h \text{ 是普朗克常数。} \tag{5-8-1}$$

定义是可以人为选择的,重要的是只有选择适当的定义方能保证能量和动量在涉及光子的相互作用中的守恒性。迄今所做的大量实验证明,只要光子的能量和动量按式(5-8-1)定义,能量和动量都是守恒的。

§5.9　质点为何必须亚光速?

"质点必须亚光速"是狭义相对论中人所共知的基本结论(信条),准确含义是:质点相对于任一惯性系的速率 u 都小于 c,用几何语言说就是"质点的世界线是类时曲线"。但是,为什么有这个结论?一个重要原因是许多公式都含有因子 $\gamma\equiv1/\sqrt{1-v^2/c^2}$。最早出现 γ 的公式是洛伦兹变换式:

$$\text{(a) } t'=\gamma(t-vx/c^2),\quad \text{(b) } x'=\gamma(x-vt),\quad \text{(c) } \gamma\equiv1/\sqrt{1-v^2/c^2}。 \tag{5-9-1}$$

因为坐标 t, x 和 t', x'都是实数,由上式(a)、(b)可知 γ 必为实数,于是由上式(c)立即看出 v 必须小于 c。这说明任一惯性系的空间坐标原点相对于任何惯性系的运动都必须亚光速,人们第一次尝到极限速率的味道。不过我们要论证的是更强的结论,即任何质点(不一定是惯性系的空间坐标原点,也不一定做惯性运动)在任何时刻都亚光速,这时用洛伦兹变换就力不从心。注意到 γ 也在质点的能量公式(5-4-5)中出现,就想到还可从能量角度考虑。质点的能量公式(5-4-5)也可表为式(5-5-2),即

$$E=\frac{mc^2}{\sqrt{1-u^2/c^2}},\tag{5-9-2}$$

其中 E, m 和 u 分别是质点的能量、(静)质量和速率。待证命题"质点的世界线是类时线"等价于"质点世界线不含类空和类光线段"。我们用反证法来证明。假若世界线上有类空线段,则在该段的任一时刻都有 $u>c$(理由见 §3.3),导致上式右边分母为虚数。然而质点的能量必须为实数,所以质点世界线不能含有类空线段。再假若质点世界线上存在类光线段,则该

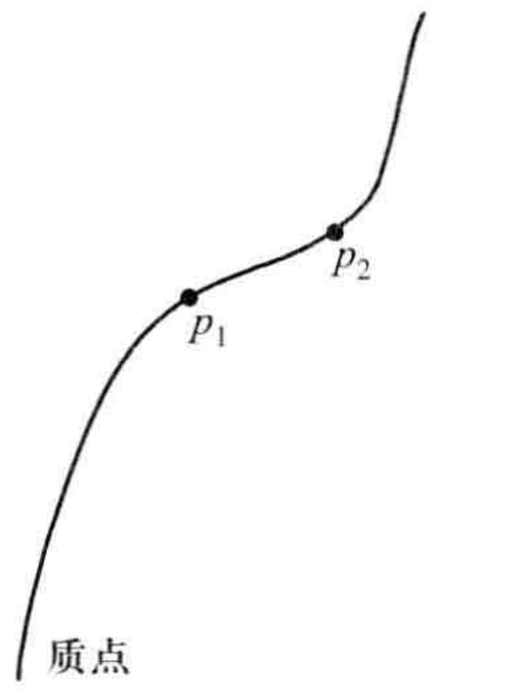


图 5-4　质点世界线含类空线段能推出因果矛盾

段的任一时刻都有 $u=c$,导致式(5-9-2)右边分母为零。分母为零的分数没有意义,但质点的能量总应该有意义,构成矛盾,所以质点世界线也不能含有类光线段。于是命题得证。

除了借用因子 γ 之外,因果关系的讨论也很有助于相信上述命题。假定质点世界线不是类时线,则它至少含有一个类空(或类光)线段,先设此线段为类空(图5-4)。在段上任取两点 p_1 和 p_2,则它们有类空的时空间隔(即 $I_{12}>0$,见§3.6)。任取一个惯性系 $\mathrm{K}\equiv\{t,\ x,\ y,\ z\}$,以 t_1、t_2 分别代表事件 p_1、p_2 在 K 系的时间坐标。不失一般性,设 $t_1<t_2$,则 K 系认为事件 p_1 早于 p_2 发生。但是,因为 $I_{12}>0$,由§3.6的定理3-2可知必定存在另一惯性系 $\mathrm{K}'\equiv\{t',\ x',\ y',\ z'\}$,它认为 p_1 晚于 p_2 发生($t_1'>t_2'$)。而 p_1、p_2 作为质点世界线上的两点,总可被看作有因果联系的事件,例如,设质点是子弹,p_1 代表子弹出膛事件,p_2 代表子弹中靶事件,则 p_1 是因,p_2 是果。然而 K′系竟然认为子弹出膛事件发生在中靶事件之后,造成因果倒置。可见,只要质点世界线上有一个类空线段,就会破坏因果关系,而这是谁都不能接受的。以上用反证法证明了质点的世界线不能含有类空线段。但是,要证明该世界线是类时线还要证明它也不含类光线段。这时用因果关系显得力不从心,所以还得求助于(刚才讲过的)用能量公式的论证。

注记2　数年前,意大利国家实验室的奥普拉(OPERA)团队在中微子震荡研究中意外发现中微子竟然超光速。他们让位于瑞士日内瓦的欧洲核子中心(CERN)发射的中微子束在地下管道中高速前进,最终到达730公里外的意大利格兰萨索(Gran Sasso)。他们在精确测量运动距离和时间后发现中微子的速率竟然比光速略大:中微子比光子少用60纳秒跑完全程。消息在2011年9月22日由《自然》杂志网站首次公布后,立即引起全球性的轰动效应。中微子是静质量非零的粒子(因而是质点,而非光子),超光速中微子的存在将动摇相对论的立论根基,因而撼动整个物理学。世界上大多数物理学家都对这一实验结果表示怀疑。五个月后,美国《科学》杂志网站援引"与实验有关的消息人士"的话报道说,中微子超光速是由GPS接收器与电脑之间的光缆连接松动(接触不良)造成的。可见中微子超光速纯属实验失误,光速作为极限速率的结论不变。当然,仍有若干项目组(包括OPERA)正在进行进一步的实验,以期取得更为明确和准确的结果。

第5章习题

1. 太阳表面的辐射功率为 3.9×10^{26} W(每秒辐射出去 3.9×10^{26} J 的能量),求:(a)太阳静质量每天的减小量;(b)太阳静质量每世纪的相对减小量。注:太阳质量 $m=2\times10^{30}$ kg。

2. 1 克 TNT 炸药爆炸时放出 4.2×10^{3} J 的能量。求炸药爆炸的相对质量减小量(即每单位质量所减小的质量)。

3. 试证 $u \ll c$ 的质点的动能总是很小于其静能。

4. 当质点的速率 u 为何值时其动能等于静能?

5. 质子和电子的静质量分别为 $m_p = 1.67\times10^{-27}$ kg, $m_e = 9.1\times10^{-31}$ kg,求两者的静能。

6. 宇宙线中高速质子的 γ 值高达 10^{11},为使静质量为 5 kg 的炮弹的动能与该质子的动能相等,炮弹的速率为多大? 提示:因 $u/c \ll 1$,可用牛顿二项式定理 $(1-x)^{-1/2}\approx 1+x/2$。

7. 加速器中的质子被加速到动能高达 2 GeV,求它的能量、动量和速率。

8. 求电子在下列两种情况下的速率 u(用几何制表出):(a)能量为 1 MeV;(b)能量为 1 GeV。提示:当 $\gamma \gg 1$ 时可用下法求 u:由 $\gamma^{-1} = \sqrt{1-u^2}$ 得

$$u = (1-\gamma^{-2})^{1/2} \approx 1 - \frac{\gamma^{-2}}{2} = 1 - \frac{1}{2\gamma^2},$$

其中第二步用到牛顿二项式定理。

9. 以 m、u、p、E、E_k 依次代表质点的静质量、速率、动量(大小)、能量和动能,试证

$$\text{(a)}\ u = \frac{pc}{\sqrt{p^2+m^2c^2}},\quad \text{(b)}\ \frac{dp}{p} = \frac{1}{1-(u/c)^2}\frac{du}{u},$$

$$\text{(c)}\ \frac{du}{u} = \frac{1}{(E/mc^2)^2-1}\frac{dE}{E},\quad \text{(d)}\ pc = \sqrt{E_k(E_k+2mc^2)}。$$

10. 试证质能动三角形(图 5-2)中的 θ 满足 $\sin\theta = u/c$。

11. (a)设光子的频率 $\nu = 8\times10^{15}$ Hz,求其能量 E 和动量 p。(b)一袋方便面(静质量 $m = 0.1$ kg)的静能等于多少个上述光子的能量之和? 注:普朗克常数 $h = 6.6\times10^{-34}$ J·s。

第6章　广义相对论初步

§6.1　引力的实质是时空的弯曲

相对性原理要求物理规律在所有惯性系中有相同的数学表达式，用于狭义相对论，就是要求物理规律的数学表达式具有洛伦兹协变性。这是一个管定律的定律。因此，在建立狭义相对论物理学时，原则上应该审查已有的物理定律，凡符合这个要求的就仍视之为定律，凡不符合的就要进行改造，直到符合要求方可作为物理定律纳入狭义相对论物理学的框架中。我们首先来审查麦氏电磁理论。麦氏方程天生就有洛伦兹协变性（见 §1.3），因而可以不经改造地纳入狭义相对论物理学框架。可以说麦氏方程是狭义相对论的“天之骄子”。再来审查牛顿的运动定律，仅以动量守恒定律为例。正如 §5.1 所指出的，牛顿的动量定义 $\boldsymbol{p}\equiv m\boldsymbol{u}$ 使动量守恒不具备洛伦兹协变性，因而必须修改。只要把动量定义改为

$$\boldsymbol{p}\equiv m\boldsymbol{u}/\sqrt{1-u^2}\,,\quad（其中\ m\ 代表静止质量）\qquad(6-1-1)$$

问题便迎刃而解，动量守恒定律便可纳入狭义相对论物理学中。第三，我们来审查牛顿的万有引力理论。牛顿引力论的基本方程是反映引力势 Φ 和质量密度 ρ 的关系的泊松方程：

$$\nabla^2\Phi=4\pi\rho\,,\qquad(6-1-2)$$

它有伽利略协变性而没有洛伦兹协变性，因而应该修改。牛顿的万有引力定律形式上与静电库仑定律很相似，既然麦克斯韦能把静电学推广改造为如此漂亮的麦氏电磁理论，看来牛顿引力论也不难被改造为狭义相对论框架内的引力理论。然而情况远非如此简单。关键在于，万有引力定律与库仑定律虽然很相似，却存在“符号差别”：电荷有正负两种，同性相斥，异性相吸；质量则只正不负，虽然同性，却只吸不斥（万有引力嘛！）仿照电磁理论，的确可以构造一个在狭义相对论框架内的引力理论，根据这个理论，在引力有变化（有扰动）时将出现类似于电磁波的引力波，而且也以光速传播。不幸的是，由于上述符号差别，由引力波带走的能量竟是负的。这意味着系统在发射引力波时自身能量不降反增，从而发射强度变大，由此又会获得更多的能量。如此循环，必然导致物理上不可接受的后果。虽然可以通过修改理论来克服这一困难，但又出现新的困难。事实

上，狭义相对论框架内的引力理论远非一个，但每个理论都有其自身问题。虽然无法绝对否定在狭义相对论框架内建立满意的引力理论的可能性，爱因斯坦还是独辟蹊径，于1915年成功地创立了革命化的、崭新的、突破狭义相对论框架的引力理论，这就是广义相对论。有趣的是，后人在克服狭义相对论框架内的某种引力理论的困难的努力中，几经修改后所得到的竟然也是与爱因斯坦广义相对论实质相同的理论！

有两个因素促使爱因斯坦创建广义相对论，它们是引力的“普适性(universality)”和马赫原理。我们只介绍前者。引力的“普适性”包含两层含义：①任何物体在引力场中都受引力(电中性的物体在电场中却不受电力，故电力无普适性)；②引力场中的任何物体，不论其质量、结构和组分如何，只要初始状态(位置和速度)相同，而且除引力外不受力，以后每一时刻的位置和速度就必定一样。这一结论已被越来越精确的实验所证实，它可以表述为：任意两个质点在引力场的同一点有相同的引力加速度。这虽然是司空见惯的事实，但为什么会这样？静电场中的两个点电荷就不这样。设点电荷 q 的质量为 m，所在点的静电场强为 $\boldsymbol{E}$，则它所受的电场力为 $\boldsymbol{f}=q\boldsymbol{E}$，获得的加速度为

$$\boldsymbol{a}=\frac{\boldsymbol{f}}{m}=\frac{q}{m}\boldsymbol{E}。\tag{6-1-3}$$

若在同一点放置质量为 m' 的另一点电荷 q'，则其加速度为

$$\boldsymbol{a}'=\frac{q'}{m'}\boldsymbol{E}。$$

$\boldsymbol{a}'$ 与 $\boldsymbol{a}$ 不等，除非两者的荷质比相同。在对引力做类似讨论时，不妨也把“荷”与“质”加以区别。质点的“引力荷”是其物质含量的量度，决定它在引力场中怎样受力，可称为**引力质量**(gravitational mass)，记作 m_{G}；质点的“质”则是其惯性的量度，决定它在力的作用下出现多大的加速度(对力的响应程度如何)，可称为**惯性质量**(inertial mass)，记作 m_{I}[即式(6-1-3)中的 m]①。仿照上述讨论不难导出质点在引力场中的引力加速度为

$$\boldsymbol{a}=\frac{m_{\mathrm{G}}}{m_{\mathrm{I}}}\boldsymbol{g},$$

其中 $\boldsymbol{g}$ 是该点的引力场强。如果不同质点有不同的引力荷质比 $m_{\mathrm{G}}/m_{\mathrm{I}}$，它们在引力场中同一点就不会有相同的引力加速度。然而无数的、一个比一个更精确

① 我们至今一直在用牛顿引力论。在牛顿引力论中，引力质量又可分为主动(active)和被动(passive)两种，前者指物体作为引力场源时的引力质量，决定它产生引力场的强弱；后者指物体作为引力场中的试探质点的引力质量，决定它在给定引力场中所受引力的强弱。正文的引力质量 m_{G} 是指被动引力质量。

的实验表明 m_G/m_I 对任何质点都相同，通过调整引力常量 G 还可使比值为 1 而简写为 $m_G = m_I$。通常把这一事实称为等效原理（理由及讨论详见 §6.3）。这是一个极其非同寻常的实验事实，应该引起深思。引力的“荷”与“质”本是两个完全不同的概念，风马牛不相及，它们为什么相等？牛顿引力论不能回答这个问题。在牛顿引力论中，这是作为实验事实（牛顿理论体系的一个公理）被承认的。然而，难道 $m_G = m_I$ 仅仅是一种巧合吗？难道就没有更深刻的原因藏在这个事实的背后吗？难道一定不存在一个更优雅的理论，在这个理论中，$m_G = m_I$ 是可用推理来证明的吗？正是对 $m_G = m_I$ 的思考，加上马赫原理的启发，促使爱因斯坦创立了他的广义相对论。

$m_G = m_I$ 的事实等价于初始位置和速度相同的、除引力外不受力的任何物体在引力场中都“齐步走”。这种毫无个性的集体行为强烈地暗示着引力是整个时空背景的内禀性质，与所有其他力有着实质性的差别。物理学是研究物理客体运动（演化）规律的学问，物理客体好比演员。正如演员的表演不能没有舞台一样，物理客体的演化也总是在某种舞台上（或背景前）进行的，这个舞台（背景）就是时空。在广义相对论创立之前，人们默认相对论的背景时空是平直的闵氏时空（见 §2.2）。闵氏时空是如此简单，以至于人们往往不注意（忘记了）它的存在。引力场中的“齐步走”现象引起了爱因斯坦对时空背景的注意。假如你看演出时发现某个演员的头顶突然下降了 20 cm，你会认为他蹲下了。然而，假如舞台上所有演员以及桌面、椅面都同时下降了 20 cm，那么最大的可能就是舞台台面由于某种原因下降所致。类似地，在引力作用下的“齐步走”现象分明强烈地暗示着引力本身是一种纯粹的时空背景效应。不妨这样猜测：引力可忽略时，时空是平直的（闵氏时空）；引力不可忽略时（例如在必须考虑地球或太阳的引力场时），时空变得弯曲，弯曲的情况取决于产生引力场的物质分布。根据这一猜测，引力非常不同于所有其他力，它特殊到这样一种程度，以至于在 4 维语言中它根本就不是什么 4 维力而是时空的弯曲！根据狭义相对论，每个 3 维力（记作 $\boldsymbol{f}$）都对应于一个 4 维力（记作 $\boldsymbol{F}$），而且 $\boldsymbol{f}\neq 0$ 当且仅当 $\boldsymbol{F}\neq 0$。（本书对此并未述及，读者只需承认这一结论。）在引力场存在时，谁都承认任何质点都要受到一个称为“引力”的 3 维力（记作 $\boldsymbol{f}_{引}$），但是按照刚才的猜测，这个 3 维力 $\boldsymbol{f}_{引}$ 竟然并不对应于任何 4 维力 $\boldsymbol{F}_{引}$，就是说，$\boldsymbol{f}_{引}\neq 0$ 而 $\boldsymbol{F}_{引}=0$！这是一种前所未有（在整个狭义相对论中都没有）的情况。可见 $\boldsymbol{f}_{引}$ 是唯一特殊的 3 维力，它特殊到这种程度，以至于其相应的 4 维力 $\boldsymbol{F}_{引}$ 竟然为零！实质上就是：虽然引力在 3 维语言中表现为一个力 $\boldsymbol{f}_{引}$，但在 4 维语言中它根本不是力而是时空的弯曲。于是，除引力外不受力的质点就应称为自由质点（已经自由到不能再自由的程度）。注意到闵氏时空中自由质点的世界线必为测地线的结论（见 §3.2），

自然进一步猜测(假定)弯曲时空中自由质点的世界线也是(该时空的)测地线。自由质点是最简单的质点,测地线是最简单的世界线,“自由质点的世界线是测地线”的这一假定非常符合美学原则。可见,引力场的存在不表现为质点受到一个称为“引力”的4维力,而是表现为时空的弯曲。有人问:既然引力不表现为4维力,自由质点在有、无引力场情况下岂非有相同的运动情况?引力场中的自由质点岂非“白受”了引力?引力场的影响体现在哪里?答案是:引力场的存在导致时空背景的弯曲,而弯曲时空与平直时空有不同的测地线,虽然自由质点都以测地线为世界线,但有无引力场时的测地线不同,所以质点有不同的运动情况。质点并不是“白受”了引力的。以上就是广义相对论的最基本的假定。由此出发就可以把 $m_G = m_I$(亦即“齐步走”现象)作为逻辑结论推导出来(可以说现在已经到了水到渠成的一步)。为了具体和形象,我们以比萨斜塔为例来讨论。设比萨斜塔顶部的一个铅球和一团棉花被释放而下落,世界线如图6-1。如果铅球所受的空气阻力可以忽略,则它可看作自由质点,其世界线 L_1 是测地线。设棉花团的空气阻力不可忽略,则它不是自由质点,其世界线 L_2 偏离测地线(见图6-1)。然而,假如根本没有空气,则 L_1 和 L_2 都是测地线。如果两者同时以相同的初始状态被下抛,则①初始位置相同(而且同时出发)表明两条世界线都发自同一时空点 p;②初始速度相同保证两者在 p 点有相同的切矢(图6-1表现的是初速为零的情形),于是由定理2-1(“一点一矢定一测”)可知测地线 L_2 和 L_1 重合。这不就是“齐步走”吗?可见,从广义相对论的上述基本假设出发的确可以把“齐步走”现象(亦即 $m_G = m_I$)作为逻辑结论推出来。广义相对论正是这样以其特有的优雅方式首次把引力解释为4维时空的几何效应(首次把引力与时空几何统一了起来)。

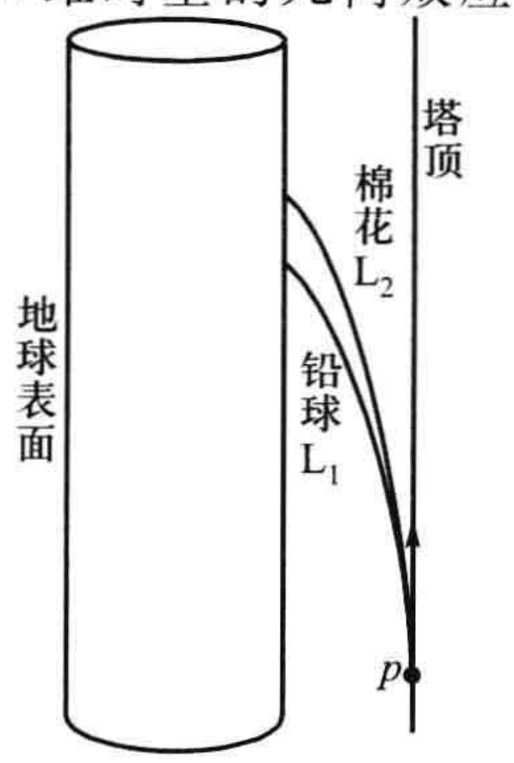


图6-1　比萨斜塔实验示意图。“地球表面”是指地球表面的世界面。“塔顶”、“棉花”、“铅球”是指它们的世界线。

上述讨论清楚地表明,广义相对论是独立于狭义相对论的物理理论,狭义相对论框架内容不下引力。反之,狭义相对论倒可以被看作广义相对论在引力场足够弱的极限情况,这时弯曲时空近似成为平直时空。

用现代语言来表述,广义相对论的最基本假设可以归纳为以下三点。

(a)3维空间中的引力实质上是4维时空的弯曲。就是说,当引力不可忽略时,时空背景不再是平直时空而是某种弯曲时空。据此,除引力外不受力的质点自然叫做自由质点。

(b)自由质点的世界线是它所在的那个弯曲

时空的测地线。

(c) 时空的弯曲情况(由时空线元描述)由物质分布决定,两者的关系由爱因斯坦场方程描述(见§6.5)。

可以证明,当引力场足够弱、质点速率足够低时,广义相对论力学的计算结果与牛顿力学的结果近似一致。所以说牛顿力学可以看作广义相对论力学的弱场低速近似[简介见§6.5末段,详细证明见梁灿彬,周彬(2006)上册小节7.8.2]。然而应该说明,尽管计算结果近似一致,两者看问题的观点却有明显区别。以树上的苹果落地为例。按照牛顿力学,这是由于苹果受地球引力而获得加速度,属于非惯性运动。但是按照广义相对论,苹果并不受到任何4维力(引力不对应于一个4维力),所以是自由质点。地球的影响在于使时空变得弯曲,苹果的世界线是这个弯曲了的时空中的一条测地线。就是说,同是苹果落地的运动,牛顿理论认为它有(3维)加速度(相对于惯性系),而广义相对论认为它没有(4维)加速度。(简单地说,存在着与 $\boldsymbol{f}=m\boldsymbol{a}$ 类似的4维等式 $\boldsymbol{F}=m\boldsymbol{A}$,所以4维力 $\boldsymbol{F}$ 为零导致4维加速度 $\boldsymbol{A}$ 为零。)反之,设苹果静止于地面,牛顿理论认为它所受的地球引力被地面的支撑力 $\boldsymbol{f}_{支}$ 所抵消,因此保持静止,其(3维)加速度为零,处于惯性运动状态;而广义相对论则认为它只受到一个4维力(与支撑力 $\boldsymbol{f}_{支}$ 相应的4维支撑力 $\boldsymbol{F}_{支}$),因此其世界线不是测地线,其(4维)加速度不为零。读者是否已经意识到,当你舒适地坐着阅读本书的时候,你的世界线并不重合于因地球的存在而弯曲了的时空的测地线,你的4维加速度不为零,因而你并不处于最自然(最简单)的运动状态?另一个例子是地球绕太阳的公转。按照牛顿力学,这是地球在太阳引力(充当向心力)的驱使下所做的加速运动(3维加速度非零);而广义相对论则认为地球并不受到任何4维力,太阳对地球运动的影响只体现为使时空变得弯曲,地球始终沿着这个弯曲时空的一条最简单曲线——测地线(绕太阳世界面的螺旋线)——做最自然的运动,根本没有(4维)加速度。

把引力的实质归结为时空的弯曲的这一认识是人类智慧的一个伟大胜利。**黎曼**(Riemann)在年仅28岁时(1854年)就提出了(内禀)曲率(又称黎曼曲率)的概念和计算方法,并在逝世(年仅40岁)前的一段时间内致力于寻求把电力和引力统一起来的某种理论。这一努力未获成功的最重要原因就是他只专注于空间及其曲率而没有注意4维时空及其曲率。直至1905年狭义相对论问世之后,时间和空间才被同等看待。(实际上,直到1908年闵可夫斯基才首次明确地提出时空这一绝对概念。)再过几年后,"引力的实质是时空的曲率"这一划时代认识才伴随着爱因斯坦对广义相对论的构思而逐渐建立了起来。

在爱因斯坦创立广义相对论的思考过程中,除了深受马赫思想的影响和启

发之外,还有两个人功不可没,值得一提。第一个人是闵可夫斯基,如果没有他首创的4维时空这一绝对的概念,就不可能有后来的广义相对论(见§3.7末);第二个人是格罗斯曼(Grossmann),他是爱因斯坦在苏黎世理工学院的老同学,后来成为该学院的数学教授,爱因斯坦是在格罗斯曼的帮助下学会黎曼几何并以此为数学工具创建了广义相对论的。

§6.2　弯曲时空的物理定律

广义相对论认为一切物理现象不过是物理客体在某种弯曲时空中的演化。因此,用广义相对论研究物理,首先就要找出各种物理客体(质点、理想流体、电磁场……)在给定的弯曲时空背景上的演化方程。由于实际生活和实验室中的引力场太弱,广义相对论同牛顿引力论的区别一般难以察觉,想通过观测或实验来归纳出弯曲时空的物理定律是没有指望的。我们能够做到的只是:根据某些基本原则,用假设的手法"猜出"弯曲时空中的各种物理定律,其正确性则有待于由它推出的各种结论的自洽性及其与实验(如果可能进行的话)结果的一致性来验证。这种"猜测"有两个重要依据(不妨称之为两根"拐棍")。第一个依据就是**广义协变性原理**(principle of general covariance),这其实是一个假设,是爱因斯坦在创立广义相对论的过程中提出的,它的实质性内容就是弯曲时空中所有坐标系都是平权的,因而物理定律的数学表达式在任意坐标变换下形式不变。第二个依据有两种提法,多数教材的提法是"等效原理"(详见§6.3),少数教材的提法是:"当时空几何是闵氏几何时,物理定律应回到狭义相对论的相应定律"[详见 Wald(1984);梁灿彬,周彬(2006)]。由于本书连狭义相对论的许多物理定律都没有给出定量表达式,所以对此只能点到为止,无意展开。想知道弯曲时空各种物理定律的定量表达式的读者可以参阅梁灿彬,周彬(2006)§7.2。

谁都知道相对论的基本假设是光速不变原理,并记住其内涵是"光速为 c ($=3\times10^8$ m/s)"(此处的"光速"一律指真空光速),于是自然会问:"广义相对论(弯曲时空中)的光速也是 c 吗?"笔者对此的回答是:速率代表单位时间所走的距离,而时间和距离都同坐标系有关,所以只记住"光速为 c"有可能导致误解,应该记住的是"惯性系中的光速为 c"(光速相对于非惯性系很可以不等于 c,§3.3 的例1就是一例)。闵氏时空中存在着一类特殊的、优越的、与众不同的坐标系,即(整体)惯性系,用3维语言讨论狭义相对论问题时除特殊情况外都选择惯性坐标系,谈到光速时自然就暗指相对于惯性系的光速。然而弯曲时空中任何一个坐标系都不比其他坐标系更优越(这正是广义协变性原理的实质性内涵),所以你只能任选一个坐标系来定义光速(虽然在某些具体场合下你会觉

得你所选的坐标系最恰当),其结果自然是千变万化的,不但可以不等于 c,而且完全可以大于 c。因此这种"坐标光速"其实是用处不大的。"光子速率"和"质点速率"的实质性问题最好用4维语言表述,这就是光子和质点的世界线问题。请回忆§3.3中关于极限速率的讨论,当时曾将狭义相对论的两个基本"信条"分别用3维和4维语言表出,复习如下。

3维语言表述:

(A) 光子相对于任何惯性系的速率 $u=1$;

(B) 质点相对于任何惯性系的速率 $u<1$。

4维语言表述:

(a) 光子的世界线是(闵氏时空的)类光曲线;

(b) 质点的世界线是(闵氏时空的)类时曲线。

3维表述必然涉及惯性系,而惯性系是闵氏时空特有的坐标系,弯曲时空中不存在(整体)惯性系,所以无法将3维表述(A)、(B)直接推广到弯曲时空。然而4维表述(a)、(b)却可以方便地推广,为此只须将括号中的"闵氏时空"改为"弯曲时空",就是说,我们默认广义相对论要遵守以下两个基本信条:

($\tilde{a}$) 光子的世界线是(所论弯曲时空的)类光曲线;

($\tilde{b}$) 质点的世界线是(所论弯曲时空的)类时曲线。

刚才已经指出弯曲时空的"坐标光速"用处不大。反之,4维表述"光子世界线是类光曲线"却可以准确地告诉你光子在弯曲时空中的运动情况。所以我们真诚地奉劝读者,在涉及光在弯曲时空的传播问题时,除了极少数特殊情况之外,不要再去津津乐道地大谈特谈坐标光速了,拿起最简洁、清晰而又保险的4维工具,从"光子世界线是类光曲线"出发吧。

在此我们还想顺便指出,不少人在弯曲时空背景中讨论问题时,总爱或明或暗地用到惯性系这一概念,殊不知在弯曲时空中不存在(整体)惯性系。弯曲时空中倒是存在许多局域惯性系,而且与闵氏时空的整体惯性系有某些类似之处(见§6.4),但两者之间毕竟存在着许多区别。不注意这些区别有时就会导致误解甚至错误。一个常见的例子是以为弯曲时空中局域惯性系的坐标光速等于 c,其实这是一种误解。

§6.3　等效原理

对经验定律 $m_G=m_I$ 的深刻思考是爱因斯坦创立广义相对论的重要动机。为了帮助理解,我们先介绍爱因斯坦根据牛顿力学借用电梯构思的一个思想实

验,即爱因斯坦电梯实验。设电梯因缆绳断裂而自由下落,其内部观者便有失重的感觉。如果他放下手中的苹果,便会发现它不像平常那样离手下落,而是处于随遇平衡状态之中。理由非常简单:电梯观者相对于惯性系(地球系)有重力加速度 $\boldsymbol{g}$,所以是一个非惯性观者(也可说电梯是个非惯性参考系),故他认为苹果除了受到重力之外还受到一个惯性力。设苹果的引力质量和惯性质量分别为 m_G 和 m_I,则它所受的重力为 $m_G\boldsymbol{g}$(此处 $\boldsymbol{g}$ 作为重力场强),所受的惯性力为 $-m_I\boldsymbol{g}$(此处 $\boldsymbol{g}$ 作为非惯性系的加速度)。因为 $m_G = m_I$(这是关键),所以合力为零,因而苹果随遇平衡,或说处于失重状态。假如电梯观者是个有经验的宇航员,他将觉得这苹果与远离各星球(因而引力可以忽略)的、做惯性运动的飞船内的苹果有相同表现。推而广之,由于 $m_G = m_I$,根据牛顿力学,自由下落电梯内部的一切力学实验都与远离星球的惯性飞船内的相应实验结果相同,因此把 $m_G = m_I$ 称为**等效原理**。由于广义相对论尚在构思过程中,以上是纯用牛顿力学得到的结论。

在构思广义相对论的过程中,爱因斯坦又大胆地、假设性地把这一原理从力学实验推广到一切物理实验,即假设自由下落电梯内的一切物理实验都与远离星球(因而时空平直)的惯性飞船内的实验结果一样,并由此推出光的引力红移、光在引力场中走曲线等重要而有趣的结论。为了区分清楚,后人把 $m_G = m_I$ 相应的原理称为**弱等效原理**(weak equivalence principle, 即 WEP),把爱因斯坦推广后的等效原理称为**爱因斯坦等效原理**(Einstein equivalence principle, 即 EEP),此外还有一个强等效原理(见后)。

现在介绍如何从爱因斯坦等效原理推出光在引力场中走曲线的结论。设自由下落电梯的左墙中心 A 有一光源向右墙发光(图 6-2)。根据爱因斯坦等效原理,电梯内的一切物理实验都与远离星球的惯性飞船内的相应实验有相同结果,而惯性飞船内的光线是直进的,所以电梯观者也应看到光走直线,因而将打在右墙的正中心 B。然而,对地面系而言,在光从左墙走到右墙的时间内电梯已下落了一段距离(而且是匀加速地下落),所以光的轨迹是从图 6-2 的 A 点到 B′点的一段抛物线(电梯观者认为 B′和 B 是同一空间点)。另一方面,地面系认为电梯及光线是处在引力场中,于是得出光在引力场中走曲线的结论。反之,电梯系认为所有物体都失重("引力场消失了"),光走直线是很自然的。

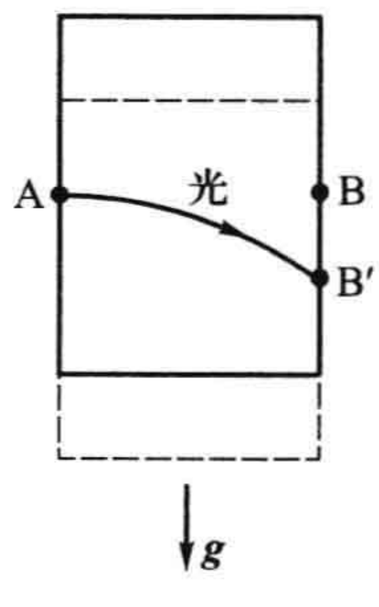


图 6-2 用爱因斯坦电梯说明光在引力场中走曲线

除了利用自由下落电梯之外,人们也常用下面的方式描述弱等效原理 $m_G =$

m_1。设飞船在远离星球的空间做匀加速运动(非惯性运动),相对于惯性系的加速度为 $\boldsymbol{a}$(不妨设 $\boldsymbol{a}$ 等于地面的重力加速度),见图 6-3。设苹果的质量为 m,则船内观者(非惯性观者)将看到离手的苹果在惯性力 $-m\boldsymbol{a}$ 的作用下做反向加速运动,就像地球表面被释放的苹果在重力作用下的自由下落运动那样。进一步也不难相信,只要飞船加速度 $\boldsymbol{a}$ 等于地面重力加速度 $\boldsymbol{g}$,飞船内的一切力学实验就都与地面附近的相应实验结果相同,这可以看作弱等效原理的另一表述。据此,人们又常说"加速飞船内的宇航员发现自己置身于引力场中","加速度与引力场等效"。对这两句话应有正确理解。根据第一句话,爱思考的初学者常会提出这样的问题:既然加速飞船内的宇航员感到有引力,而引力就是时空的曲率,船内宇航员岂非觉得自己置身于弯曲时空中?答案当然是否定的:因为早已约定飞船远离各星球,它所在的时空必然(近似)平直,而时空曲率是绝对的,谁看都一样,决不能得出"加速宇航员觉得时空弯曲而惯性观者觉得时空平直"这样一个荒唐的结论。导致上述错误结论的关键在于推理过程中两次用到引力一词,而两次含义不同(犯了"偷换概念"的逻辑错误)。加速宇航员所感到的"引力"其实只是非真实的**表观引力**,它不由物质产生(在这一点上与真引力根本不同),不对应于时空弯曲(在这一点上也与真引力根本不同),只因加速宇航员的感觉而得名。可见,对于"加速度与引力场等效"一类的提法应该格外小心,必须清醒地知道与加速度等效的"引力"只是表观引力(假引力)而不是真引力。根据笔者所见,与此类似的错误实在是屡见不鲜,可以说是在对广义相对论半懂不懂的知识人群中的常见病和多发病。

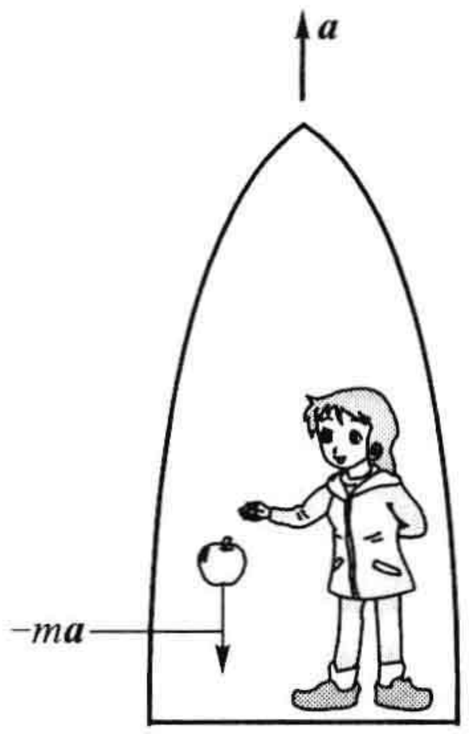


图 6-3　远离星球的匀加速飞船。船内观者发现苹果"自由下落"。

爱因斯坦的等效原理对广义相对论的创立无疑起到过重要作用。然而,在有了广义相对论之后,人们逐渐发现爱因斯坦等效原理可以从广义相对论的理论推出[见梁灿彬,周彬(2006)]。关于等效原理的含义、地位和作用,不同学者有不同看法。下面列举两种有代表性的极端看法,双方都是国际上非常杰出的相对论学家。大力推崇等效原理的是**密斯纳**(Misner),**索恩**(Thorne)和**惠勒**(Wheeler),他们在合著的书[Misner et al. (1973) P. 386]中说:"*等效原理功能强大,用它可把狭义相对论的所有物理定律推广到弯曲时空。*"而**辛格**(Synge)对等效原理则持极端相反的观点,他在书[Synge(1960)]的序言中写道:"*我从未懂过这一原理。……它意味着引力场的效应与一个观者的加速度的效应不可*

分辨吗？如果这样，那是错的。在爱因斯坦理论中，要么存在一个引力场，要么不存在，取决于黎曼张量是否为零。这是一个绝对的性质，与任何观者的世界线都毫无关系。时空要么平直，要么弯曲，在本书的若干地方我都不得不煞费苦心地把由时空曲率导致的真实引力效应与那些由观者世界线的弯曲导致的效应区分开来（在多数通常情况下以后者为主）。等效原理在广义相对论的诞生过程中实质上起着接生婆的作用，……我建议我们以适当的荣誉埋葬掉这位接生婆而正视绝对时空这一事实。”对等效原理的这一看法也许偏激，但上引段子中的一些提法不失为防止误解的清醒剂。例如，他关于分清由时空弯曲导致的真引力与平直时空中由于观者世界线的弯曲（非测地线）导致的表观（假）引力的警告就是十分必要的。我们衷心地希望他所煞费的苦心没有白费，希望有更多读者从此对分清真假引力的重要性留下深刻印象。

下面简述笔者对等效原理的肤浅认识。第一，爱因斯坦等效原理是爱因斯坦在酝酿广义相对论的过程中对弱等效原理的假设性推广，对广义相对论的诞生起过重要的“接生婆”作用。这是连辛格也同意的。第二，§6.2 讲过，等效原理是从闵氏时空的物理定律“猜出”弯曲时空的物理定律的两个重要依据（拐棍）中的一个（另一个是广义协变性原理），因此可以说等效原理是“由狭义相对论通往广义相对论的桥梁”。不过，正如 §6.2 所说，也可不用等效原理而改用“当线元为闵氏时回得去狭义相对论的相应定律”，同样可从闵氏时空的定律“猜出”弯曲时空的相应定律。而一旦接受这些定律（从而建立起广义相对论），讨论问题时原则上就可以完全不用等效原理（虽然许多作者在许多问题上喜欢用等效原理）。因此，从这个角度来看，“埋葬掉接生婆”对广义相对论似乎没有实质性影响。第三，对于某些较为复杂的情况，“等效原理是否被违反”是长期以来有争论的问题。这里的关键问题是等效原理在这些具体情况下的准确含义尚待澄清。据笔者所知，不同作者对等效原理有不同的表述方式，但似乎很难说有哪一种方式是百分之百地清晰明确的。在这个意义上，辛格说“我从未懂过这一原理”也许未必过分。第四，文献中存在着许许多多互不相同的引力理论［广义相对论只是其中之一，详见 Will (1993)］，所有引力理论可以分为两大类，即度规理论和非度规理论。度规是一个张量，选定坐标系后就可表为一个线元，其中各项的系数就是度规张量在该坐标系的分量。（本书出现的各种线元其实代表各种度规张量，只因不想讲张量而一直避开度规、使用线元。）在广义相对论中，时空就是某个背景空间上配以适当的线元（度规），所以广义相对论是当然的度规理论。此外还有许多种度规理论，包括著名的、很有竞争力的**布兰斯 - 迪克 (Brans - Dicke) 理论**。（以下简称**布 - 迪理论**）。判断哪种引力理论是正确理论的标准当然是实验。为此需要一个关于引力实验的理论。**迪克** (Dicke) 从

20世纪60年代开始从事这种理论的研究,他的开创性工作使人们对等效原理及其意义的理解逐渐加深,并终于意识到应把等效原理摆在考察引力理论(而不只是限于广义相对论)的基础这一重要位置上。等效原理可分为三个层次,即**弱等效原理**(weak equivalence principle,即WEP)、**爱因斯坦等效原理**(Einstein equivalence principle,即EEP)和**强等效原理**(strong equivalence principle,即SEP),SEP与EEP的区别在于:EEP(及WEP)只考虑系统(如电梯)所处的外引力场而不考虑系统中的物体所激发的自引力场,即在引力上只考虑它们的被动方面而忽略其主动方面,而SEP对主动、被动方面都做考虑,讨论对象是"自引力系统",大到恒星内部的自引力,小到卡文迪许实验(测量引力常数G)中两个铅球间的引力都在考虑之列。EEP可以看作SEP在自引力可忽略情况下的特例。这三个等效原理的实验检验对选择(淘汰)引力理论有重要意义。任何引力理论都服从WEP(因为有越来越精密的实验证实WEP,无人愿意创造一个不服从WEP的引力理论),但对EEP和SEP则不然。讨论表明[见Will(1993),Will(2001)],如果EEP成立,则只有度规理论才可能正确。这就表明,如果有越来越精密的实验证实EEP,那么非度规理论就将越来越没有立锥之地(事实上已经如此)。进一步的讨论还表明[见Will(1993),Will(2001)],广义相对论满足SEP而所有其他已知的度规理论(包括布-迪理论)都不满足。[可惜这一讨论还不等于完全严格的证明,因此上述结论至今仍被称为猜想(conjecture)。但至少可以肯定地说,除广义相对论外的多数度规理论不满足SEP。]所以,如果有越来越精密的实验证实SEP,那么广义相对论很可能就是唯一正确的引力理论。可见三个等效原理的实验检验有着非常重要的理论意义,这些实验正以越来越高的精度在不断进行中(可参阅§8.5后半部分)。

§6.4　潮汐力与测地偏离

上节从牛顿力学推出一个结论:自由下落电梯内的一切力学实验都与远离星球的惯性飞船内的相应实验结果相同。本节的要旨在于说明,由于地球引力场的非均匀性(离地心越远引力越弱),这一结论其实只能近似成立,就是说,电梯内的一个足够精确的力学实验与惯性飞船中的同样实验的结果并不完全相同。先用一个理想实验揭示这一问题。

在自由下落电梯内用8个小球摆成一个圆形花样(圆平面与地面垂直),如图6-4左。先用牛顿力学讨论。设球1,2连线的延长线正好过地球中心,开始时各球相对静止。由于地球的引力场强在球1处比球2处略大,球1的重力加速度就略大于球2,于是两者的距离渐增,站在球1的立场就会觉得球2与自己

渐行渐远。另一方面，球 3 及 4 与地心的连线与竖直方向略有偏离（但两球的偏离方向相反），于是球 3，4 的距离渐减。一段时间后的整个花样将变成图 6－4 右的卵形。然而，如果无引力地区的惯性飞船内也用 8 个小球摆成圆形花样，而且开始时各球之间相对静止，则球 1 不会发现球 2 与自己的距离有任何变化，整个花样一直保持圆形。可见，即使就力学实验而言，电梯与飞船也并非完全等效。

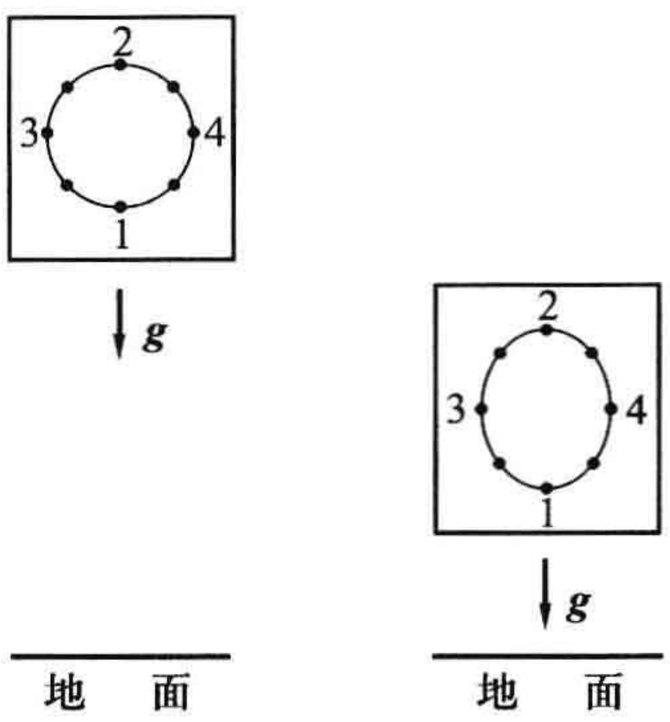
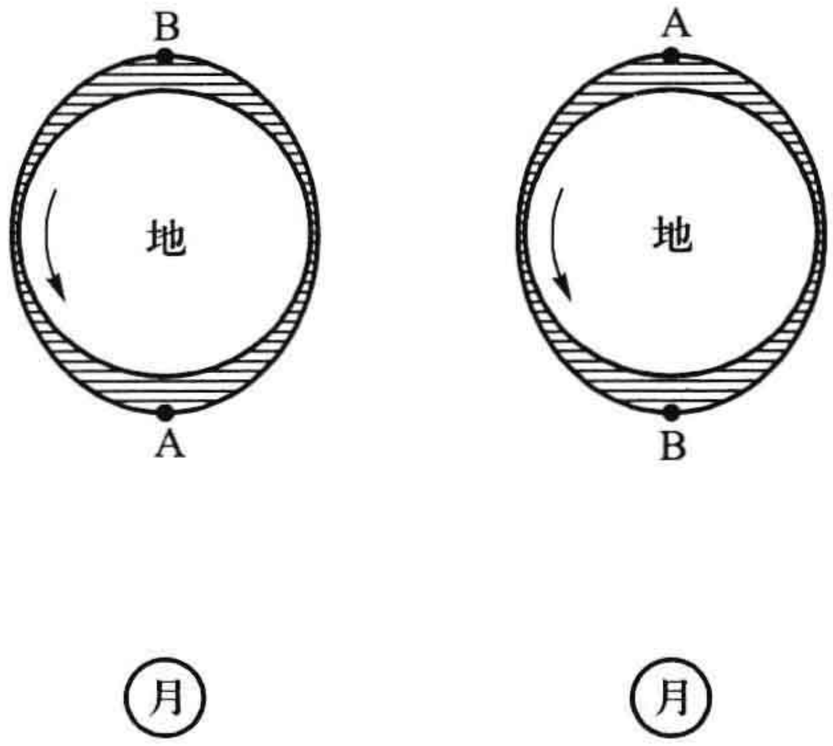


图 6－4　电梯内小球花样
在自由下落中变形



（a）自转半周前　　（b）自转半周后

图 6－5　潮汐现象示意图
（地球自转半周前后）

以上是假想实验，但道理与此类似的现象却不难在日常生活中找到，海水的潮汐现象就是一例。下面是用牛顿力学对这一现象所做的简化分析，旨在突出本质。潮汐现象的主要起因是月球（“罪魁祸首”），其次才是太阳。忽略太阳的影响可使问题简化而不影响对实质的理解。地球，作为一个物体，处于月球的引力场中。假定地球表面被一层海水覆盖，考虑水面上的 A，B 两点，其连线经过地心（图 6－5）。设某一时刻 A 点离月球最近，则 B 点离月球最远。月球在 A，B 两点的引力场强（大小）不同，两点就要相互远离，于是 A，B 附近的水面向外鼓起［图 6－5（a）］①。随着地球的自转，A 点不再正对月球，水面高度降低。地球自转半周后，A 点离月球最远［图 6－5（b）］，水面再度鼓起。这就解释了海水为什么每天两度涨潮两度退潮。这种“潮汐现象”其实也不限于海水。对地面附近自由下落的人（蹦极运动者）来说，头顶和脚底与地心距离不同，也存在着把人拉长的力（只考虑地球引力场），不过这种“潮汐力”很小，不会造成任何感觉。但是，如果你在中子星表面附近自由下落，潮汐力可以大到10^{11} N（牛顿）的程度，你将被撕裂而丧生。注：①中子星是一种主要由中子组成的、密度高达水

① 从地球观者看来，A 点鼓起的原因是两个力的合力：（a）月球引力，（b）地球绕地月共同质心做圆周运动导致的惯性离心力。两力的合力叫**生潮力**（或**起潮力**）。

密度的10^{14}倍(!)的致密星体,高密度导致表面引力场强的变化率奇大,见小节7.2.3。②按照估算,人体能经受的临界压力或拉力(超过此限则人体破裂)约为$10^7 N/m^2$。

以上讨论说明地球和月球引力场中的自由下落物体都要受到潮汐力。其实潮汐现象是引力场的普遍性质。应该指出,至此我们一直在用牛顿引力论,而且已经感觉到引力场的非均匀性是潮汐现象的根源。既然牛顿引力论是广义相对论在弱引力场条件下的近似,可以预期,潮汐现象也应该是广义相对论的一种效应。于是不免要问:在广义相对论范畴内,潮汐现象的根源又是什么?

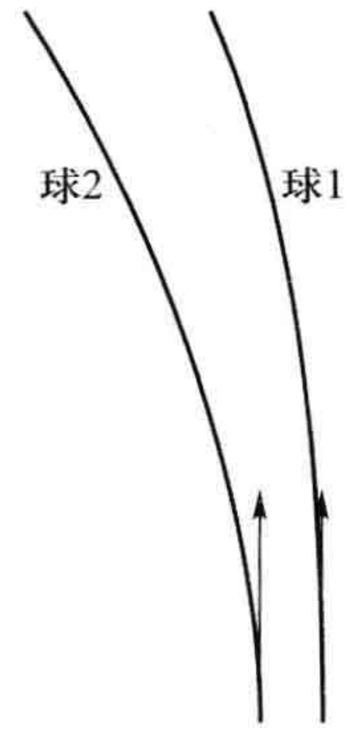


图6-6　地球附近自由下落小球(存在测地偏离,即潮汐效应)

我们早已知道,①根据广义相对论,3维空间中测得的引力其实是4维时空的弯曲;②自由质点在弯曲时空中的世界线是该时空的类时测地线。由此可知,图6-4中8个自由下落小球的世界线必定是因地球存在而弯曲了的时空中的8条类时测地线,图6-6示出其中的两条(图6-4的小球1和2)。计算发现,由于时空曲率非零,两线之间的空间距离(即球1和2的距离)会随着时间而加速增大(与牛顿力学结论一致),在广义相对论中称此现象为"**测地偏离**"(geodesic deviation),由一个方程定量描述(称为**测地偏离方程**),该方程左边代表两线距离加速增大的加速度(称为**潮汐加速度**),右边则涉及时空曲率(黎曼张量)。在时空曲率点点为零的情况下(平直时空),方程右边为零,故左边的潮汐加速度也为零,所以平直时空(亦即无引力场时)不存在潮汐现象。反之,从测地偏离方程可以证明,只要时空曲率非零,就总可找到这样两个自由下落的相邻质点(两条类时测地线),它们之间的潮汐加速度非零。可见,从广义相对论的角度看来,潮汐现象的"罪魁祸首"(根源)就是时空曲率:当且仅当时空曲率非零时存在潮汐效应。

特别值得强调的是,虽然自由下落电梯内的观者会感到物体失重(似乎消除了引力场),但是电梯内8个小球的假想实验表明由引力场导致的潮汐效应是无法消除的。(因为曲率是绝对的,不因参考系或坐标系的改变而改变,所以不可能通过选用电梯系来消除潮汐效应。)从广义相对论的角度看来,选择自由下落电梯无非是选择了与电梯观者相应的一个坐标系(称为**局域惯性系**),虽然引力效应在此系中得以表观地消除(感到失重),但是,时空曲率是个绝对量("硬邦邦"的张量),是不可能通过任何坐标变换消除的(无人能选择一个坐标系使得非零张量在该系的全部分量为零),它通过潮汐效应顽强地表现自己。

因此,在学习和讲解等效原理的时候,不应忽视(相反地,应该强调)潮汐现象。爱因斯坦是在1912年意识到潮汐力的实质是时空曲率的,尽管他当时的理解还很不清晰、不准确。毕竟那时广义相对论尚未建立起来。

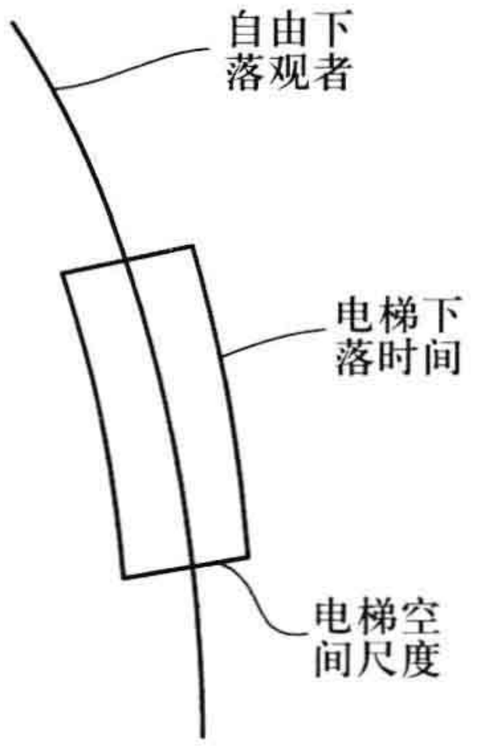


图6-7　电梯观者做实验涉及的时空邻域

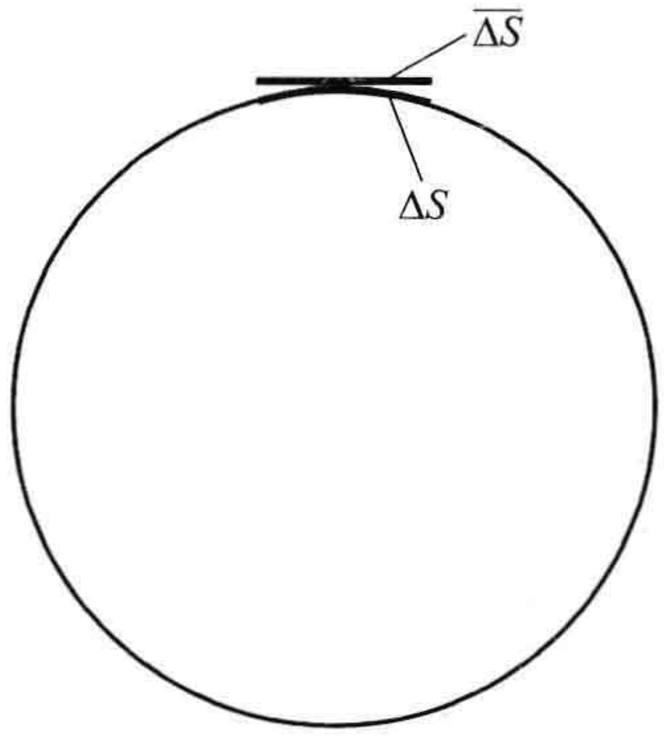


图6-8　一小片球面可用切平面的一小部分代替

小结　牛顿引力论和广义相对论都是关于引力的理论。在这两种理论中都存在潮汐效应。牛顿引力论认为潮汐效应的根源是引力场的非均匀性;广义相对论认为潮汐效应的根源是时空曲率。

严格说来,自由下落观者的失重感(感到引力被"消除")只发生在该观者身上,只要离开他的世界线就会观察到潮汐现象。偏偏做实验时往往要涉及观者世界线周围的一个小的时空区域(例如图6-7),于是"失重"问题就变得不那么简单。幸好时空曲率的效应只有在足够大的时空范围内才会有所显示(被实验测出),只要实验涉及的时空区域足够小(对电梯而言,只要其空间尺度及下落时间足够小),实验结果便与平直时空的相应实验近似地不可区分。这同以下的简单例子类似:2维球面上每点的曲率张量都不为零,但若只关心某点附近一小片球面 ΔS 的情况,就不妨用该点的切平面的一小部分 $\overline{\Delta S}$ 近似代替(见图6-8)。例如,为测定地球两条经线在北极的夹角,不妨把每条经线的一小段看作直线。在任意弯曲时空中,如果只关心某个时空点的一个足够小的邻域,就可近似使用狭义相对论物理定律。这也可看作是等效原理的某种提法。

[选读6-1]

甲　我对均匀引力场一直存在一个问题。一方面,根据牛顿引力论,均匀引力场不存在潮汐效应;另一方面,广义相对论认为只要有引力就有时空曲率,根据正文的讲法,就必定有潮汐效应。于是我要问:均匀引力场中到底有没有潮汐

效应？

乙　“均匀引力场”一词在牛顿引力论中意义明确(引力场强为常矢量),但在广义相对论中尚未定义。只有先下定义方可讨论你的问题。

甲　想来想去,我觉得比较合理的定义是:如果任意两个自由下落的相邻质点之间都没有潮汐加速度,所在的引力场就称为均匀引力场。

乙　但是,按照这一定义,广义相对论中根本就不存在均匀引力场,因为有引力场就有时空曲率,就必然有潮汐效应。

甲　然而在牛顿引力论中是有均匀引力场的。例如,我们熟知均匀带电无限平面两侧有均匀静电场,由牛顿万有引力定律与库仑定律的形式相似性可知有均匀质量密度的无限平面两侧应有均匀引力场。难道广义相对论不认为这是均匀引力场吗?

乙　按照你刚才下的定义,它当然不是均匀引力场。事实上,这个例子在广义相对论中是这样一个时空,其几何由一个称为 Taub 线元的弯曲线元描述,线元在有均匀质量密度的无限平面上存在奇性[见梁灿彬,周彬(2006)§8.6]。完全可以从这一线元出发用计算证明它有潮汐效应。　**[选读6-1完]**

§6.5　爱因斯坦场方程

牛顿力学认为引力场由物质分布产生,由引力势 Φ 描述,它与物质的质量密度 ρ 的联系体现在泊松方程

$$\nabla^2\Phi=4\pi\rho \tag{6-5-1}$$

之中。广义相对论对引力的看法虽然与牛顿力学很不相同,但同样认为引力由物质分布产生。既然引力是时空弯曲的体现,就应该存在一个把时空曲率与物质分布联系起来的关系式。爱因斯坦经过几年努力,终于找到这一公式,后来称为**爱因斯坦场方程**(是广义相对论的另一重要假设),是一个相当复杂的张量方程。由于本书不讲张量,此处只能把这一方程半定性地表为

$$\text{爱因斯坦张量}=8\pi\times\text{物质的能量、动量密度及应力张量。} \tag{6-5-2}$$

其中“爱因斯坦张量”是由黎曼张量派生出来的张量,也反映时空的弯曲情况。上式表明,不但物质场的能量密度,而且它的动量密度以及应力张量也对时空弯曲(因而对引力场)提供贡献。

首先可以提出这样一个问题:平直时空的曲率(黎曼张量)点点为零,导致爱因斯坦张量为零,于是方程(6-5-2)意味着物质的能量、动量密度及应力张量都为零,也就是根本不存在物质。然而,平直时空是狭义相对论的背景时空,难道狭义相对论根本不涉及物质?物理学是研究物质运动(演化)规律的学问,

没有物质还有物理吗？这个表面上的矛盾是由物理学中的近似处理方式带来的。实际上，狭义相对论研究各种物体和场的运动及其相互作用，单单忽略它们之间的引力作用，亦即忽略所有这些物质产生的引力场。如果引力不能忽略，时空就不平直，就超出了狭义相对论的范畴，就必须使用广义相对论。我们生活在一个充满物质的宇宙中，整个宇宙（看作一个无所不包的最大的4维时空）或其中任一部分都是弯曲时空。只当我们所关心的是远离各大星体（因而引力场足够弱）的区域时才可以近似地使用狭义相对论。可见，虽然狭义相对论本身在理论上非常严密准确，它也不过是一个理论模型，一旦用于客体就必然出现近似性。不是理论本身有近似性，而是这一理论只近似地适用于所涉及的客体。

根据“线元决定几何”的提法（见§2.1），时空的所有几何性质（包括黎曼张量和爱因斯坦张量）都由该时空的线元决定。所谓求解爱因斯坦方程，其实就是要找到某一线元，使得由它决定的爱因斯坦张量正好等于方程（6－5－2）右边的张量。

虽然我们生活在无所不包的宇宙时空中，但往往只关心其中的某一局部时空区域。如果区域中没有任何物质（真空区域），爱因斯坦方程（6－5－2）的右边就等于零，方程就简化为

$$\text{爱因斯坦张量}=0。\tag{6-5-3}$$

［请注意爱因斯坦张量为零不意味着黎曼张量为零。］上式称为**真空爱因斯坦方程**。例如，天文学家常常关心某一恒星（尤其是太阳）周围的情况。由于其他天体与该恒星（记作H）相距甚远，通常可以采用孤立体系模型，即认为除此恒星外一无所有——恒星H是真空中的孤立物体。从广义相对论的角度说，应该关心由该恒星的物质分布造成的星外真空区域的时空弯曲情况。定量地说，就是希望求解爱因斯坦方程以得到星外时空的线元表达式。取定坐标系后，张量方程就等价于一组分量方程，构成关于待求量（时空线元）的偏微分方程组，如能求解，便可得到待求线元。然而这是一个高度非线性的偏微分方程组（可被列为宇宙中的极端非线性现象之一），求解通常非常困难。幸好物理学家有自己的高招：充分利用物理问题本身的各种对称性。外部时空的弯曲是由恒星造成的，恒星自身的对称性也就决定了外部时空线元的对称性。我们所关心的恒星有两个对称性：①球对称性；②静态性，是指恒星（在很长的时间内）的恒定不变性[①]。这两者就导致外部时空线元既有球对称性又有静态性（通俗地说，线元的

① 恒星是靠内部烧氢变氦的核聚变反应维持自身的稳定性的。核燃料总有用完的一天，那时恒星将在自身引力作用下出现剧烈的坍缩（详见第7章）。但是现在关心的只是恒星在其（很长的）稳定期内的情况，故言有静态性。

静态性就是其各项的系数不随时间而变①)。利用线元的这两个对称性可使真空爱因斯坦方程(6-5-3)大为简化。**史瓦西**(Schwarzschild)在爱因斯坦方程发表后不到一个月就求得了这个解,其线元在"球坐标系"$\{t, r, \theta, \varphi\}$的表达式为

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2), \tag{6-5-4}$$

其中 M 是常数,代表恒星的质量。后人称上式为**史瓦西真空解**(Schwarzschild vacuum solution),称史瓦西所用的坐标系$\{t, r, \theta, \varphi\}$为**史瓦西坐标系**。上式是史瓦西真空解在几何单位制的表达式。几何单位制在涉及引力时除了令 $c=1$ 外还令引力常数 $G=1$,于是公式大为简化。史瓦西真空解在国际单位制的表达式则要复杂得多,见式(8-1-1′)。

当 $M=0$ 时,上式归结为

$$ds^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2), \tag{6-5-5}$$

这无非是闵氏线元在球坐标系$\{t, r, \theta, \varphi\}$的表达式。这是很自然的,因为 $M=0$ 意味着恒星根本不存在,时空当然平直。

史瓦西是20世纪最有名的德国天体物理学家之一。爱因斯坦方程发表时(1915年11月)正值第一次世界大战,那时史瓦西正在对俄战争前线的德军服役,他是一名志愿入伍的爱国者。当他在《普鲁士科学院会议报告》中读到爱因斯坦的广义相对论时,寻求爱因斯坦方程解的强烈欲望使他在几天之内就求得了静态球对称真空解(即史瓦西真空解)。爱因斯坦收到史瓦西的论文时大喜过望,回信说:"**我没有预料到能得出方程的精确解②,您对问题的解析处理令我极为满意**。"1916年1月13日,爱因斯坦代表史瓦西在柏林普鲁士科学院的会议上作了报告。虽然物理学家对这个解很重视,但是连史瓦西和爱因斯坦都没想到它还包含了描述黑洞的重要内容(见第7和第8章)。

史瓦西真空解只是真空爱因斯坦方程众多的精确解之一,但至今仍不失为爱因斯坦方程很重要的一个解。如果所关心的时空区域存在物质场(非真空),时空线元就应是**有源爱因斯坦方程**[式(6-5-2)]的解。这种方程的求解更为复杂。史瓦西在求得真空解后仍不满足,他还想知道恒星内部时空的弯曲情况。为了便于求解,他只考虑星体内部密度 ρ 为常数(均匀密度星)这一最简单情况。这虽然不是一个很好的恒星模型,但仍不失为压强不太大的小恒星的一级

---

① 静态性的定义见选读7-1。

② 当时的爱因斯坦连方程的一个近似解(在弱引力场近似下的解)也没有求得过。

近似。史瓦西求得了均匀密度星内部的线元表达式[详见梁灿彬,周彬(2006)P.326],现在称为**史瓦西内解**。在爱因斯坦向科学院报告了史瓦西真空解的短短几周之后,他又向科学院报告了史瓦西的第二篇论文,即史瓦西内解。非常不幸的是,史瓦西在寻求他的真空解的同时已经身患不治之症天胞疮。他不久便被送回国内,并于1916年5月英年早逝。6月19日,爱因斯坦悲痛地向科学院报告了史瓦西因病辞世的噩耗。

史瓦西的全名是卡尔·史瓦西,其子马丁·史瓦西在父亲去世时年仅四岁,后来也成为天文学家,1955年同霍伊尔合作发表的重要论文首次算出普通恒星如何演化为红巨星(见小节7.2.1)。

爱因斯坦方程的高度非线性给求解以及整个广义相对论研究带来许多困难。在大多数情况下引力场很弱,可用近似处理手法把爱因斯坦方程变为线性方程,从而使问题大为简化。在4维语言中,弱引力场意味着时空线元 ds^2 接近闵氏线元 $(ds^2)_{闵}$,可以表为 $(ds^2)_{闵}$ 加上一个"很小"的微扰线元 $(ds^2)_{微}$,即

$$ds^2 = (ds^2)_{闵} + (ds^2)_{微}, \tag{6-5-6}$$

微扰线元 $(ds^2)_{微}$ 与闵氏线元 $(ds^2)_{闵}$ 相较是如此之小,以至于它的2阶和高阶项都可忽略。在此基础上,经过一番计算就可求得爱因斯坦方程(6-5-2)的线性近似表达式,称为**线性爱因斯坦方程**,它可以看作是关于待求线元 $(ds^2)_{微}$ 的线性的偏微分方程组。这一线性化使得求解和讨论都简单得多。进一步说,如果问题所涉及的物质运动的速率都很小于光速(因而满足弱场低速条件),从线性爱因斯坦方程出发还可推出两个简单的方程,第一个(在实质上)就是牛顿引力论的泊松方程(6-5-1),第二个则是自由质点的加速度表达式

$$\boldsymbol{a} = -\nabla\Phi, \tag{6-5-7}$$

而这正是牛顿引力论中自由质点的运动方程(右边无非是该质点单位质量所受的引力)。上述结果表明牛顿引力论可以看作广义相对论的弱场低速近似。读者至此是不是倍感欣慰?试想,假若广义相对论在弱场低速条件下与牛顿引力论竟然还有比较明显的岐离,它还能站得住脚吗?须知在弱场低速情况下牛顿引力论几乎是战无不胜、攻无不克的!事实上,用广义相对论进行的计算通常都很复杂,"杀鸡焉用牛刀",在满足弱场低速条件时完全可以放心使用牛顿引力论。太阳系内的引力场对于我们关心的绝大多数问题都是足够弱的,牛顿引力论仍然大有用武之地。[不过,太阳系中也有少数天文观测数据与牛顿引力论的理论结果不完全吻合(例如水星近日点的进动和星光在太阳附近的偏转),但使用广义相对论就能得到与观测非常吻合的结论(详见§8.4和§8.5)]。反之,黑洞附近的引力场则强得离奇(见§7.3、§7.5及§8.9),这时就非用广义相对论不可。

第7章　恒星演化和黑洞

§7.1　引力是天体物理学的主角

我们生活在太阳系，其主要成分是太阳和八大行星①。太阳只是一颗非常普通的恒星。由于万有引力，宇宙中的恒星存在着“结团”现象：这里聚集着一大堆恒星，那里也有一大堆恒星。由十颗以上的恒星组成的、被引力束缚在一起的恒星群叫做一个**星团**(star cluster)；由数十亿至数千亿颗恒星(连同星际气体和尘埃)组成的系统则称为**星系**(galaxy)。太阳系所在的星系就是**银河系**(Milky Way)，它包含大约一千亿颗恒星。其他星系一律称为**河外星系**。

站在地球上仰望星空，看到的星星好像离我们都一样远，就像镶嵌在一个球形天幕上的宝石。这个球面(球形天幕)就叫做**天球**。天体在天球上的位置只反映它们的方向(角位置)，不反映它们与地球的距离。为了便于认星，人们按恒星在天球上的分布划成若干区域，每个区域称为一个星座。用线条连接同一星座内的亮星，形成图形，便可据此给星座命名，例如金牛座、天鹅座和天蝎座等。由此可知星座与星系是非常不同的概念：星系中的恒星从天文尺度来看相当靠近(一个星系所占的空间尺度在几千至几十万光年之间)，而位于同一星座的两颗恒星的距离却可以很远，只不过它们在天球上的投影互相挨近而已。

天文学还有一个术语叫**星云**，是指银河系中在太阳系以外一切非恒星状的气体尘埃云。后来发现某些被称作星云的对象其实是星系。

浩瀚无垠的宇宙含有无数的微观粒子，粒子之间存在着相互作用(力)。基本的相互作用力只有四种，即电磁力、万有引力、强相互作用力和弱相互作用力。强力和弱力的作用范围(“力程”)很小，大约在小于 10^{-15} m 的距离内才会有所显示。反之，电磁力和引力的力程则是无限大。四种力的强度大小不一，差别悬殊。表7-1给出了这四种力的相对强度(以引力的强度为1)和力程②。

---

① 还包括各行星的卫星、小行星、彗星、流星体和行星际物质。

② 不同文献给出的强度和力程不尽相同，这是因为比较方法不同，而且这两个概念只在某种相当复杂的理论框架内才有意义，而理论的研究仍在不断进展中。表7-1的数据取自互联网上的维基百科(Wikipedia)。

表 7－1　四种基本相互作用力的比较

力	相对强度	力程
强力	10^{38}	10^{-15} m
电磁力	10^{36}	∞
弱力	10^{25}	10^{-18} m
引力	1	∞

由于强力和弱力的力程是如此之短，在宏观现象中起作用的几乎只有电磁力和引力。电磁力在强度上是引力的 10^{36} 倍（一万亿亿亿亿倍），但并非处处都占上风。事实上，只有引力才是天体物理学（乃至于宇宙演化）的主角。关键在于电荷有正负两种，同性相斥而异性相吸，如果物体所含的正、负电荷绝对值相等，它在电场中就不受力。天体虽然由无数带点粒子（质子和电子）组成，但通常天体的净电荷数甚小，所以天体之间的电磁力小得可怜。反之，引力只吸不斥，“团结就是力量”，其效应变得相当可观。这正是引力在天体物理学中充当主角的根本原因。在本章中我们处处都要同引力打交道。

先不说恒星这样大的天体，就以地球来说，引力的重要性也是尽人皆知的。你之所以不敢从五楼跳出窗外，不就是因为你知道引力的厉害吗？不过，在地球的内部，电磁力也起着非常重要的作用。为了帮助理解，先讲一个简例。放在桌面上的书受到地球的引力而不往下掉（没有向下的加速度），是因为它被桌面阻挡。从力的角度来说，就是因为书所受的引力被桌面的支撑力所平衡。若问这个支撑力的来源，就要谈到电磁力。书和桌子都由原子组成，原子的外层是电子。不妨认为书的下侧和桌面各有一个电子薄层，书由于受到地球引力而倾向于向下运动，当两个电子薄层足够靠近时，两者之间相斥的电力就表现为支撑力。其实，把原子（以及分子）结合在一起的所谓化学力也是电力的表现。把地球表面的一层看作刚才的那本书，其他部分看作桌子，该表面层所受到的（来自地球的）引力遭到了电磁力的抵抗，于是就保持平衡。粗略地说，行星的稳定就是靠电磁力顶住了引力而得以维系的。质量越大的星体引力越强，靠电磁力维持平衡就越不容易。当质量大到太阳（恒星）质量的数量级时，情况就根本不同。恒星与行星的区别不仅在于质量，而且在于恒星会发出炽热的光芒，而且“质量大”和“会发光”是相辅相成的。历史上的天文学家一直在思考恒星发光的机制和能源，曾经出现过若干不正确的猜想。1920 年，英国著名的物理和天文学家爱丁顿（Eddington）首次提出恒星的能源来自核聚变的正确想法，并为此

与金斯爵士进行过一场旷日持久的辩论。

§7.2　恒星演化

黑洞是广义相对论和近代天体物理学的一个非常重要、非常诱人的词汇。要了解黑洞的形成，应该具备一些恒星演化的知识。本节介绍球对称恒星的演化过程。当密度不十分高时，引力场不十分强，牛顿引力论近似适用。广义相对论只在演化的最后期间才是不可或缺的。

7.2.1　恒星的形成　红巨星

顾名思义，“恒星”一词使人觉得它是永恒不变的。但是，恒星也像人一样要经历生、老、病、死各个阶段，只不过时间尺度长得多而已。恒星的前身是一大堆密度略有不匀的气体，主要是氢。密度较大处有较强引力，吸引来更多气体，逐渐形成一个球状气团。由于球对称性，气团中每一薄球层内的每一体元所受的来自气团的引力都指向球心（图7-1），所以整个气团就在自身引力作用下收缩（各层都缩），简称**引力收缩**（gravitational contraction）。收缩过程实质上是引力势能转化为热能的过程，于是气团温度越来越高。温度的分布是从球心到球面逐渐下降的。由于压强随温度增高而增大[①]，任一薄球层内壁的压强都大于外壁的压强（见图7-2），因此内外壁所受的压力之差向外，形成对引力的一种抗衡作用。看来当温度足够高（因而压强差足够大）时有可能遏制收缩。然而在没有能源的情况下这是不可能的：由于气团温度比周围高，它不断向外辐射能量。如果收缩竟然停止，温度（因而压强差）就要下降，薄球层内外壁的压力差也就抗衡不了引力。从能量角度来看它也必须不断收缩，以使引力势能的一部分可以不断转化为辐射出去的能量。经过长时间的缓慢收缩，气团中心附近区域（称为**中心球**或**星核**）的温度和密度终于高到足以点燃热核反应的程度——中心球内的氢原子核（质子）经热核聚变而成氦原子核（与氢弹的反应相同[②]）。核聚变释放的核能使气团内部的高温得以维持，相应的压力差也就足以抗衡住引力，气团就不再收缩而达到平衡，从此就成为一颗恒星（star）。事实上，所有（普通）恒星都是依靠烧氢变氦来抗衡引力的，其中心球可以视为一颗超巨型的

---

① 经典理想气体的压强公式为 $p = nkT$，其中 p 是压强，T 是热力学温度，n 是粒子的数密度，k 是玻耳兹曼常数。

② 原子弹爆炸靠的是**核裂变**（较重的核分裂为较轻的核）；氢弹爆炸靠的是**核聚变**（较轻核聚合为较重核）。

炸弹。太阳只是恒星的一个普通例子,它已在这种稳定状态中度过了大约 50 亿年。

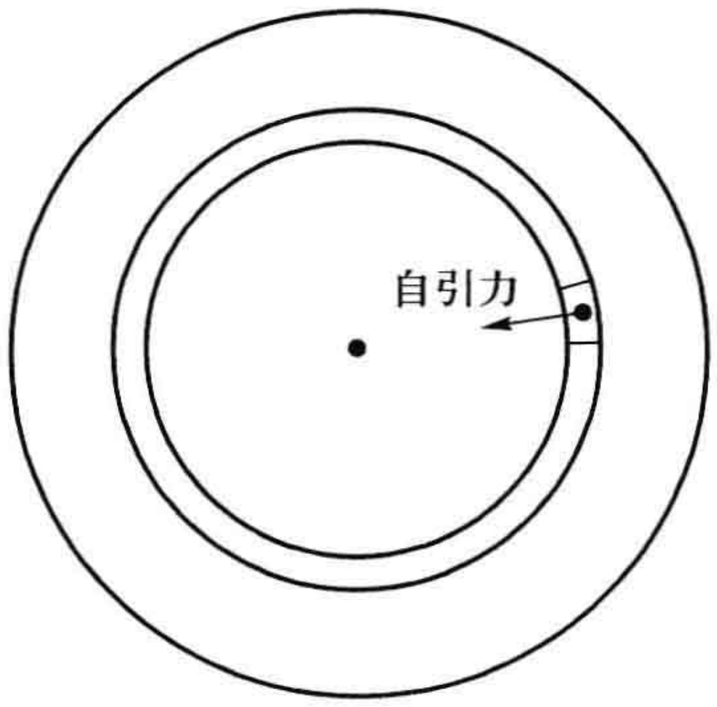


图 7 – 1　气团各体元受到的自引力都指向球心

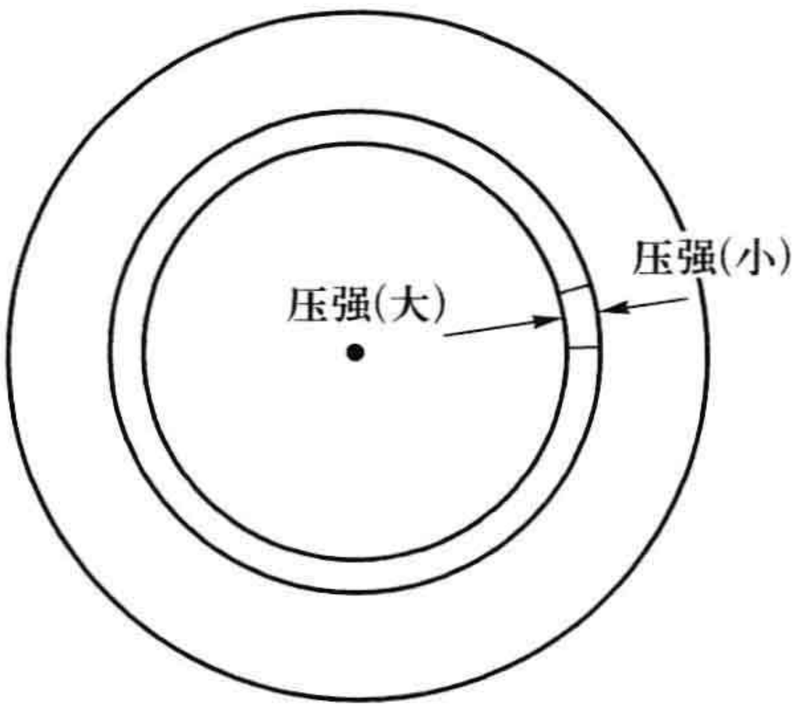


图 7 – 2　薄球层内外壁压力之差是外向的

然而中心球内部的氢总有烧完的一天,烧完后的命运又将如何? 估算表明,太阳中心球内的氢大约还能再烧 50 亿年,所以太阳正值它的“中年”阶段。想一想,50 亿年! 从实用角度来说,我们大可不必“杞人忧天”。然而,从理论上我们当然非常关心太阳(以及其他恒星)在中心球内部的氢燃烧殆尽之后的命运。大致地说,这时的中心球已变成一个氦球,但周围还有一个薄层正在烧氢变氦,薄层以外由于温度不够而仍然是氢,星体内部的情形可由图 7 – 3 粗略表示。当中心球的温度尚不足以点燃氦的核聚变时,情况与先前(当中心球温度尚未达到点燃氢时)的情况类似:氦球在引力作用下收缩,同时变热。这又加剧周围薄层的氢燃烧,从而导致星球外部物质的膨胀和冷却,最终变成一颗**红巨星**(red giant)。“红”是由于表面温度降低,“巨”则因膨胀得名。红巨星的表面温度虽然很低,但因为有巨大的体积,所以显得很亮,肉眼所见的最亮的星中就有不少

红巨星。可见,太阳在大约50亿年后将逐渐变成一颗红巨星,八大行星中最靠近太阳的水星和金星都将被“吃”进去,地球(“老三”)的海洋将被煮沸,继而整个地球也将被膨胀到最大半径的太阳所吞没并烧焦,但其残骸仍将绕日公转。

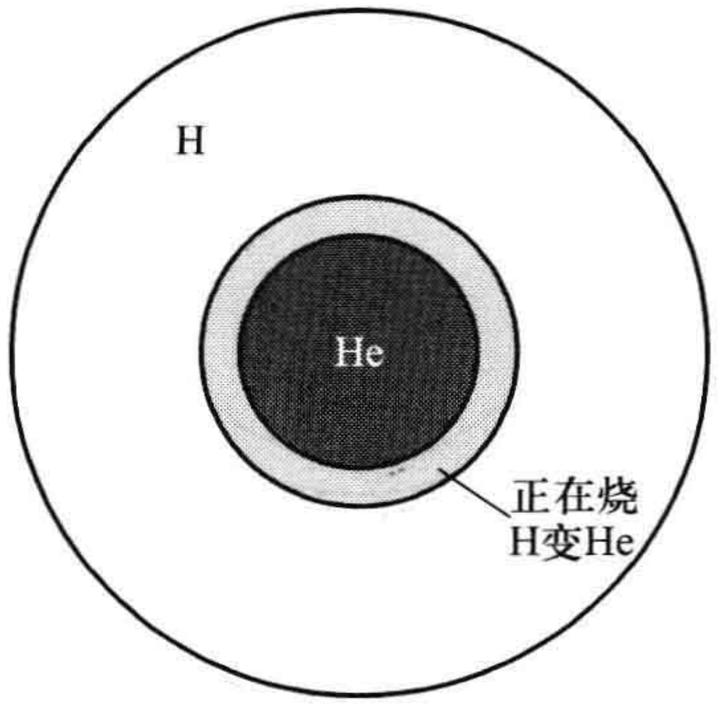


图7-3　中心球内部氢烧完后的略况(不按比例)

讨论红巨星时要分清中心球及其外部。随着中心球(氦球)的收缩和变热,球内的某个(较小的)中心球的高温足以点燃氦的核聚变——烧氦变碳(及氧),直到烧成碳-氧中心球时核反应再度停止,中心球再度在引力作用下收缩。另一方面,中心球外部的烧氢范围逐渐外延至外壳,外部物质不断脉动式地脱落,整个星球最终只留下一个碳-氧中心球。

7.2.2　白矮星

现在的问题是:这个碳-氧中心球还会进一步演化吗?恒星的晚年命运因其质量而异。对于质量不很大的恒星(包括太阳),引力不够强,中心球的引力收缩不能提供足够温度使碳和氧发生核聚变,靠原子核能维持平衡就不再可能。还有没有什么力量足以抗衡引力?在经典物理学中找不到这样的力量(无法造成足够的外向压力以抵消内向的引力)。按照经典物理学,要有高压就必须有高温(不妨称之为“热压强”),由于星体不断辐射(损失)能量,除核反应外没有任何机制可以提供能量以维持高温(没有能源)。然而,根据量子物理学,即使在绝对零度时也可能存在可观的压强。由于高温,中心球内的原子早就处于电离状态,里面存在大量电子,它们的行为很像理想气体(称为“电子气”)。在经典物理学中,电子的平均动能与温度 T 成正比①,故 $T=0$ 时平均动能为零,电子气处于能量为零的状态。然而,根据量子物理学,电子服从“泡利不相容原理”,

① 平均动能等于 $3kT/2$,其中 k 为玻耳兹曼常数。

很粗略地说,当它们被挤得太紧时就有向外扩张的趋势,相应于某种外向压强,称为电子**简并压**(degenerate pressure),它不是一种热压强("完全简并电子气"指的就是绝对零度下的电子气),不妨称之为"冷压强"。简并压在普通密度下微不足道,但在密度很高的中心球内部却有可能与引力抗衡。质量较小的中心球内部引力较小,简并压足以抵消引力,遏止收缩,使中心球成为一个永不收缩的、稳定的星体,称为**白矮星**(white dwarf)。"矮"是因为比普通恒星小得多,"白"则因表面温度很高(达到白热化)而得名。恒星一旦演化成白矮星,就不再有任何重要的演化过程,可以说已经死亡(尸体就是白矮星),只因温度高于外界而继续向外辐射能量。由于没有能源,辐射导致温度下降,直至温度低到不再辐射,最终只能变成一颗黑矮星。

人类观测到的第一颗白矮星称为天狼 B 星。天狼星(现在叫做天狼 A 星)是夜空中最亮的恒星。德国天文学家**贝塞尔**(Bessel)从 1834 年起对天狼星进行观测,发现其运动轨迹与众不同。大多数恒星在天球上的轨迹都是一个大圆弧,但天狼星却在大圆弧上叠加上一个微小的波浪形,说明其运动受到周期性的干扰。贝塞尔认定这种干扰只能起因于引力,而且能够提供如此巨大引力的物体只能是附近的一个星球,一个与天狼星为伴的"伴星"。然而这是一个黑暗的伙伴,尽管贝塞尔付出了异乎寻常的努力,至死仍未看见这颗伴星(现在我们知道这是因为它太小的缘故)。28 年后的 1862 年,美国天文学家兼望远镜制造匠**克拉克**(Clarke)在测试自己刚造好的优质高倍望远镜时发现天狼星附近有一个很暗的亮点,后来肯定这就是贝塞尔当年预言而看不见的伴星,于是"黑暗的伙伴"升格为"昏暗的伙伴"。从此,天文学家才认识到天狼星其实是由两颗子星组成的一个双星系(两者互为伴星,围绕着它们的共同质心转动),并把其中明亮的和亮度低得多的两颗子星分别称为天狼 A 星和天狼 B 星。后来知道天狼 B 星就是一颗白矮星,其质量为 $0.85M_\odot$($M_\odot$代表太阳质量),而体积比地球还小,所以密度异乎寻常地高,约为水的密度的 6×10^4 倍。

英国著名物理和天文学家**爱丁顿**(Eddington)在他 1926 年出版的《恒星的内部结构》一书中写道:"白矮星可能很多,我们只是有限地知道 3 个,它们都离太阳很近……其中最著名的是天狼星的伴星。"书中还介绍了利用望远镜观测结果估算天狼 B 星的质量和周长的方法。利用质量和周长便可求得密度,发现天狼 B 星的密度大得惊人(大到当时的天文学家认为荒谬的地步),但却是必须承认的事实。爱丁顿的书也介绍了恒星内部热核反应所保持的高温高压足以抗衡引力,这种高压就是与高温相应的热压强。但是,内部没有核聚变的星体(例如白矮星)辐射出去的能量得不到补充,温度和压强不断降低,无法抗衡引力,又怎么能够稳定?(而白矮星却是稳定的!)当时的天文学家尚未认识到除了热

压强之外还有由纯量子起因的冷压强(简并压),所以爱丁顿无法解释白矮星的平衡机制,在书中表现出了困惑。虽然难以解释,但是白矮星的存在却是观测事实。而且,爱丁顿以及当时几乎所有天文学家和物理学家(包括爱因斯坦)都默默地持有一个信念,认为所有恒星在核燃料耗尽后的最终归宿只能是白矮星,而不是其他更加难以想象的"怪物"。特别是,他们非常不愿意相信某些恒星由于不断收缩而最终缩为一点(密度、温度都无限大的点)这样一种离奇的结果。当然,今天我们早已知道他们的想法是不对的,恒星演化的最终产物("尸体")不但可以是白矮星,还可以是中子星,甚至黑洞,详见后面的介绍。

最先想到在白矮星中抗衡引力的力可能来自简并压的人是英国物理学家**否勒**(Fowler),他在1926年底发表的文章中提出了这一观点。这就从根本上克服了爱丁顿的困惑。然而否勒并未就简并压与引力平衡的细节做过探讨和计算。卓有成效的、千古流芳的计算是印度天体物理学家**钱德拉塞卡**(Chandrasekhar)在19岁时做出的。钱氏1910年生于印度的一个中产阶级家庭,从小就有神童的美誉。读大学时由于出类拔萃而获得政府为他赴英深造的特设奖学金。那时他心目中的偶像是爱丁顿和否勒。1930年,19岁的钱氏登上了赴英的轮船,脑子里充斥着爱丁顿书中表达的困惑和否勒用简并压的解决方案。他确信只有电子简并压才有可能充当白矮星内抗衡引力的压强。利用船上的18天航行时间,钱氏对于电子简并压与引力的消长情况以及白矮星内部结构的细节做了详细计算,得到一个令人吃惊的结论:只当恒星中心球的质量 M 小于太阳质量 $M_\odot$ 的1.4倍时,即只当 $M<1.4M_\odot$ 时,引力才可能被电子简并压平衡,中心球才成为一颗白矮星。质量大于 $1.4M_\odot$ 的任何中心球都不能以白矮星为演化的归宿,它们将在引力作用下继续收缩下去。至于其最终产物,当时的钱氏当然也不得而知。后来人们把 $1.4M_\odot$ 称为**钱氏质量上限**(Chandrasekhar limit)。这一上限的推导思想和过程见梁灿彬、周彬(2006)选读9-3-3。

钱氏在船上把他的研究结果写成两篇文章(第二篇是关于白矮星质量上限的计算),到英国剑桥大学后就交给了否勒教授。否勒只同意发表第一篇。由于对结果感到迷惑不解,他不同意发表第二篇论文。钱氏只好将它改投美国的《天体物理学杂志》,一年后被发表,可惜没有引起天文学界的注意。剑桥大学的爱丁顿一直注视着钱氏的研究进展,但从不明确表态。1935年,在爱丁顿的安排下,钱氏终于有机会在英国皇家天文学会报告自己的工作。在他结束报告并且特别强调了他的白矮星质量上限后,主席请爱丁顿讲话。爱丁顿出其不意地指出,如果钱氏的质量上限果真存在,那些质量足够大的恒星在核燃料用尽后就不会像太阳那样平静地死亡,它们将不断收缩,其结果将不堪设想(缩为一点?!)。他严厉地当众宣称钱氏的计算和结果全盘皆错,甚至还将钱氏的讲稿

撕成两半。由于相信(甚至迷信)爱丁顿的权威,与会者都以为钱氏真的犯了一个严重错误。这个突如其来的打击不但给钱氏的心灵造成巨大伤害,而且断送了他在英国发展的前途,使他无法在英国谋得任何合适的职位。他只好暂时不再发表白矮星方面的工作,把注意力转移到其他研究领域,后来陆续取得出色成果,终于成为天体物理学中涉猎甚广的大师。爱丁顿的打击也许正是造就这位大师的诱因。1937 年,他受到芝加哥大学叶凯士天文台的聘任赴美工作,直至 1983 年荣获(迟到的)诺贝尔物理学奖以及 1995 年与世长辞。

爱丁顿是 20 世纪初期的英国杰出天文学家和物理学家,科研成果丰硕卓著。他是第一个坚信恒星的能源是核能的人。广义相对论面世时适逢第一次世界大战,德英两国又是交战国,但爱丁顿还是很快就弄到了爱因斯坦的论文,读后欣赏有加,成为第一个用英语宣讲广义相对论的人。战后的 1919 年,为了验证爱因斯坦关于恒星发到地球的光线会受太阳引力偏转的预言,他带队到巴西进行日全食观测,取得了支持爱因斯坦预言的证据(详见小节 8.5.1 和 8.5.3)。他与爱因斯坦结下了深厚的情谊。传说在他带队观测日食后不久,有记者向他提问:“听说当今世界只有三个人懂得相对论,对吗?”他沉思后反问道:“那第三个人是谁?”总之,爱丁顿是一个智商超群又极度自信的伟大科学家,但是,不得不说他对钱氏质量上限的武断否定是他在学术生涯中犯下的一个严重错误。

7.2.3 中子星

既然白矮星有个质量上限 $1.4M_\odot$,自然要问:如果中心球的质量大于 $1.4M_\odot$,情况又如何? 我们从小节 7.2.2 的末段出发继续讨论。对于质量不很大(大约在 $8M_\odot$ 以下)的恒星,由于演化过程会丢失大量物质,其碳 - 氧中心球的质量将小于钱氏上限 $1.4M_\odot$,最终成为白矮星(恒星死亡后的尸体)。但是,如果碳 - 氧中心球的质量大于 $1.4M_\odot$,由引力收缩所导致的高温足以点燃碳的核聚变(烧碳变氖等),释放出的核能所维持的高温高压足以再次抗衡住引力,中心球得以再度平衡。由于温度越高的核聚变反应越快,这个中心球的碳将很快耗尽,于是再次出现引力收缩。以后还会几度重复如下的循环:

“收缩—核聚变—中心球形成—收缩”,

其中每个中心球内都含有前一个(较大的)中心球内所不含有的元素(的原子核)。仿照图 7 - 3[那里把星体分为三“层”(三个区域)],不难理解现在的星体被分成许多层,形成所谓的“洋葱头结构”,粗略地如图 7 - 4 所示。请特别注意最里面的一“层”(最小的中心球),其主要成分是铁,所以称为中心铁球。这是密度甚高的气态铁球(而且高度电离化),与常人脑子里的普通密度的固态铁球非常不同。铁核是结合得最紧的原子核(核子的平均结合能最大),不可能因核

聚变而放能。所以,一旦烧成中心铁球后,球内核燃料就告用光。刚开始形成时,铁球内的电子简并压尚可抗衡引力,但质量一旦超过 $1.4M_{\odot}$ 就会突然剧烈收缩,其猛烈程度根本无法用通常语言形容,只能称之为**引力坍缩**(gravitational collapse),其温度在 1/10 秒内竟能猛增至 50 亿摄氏度! 这种甚高密度导致很强的引力,所以从这一阶段开始必须用广义相对论代替牛顿引力论,结果是:为达到平衡所需的中心压强更大,因此平衡更难实现。在如此高温高密下,铁球内的光子有异常高的能量,足以将铁原子核打碎成中子和质子("光分裂"),而质子又将同电子结合而成中子①,于是中子充满了整个中心球。这一过程称为"中子化"。中子和电子一样也服从泡利不相容原理,一样也有简并压。只要中心球质量不太大,中子简并压也能抗衡引力而永葆平衡。这种依靠中子简并压支撑的稳定星体称为**中子星**(neutron star)。与白矮星类似,中子星也可看作恒星死亡后的尸体。中子星内的密度高达水的密度的 10^{14} 倍(!!!),等于原子核内的密度,因此不妨认为中子星是一个超巨型的"原子核"。物理学家对于如此高密下的原子核的行为(因而对中子星的状态方程)的了解还很不确切,这给中子星质量上限的估算带来不确定性。不同文献给出不同结果,只能大致说中子星的质量上限约为 $2M_{\odot}$(或说介于 $1.5M_{\odot}\sim3M_{\odot}$ 之间)。中子星比白矮星还要小得多,典型的中子星半径只有 10 km 的量级,而白矮星的半径约在 3 千至 2 万公里之间。中子星是一种非常特别和复杂的天体,它有各种极端(超常)的表现:高达核密度的密度、异乎寻常的强磁场(高达 10^{8} 特斯拉)、甚高速的旋转(频率从 1 Hz 至 1 000 Hz)、离光速不远的高声速、超流的内部……人们至今还难于对它了解得非常透彻。

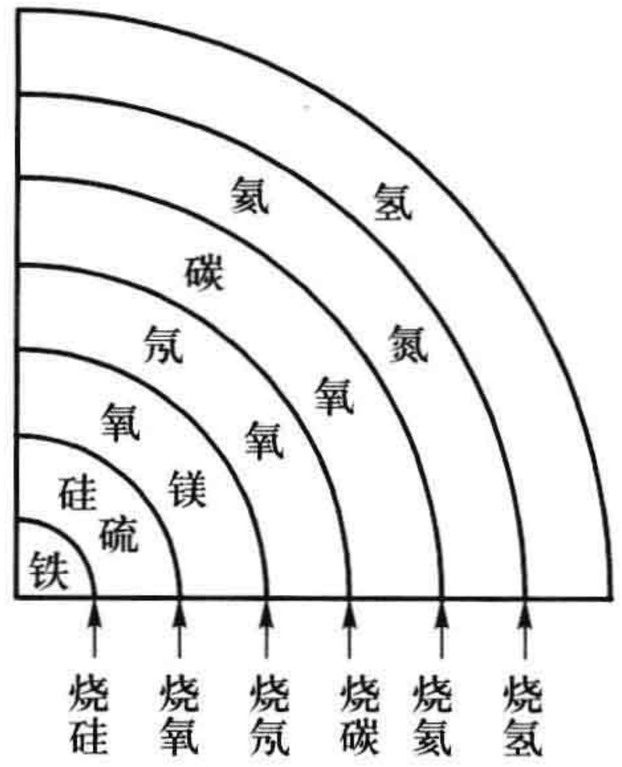


图 7-4　中子星形成前的"洋葱头结构(不按比例)"。温度和密度越近中心越高。

中子星的第一个理论模型是**奥本海默**(Oppenheimer)和他的学生**沃尔科夫**(Volkoff)于 1939 年发表的②。由于文章没有给出可观测的物理效应,长期以来

① 同时放出中微子(neutrino)。

② 苏联杰出的理论物理学家朗道于 1938 年 2 月受政治迫害被捕入狱。被捕前不久他写了一篇文章,认为恒星内可能存在"中子核",被《自然》杂志于 1938 年 2 月发表。奥本海默曾经认真研读过这一文章并受到重要启发。

被天文学家束之高阁,无人问津。当时正值第二次世界大战,文章发表后不久,珍珠港事件爆发,美国对日宣战,奥本海默本人也忙于研制原子弹而无暇顾及天体物理学。(他是美国研制原子弹的曼哈顿工程的领导人,享有原子弹之父的称号。)对中子星的研究阴差阳错地被冷落了28年,直到1967年才因为某种偶然发现而成为重要的研究课题(详见下一小节)①。奥本海默在二战后任美国原子能委员会的首席顾问,他为防止核扩散和停止美苏军备竞赛而向美国政治家大胆直言,1954年(当时美国盛行麦卡锡主义)在受到关于秘密工作人员的忠诚调查后被撤职,继续进行研究、写作和授课(他是个讲课极其出色的教员)。大约十年之后,美国总统肯尼迪对他授予费米奖金,以示给他恢复政治名誉。

7.2.4 超新星

恒星的晚期坍缩还会出现戏剧性的一幕:在中子星即将形成的瞬间,整个星体的非中子化外层在引力作用下以极高的速率落到中子化中心球的表面,遇到一堵由中子简并压支撑着的坚硬的"墙壁",巨大的能量被迫反向释放,使外层物质向外反弹,形成威力无比的冲击波,整个外层向四面八方爆炸出去,放出的大量光子使得星体的光度陡增至数十亿倍之多(可以比它所在的整个星系还亮)②,这一壮观称为**超新星爆发**(supernova explosion)。不妨形象地说超新星爆发是宇宙中最为壮丽辉煌的"焰火秀"。顾名思义,"超新星"似乎代表一个新生婴儿的诞生,但只要你明白了现象的实质,不如更为恰当地把它比喻为一个垂死老人在死亡前的回光返照,而死亡后的尸体正是位于中心的那颗中子星③。

我们的宇宙在历史的长河中已经有过无数次超新星爆发事件,但只有发生在银河系内或者银河系紧邻处的超新星爆发才有可能被地球人用肉眼看到。我国古代连续两千多年的天象记录中就有多次超新星的记载。由于超新星在天空某处突然出现又渐消失(看不见),有如过客一般,所以我国古代天文学家称之为"客星"。公元1006、1054、1181、1572和1604年的客星记录都是银河系的超新星事件,其中最著名的是1054年(我国北宋时期)观察到的超新星。当时的皇家天文学家杨惟德对此做了详细记录,其中写道:"**昼见如太白,芒角四出,色赤白,凡见二十三日**"。解释:"太白"是金星的别名,是夜空中最亮的行星。"昼

---

① 此前也有人寻找过中子星,但是未获成功。中子星半径太小导致发光面积太小,发出的可见光只有太阳光的百万分之一。就算你看到它,你也说不清这是中子星还是普通的暗星。其他探寻中子星的方法也均无所获。

② 虽然这些光子携带出去巨大的能量(光能),但是超新星爆发所释放的能量主要是由中微子携带出去的。中微子在短短几秒钟内就带走了绝大部分能量,耗尽能源。

③ 也可能是白矮星或黑洞(见小节7.2.5末)。

见如太白”是说该超新星白天看来也像晚上看到的金星一样明亮,“凡见二十三日”是指白天如此明亮的情况持续了23天。晚上能看见的时间当然更长,大约持续了两年。这颗超新星位于金牛星座,离地球约5 000光年(以天文尺度衡量,这已经很近了),爆发产生的光波要走5 000年才到达地球,所以不应说“金牛座在1054年爆发了一颗超新星”,而应说“地球人在1054年看到了金牛座中一颗超新星传来的光”。

离今天最近的、肉眼可见的超新星是在1987年被地球人看到的,编号是SN1987A[SN是“supernova”(超新星)的简写,1987A代表1987年被发现的第一颗],它在爆发前是大麦哲伦星系(银河系的近邻星系)的一颗普通恒星,距地球约17万光年,是望远镜问世后看到的离地球最近的第一颗超新星。1987年2月24日凌晨,在智利某天文台①工作的加拿大天文学家**谢尔顿**的助理首先以肉眼看到了这颗超新星(最亮时为4等星)。一时间,全世界的许多天文学家纷纷赶赴南半球去搜集第一手资料。这是1987年轰动国际物理学界、天文学界的两大事件之一②,另一事件是高温超导材料研究的突破性进展。

7.2.5　脉冲星

脉冲星的发现与射电望远镜密不可分。从光学(可见光)望远镜到射电望远镜是天文观测手段的一大进步。“射电”是“无线电波”的简称,是电磁波中波长最长的一段($>10^{-3}$ m)。由于波长比可见光波要长得多,为了提高分辨率,射电望远镜的天线口径要大得离奇。1962年,英国剑桥大学的**赖尔**(Ryle)利用干涉原理在免除大口径天线的前提下大大提高了射电望远镜的分辨率,为此获得1974年诺贝尔物理学奖。从此出现了射电天文热。

1967年,剑桥大学的女博士生**贝尔**(Bell)在仔细查看射电望远镜的长长的记录纸带时注意到一种奇怪的周期性脉冲信号,虽然脉冲的形态和高度各不相同,但周期(相邻脉冲的时间间隔)却保持常数(等于1.33秒)。她马上向导师**休伊什**(Hewish)汇报,导师也不知道这一信号来自何方、意义何在。莫非这是地外生命发出的信号?在内部学术讨论会上的确有人持此猜测,而且还冠以LGM(the“Little Green Man”,即“小绿人”)的称号。不久之后,贝尔又连续发现其他三个类似的脉冲信号,周期都是秒的量级。四种“小绿人”频繁光顾的可能性不大,所以他们逐渐倾向于相信这是来自某种(尚不清楚的)天体的脉冲,并

---

① 拉斯康帕纳斯天文台。

② 后来西班牙天文爱好者又发现了一颗引起轰动的超新星,此即SN1993J,最亮时的视星等为10.8,离地球比SN1987A远6倍,也算是地球的近邻。

称这种天体为**脉冲星**(pulsar)。他们在1968年1月就第一颗脉冲星向英国的《自然》杂志投稿,同年2月9日,全世界都获知他们发现了脉冲星的重要消息。由于文中提到剑桥的天文学家在一个时期内也曾考虑过收到天外来客所发信号的可能性,而且其中还涉及一位年轻女性,新闻媒体胃口大开,加油添醋地抓紧做足了宣传报道。后来休伊什由于发现脉冲星而与赖尔分享了1974年的诺贝尔物理学奖。人们对贝尔小姐未能获奖颇有微词,她的第一手发现只在她的博士论文的一个附录里被提及。

然而,脉冲星究竟为何物?虽然企图解释脉冲星的理论文章一时间充斥于各种学术刊物,但只有天文学家**戈尔德**(Gold)在1968年发表的观点很快被学术界普遍接受。他认为脉冲星无非就是快速自转的中子星。中子星有强磁场,以接近光速运动着的电子会发出一种称为"同步辐射"的电磁波,形成随中子星自转而一同旋转的射电波束。由于自转轴与磁轴不重合(这是天体的常见现象,地球也如此),中子星就像旋转着的灯塔那样(见图7-5),每当射电波束扫过地球时我们就接收到一个脉冲。

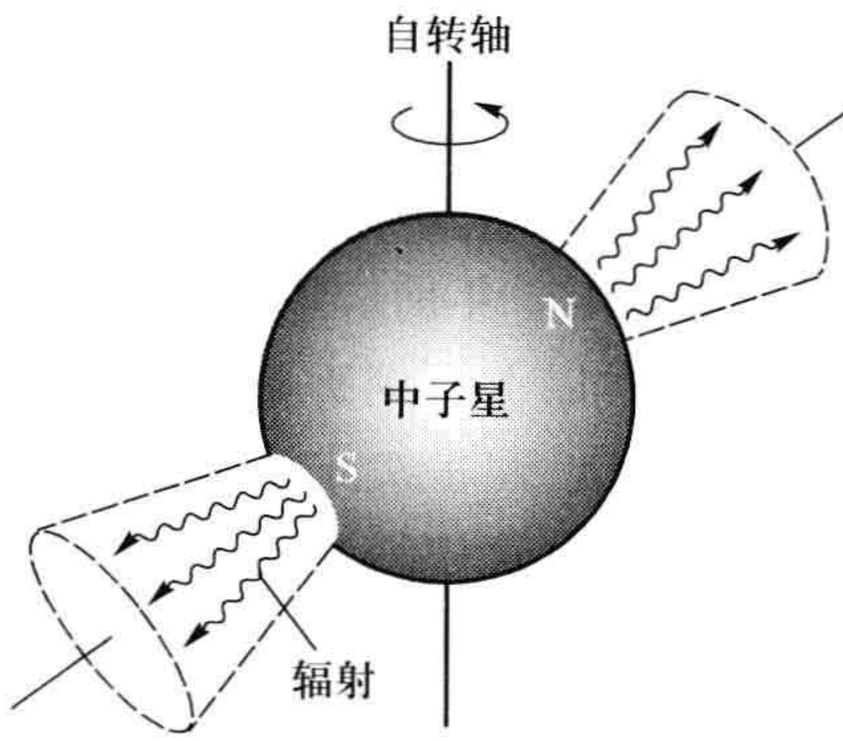


图7-5 波状箭头代表电子发出的电磁辐射,形成的两个"辐射锥"随中子星自转而旋转,产生"灯塔效应":地球人只当被锥扫过才收到辐射(脉冲),相邻脉冲的时间间隔就等于自转周期。

值得强调的是,虽然奥本海默与沃尔科夫早在1939年就预言了中子星的可能存在性,但直到脉冲星发现后才有人想到中子星。如果射电波果真像"脉冲星"的名称那样来自某颗星体,脉冲的形成很可能就是因为星体有自转。然而,以秒为量级的自转周期未免也转得太快了吧(地球的自转周期是24小时)?星体表面的物质岂不是会在(惯性)离心力作用下四散纷飞?星体岂不是会被撕碎?于是想到致密星体(半径小、密度高的星体)。在相同的角速率下,半径越

小的星体的表面线速率越低;更重要的是,密度越高的星体表面引力场越强,有可能使星体不被撕碎。当时天文学家熟知的致密天体就是白矮星,但是估算表明也仍然不足以防止在如此高速自转下散架。于是才想到奥本海默在1939年预言的、比白矮星还要致密得多的中子星。中子星从被预言开始,沉默了28年(直到1967年)方才露出其"庐山真面目",这个"天机"完全是靠中子星既有自转又有射电辐射而"泄漏"出去的。电磁波带走的其实是中子星的自转能,所以辐射会导致自转变慢,即自转周期变长。

小节7.2.4开头关于"回光返照"和"尸体"的比喻很形象地描写了超新星爆发与中子星的关系。读者在得知脉冲星就是中子星之后,不免会好奇地问:超新星爆发的遗迹(残渣)中是否存在中子星?答案是有趣的。最著名的超新星遗迹就是位于金牛星座的蟹状星云(由气态物质组成)。根据星云的大小和气态物质相互离开的速度可以反推出爆发时间,结果是:爆发产生的光子约在公元1000年到达地球。最后确认这正是SN1054的遗迹。美国某射电天文台于1968年底果然在蟹状星云里找到一颗脉冲星(中子星),脉冲周期为0.033秒(每秒转33圈,非常快),这就是SN1054爆发后留下的尸体。这个周期大约仅为贝尔首次发现的脉冲星的周期的1/40倍,说明它比贝尔脉冲星要年轻得多。此外,悉尼大学的天文学家于1968年在船帆星座中也发现了一颗脉冲星(离地球只有800光年),并证实它是人类在1万1千年前应该看到的一次超新星爆发所留下的尸体(可惜当时不可能有记录)。以上两例听起来都很舒服,因为它们似乎说明超新星遗迹必含中子星。然而问题远非如此简单:超新星与中子星并非总是这样"成协"地出现的。SN1987A的遗迹中一直找不到中子星就是一例,还有许多其他例子。天文学家对此曾做过各种解释,但这个"成协"问题仍是一个不断探讨的课题。

§7.3　奇点、视界和黑洞

如前所述,恒星在演化过程中会大量丢失质量。大致地说,原来质量小于$8M_\odot$的大多数恒星的晚期质量都小于$1.4M_\odot$,最终成为白矮星;原来质量在$8M_\odot \sim 20M_\odot$范围内的大多数恒星最终成为中子星。但是,质量大于$20M_\odot$的恒星的晚期归宿又是什么呢?还有什么力量可以抗衡引力以使星体保持平衡吗?答案是:没有了!这意味着星体会在引力作用下不断地、剧烈地、灾难性地坍缩下去,最终只能缩为一点!这是密度、温度以及时空曲率都无限高的点,称为**奇点**,有极度离奇的性质,称为**奇性**(奇点和奇性在英语中都是singularity)。物理学家(包括爱因斯坦在内)长时期来都无法接受奇点的存在。奥本海默是第一

个证明恒星坍缩会产生奇点(以及黑洞)的人,他与学生**斯奈德**(Snyder)对高度理想化的球对称恒星①的晚期径向坍缩做了定量计算(求解爱因斯坦方程),并于1939年发表文章公之于众。结论是,对这种高度简化的恒星而言,① 最终必定坍缩为奇点;② 以奇点为球心、以某个适当长度为半径的球状区域内的引力场是如此之强,以至于任何粒子(包括光子)都会被引力死死地拉向奇点。这个区域的表面(球面)就成为一个有进无出的"单向膜"——任何粒子一旦从外部穿越球面进入球内,就再也不能出来,连光子也不例外。于是,这个区域很自然地被称为**黑洞**(black hole),而黑洞的表面(那个有进无出的单向膜)则被称为**视界**(horizon),代表"视觉的分界面"。"black hole"一词是美国大物理学家**惠勒**(Wheeler)于1967年首次提出的(以前别人用过不同称谓),他是个很善于为物理对象起名的人②。"黑洞"是对这一"怪物"的最确切不过的称谓,因为它不仅仅黑,而且还是一个洞:任何东西一旦掉了进去就再也不能出来。

奥本海默和斯奈德的上述论文不但写作精美,结论明确,而且在短短几页的篇幅中就把恒星晚期坍缩的所有细节都用公式表达了出来。你可以在这些公式中找到例如以下问题的答案:远方观者看到的表现是什么?坍缩星内部看到的是什么?坍缩星表面看到的又是什么?然而,由于奇点过于离奇,尤其是许多物理量在趋近奇点时都趋于无限大,许多物理学家当时都无法相信奇点的物理存在性。反对声浪来自美国、西欧和苏联的一批大物理学家(不过朗道从一开始就是支持的),反对的理由不止一个方面,很重要的一点就是奥本海默等所用的恒星模型过于理想化。单是球对称性一条就招致不少反对意见:任何恒星都不可能具有理论上所假定的那种完美无缺的球对称性。通俗地说,他们认为不应排除"差之毫厘,失之千里"的可能性,就是说,虽然恒星的初始状态与球对称性只有微小的偏离,但晚期坍缩的结果却可能明显不同,最终的奇点就有可能避免。一种强有力的反对意见来自苏联的**卡拉特尼科夫**和**栗弗席兹**,他们认为奇点正是由奥本海默的理想化条件导致的,真实恒星的形状、密度等等总有微弱的随机变化,这种随机变化在坍缩过程中会被放大,并在奇点形成之前遏止坍缩(通俗地说就是"差之毫厘,失之千里")。他们在文章中还用爱因斯坦方程在奇点附近的幂级数解来支持上述观点。

就在多数物理学家和天文学家不相信奇点存在的时候,英国的**彭罗斯**(Pen-

---

① "高度理想化"是指恒星内部压强为零、密度是常数[专业术语是"均匀密度尘埃球(dust ball)"],等等。当时的物理学家和计算机都不可能对更接近真实的恒星模型进行计算。即使是这样高度理想化的模型,计算仍然十分困难,不过都被他们巧妙地克服了。

② 惠勒过去曾是奥本海默黑洞的激烈反对者,从20世纪60年代开始转变为热情的支持者。

rose)和**霍金**(Hawking)却在 1965—1970 年借助于微分拓扑学(differential topology)独辟蹊径地证明了一系列定理(后来称为**奇性定理**),大意是:无需依赖对称性假设,大质量恒星晚期坍缩和宇宙原初的"大爆炸"奇性在一定条件下都是不可避免的。彭罗斯在大学和读博期间主修数学,本科四年级(1952 年)时因为受到广义相对论学家**沙马**(sciama,后来是霍金的博士生导师)的影响而立志将数学用于物理学研究。虽然大多数物理学家都很重视数学,但是此前的物理学家完全想不到在物理学中应用微分拓扑学。由于具有坚实的微分拓扑功底,彭罗斯在思考恒星坍缩问题时突然产生一种灵感,逐渐意识到可以用定理的形式论证奇点的存在性,后来便有了他在 1965 年发表的开创性论文,其内容被公认为第一个奇性定理。定理断言,只要时空满足某些不很苛刻的条件,就必然存在奇性。重要的是这个定理不以球对称性为前提,该文的原话是:"**偏离球对称性不会防止时空奇点的出现**"。这篇文章不仅证明了奇点在一定条件下的存在性,而且向物理学家显示了微分拓扑的重要性。它的问世立即引起一小批物理学同行的注意,他们纷纷尽快熟悉有关理论工具并寻找能做出成果的课题。霍金便是其中的佼佼者,他马上意识到彭罗斯的证明可以被"底朝天"地用到宇宙的研究中,很快(同年)便发表了自己的奇性定理,大意是:只要满足某种称为"能量条件"的条件,宇宙[①]的原初必定存在大爆炸奇点(原初奇点),即使宇宙存在局域不均匀性也不妨碍原初奇点的存在。此后,一系列论证奇性存在的文章的预印件相继涌现。由于所需的一些工具当时尚不够成熟,某些结果或证明并不正确(后来有些得到了改正)。所有奇性定理都离不开以下三个前提条件:① 重力是吸引力而非排斥力(这只是粗略的物理提法,专业提法是满足不同层次的"能量条件");② 时空中某些区域的引力强到任何物质一旦被俘获就无法逃逸的程度(专业提法是存在"陷俘面"或其他);③ 有合理的因果关系(专业上又分成几个层次)。各种奇性定理无非是说,只要上述三类条件的某种适当组合成立,奇性就必定存在。如果要弱化三类条件的某一类(对它降低要求),就要适当强化其他两类。对奇性定理的研究大约持续了 5 年,霍金与彭罗斯在 1970 年合作发表的文章可以被视为最后的(也是最强的)一个奇性定理,它比其他几个奇性定理有宽广得多的适用性,特别是,它给出了强烈的理由使人相信我们的宇宙的原初确实存在大爆炸奇点。霍金本人非常重视这一成果,他说过[②]:"**现在几乎人人都相信,宇宙以及时间本身在大爆炸处有一开端。这个发现比发现各种非稳定的粒子重要得多了,但是它还没重要到能赢得诺贝尔奖金的青**

---

① 宇宙可分为三种,分别由 $k=0$,$k=-1$ 和 $k=1$ 标志。该文只涉及 $k=0$ 和 $k=-1$ 的宇宙。

② 引自霍金、彭罗斯著,杜欣欣、吴忠超译,《时空本性》第 18 页。

昧。”虽然当时许多物理学家和天文学家都看不懂奇性定理的证明，但都相信证明是对的，于是出现有趣的“一边倒反向”现象：从 1965 年前的“一边倒”（普遍不相信奇点存在）变为 1970 年后的、倒向另一边的“一边倒”（普遍接受奇点）。于是对奇性的研究应运而生。然而这种“奇性研究热”只持续了大约十年，这与“量子引力论”的研究进展密切相关，见下面的三段。

奇点的存在性虽然已经以定理的方式被确定下来，奇点处的各种无限大量仍然使物理学家深受困扰。物理学家普遍相信真实世界的物理量不会变得无限大，而奇点定理的证明又无懈可击，就只有再仔细审查它所依赖的前提条件。奇点定理的前提是默认广义相对论的完全正确性。既然物理量（包括时空曲率）在趋近奇点时趋于无限大，我们不免要问：在时空曲率出奇地大的时空域内广义相对论还成立吗？估算表明，“曲率出奇地大的时空域”是指与奇点相距小于 10^{-35}m（这一数值称为**普朗克长度**）的区域。物理学家普遍相信，对于这样一个非常微观的领域，量子力学的效应必须考虑（量子力学本来就是因为微观粒子不服从经典力学而出现的）。由于这一区域既涉及强引力又涉及量子力学，要弄清奇点的情况就必须依靠一个新的理论，即“量子引力论”。下面先插入一段关于量子引力论的简介。

物理学从 19 世纪末开始出现危机，导致 20 世纪初的两大革命——相对论（狭义和广义）的创立和量子论的诞生。当时人们把相对论和量子论一同称为“现代物理学”，而把此前的物理学称为“经典物理学”。相对论虽然在时空观以及引力的实质等问题上十分革命，但在一个重要方面却非常“不革命”——它坚持非量子论的观点。例如，稍有量子力学知识的人都知道，微观粒子在一个时刻并无确定的位置，它既可能在这里，又可能在那里，对粒子在某时刻的位置只能做概率性的预言。然而相对论却用世界线描述粒子的历史，这分明是认定粒子在某一时刻有一个确定的位置。种种迹象告诉物理学家（特别是相对论学家），无论狭义和广义相对论在理论上多么优雅，只要不是有机地与量子论结合，就不会是完全正确的理论。奇点之所以有如此难以接受的奇性，正是由于奇点紧邻处的时空曲率异常之大，大到非考虑量子效应不可。只有与量子论有机结合才有可能避免奇性。从 20 世纪 30 年代开始，某些先驱物理学家就致力于将广义相对论“量子化”，用现代语言来讲，就是企图创建一个前所未有的**量子引力论**（quantum gravity）。一大批国际一流的相对论学家、场论（及弦论）学家以及数学家多年来都在为相对论与量子论的“联姻”争当“红娘”。随着研究的深入，为了同量子引力论相区别，国际理论物理学界早就把不与量子论结合（没有“量子味”）的广义相对论称为**经典广义相对论**，把“经典”一词用作“非量子”的同义词。于是爱因斯坦的狭义和广义相对论都属于经典物理学的范畴。“经典物理

学”的这一界定标准早已相当普遍地在国际理论物理界得到公认。可惜至今尚有太多的人不知道这一点,例如他们仍然喜欢说“经典力学是狭义相对论的低速近似”(分明认为相对论不属于经典物理学),而我们则说“牛顿力学是狭义相对论的低速近似”。我们不禁要问他们:你们对“经典”一词的理解未免也太“经典”(古典)了吧?

虽然量子引力论至今尚未完全创建成功,但多种途径的研究都在不断地取得可喜的进展。早在20世纪80年代初期,霍金沿着自己的途径通过对某些“玩具模型”的研究就逐渐建立起一种信念(但不是证明):只要考虑量子效应,黑洞内的奇点是可以避免的。你也许会问:他用奇性定理证明了奇性的存在性,现在又说奇性可以避免,这不是自打嘴巴吗?非也。这种“有也霍金,无也霍金”的现象其实是科研的一大进步。奇性的存在性是纯经典广义相对论(一点量子味都没有)的结论,后来加上量子理论的考虑,结论当然可以不同。现在的公认看法是:与其说奇性定理证明了奇性的存在性,不如说它表明了经典广义相对论在奇点附近的不适用性。这再次说明经典广义相对论在时空曲率甚大的区域必须修改。所谓修改,其实就是要创建一个量子引力论。不同学者有不同的创建途径,此处只简介其中之一,即所谓的“圈量子引力论”。根据这一理论,不但物质场(例如电磁场)需要量子化,而且奇点附近的时空几何也应该(而且可以)量子化。他们早已在运动学的层面上得到了一个数学上严格的量子几何理论(对几何实行量子化),在此理论中长度和面积等几何量都取分立的本征值,表明“时空连续性”这一经典图像在普朗克尺度下不再成立。做出这一突破之后,圈量子引力论近年来在宇宙论和黑洞方面都取得了重要进展。例如,通过对某些高度简化了的模型的研究,发现宇宙的大爆炸奇点(原初奇性)和黑洞内的奇点(终结奇性)都可以避免,代之而存在的是一个以普朗克尺度计的很小的时空区域(称为**普朗克小区**),经典奇点的各种无限大问题原则上就不复存在。更有趣的是,以圈量子宇宙论为例,经典意义下的“奇点”(现在的普朗克小区)还存在“另一侧”,它也是一个与“这一侧”类似的很大的经典宇宙(不过是在收缩),两个相邻宇宙(一个在收缩一个在膨胀)由普朗克小区相连接,正是小区中的量子几何充当了连接两大经典宇宙的桥梁。虽然这些结论只是对高度简化模型研究的结果,但它能加强信心,鼓舞士气:看来奇点问题远不如经典广义相对论认为的那样可怕!

大质量球对称恒星的晚期坍缩形成的黑洞是最简单的黑洞,洞外(视界以外)的时空几何由史瓦西线元描述,所以称为**史瓦西黑洞**。

§7.4　黑洞无毛定理

史瓦西真空解求得后不久，德国的**莱斯纳**(Reissner)和荷兰的**诺斯特朗**(Nordström)分别在1916年和1918年发现另一个静态球对称解，称为**莱斯纳－诺斯特朗解**(本书简称**莱－诺解**)，它描述静态球对称带电恒星外部的时空几何。后来发现，如果静态球对称带电恒星坍缩为黑洞，其外部时空几何就由莱－诺线元描述，因此被称为**莱－诺黑洞**。与史瓦西黑洞不同，莱－诺黑洞外部存在静电场①，其电场线是从洞心向外辐射的直线。

先讨论不带电(而且无自转)的恒星。球状恒星坍缩成的黑洞一定是史瓦西黑洞，其视界为球面，这是容易接受的结论。但是，如果恒星是扁平状的，它坍缩成的黑洞的视界也会是扁平状的吗？(对其他形状，例如长条状或立方体状，也可提出类似问题。)苏联的**诺维科夫**等三人大约在1964年开始研究这一问题，为了能用微扰法计算，他们从一个简单问题入手：表面隆起一座小山的近球状恒星坍缩成的黑洞视界面上也有小山吗？计算的结果是：黑洞视界仍是球面，没有小山②。他们坚信，这是因为黑洞形成的过程必定涉及引力波的辐射，所有偏离球状的因素都会被辐射带走。诺维科夫在国际会议的精彩报告("小山"问题只是其中一部分)受到西方同行的高度重视和好评。生于柏林的**伊斯雷尔**(Israel)在1967年对此又做了重大改进，他彻底突破了"小山"的条件，给出了强得多的结论：不带电、无自转的任意形状恒星的坍缩如果产生稳态黑洞，就一定是完全球形的黑洞，即史瓦西黑洞。无论各个恒星原来的性质存在怎样的千差万别(除形状外当然还允许其他各种差别)，坍缩成黑洞后的唯一差别只能是质量。

但是，如果恒星带电，结论就稍有不同。研究表明，恒星所带电荷与恒星的其他性质不同，它不会被辐射掉(引力波不会带走任何数量的电荷)。因此，无自转的带电恒星如果坍缩成稳态黑洞，就一定是莱－诺黑洞(伊斯雷尔1968年证明)。莱－诺黑洞与史瓦西黑洞不同，它由两个参数——质量 M 和电荷 Q——决定。

不能被辐射掉的第三个(最后一个)参数来自恒星的自转。1963年，新西兰

---

① 是指静态观者所在参考系测得的电场是静电场。

② 他们的计算实际上没有考虑坍缩过程(动态过程)本身，而只是研究了一个序列的静态情况，即只考察了静态黑洞的形态问题。本节关于黑洞无毛的结论其实只涉及静态黑洞(属于静态黑洞的唯一性问题)，涉及动态演化的黑洞无毛问题实际上至今还远未得到圆满解决，见选读7－1。

数学家**克尔**(Kerr)找到爱因斯坦方程的一个新解,它描述某种特殊的自转恒星的外部时空几何。史瓦西线元只涉及一个参数(质量 M),克尔线元则涉及两个参数(质量 M 和角动量 J)。鉴于所有静态球对称带电恒星的外部引力场都由莱-诺线元描述,你可能以为凡是自转恒星的外部引力场都由克尔线元描述,然而这是误解。除了质量和角动量之外,自转星体还可以存在"千姿百态"(例如表面的隆起和湍流),克尔线元只适用于一类特别的自转星体的外部引力场①。所以"克尔解描述自转星体外部引力场"的提法一般而言是不对的。但是,有趣的是,后来英国相对论学家**卡特**(Carter)发现,如果自转星体最终坍缩为旋转黑洞,其外部引力场就一定由克尔线元描述。所以,与其说"克尔解描述自转星体",不如说"克尔解描述旋转黑洞"。进一步说,如果自转星体带有电荷,坍缩成黑洞后,其外部引力场就要用一个比克尔线元还要复杂少许的线元描述,这个线元叫做**克尔-纽曼线元**,也是爱因斯坦方程的解,它由质量 M、角动量 J 和电荷 Q 等三个参数决定。这样的黑洞称为**克尔-纽曼**(Newman)**黑洞**。还有更复杂的稳态黑洞吗?答案是:没有了!晚期坍缩的大质量星体带着自己的全部信息穿越视界进入黑洞内部,它们对视界外部的影响只体现为三个参数 M、J 和 Q 的影响。($J=Q=0$ 是史瓦西黑洞,$Q=0$ 是克尔黑洞,$J=0$ 是莱-诺黑洞,M,J,Q 都非零时是克尔-纽曼黑洞,与克尔黑洞合称**旋转黑洞**。)这是一个来之不易的结论,是由许多相对论学家在近十年的不懈努力中逐步取得的重要成果。一个星体的性质本来多如牛毛(例如形状和气体运动的复杂性),数不胜数,但是一旦坍缩为黑洞,能够留下的就只有质量 M、角动量 J 和电荷 Q 等三个参数,变得异常简单!善于命名的惠勒在1971年把这一结果形象地说成是"黑洞没有毛发"(black holes have no hair)。后来又被广泛地称作黑洞的**无毛定理**(no hair theorem)。然而,如果把一个参数称作一根"毛"的话,考虑到我国的已故著名漫画家张乐平先生笔下的三毛形象(图7-6),对国人而言改称"三毛定理"或许更为合适。

由于坍缩过程所发射的引力波和电磁波会带走能量和角动量(但不会带走电荷),星体的 M 和 J 在坍缩过程中也会改变,不过很快就会达到最终的稳定状态,这时的 M 和 J 将与原来的不同,因而终态黑洞及其外部引力场由这终态 M、J 和 Q 唯一决定(除此以外的所有性质都被引力波带走或被埋葬在黑洞中)。这个黑洞与它的"前身"——原来的星体——的性质(例如物质结构、内部原子结构、由什么化学元素组成、涉及怎样的场)完全无关,从这一角度来看,黑洞是

---

① 外部线元与星体形状及诸多因素有关,即使离星体很远处的线元也要取决于由星体质量分布决定的质量和角动量多极矩。

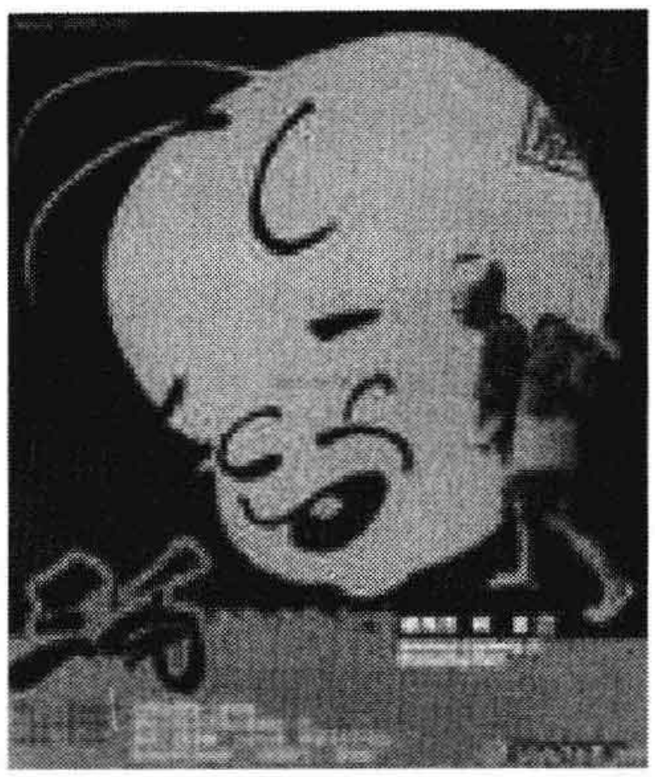


图 7-6　张乐平先生笔下的三毛形象

最简单不过的东西,简单到你难以想象的地步。美国相对论学家**索恩**(Thorne)曾做过意思如下的比喻:如果你仅从一位女士的体重和头发的颜色就能决定这位女士,你不觉得太简单了吗?

下面简介黑洞这三根毛对黑洞外部场的影响。质量 M 使外部存在引力场;电荷 Q 使外部存在静电场;角动量 J 则使外部引力场具有某种漩涡性,这种漩涡性的重要表现就是它会迫使引力场中的物体围绕黑洞转圈(有点像行星围绕太阳公转),而且越是接近视界就转得越快。如果黑洞没有角动量($J=0$),物体在洞外引力场中的自由下落是沿径向的,但是,在有角动量($J\neq 0$)的黑洞外部的物体将被迫做旋转式的"自由下落"运动(转着圈地掉入黑洞)。不妨称这种有漩涡性的引力场为**漩涡引力场**。在 4 维语言中,任何引力场都是某种时空线元的 3 维表现(体现为一个 3 维引力,虽然在 4 维语言中引力不是力而是时空的弯曲)。漩涡引力场无非就是克尔线元或克尔-纽曼线元相应的 3 维引力场。

正如地球的自转会使地球表面变成扁球状那样,自转也会使旋转黑洞的表面变成扁球状,不过这个"表面"并不是指黑洞的视界(视界面仍然保持球状),而是指视界之外的一个非常有特征意义的面,称为旋转黑洞的**静界**(static limit surface),它的形状以及它同视界的关系示于图 7-7 中。静界与视界之间的区域称为**能层**(含义稍后解释)。视界的定义仍然是有进无出的单向膜,而静界在原则上可进可出:就算你置身于静界之内(指能层之中),只要你所在的飞船功率够大,你总可以使自己逃离黑洞(径向坐标 r 增大),飞到静界之外。既然如此,静界又有什么特殊之处呢?它是什么与什么的分界面呢?"落入黑洞"和"逃离黑洞"所形容的只是径向运动,但对旋转黑洞来说还有个围绕黑洞公转(角向坐标 φ 随时间增大)的问题。位于视界以外的任何物体在漩涡引力场的

作用下都有边公转边下落的倾向。假定你驾驶着功率够大的飞船，你一定可以使飞船不绕黑洞公转吗？计算表明（详见梁灿彬、周彬中册 13.4.1），如果你在静界之外，一定可以；如果你在静界与视界之间（亦即在能层内），无论你的飞船功率如何强大，你都不得不被漩涡引力场拖着绕黑洞公转（你的角向坐标 φ 不得不随时间而变），也就是说，你只要在静界以内就不能静止。所以静界就是“允许静止”和“不允许静止”的分界面。

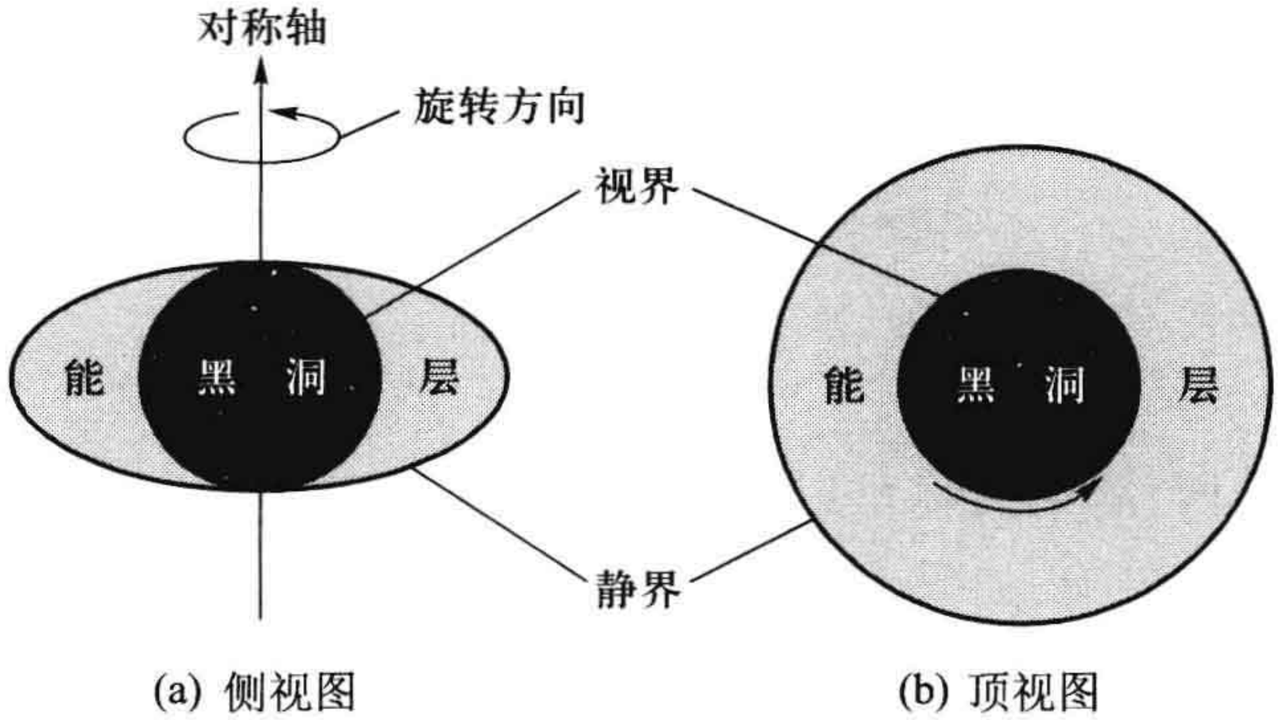


图 7－7　旋转黑洞的静界、视界和能层

再说能层。能层是个非常特别的时空区域，其中存在特别的类时曲线，以此为世界线的质点竟然具有负的能量，因此称这种曲线为**负能轨道**。1969 年，彭罗斯利用这一特点提出了从旋转黑洞提取能量的有趣想法，后来称为**彭罗斯过程**。设能量为 $E(>0)$ 的质点从“无限远”实验室向旋转黑洞自由下落，由于自由下落保证能量守恒，所以 E 为常量。质点备有定时装置，以使其在能层内某时空点 x 自动爆裂为两块（图 7－8），能量各为 E_1 和 E_2。由能量守恒可知 $E=E_1+E_2$。原则上总可这样安排爆裂事件，使碎块 1 以负能轨道为世界线（因而 $E_1<0$ 且 $E_2>E$）而碎块 2 沿外向测地线返回远方实验室。可以证明碎块 1 最终必然落入黑洞。整个过程的净结果为：① 远方实验室得到能量 $E_2-E=-E_1>0$；② 由于“吃进”碎块 1，黑洞能量由 M 减为 $M-|E_1|$。这就是“从黑洞提取能量”的实质，“能层”也由此得名。表面看来，这一过程好比“从一毛不拔的铁公鸡”身上拔毛，其实被提取的只是能层中（视界外）的能量，只不过手法巧妙而已。存储于能层内的能量其实是旋转黑洞的转动能，反复提取总有取完的时候，那时黑洞就从旋转黑洞变为史瓦西黑洞，能层不复存在。可以证明（详见梁灿彬、周彬中册 §13.5），反复利用这一过程可以提取的能量极限值（全部转动能）为黑洞总能量的 29%。在把质量（能量）转化为其他能量形式的各种机制中，这是一个异常巨大的转化百分数。烧氢变氦的核聚变是实验室中可实现的

转化率最高的过程，其转化百分数也不到1%。彭罗斯过程可以付诸实用吗？设想图7-8的远方实验室是一座城市，卡车满载垃圾向能层自由下落，到 x 点把垃圾抛出并使其沿负能轨道进入黑洞，能量变大了的卡车沿外向测地线返回城市，把获得的动能转给一个巨大的飞轮并使其推动发动机，然后卡车再次满载垃圾进入能层，……这似乎为一箭双雕地解决能源和垃圾处理问题提供了一种原则上的可能性，然而付诸实践却远非如此简单。纵使有朝一日能在足够近处发现一个可供利用的旋转黑洞，上述过程所要求的非常高的爆后速率（为使能量为负，碎块1的速率须大于 $c/2$）以及极端精准的定时技术，特别是对两个碎块的精确去向的控制都是远非目前人类技术所能达到的。有趣的是，目前人类不能做到的这件事情大自然早已做到了，虽然是以一种不太一样的方式（见选读7-2前关于类星体能源的一段）。

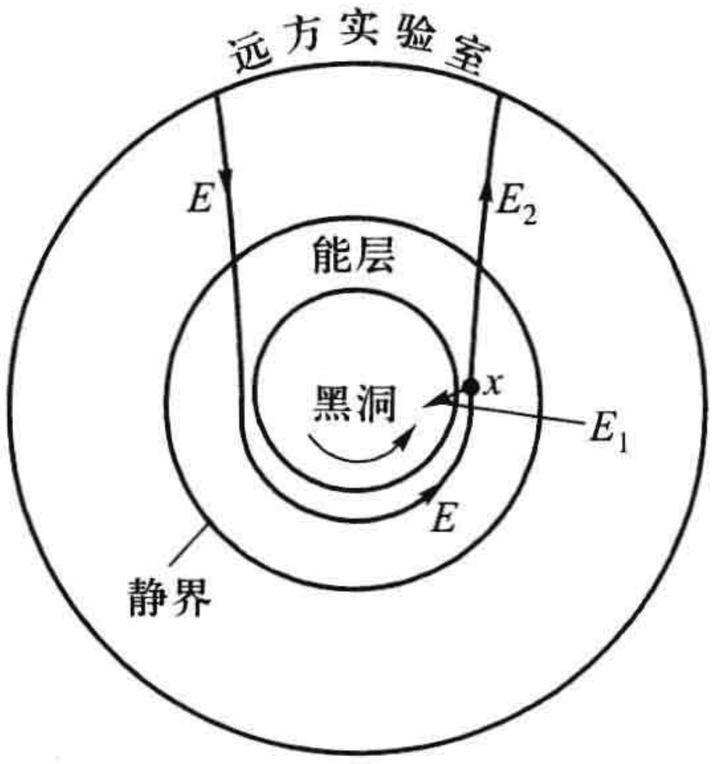


图7-8　从旋转黑洞提取能量示意图

[选读7-1]

本节的正文部分侧重从物理图像的角度介绍黑洞的无毛问题，不太注重数学的严格性。黑洞无毛定理的证明涉及非常复杂的数学问题（统称为黑洞解的唯一性问题），是许多学者多年来前赴后继的研究结果，虽然取得很大成功，仍然遗留某些非常棘手的数学问题有待探讨。本选读打算分三个层次粗略介绍黑洞解的唯一性问题。

第一个层次是稳态电真空黑洞的唯一性定理。先解释两个术语。“电真空”是指时空中的物质场只能是电磁场（“除电场之外就是真空”）。“稳态黑洞”是指黑洞所在时空是稳态时空，它比静态时空要求略低。静态时空的充要条件是：存在这样的坐标系 $\{t,x^1,x^2,x^3\}$，线元在该系的表达式满足① 各项的系

数与 t 无关;② 不含时空交叉项(即形如 $\mathrm{d}t\mathrm{d}x^i$ 的项)。稳态时空则只保留条件①,所以静态必定稳态,稳态未必静态。克尔时空和克尔-纽曼时空都是稳态(而非静态)时空。现在就可以介绍稳态电真空黑洞的唯一性问题。**伯克霍夫**(Birkhoff)早在1923年就证明了一个定理:真空爱因斯坦方程的球对称解必为史瓦西解(详见§8.8)。仿照他的方法,不难证明电真空爱因斯坦方程的球对称解必为莱-诺解。这一结论后来被伊斯雷尔强化,他在1967年证明了如下定理:真空爱因斯坦方程的任意(而不只是球对称)静态单一黑洞解必为史瓦西解(真空静态黑洞一定是史瓦西黑洞)。稍后(1968年),他又证明了上述定理的推广:任意电真空静态单一黑洞一定是莱-诺黑洞。后来经过许多相对论学家近十年的不懈努力,终于得到以下的唯一性定理:任意电真空稳态单一黑洞都是克尔-纽曼(Newman)黑洞。这就是所谓的**黑洞无毛定理**。但这个定理只适用于单个黑洞,即附近没有其他黑洞的情形。在多个黑洞并存时,由于相互影响,问题还要复杂得多。1987年,英国数学家**斑庭**(Bunting)和巴基斯坦数学家**马苏德**(Masood-ul-Alam)证明了如下定理:静态真空情况下不可能有多个黑洞出现。虽然数学证明比较复杂,但结论本身从直观上不难理解:两个处于静态的黑洞,既无转动又缺乏支撑,单从万有引力的角度考虑,它们也会落到一起成为一个黑洞。困难的是稳态(而非静态)的情况(即允许黑洞彼此绕转的情况)。一方面,牛顿引力的经验让我们觉得这种情况似乎可能存在;但另一方面,绕转可能会产生引力波辐射,导致系统能量损失,从而破坏稳定性。因此单凭直观考虑,哪种可能性都无法排除。事实上,这个问题至今仍是个未解难题,目前已有的一些数值计算结果比较支持后一种猜测。在电真空情况下,上述结论有了例外:如果两个黑洞带了足够多的同号电荷,静电斥力可以平衡掉引力,使得两个黑洞不会掉到一起。但实际计算表明要实现这种平衡所需的电荷量极大,一般认为天体不太可能带有如此多的电荷,因此这样的黑洞不太可能由天体的引力坍缩产生,但不排除在宇宙极早期由于大爆炸会产生这样的黑洞。

第二个层次是,放弃电真空条件,考察带任意物质场的稳态黑洞的唯一性问题。提出这一问题的一个重要动机是:“黑洞无毛”真是定理吗?换句话说,稳态黑洞除了质量 M、电荷 Q 和角动量 J 外果真不能有其他参数吗?自从无毛定理提出以来,对黑洞是否有毛的问题的探讨就一直没有停止过。既然把只有 M、Q、J 三个参数的黑洞称为无毛黑洞,黑洞“毛”就应定义为存在于黑洞之外而又不是电磁场的任何物质场。**蔡斯**(Chase)在1970年发现只带标量场的静态黑洞的标量场必须为零,也就是说标量场这根毛不会出现。但从20世纪80年代开始出现的若干文章指出,在稳态黑洞外部除电磁场外还可存在其他场,例如非阿贝尔规范场(电磁场是阿贝尔规范场)。由此引发了对“有毛黑洞”的研究热潮,

人们在问:哪些毛会有,哪些毛不能有?到目前为止这仍是国际相对论界的一个研究课题。

第三个层次的问题更要复杂得多,它涉及黑洞形成的动态过程。第一个层次证明的仅仅是稳态黑洞的唯一性定理。然而,从各种不同的初态出发,经由引力坍缩过程的演化结果一定是黑洞吗?就算是黑洞,一定是稳态黑洞吗?如果不是稳态黑洞,所形成的黑洞可能有多少种?前面提到的奥本海默-斯奈德解是第一个关于星体引力坍缩形成黑洞的具体物理模型,虽然高度简化,但仍能提供一个最基本的物理图像。之后,关于球对称条件下的引力坍缩过程有很多研究,但伯克霍夫定理保证这样形成的黑洞一定是莱-诺黑洞。然而恒星可以有各种形状,比如太阳就不是一个标准球体,所以非球对称引力坍缩过程的研究更有物理重要性。那么,非球对称的引力坍缩过程所形成的黑洞也可以有各种形状吗?在这方面,人们首先考虑对黑洞微扰的演化问题。美国的**普莱斯**(Price)第一个给出非常小的线性微扰演化结果,得出史瓦西黑洞的任意微扰都会以幂律形式消失的结果。当线性微扰比较大的时候,人们发现微扰会带着黑洞的特征频率震荡式地指数衰减,最终仍然得到克尔黑洞。如果考虑的不是线性微扰,而是完全爱因斯坦方程的非线性微扰,目前最好的结果是1993年**克莱纳曼**(Klainerman)和**克里斯托度鲁**(Christodoulou)关于闵氏时空的非线性微扰的稳定性结果。关于黑洞的稳定性,目前几乎还没有任何结果。关于黑洞形成问题,目前最好的结果是2008年克里斯托度鲁得到的、由脉冲引力波形成黑洞的结果(这个结果让他获得了2011年的邵逸夫数学科学奖)。最近,清华大学的年轻学者**于品**推广这一结果得到脉冲电磁波形成黑洞的结果。要一般地研究这个涉及黑洞形成过程的黑洞唯一性问题,需要考虑爱因斯坦方程与流体力学方程等多种物理方程组成的耦合系统,这样的系统在解析上处理起来非常困难,目前已知的结果主要来自数值计算,虽然有许许多多的数值研究结果显示最终演化得到的黑洞会很快变成克尔黑洞,但目前对这一问题还远没有一个明确肯定的结果。

[选读7-1完]

§7.5　黑洞的搜寻

广义相对论在1915年诞生后,经历了近半个世纪的“冬眠期”,主要原因是理论自身的深奥以及应用价值尚无体现。虽然奥本海默和斯奈德早在1939年就揭示了黑洞存在的可能性,但绝大部分天文学家都不为所动,无人问津。20世纪60年代是广义相对论和天体物理学的转折期,这一时期之所以不平凡,是因为天文观测有了许多重大发现,例如活动星系核、类星体(其发射功率竟比亿

万颗恒星的功率还大,见本节末)、宇宙微波背景辐射以及脉冲星等等。这里的许多现象很可能与广义相对论密切相关,广义相对论正好在这一领域找到自己的用武之地。这些发现迫使天文学家改变态度,他们逐渐认识到广义相对论对天体物理研究的重要性。于是,广义相对论预言的黑洞的天文观测(搜寻)也就被提上了日程。

观测黑洞?谈何容易!首先,它是黑的;其次,它非常小;第三,它离我们很远。作为重星坍缩的产物,最靠近地球的黑洞也有数十光年之遥,对我们的张角只有 10^{-8} 角秒的量级。用望远镜直接观测这么一个小黑点是毫无指望的。要想找到黑洞,必须设法观测那些既与黑洞有关又会发光(因而可被观测)的对象。首先想到的对象是双星系的可见伴星。前面讲过,天狼星其实是双星系,其中的不可见伴星(天狼 B)是白矮星。我们知道它的存在,是因为可见伴星(天狼 A)不断围绕它转(更准确地说是两者围绕共同的质心转)。如果一个双星系的不可见伴星是黑洞,同样可以从其可见伴星围绕它转来判断它的存在性。1964 年,苏联天体物理学家**捷尔多维奇**在研究生**古赛诺夫**的协助下,在浩繁的双星系编目表中选了 5 个有希望的黑洞候选者。1968 年,这份候选者名单又被美国相对论学家**索恩**(Thorne)扩充到 8 个双星系。如果其中某个双星系含有一个不可见伴星,而且质量远大于中子星质量上限($3M_{\odot}$),它就很可能是黑洞。虽然苏美天文学家都用这种方法搜寻过,但是都不成功,因为每个"疑似黑洞"的不可见性都可以用传统的、更易接受的非黑洞方法加以解释,实在没必要认为它是黑洞这个怪物。

上述搜寻方法只涉及黑洞的一个性质,即它在外部产生引力场并导致可见伴星围绕它转。但是可见伴星离它还不够近,还未感受到黑洞视界紧邻处的超强引力场。应该充分利用黑洞超强引力场造成的效应。苏联的捷尔多维奇和**诺维科夫**于 1966 年想到,如果黑洞外面存在气体(这常会有),就会被黑洞吸引到视界附近,超强引力场就会将它加速而达到高温并发光。但是,孤立黑洞周围的气体密度太低,流向黑洞的气体太少,发光效应很不明显。他们进一步设想,如果黑洞与一个紧邻的、质量很大的可见恒星结伴组成双星系,恒星表面的大量气体就会源源不断地流进黑洞。略去其中的物理细节,其结果是:气层之间的摩擦将把进入黑洞前的气体加热到 1 千万度的高温,以至可以发射 X 射线[电磁波谱中频率很高(因而光子能量很大)的波段]。这种发光效应(指发出 X 射线)应该是含有黑洞的双星系的重要特征。天文学家把能发射 X 射线的星体称为**X 射线星**。不过,中子星充当不可见伴星时的引力场也很强,也能发射 X 射线。为了最后确定所探测到的 X 射线星的确是黑洞,就要以其质量为判据。如果质量比中子星的质量上限 $3M_{\odot}$ 还大得多,它就应该是黑洞了。请注意这种搜寻方

法仍然沿用上面所讲的思想——观测那些既与黑洞有关又会发光的对象，只不过现在的光是 X 光，发光的对象则是高速进入黑洞的气体。

于是，搜寻黑洞的重任就落在了 X 射线天文学家的肩上。X 射线天文学始于 20 世纪 40 年代。由于 X 射线不能穿透地球大气层，当时只能借火箭把探测仪带到大气层以外。在有了人造地球卫星之后，探测仪自然就可以搭乘卫星做长时间的飞行。1970 年 12 月，第一颗 X 射线探测卫星“自由号”（音译名为“乌呼鲁”）在肯尼亚发射升空，1973 年 3 月停止工作。飞行期间做出了里程碑式的贡献，一共探测到 339 个 X 射线源（除 X 射线双星系之外还有超新星遗迹及星系团等等），其中有几个可称为黑洞候选星，最可信的一个位于天鹅座，称为**天鹅 X－1**。这是一个双星系，其中一个伴星是光学明亮而 X 射线暗淡的普通恒星，质量约为 $20M_{\odot}$；另一伴星则相反，是一个光学上暗淡无光但 X 射线明亮的星体，其质量约为 $8.7M_{\odot}$。根据这一数据，加上后来进一步研究发现的种种迹象，人们普遍相信它的确是一个黑洞。这是人类发现的第一个黑洞。即使如此，仍然有极少数专家对这个不可见伴星给出过非黑洞的种种解释，不过都很牵强，信者寥寥。时至今日，主流专家的意见是一致的：不仅天鹅 X－1 是黑洞，而且银河系中存在着大量黑洞，少说也有数百万，也许达数十亿之多。

随着时间的推移，黑洞搜寻技术也在不断进步。美国宇航局发射的、与黑洞观测关系密切的就有三个空间望远镜：① 1990 年发射的**哈勃空间望远镜**，主要观测范围是紫外、近红外和可见光波段，它弥补了地面观测的不足，当时被认为是“全能冠军”，是天文史上最重要的仪器；② 1999 年发射的**钱德拉 X 射线天文卫星**，是 X 射线天文学上具有里程碑意义的空间望远镜；③ 2003 年发射的**斯皮策空间望远镜**，主要观测红外波段。银河系中心存在大量尘埃和气体，可见光难以透过，所以红外观测对了解银河系中心以及太阳系外恒星有重要帮助。除美国之外，欧洲和日本的天文卫星也不逊色。这些先进的观测手段使得黑洞搜寻与其他天文观测互相结合，近二十多年来探测到大量的黑洞候选星。此处应该说明，前面关于黑洞形成的理论只涉及由一颗重星坍缩而成的黑洞，称为恒星级黑洞，其质量至多只有太阳质量的数十倍。从天鹅 X－1 算起，至今找到的黑洞候选星已达数十个。但是，越来越多的迹象和刻意观测的结果表明，宇宙中还存在着另一种类型的黑洞，它们的质量比恒星级黑洞要大得多，最小的也有太阳质量的数十万倍。这种黑洞的形成过程与恒星级黑洞非常不同，简介如下。同一星系中的两颗恒星如果在运动中互相靠近，双方都会在引力作用下偏离原定轨道，其中之一通常会偏向星系中心（另一个则偏离中心）。这种过程的积累效应是：星系中的若干恒星会奔向星系的中心。此外，引力也会驱使恒星之间的气体向星系中心运动。这些奔向中心的恒星和气体逐渐积聚成一个超大质量的集

团,集团内部的压力将抗衡不住引力,因而最终会坍缩成一个质量巨大的黑洞。这种黑洞称为**超大质量黑洞**(supermassive black hole)①,其质量约在太阳质量的数十万倍至数十亿倍之间,可以说是宇宙中的"巨无霸"。有趣的是这种巨无霸竟然充斥于全宇宙之中,因为有大量迹象表明,几乎每一个星系的中心都有一个超大质量黑洞。

由于引力超强,超大质量黑洞无时无刻不在吞噬着周围的气体和恒星,因而质量不断增大。由于几乎所有黑洞都是旋转黑洞,被吞噬的物质围绕着黑洞一边公转一边下落,在公转中结合而成一个盘状物,称为**吸积盘**(被黑洞吸引而积聚成的圆盘)。黑洞附近的吸积盘总在黑洞的赤道面上。图7-9是一个国际小组在2008年首次拍摄到的吸积盘照片。吸积盘中的物质在高速下落的过程中由于相互强烈摩擦而获得高温,因而向外发出遍及各个波段的电磁辐射,人类正是因为观测到这些辐射才发现超大质量黑洞的(见图7-10)。日常遇到的摩擦生热效果轻微,而黑洞的超强引力场能把吸积盘加热到这样的程度,使它的亮度竟然达到最亮星系的100倍。

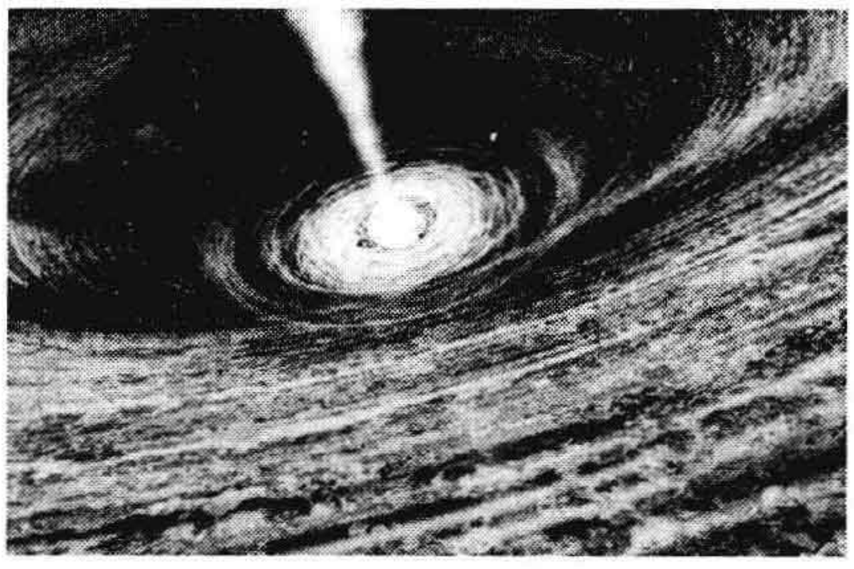


图7-9　吸积盘照片

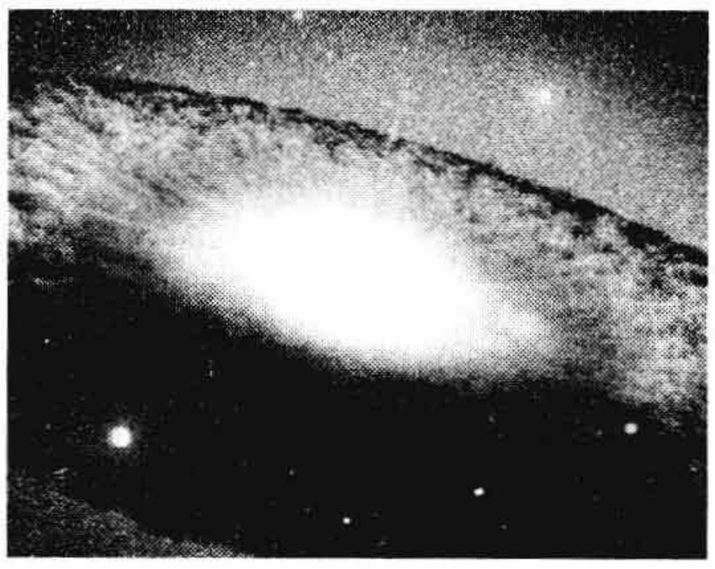


图7-10　超大质量黑洞

以下是近年来对超大质量黑洞的几个重要观测事例。

(1) 从1997年起,利用位于北半球的、由十个射电望远镜组成的阵列,中科院上海天文台沈志强研究员领导的一个国际小组进行了大量观测,在5年中把射电波的"视线"步步逼近银河系中心的超大质量黑洞,最终获得了世界上第一张3.5毫米波长的高分辨率图像,成为当时确认该黑洞的最令人信服的证据。该黑洞位于人马座方向,距太阳系2.6万光年,按目前的估计,质量达到太阳质量的300万倍以上。

(2) 2004年5月,Padovani等天文学家发表文章,表明他们发现了银河系外的30个超大质量黑洞。

---

① 超大质量黑洞还有其他的形成途径,略。

（3）斯皮策空间红外望远镜在2005年发现了一群超大质量黑洞，位于遥远星系的中心，每年要吞噬掉相当于1千颗恒星的质量。这群黑洞由于被浓厚的尘埃和气环所围绕而未被X射线探测仪发现，却逃不过斯皮策望远镜那双“明亮的红外眼睛”。

（4）天文学家长期来相信M87星系的中心存在一个超大质量黑洞（离地球约5 000万光年），但对其质量一直估计不足，认为大约是太阳质量的30亿倍。2009年，美国天文学家吉布哈特（Gebhardt）用精确的观测数据和计算断言该质量应为太阳质量的64亿倍。后来，更精确的数据和计算表明它是太阳质量的66亿倍。吉布哈特在一次发布会上说：“在现阶段，为了确认黑洞的存在，我们采用的是排除法。也就是说，对于某些天文现象，除了黑洞，其他解释真的站不住脚。”所以，慎重的天文学家仍然爱把黑洞称为“黑洞候选星”。

（5）天文学家早就知道位于长蛇座方向的星系NGC 3393中心存在超大质量黑洞。利用钱德拉X射线望远镜长期观测后，2011年确认这其实是一对（两个）超大质量黑洞，离地球约1.6亿光年，两者距离仅490光年，每个黑洞的质量都超过$M_\odot$的100万倍。

前面提到，20世纪60年代发现的类星体的怪异表现曾经令天文学家瞠目结舌。后来超大质量黑洞的发现使之得到合理解释。1960年，美国天文学家桑德奇在用5米口径的光学望远镜观测编号为3C48的射电源时，发现它不是星系而是一个蓝色亮点，很像一颗恒星，但光谱极其特别，含有一些无法弄清意义的发射线。1963年，荷兰天文学家施密特发现射电源3C273的光谱也很特别，仔细研究后认出这些谱线无非是人们非常熟悉的氢的发射线，只不过向红色方向有很大的移动（谱线的多普勒红移）。不久，又有人认出3C48的光谱线也是熟悉谱线（发自镁、氧和氖）的红移结果，而且红移量更大。由于这些射电源很像恒星，所以起名为**类星体**（quasar）。类星体最初是（作为射电源）在射电波段上被发现的，其实它在光学波段、紫外波段和X射线波段都有很强的辐射，射电波段只占辐射中的很少一部分。类星体至少有以下三个异乎寻常之处：

（1）类星体与地球的距离极其遥远。我们的宇宙现在正在膨胀之中（可参阅选读3－2），所以任意两个星系都在互相退行（渐行渐远），就像膨胀中的气球表面上的蚂蚁。我们收到的来自其他星系的光波因此就会红移（多普勒频移，见小节4.2.4）。根据哈勃定律，退行速率正比于星系间的距离。类星体谱线的巨大红移表明它们离我们甚远。例如，3C273和3C48离我们竟然分别达到20亿光年和45亿光年！

（2）类星体发光的功率极其巨大。仍以3C273和3C48为例，它们是如此遥远，为了解释它们在望远镜中的亮度（视亮度），必须认为它们的实际亮度（发

射的功率)大得离奇:比最亮的星系还要高出100倍。如何解释它的能量来源?

(3) 类星体的尺度小得离奇。例如,观测和讨论表明(见选读7-2),3C273的发光区域的尺度只有1“光月”(光在一个月内走的距离),作为对比,请注意银河系(简化看作一个圆盘)的直径就有10万光年!从尺度为1光月的类星体发出的光能竟然大于100个银河系所发的光能,这是当时的天文学家根本无法接受但又不得不承认的结论①。

类星体实在太不可思议了!在它那小小的地盘上制造出这么大的功率的“发动机”究竟是何方神圣?当时的天文学家虽然努力做出种种解释,但是除了黑洞解释(以引力作为能源)外都站不住脚。过去的天文学家认为广义相对论太过抽象而无用,正是为了解释类星体使他们第一次觉得广义相对论在天体物理学中大有用武之地。1963年的首届德克萨斯相对论天体物理学会议促成了天文学家与相对论学家的史无前例的互相学习和交流。此后的一系列深入的理论研究和天文观测使得学者们普遍接受如下结论:类星体的超大功率发动机就是藏在星系中心的超大质量黑洞,具体地说,① 类星体与恒星有完全不同的发光机制:恒星的光(辐射)来自电子、原子和分子的无规热运动;类星体的光则是高速电子在磁场中的螺旋运动所发出的同步辐射[类似于脉冲星的同步辐射(见小节7.2.5),但强烈得多]。② 类星体发光与恒星发光的能源根本不同:恒星依靠核聚变提供核能;类星体的能源是超大质量黑洞的引力能,更具体说就是存储在旋转黑洞能层内的转动能。目前的人类技术尚不足以实现彭罗斯的能量提取过程,但大自然早就这样成功地运作了!

[选读7-2]

怎么知道3C273的发光区域最多只有1光月的尺度?天文学家很早就发现3C273的视亮度在一个月中就有明显的变化,假若它的尺度大于1光月,用时空图(图7-11)不难推出矛盾。竖直线G代表地球观者,A和B代表类星体发光区域的两个边缘,水平直线段ab的线长就代表发光区的尺度。在G上取时空点(事件)g作为观察时刻,过g作类光测地线交A,B线于a和b'点。bb'段与ab段有相等的线长。类星体的实际亮度随时间而变,假若bb'段的线长竟然大于类星体亮度变化的周期(1光月),则bb'(因而ab')段内将有明暗交替现象(不同点亮度不同),由于ab'段内各点发出的光都同时进入观者G的眼睛(事件g),眼睛感觉到的是某种平均亮度。G线上的其他点也如此,可见观者G测得的视

① 多数类星体的发光区域尺度为1光年,3C273比较特别,只有1光月。

亮度基本不随时间而变，与上面提到的观测事实矛盾。这一矛盾来自“bb'段的线长大于亮度变化的周期”的假定，所以 bb'段（因而 ab 段）的线长最多只有 1 光月。

［选读 7－2 完］

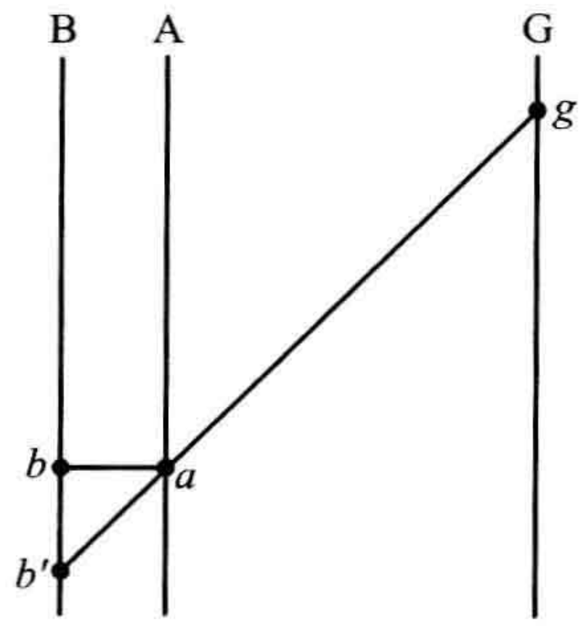


图 7－11　3C273 发光区域最多只有 1 光月的说明

第8章　史瓦西时空

§8.1　史瓦西真空解

史瓦西真空解是爱因斯坦方程的第一个精确解(§6.5),描述一个静态的、球对称的物质分布(通常是天体)在其外部(真空区域)造成的时空弯曲,这一弯曲线元可用史瓦西坐标系$\{t,r,\theta,\varphi\}$表为

$$\mathrm{d}s^2 = -\left(1-\frac{2M}{r}\right)\mathrm{d}t^2 + \left(1-\frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2), \tag{8-1-1}$$

其中M代表天体的质量。上式是史瓦西线元在几何单位制的形式,相应的国际制形式要复杂一些:

$$\mathrm{d}s^2 = -\left(1-\frac{2GM}{c^2r}\right)(c\mathrm{d}t)^2 + \left(1-\frac{2GM}{c^2r}\right)^{-1}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2), \tag{8-1-1'}$$

其中G为引力常数,c为真空光速。式(8-1-1)比式(8-1-1′)简单,适宜于纯理论讨论。涉及数值计算时最好用式(8-1-1′)。

线元(8-1-1)右边各项的系数不含时间坐标t,这是静态性的反映。另一方面,右边的后两项取$r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2)$的形式,表明线元有球对称性。所以说史瓦西线元是一种静态球对称线元。

当径向坐标r很大时,$\mathrm{d}t^2$和$\mathrm{d}r^2$的系数$1-2M/r$和$(1-2M/r)^{-1}$都很接近于1,式(8-1-1)近似成为

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2)。 \tag{8-1-2}$$

这无非是平直线元在球坐标系$\{t,r,\theta,\varphi\}$的表达式[即式(2-1-16)],可见史瓦西线元在r越大时越接近于平直,所以是一种**渐近平直线元**,也说史瓦西线元具有渐近平直性。

两个史瓦西线元的差别只体现在它们的质量参数M上。作为天体的质量,M可取任何正数。极端情况$M=0$代表天体不存在,故时空应该平直。事实的确如此:把$M=0$代入式(8-1-1)果然得到平直线元式(8-1-2)。

如果时空的线元是史瓦西线元,它就称为**史瓦西时空**。平直时空可看作史

瓦西时空在 $M=0$ 的特例。前面关于闵氏时空的讨论(见 §3.7 例 2)已经讲过,时空是绝对的,但"每一时刻的全空间(同时面)"则是相对的。闵氏时空存在着一类特殊的坐标系,即惯性坐标系,任选一个惯性系 $\{t,x,y,z\}$ 都可对闵氏时空做"3 +1 分解",每一张等 t 面 $\sum_{\hat{t}}$(其中 $\hat{t}$ 为常数)就代表 $\hat{t}$ 时刻的全空间。躺在 $\sum_{\hat{t}}$ 上的任一曲线的任一元段都有 $\mathrm{d}t=0$,代入闵氏线元

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$$

便诱导出 3 维欧氏线元 $\mathrm{d}\hat{s}^2 = \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$,可见闵氏时空被任一惯性系分解所得的空间都是 3 维欧氏空间(有 3 维平直几何)。下面仿此讨论史瓦西时空的"空间几何"。

仿照闵氏时空用惯性坐标系 $\{t,x,y,z\}$ 做"3 +1 分解"的做法,我们用史瓦西坐标系 $\{t,r,\theta,\varphi\}$ 对史瓦西时空做"3 +1 分解"。以 $\sum_{\hat{t}}$ 代表等 t 面(面上各点有 $t=\hat{t}=$ 常数),则躺在 $\sum_{\hat{t}}$ 上的任一曲线的任一元段都有 $\mathrm{d}t=0$,代入式(8 -1 -1)得

$$\mathrm{d}\hat{s}^2 = \left(1-\frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\varphi^2)。\qquad (8-1-3)$$

上式称为由史瓦西线元在 $\sum_{\hat{t}}$ 上的**诱导线元**,右边第一项系数 $(1-2M/r)^{-1}$ 的存在使得这个 3 维线元并不平直(除非 $M=0$)。(由黎曼张量的定义可以求得一个非零的黎曼张量,不过本书并未给出这一定义。)可见史瓦西时空的任一等 t 面都是一个弯曲的 3 维空间,物理地说,就是天体质量 M 的存在使天体外部的空间变得弯曲,成为非欧空间。图 8 -1 是时空图,圆柱面代表天体表面的世界面,水平平面代表等 t 面 $\sum_{\hat{t}}$,它与圆柱面的交集 S 代表天体表面在时刻 $\hat{t}$ 的表现。天体表面本是个 2 维球面,只因画时空图时必须压缩掉一个空间维,所以 2 维球面才被画成 1 维圆周 S。

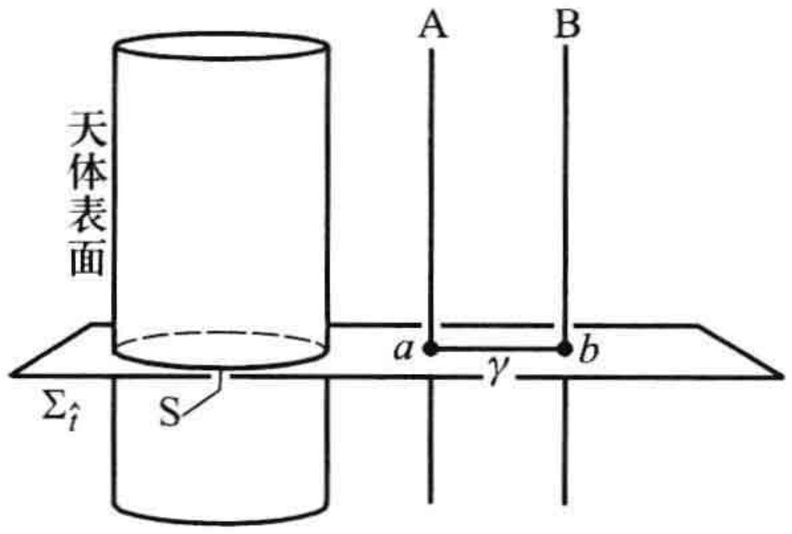


图 8 -1　星体外部时空图(压缩掉一维)。$\sum_{\hat{t}}$ 代表时刻 $\hat{t}$ 的全空间。

作为静态时空,史瓦西时空存在着一类特殊的观者,称为**静态观者**(static observer),其世界线上的空间坐标 r,θ,φ 都是常数(只有 t 沿线改变)。例如,把地球近似看作静态物质分布(忽略其自转),则静止于地面上高楼每一层的每一住户都是一个静态观者。图 8 -1 的两条竖直线 A 和 B 就是两个静态观者的世界线,两线与 $\sum_{\hat{t}}$ 的交点 a,b 就是这两个观者在时刻 $\hat{t}$ 的表现。设 $a\equiv(r_{\mathrm{A}},\theta_{\mathrm{A}},\varphi_{\mathrm{A}})$,$b\equiv(r_{\mathrm{B}},\theta_{\mathrm{B}},\varphi_{\mathrm{B}})$,而且 $\theta_{\mathrm{A}}=\theta_{\mathrm{B}}$,$\varphi_{\mathrm{A}}=\varphi_{\mathrm{B}}$。我们来关心 a,b 之间的距离。不论 $\sum_{\hat{t}}$ 是平直还是弯曲,两点的距离都定义为它们之间的(类空)测地线段的线长。

可以证明(直观上也不难相信),图 8 -1 中 a,b 之间的径向线 γ 正是测地线,所以 γ 的线长 l_{ab} 就是 a,b 之间的距离。由线长公式(2 -2 -3)得

$$a,b \text{ 的距离} = l_{ab} = \int_a^b \sqrt{\mathrm{d}\hat{s}^2 \mid_\gamma}。 \tag{8-1-4}$$

把式(8 -1 -3)用于 γ 的任一元段,注意到 γ 线上有 θ = 常数及 φ = 常数(因而 $\mathrm{d}\theta = 0, \mathrm{d}\varphi = 0$),便得

$$\mathrm{d}\hat{s}^2 \mid_\gamma = \left(1 - \frac{2M}{r}\right)^{-1} \mathrm{d}r^2, \tag{8-1-5}$$

代入式(8 -1 -4)给出

$$a,b \text{ 的距离} = l_{ab} = \int_a^b \left(1 - \frac{2M}{r}\right)^{-1/2} \mathrm{d}r > r_{\mathrm{B}} - r_{\mathrm{A}}。 \quad (\text{除非 } M = 0) \tag{8-1-6}$$

可见 a,b 的距离大于 a,b 的径向坐标差(亦称**坐标距离**)$r_{\mathrm{B}} - r_{\mathrm{A}}$,除非 $M = 0$。(当 $M = 0$ 时 $l_{ab} = r_{\mathrm{B}} - r_{\mathrm{A}}$,这正是人们熟知的欧氏情况。)为了强调与坐标距离的区别,也常把距离 l_{ab} 称为**固有距离**(proper distance)。固有距离不等于坐标距离是空间弯曲性的重要表现。物理上更应重视的是固有距离而不是坐标距离。

注记 1　在欧氏空间中,球面的半径 r 有两个性质:① 等于球面与球心的距离;② 与球面积 A 的关系为 $A = 4\pi r^2$,即 $r = \sqrt{A/4\pi}$。但在弯曲空间中这两个性质互不等价:若以 r 代表球面与球心的距离,则 $r = \sqrt{A/4\pi}$ 不再成立。一种方便的做法是用球面积定义半径,即把半径定义为 $\sqrt{A/4\pi}$,代价是半径不再等于球面与球心的距离。这就是"径向坐标距离 $r_{\mathrm{B}} - r_{\mathrm{A}}$ 不等于固有距离"的原因。

与此类似,在史瓦西时空中还应分清坐标时间 Δt 和固有时间 $\Delta\tau$。设 $\sum_{\hat{t}'}$ 是另一张等 t 面,我们来关心静态观者 A,B 的世界线上介于 $\sum_{\hat{t}}$ 和 $\sum_{\hat{t}'}$ 之间的两条线段,分别记作 μ_{A} 和 μ_{B},见图 8 -2。以 Δt_{A} 和 Δt_{B} 分别代表两段所经历的坐标时间,因 $\sum_{\hat{t}}$ 和 $\sum_{\hat{t}'}$ 都是等 t 面,当然有

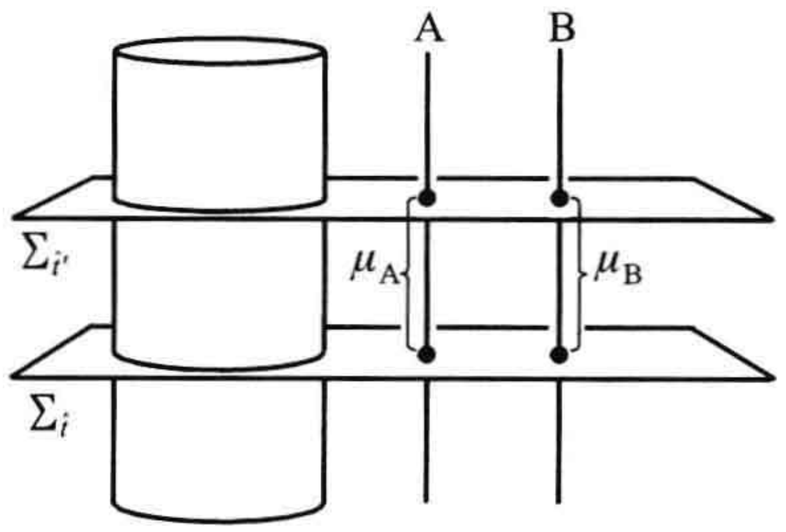


图 8 -2　静态观者世界线介于两个等 t 面之间的两段有不同固有时间

$$\Delta t_{\mathrm{A}} = \hat{t}' - \hat{t} = \Delta t_{\mathrm{B}}。 \tag{8-1-7}$$

问题在于这两段所经历的固有时间 $\Delta\tau_{\mathrm{A}}$ 和 $\Delta\tau_{\mathrm{B}}$ 是否也相同。注意到 r,θ,φ 在每条线上都是常数,把式(8 -1 -1)用于每条线的任一元段得

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2, \tag{8-1-8}$$

沿线段 μ_A 和 μ_B 积分给出[第一步用到式(3-2-1)]

$$\Delta\tau_A = \int_{沿\mu_A}\sqrt{-ds^2} = \left(1-\frac{2M}{r_A}\right)^{1/2}\Delta t_A, \tag{8-1-9a}$$

$$\Delta\tau_B = \int_{沿\mu_B}\sqrt{-ds^2} = \left(1-\frac{2M}{r_B}\right)^{1/2}\Delta t_B。 \tag{8-1-9b}$$

注意到 $\Delta t_A = \Delta t_B$ 及 $r_A \neq r_B$,便知

$$\Delta\tau_A = \left(1-\frac{2M}{r_A}\right)^{1/2}\left(1-\frac{2M}{r_B}\right)^{-1/2}\Delta\tau_B \neq \Delta\tau_B。 \tag{8-1-10}$$

可见线段 μ_A 和 μ_B 经历的固有时间并不相同。在闵氏时空中,“等 t 面”就是“同时面”,但现在看到这两个词汇对史瓦西时空有不同含义:等 t 面上各点的坐标时 t 相同而固有时 τ 一般不同。如果仍称之为“同时面”,只能理解为坐标时间相同的面。

既然现在提到固有时和坐标时,我们想借此机会谈一谈相对论中的时间概念。最有物理意义的时间就是观者的固有时 τ,它是观者携带的标准钟的读数,而且也是他的生物钟的读数,是他感觉到的实实在在的时间。除了固有时 τ 之外,τ 的任一常增函数都可以在一定程度上被解释为该观者的时间。例如,设 p 是星体外部史瓦西时空的一点,G 是一个观者,其世界线经过 p 点。一方面,作为 G 的世界线的一点,p 有一个固有时 τ_p;另一方面,作为时空中的一点,p 点又有一个坐标时 t_p(史瓦西坐标系 $\{t,r,\theta,\varphi\}$ 的第 0 坐标 t 在 p 点的值),于是对 G 线而言就有一元函数 $t(\tau)$。图 8-3 示明这是个常增函数,所以坐标时 t 在一定程度上也代表 G 的时间,称为**坐标时间**(coordinate time)。为说话方便,不妨想象观者 G 除标准钟外还携带另一个钟,其走时率很特别——它在 G 线上的读数恰好等于该点的坐标时 t。这样的钟称为**坐标钟**(coordinate clock)。

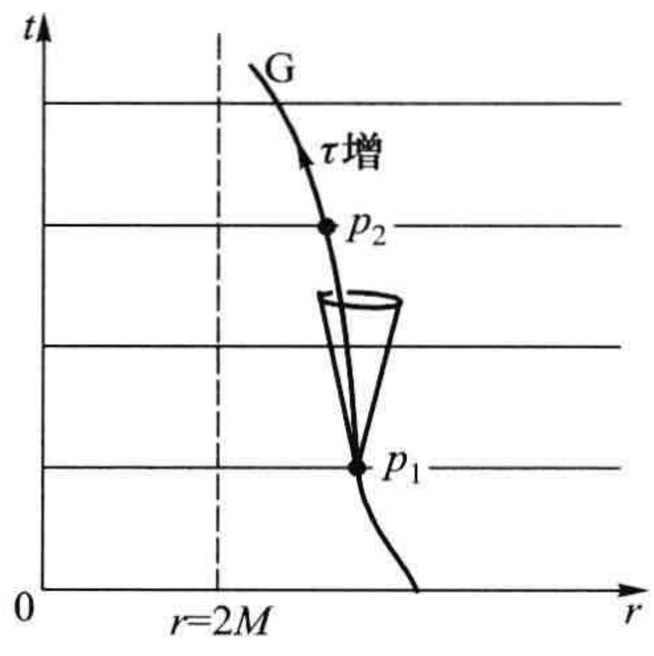


图 8-3　观者世界线 G 上任取 p_1,p_2 点。因 p_1 点的未来光锥朝上张开,故 p_2 在 p_1 的未来,即 $\tau_2 > \tau_1$。另一方面,从等 t 线(水平线)可知坐标时 $t_2 > t_1$。可见 $t(\tau)$ 是常增函数。

下面讨论 4 个史瓦西坐标的取值范围。作为角度坐标,θ,φ 的范围是

$$0 \leqslant \theta < \pi,\ 0 \leqslant \varphi < 2\pi。$$

时间坐标 t 的取值范围可达最大,即 $-\infty < t < \infty$,因为静态性意味着“永恒不

变”。[实际星体其实并不永恒而是有演化(见§7.2),但在理论模型中默认静态即永恒不变。] 径向坐标 r 的范围虽无上限,却有下限,这是因为当 r 小到等于 $2M$ 时 $1-2M/r=0$,导致式(8-1-1)的第二项发散[以0为分母的分数 $(1-2M/r)^{-1}$ 无意义]。所以 r 值要以 $2M$ 为下限,即 $r>2M$。数值 $2M$ 称为该星体的**史瓦西半径**(Schwarzschild radius),记作 r_S,即 $r_S\equiv 2M$。但在国际单位制中则应改为 $r_S\equiv 2GM/c^2$(理由可参阅附录A例3),其中 G 和 c 的国际单位制数值为 $G\approx 6.7\times 10^{-11}$, $c\approx 3\times 10^8$。太阳质量的国际单位制数值为 $M_\odot\approx 2\times 10^{30}$,故 $GM_\odot/c^2\approx 1.5$ km,因而太阳的史瓦西半径为

$$r_{S\odot}\equiv 2GM_\odot/c^2\approx 2\times 1.5\ \text{km}=3\ \text{km}。$$

类似地可求得地球的 $r_S\approx 8.8\times 10^{-3}$ m,质子的 $r_S\approx 2.4\times 10^{-54}$ m。

径向坐标 r 存在下限还有一个更为实际的原因:式(8-1-1)只是真空爱因斯坦方程的解(称为史瓦西真空解或史瓦西外解),只能描述天体外部(真空处)的时空弯曲,因此只当 $r>R$(天体半径)才有意义。常见的球对称天体半径 R 都远大于其史瓦西半径 r_S,故常见的比值 R/r_S 都很大于1。例如,这一比值的量级对太阳为 10^5,对地球为 10^9。这就是说,对常见的球对称物质分布而言,其史瓦西半径 $r_S\equiv 2M$ 深深地埋藏在物质的内部,而史瓦西真空解对物质内部根本不适用(不是解),所以完全不必理会。由此可知坐标 r 的取值范围是 $R<r<\infty$。

R/r_S 比1大得不多的例子是中子星。由于非常致密,中子星的半径比相同质量恒星的半径要小得多,所以 R/r_S 也小得多。典型中子星的半径 $R=10$ km,质量 $r=2M_\odot$,故其

$$R/r_S=\frac{10\ \text{km}}{4GM_\odot/c^2}=\frac{10\ \text{km}}{4\times 1.5\ \text{km}}\approx 1.7。$$

比1大得不多(这是可以预期的)。

§8.2　广义相对论的实验验证

爱因斯坦创立广义相对论的原始动机是纯理论性的。然而,任何物理理论问世后都要面对实验验证的问题。史瓦西真空解适用于太阳周围的引力场,它的出现使得太阳系中某些与牛顿引力论不协调的观测数据有可能用广义相对论加以解释。爱因斯坦很早就从广义相对论出发做了三个有可能与实验对比的预言(后人称之为三大经典验证),它们分别是光波的引力红移、水星近日点的进动以及星光在太阳引力场中的偏转。近日点进动的计算结果与早已存在的观测数据吻合,星光偏转的预言很快也取得观测的支持。然而,由于缺乏精度足够的实验技术来测量极端微弱的广义相对论效应(包括引力红移),广义相对论的实

验研究从20世纪10年代末期开始的45年中进展缓慢,甚至几乎止步不前。大约从20世纪60年代开始,由于理论和科技的进步以及天文观测的新发现,广义相对论的实验验证才进入全盛时期,既有对星光偏转和引力红移的精度越来越高的验证,又有一系列全新的实验。可以说,广义相对论很好地通过了迄今的所有实验检验,虽然精度更高和难度更大的许多实验还有待进行。下面的三节将分别介绍爱因斯坦提出的三个经典实验验证。关于广义相对论的实验验证的过去、现在和未来,可参阅Will(1993)和Will(2005),后者是Will在一本纪念相对论百年诞辰的文集①中的一篇文章,此处引该文末尾的一段译文作为本节的结束语:

"爱因斯坦1905年的相对论及其1915年的后续内容(指广义相对论)的巨大成功改变了科学的进程。这两者都是想象力和理论方面的胜利,实验只扮演次要的角色。在过去的40多年中,我们又见证了爱因斯坦的第二个胜利,那就是他的理论所获得的、系统的、高精度的实验验证。相对论已经极其漂亮地通过了每一个实验的检验,大获全胜。然而事情并未就此结束。在黑洞和中子星附近的强引力场中的实验仍然有待进行。人们对广义相对论的这一层面的勘探几乎从未进行过,伽马射线、X射线和引力波天文学在这一探索中将起到决定性的作用。"

§8.3　引力钟慢和引力红移

8.3.1　引力钟慢效应

我们已经熟悉闵氏时空(无引力场)中的钟慢效应,而且知道比钟结果取决于比钟方式。本节要讨论的是史瓦西时空(有静态引力场)中的钟慢效应。

设G_1,G_2是两个静态观者,他们有相同的θ坐标和φ坐标,径向坐标分别为r_1和r_2。再设$r_2<r_1$,则G_2比G_1更靠近恒星,所以位于引力场强较大处(图8-4)。现在要对G_1和G_2的标准钟(仍记作G_1和G_2)进行比对。首先约定如下的比钟方式。以p_1和p_1'代表G_1世界线上的两点,固有时分别为τ_1和τ_1',时间间隔$\Delta\tau_1\equiv\tau_1'-\tau_1>0$。设$G_1$在$p_1$时刻用左眼看钟$G_2$,所得读数为$\tau_2$[这表明图8-4的$p_2$点(作为$G_2$线的一点)的固有时为$\tau_2$],同时用右眼看自己的

① 文集由Ashtekar主编,书名为100 Years Of Relativity:Space-time Structrue:Einstein And Beyond,由Biwas Hopr Press出版社出版(2005),其中第8章是Will的文章,题为"Was Einstein Right? Testing Relativity At The Centenary".

钟，读数自然是 τ_1。一段时间 $\Delta\tau_1$ 后，G_1 再次用左右眼分别看钟 G_2 和 G_1，右眼读数自然为 τ_1'，设左眼读数为 τ_2'。记 $\Delta\tau_2 \equiv \tau_2' - \tau_2$，则 G_1 会认为在他的钟走了 $\Delta\tau_1$ 这段时间里 G_2 钟走了 $\Delta\tau_2$。如果 $\Delta\tau_2 = \Delta\tau_1$，他会认为两钟的表观走时率相同，否则就可分出快慢。请注意以上讨论旨在明确约定一种比钟方式，虽然你会觉得这种方式很合理，但我们对“合理性”不必太过在意，因为只要明确约定，任一比钟方式都（至少在某种程度上）有其合理性。

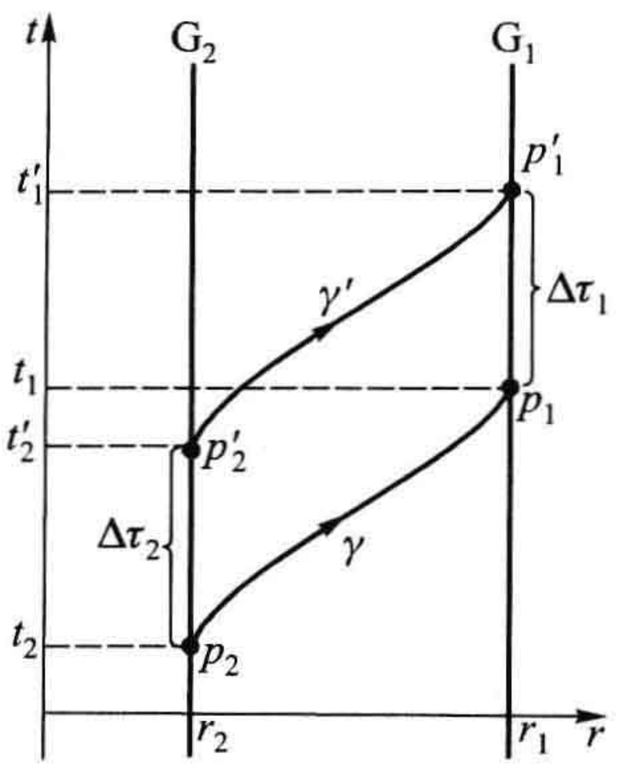


图 8－4　静态引力场的钟慢效应，曲线是光子世界线。

因为 G_1 线上 r,θ,φ 为常数，故其线元

$$ds^2\big|_{G_1} = -\left(1-\frac{2M}{r_1}\right)dt^2 + \left(1-\frac{2M}{r_1}\right)^{-1}dr^2 + r_1^2(d\theta^2+\sin^2\theta d\varphi^2) = -\left(1-\frac{2M}{r_1}\right)dt^2。 \tag{8-3-1}$$

从 p_1 积分到 p_1' 给出

$$\Delta\tau_1 = \int_{p_1}^{p_1'}\sqrt{-ds^2\big|_{G_1}} = \int_{t_1}^{t_1'}\sqrt{1-\frac{2M}{r_1}}dt = \sqrt{1-\frac{2M}{r_1}}\Delta t_1, \tag{8-3-2}$$

其中 $\Delta t_1 \equiv t_1' - t_1$。类似地，对 G_2 线也有

$$\Delta\tau_2 = \sqrt{1-\frac{2M}{r_2}}\Delta t_2, \tag{8-3-3}$$

其中 $\Delta t_2 \equiv t_2' - t_2$。故

$$\frac{\Delta\tau_1}{\Delta\tau_2} = \sqrt{\frac{1-2M/r_1}{1-2M/r_2}}\frac{\Delta t_1}{\Delta t_2}。 \tag{8-3-4}$$

下面证明 $\Delta t_1/\Delta t_2 = 1$。图 8－4 的光子世界线（径向类光测地线）$\gamma$ 上的 θ 和 φ 为常数，故其线元

$$ds^2\big|_{\gamma} = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1}dr^2。 \tag{8-3-5}$$

γ 的类光性保证 $ds^2|_{\gamma} = 0$，所以上式归结为

$$dt = \left(1-\frac{2M}{r}\right)^{-1}dr。 \tag{8-3-6}$$

对上式从 p_2 沿 γ 线积分到 p_1 给出

$$t_1 - t_2 = \int_{r_2}^{r_1}\left(1-\frac{2M}{r}\right)^{-1}dr。 \tag{8-3-7}$$

类似地，对光子世界线 γ' 有

$$t_1' - t_2' = \int_{r_2}^{r_1} \left(1 - \frac{2M}{r}\right)^{-1} \mathrm{d}r。 \tag{8-3-8}$$

对比以上两式得 $t_1' - t_2' = t_1 - t_2$，因而 $t_1' - t_1 = t_2' - t_2$，此即 $\Delta t_1/\Delta t_2 = 1$。代入式(8-3-4)便有

$$\frac{\Delta\tau_1}{\Delta\tau_2} = \sqrt{\frac{1-2M/r_1}{1-2M/r_2}} > 1 \quad (\because r_2 < r_1), \tag{8-3-9}$$

即 $\Delta\tau_1 > \Delta\tau_2$，所以观者 G_1 认为观者 G_2 的钟较慢，就是说：引力场强越大处的钟越慢。由于 G_1 和 G_2 都是静态观者，各自静止于空间点(r_1,θ,φ)和(r_2,θ,φ)上，可以认为两者之间没有相对运动(哪个也不是“动钟”)，所以说这种钟慢效应纯粹起因于引力场，因而称为**引力钟慢效应**(英语文献也称之为**引力场中的时间延缓效应**)。我们刚才只就史瓦西时空对此效应给了证明，其实还可证明这一效应对任何静态时空都成立，只不过式(8-3-9)要适当修改。

为了对式(8-3-9)有一个定量的、联系实际的理解，我们来比较位于地球表面不同高度的两只原子钟的表观走时率。第一只钟保存在美国国家标准局(National Bureau of Standards)里，该局位于美国科罗拉多州的博尔德市(Boulder, Colorado)，海拔 5 400 英尺(≈1 645.9 m)；第二只钟保存在英国皇家格林尼治天文台(Royal Greenwich Observatory)，海拔 80 英尺(≈24.4 m)。由于两钟存在高度差 $\Delta h \approx 1\ 621.5$ m，每年读数差竟达 5.6 μs。下面是求得此值的计算过程。

为便于数值计算，先将式(8-3-9)(在几何制中成立)改为国际单位制形式[改法可参见附录 A 例 3]：

$$\frac{\Delta\tau_1}{\Delta\tau_2} = \sqrt{\frac{1-2GM/c^2r_1}{1-2GM/c^2r_2}}, \tag{8-3-9'}$$

其中 G 和 c 分别是引力常量和真空光速在国际单位制的数值(将上式看作数的等式)。用于刚才的问题，则 M 代表地球质量，r_1 和 r_2 分别代表上述两钟与地心的距离。令

$$\varepsilon_1 \equiv GM/c^2r_1, \quad \varepsilon_2 \equiv GM/c^2r_2, \tag{8-3-10}$$

则式(8-3-9′)成为

$$\Delta\tau_1/\Delta\tau_2 = \sqrt{(1-2\varepsilon_1)/(1-2\varepsilon_2)} = (1-2\varepsilon_1)^{1/2}(1-2\varepsilon_2)^{-1/2}, \tag{8-3-11}$$

由数值估算可知 $2\varepsilon_1 \ll 1$，$2\varepsilon_2 \ll 1$，由牛顿二项式定理得

$$\Delta\tau_1/\Delta\tau_2 \approx (1-\varepsilon_1)(1+\varepsilon_2) \approx 1 + \varepsilon_2 - \varepsilon_1, \tag{8-3-12}$$

由式(8-3-10)又得

$$\varepsilon_2-\varepsilon_1=\frac{GM}{c^2}\left(\frac{1}{r_2}-\frac{1}{r_1}\right)=\frac{GM}{c^2}\frac{r_1-r_2}{r_2r_1}\approx\frac{GM}{c^2}\frac{\Delta h}{r_2^2}。$$

借用牛顿力学不难看出上式右边的 GM/r_2^2 正是地面的重力加速度 g，因而

$$\varepsilon_2-\varepsilon_1=\frac{1}{c^2}g\Delta h。\qquad(8-3-13)$$

于是式(8－3－12)成为

$$\Delta\tau_1=\left(1+\frac{1}{c^2}g\Delta h\right)\Delta\tau_2,$$

故

$$\Delta\tau_1-\Delta\tau_2=\frac{1}{c^2}g\Delta h\cdot\Delta\tau_2。\qquad(8-3-14)$$

可见两钟的读数差 $\Delta\tau_1-\Delta\tau_2$ 随 $\Delta\tau_2$ 的增大而正比地增大。取

$$\Delta\tau_2=1\text{ 年}=365\text{ 天}=365\times(24\times3\,600\text{ s})=3.15\times10^{13}\ \mu\text{s},$$

再利用 $g=9.8\ \text{m/s}^2$，$\Delta h=r_2-r_1=1\ 621.5\ \text{m}$ 以及 $c=3\times10^8\ \text{m/s}$，代入式(8－3－14)便得

$$\Delta\tau_1-\Delta\tau_2=\frac{1}{(3\times10^8\ \text{m/s})^2}\times9.8\ \text{m/s}^2\times1\ 621.5\ \text{m}\times(3.15\times10^{13}\ \mu\text{s})=5.56\ \mu\text{s}。$$

上述结论(两钟读数差为每年 5.56 μs)早已被实际测量所证实。然而理论上存在一个问题：式(8－3－14)来自式(8－3－9′)，推导该式时曾约定两钟都是静态钟(其 r,θ,φ 为常数)。但是地球有自转，严格说来上述两钟不是静态钟(φ 随 t 变)，式(8－3－14)还适用吗？答案是肯定的，理由是非常有趣的，我们将在选读 8－2(在 §8.6)末段介绍。

[选读 8－1]

小节 4.3.4 曾介绍过验证双子效应的原子钟环球飞行实验，结论是：飞行钟环球一周的时间 τ 不等于地面钟的经历时间 τ_0，差值为[见式(4－3－16)]

$$\tau-\tau_0=\frac{1}{2c^2}[2gh-(2R\Omega v+v^2)]\tau_0,\qquad(8-3-15)$$

其中 R 和 Ω 分别是地球赤道半径和自转角速率，v 是飞机(世界线 C)的东飞速度，h 是飞行高度，g 是地面重力加速度。上式既包含了运动学效应(狭义相对论的双子效应)，又包含了地球引力场的影响(广义相对论效应，体现为 $2gh$ 所在项)。由于当时尚未讲到广义相对论，小节 4.3.4 未能介绍添加该项的推证过程。现在，在懂得了史瓦西线元和引力场的钟慢效应之后，就有条件详加补证。

地球引力场使地球附近的时空弯曲，式(4－3－1)的闵氏线元应被史瓦西线元取代：

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2),①$$

(8－3－16)

其中 M 是地球质量。上式适用于几何单位制，为便于数值计算，宜改用国际单位制形式：

$$ds^2 = -\left(1-\frac{2GM}{c^2r}\right)c^2dt^2 + \left(1-\frac{2GM}{c^2r}\right)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2),$$

(8－3－16′)

其中 G 和 c 分别是引力常量和真空光速在国际单位制中的数值。仿照牛顿引力势的概念，用下式定义引力势 $\Phi(r)$：

$$\Phi(r) \equiv -\frac{GM}{r}, \qquad (8-3-17)$$

则

$$ds^2 = -\left(1+\frac{2\Phi}{c^2}\right)c^2dt^2 + \left(1+\frac{2\Phi}{c^2}\right)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2),$$

(8－3－18)

先讨论静止于赤道上的钟 C_0(相应于 $v=0$，仍见图4－21)。因赤道上有 $r=R$，$dr=0$，$\theta=\pi/2$，$d\theta=0$ 及 $d\varphi=\Omega dt$，故式(8－3－18)用于 C_0 线的任一元段给出

$$ds_0^2 = -\left(1+\frac{2\Phi_0}{c^2}\right)c^2dt^2 + R^2d\varphi^2 = -\left(1+\frac{2\Phi_0}{c^2}\right)c^2dt^2 +$$

$$R^2\Omega^2dt^2 = -\left(1+\frac{2\Phi_0}{c^2}-\frac{u_0^2}{c^2}\right)c^2dt^2, \qquad (8-3-19)$$

其中

$$\Phi_0 \equiv \Phi(R) = -GM/R \qquad (8-3-20)$$

是赤道上的引力势，$u_0 = R\Omega$ 是 C_0 钟(随地球自转)的线速率。于是元段的固有时间为

$$d\tau_0 = \sqrt{\frac{-ds_0^2}{c^2}} = \left(1+\frac{2\Phi_0}{c^2}-\frac{u_0^2}{c^2}\right)^{1/2}dt。 \qquad (8-3-21)$$

---

① 所选的史瓦西坐标要满足：①在地心有 $r=0$；②不随地球自转而转动(相应的 x,y,z 轴分别指向远方固定恒星)。

式(4-3-5)已证明 $u_0^2/c^2 << 1$,利用地球数据及物理常数(都指各量在国际单位制中的数值)

$$M = 6 \times 10^{24},\ R = 6.4 \times 10^{6},\ G = 6.7 \times 10^{-11},\ c = 3 \times 10^{8} \qquad (8-3-22)$$

易知赤道引力势 Φ_0 满足

$$\frac{|\Phi_0|}{c^2} = \frac{GM}{c^2 R} = \frac{(6.7 \times 10^{-11}) \times (6 \times 10^{24})}{(3 \times 10^{8})^2 \times (6.4 \times 10^{6})} \approx 7 \times 10^{-10} << 1, \qquad (8-3-23)$$

令 $x \equiv \dfrac{2\Phi_0}{c^2} - \dfrac{u_0^2}{c^2}$,则 $x << 1$,用牛顿二项式定理 $(1+x)^{1/2} \approx 1 + x/2$ 可将式(8-3-21)简化为

$$\mathrm{d}\tau_0 \approx \left(1 + \frac{\Phi_0}{c^2} - \frac{u_0^2}{2c^2}\right)\mathrm{d}t。 \qquad (8-3-24)$$

将上式沿 C_0 线(见图4-22)全程做积分便可求得 τ_0:

$$\tau_0 = \int_0^{t_q}\left(1 + \frac{\Phi_0}{c^2} - \frac{u_0^2}{2c^2}\right)\mathrm{d}t = \left(1 + \frac{\Phi_0}{c^2} - \frac{u_0^2}{2c^2}\right)t_q。 \qquad (8-3-25)$$

再讨论赤道上空的任一C线(v 值任意),它与 C_0 线有两点不同:① C_0 线在赤道上,引力势为 $\Phi_0 \equiv \Phi(R)$,而C线(飞机)在赤道上空离地面 h 米处,引力势为 $\Phi \equiv \Phi(R+h)$;② C_0 线代表固结在赤道某点的钟,满足 $\mathrm{d}\varphi = \Omega\mathrm{d}t$,其线速率 $u_0 = R\Omega$,而C线是赤道上空的飞行钟,其线速率 $u = (R+h)\mathrm{d}\varphi/\mathrm{d}t$。因此,式(8-3-18)用于C线的任一元段给出

$$\mathrm{d}s_0^2 = -\left(1 + \frac{2\Phi}{c^2}\right)c^2\mathrm{d}t^2 + (R+h)^2\mathrm{d}\varphi^2 =$$

$$-\left(1 + \frac{2\Phi}{c^2}\right)c^2\mathrm{d}t^2 + u^2\mathrm{d}t^2 = -\left(1 + \frac{2\Phi}{c^2} - \frac{u^2}{c^2}\right)c^2\mathrm{d}t^2, \qquad (8-3-26)$$

用 Φ 和 u 分别替换式(8-3-25)的 Φ_0 和 u_0 便得C线的环球飞行时间 τ:

$$\tau = \left(1 + \frac{\Phi}{c^2} - \frac{u^2}{2c^2}\right)t_q。 \qquad (8-3-27)$$

式(8-3-27)与式(8-3-25)相减给出

$$\tau - \tau_0 = \left(\frac{\Phi - \Phi_0}{c^2} - \frac{u^2 - u_0^2}{2c^2}\right)t_q, \qquad (8-3-28)$$

式中的 $\Phi - \Phi_0$ 可由式(8-3-17)及式(8-3-20)求得:

$$\Phi - \Phi_0 = -\frac{GM}{R+h} - \left(-\frac{GM}{R}\right) = \frac{GMh}{R(R+h)} \approx \frac{GMh}{R^2} = gh, \qquad (8-3-29)$$

其中 $g = GM/R^2$ 是地面的重力加速度。将上式代入式(8-3-28)给出

$$\tau-\tau_0=\left(\frac{gh}{c^2}-\frac{u^2-u_0^2}{2c^2}\right)t_q=\frac{1}{2c^2}[2gh-(u^2-u_0^2)]t_q。\quad(8-3-30)$$

选读4-5开头很长一段一直忽略飞机的飞行高度h,即默认飞机的径向坐标$r=R$,所以有$u=R\Omega+v$[式(4-3-9)]。如果不做忽略,此式应为

$$u=(R+h)\Omega+v=R\Omega+h\Omega+v,$$

右边的$h\Omega$当然可被忽略,问题在于它是否也能被v忽略。Ω是地球自转角速率,其值为

$$\Omega=\frac{2\pi}{\text{天}}=\frac{2\pi}{24\times3\ 600\ \text{s}}\approx7.3\times10^{-5}/\text{s}。$$

飞机飞行高度h的估算值为$h=10^3$ m,由此算得$h\Omega\approx0.07$ m/s,只有声速的2×10^{-4}倍,故$h\Omega<<v$,因而确有$u=R\Omega+v$。由此式及$u_0=R\Omega$又得

$$u^2-u_0^2=2R\Omega v+v^2。\quad(8-3-31)$$

代入式(8-3-30)便有

$$\tau-\tau_0=\frac{1}{2c^2}[2gh-(2R\Omega v+v^2)]t_q。\quad(8-3-32)$$

前已证得$u_0^2/c^2<<1$[式(4-3-5)]及$|\Phi_0|/c^2<<1$[式(8-3-23)],故式(8-3-25)近似给出$t_q=\tau_0$,于是

$$\tau-\tau_0=[2gh-(2R\Omega v+v^2)]\tau_0/2c^2。$$

此即待证的式(4-3-16)。　　[选读8-1完]

8.3.2　引力红移

引力钟慢效应的另一表现形式是引力场中的红移效应。设观者G_2携带一个简谐光源,所发的光的两个相邻波峰(正最大值)出现在p_2和p_2'(仍见图8-4),则$\Delta\tau_2$就是振动的周期。因为这两个波峰被观者G_1在p_1和p_1'时刻收到,所以G_1测得的周期是$\Delta\tau_1$。以λ_2和λ_1分别代表G_2和G_1测得的光波波长,由波长与周期成正比可知

$$\frac{\lambda_1}{\lambda_2}=\frac{\Delta\tau_1}{\Delta\tau_2}=\sqrt{\frac{1-2M/r_1}{1-2M/r_2}}>1。\quad(8-3-33)$$

[其中第二步用到式(8-3-9)。] 可见观者G_1接收到的光波波长大于光源G_2测得的波长,即光波出现红移。又因为G_1与G_2之间没有相对运动(都是静态),所以这种红移纯粹起因于引力场(时空弯曲),因此称为**引力红移**(gravitational redshift)。引入**红移因子**

$$\chi\equiv\sqrt{1-\frac{2M}{r}},\quad(8-3-34)$$

则

$$\frac{\lambda_1}{\lambda_2}=\frac{\chi_1}{\chi_2}。\tag{8-3-35}$$

引力红移的程度可用相对红移量(简称**红移**)$z\equiv(\lambda_1-\lambda_2)/\lambda_2$ 描写。计算表明,从太阳表面发出的光到达地球时(把太阳作为引力场源),相对红移量约只有 2×10^{-6}. 为了加大红移量,可测量从白矮星发来的光。由于白矮星的密度比普通恒星高得多(详见小节 7.2.2),其周围的引力场比太阳周围的引力场要强得多。白矮星来光的红移可达太阳来光红移的几十倍。在广义相对论发表后,人们曾几次测过白矮星来光的红移,但结果还不足以确证理论的预言。第一次成功的高精度引力红移实验是 Pound 和 Rebka 等在 1960 年利用穆斯堡尔效应完成的。穆斯堡尔在 1960 年发现,某些原子核(如^{57}Fe)在特定条件下可以发出谱线宽度很窄(很尖锐)的 γ 射线,含有这种原子核的晶体又能对这种频率的 γ 射线作选择性甚高的共振吸收。假定这种 γ 射线的频率不论由于什么原因而有微小变化,它被这种晶体吸收的程度就显著降低。这就为测出由地球引力场造成的极其微弱的引力红移提供了强有力的手段。把两块这样的晶体分置于地球表面的不同高度处,较低的一块(图 8-5 中的 E)作为发射体,较高的一块(图中的 A)作为接收体。虽然根据两者高度差(12.5 m)算得的红移只有 1.36×10^{-15},但 A 对 E 所发 γ 射线的吸收率仍然由于 γ 射线的微弱引力红移而有所下降。为了确认这一下降并测出下降的数量,可令 A 以某一常速率向 E 运动,利用由于多普勒效应出现的"蓝移"(波长减小)抵消引力红移。当速率调至某一适当值时(仅为 3×10^{-7} m/s),吸收率达到最大值。由此便可测得引力红移的数值。本实验的精确度在当时的条件下是很高的(相对不确定度约为 1%),所得结果同理论值符合得很好。后来还有更精确的实验验证,可参阅 Will(1993)。

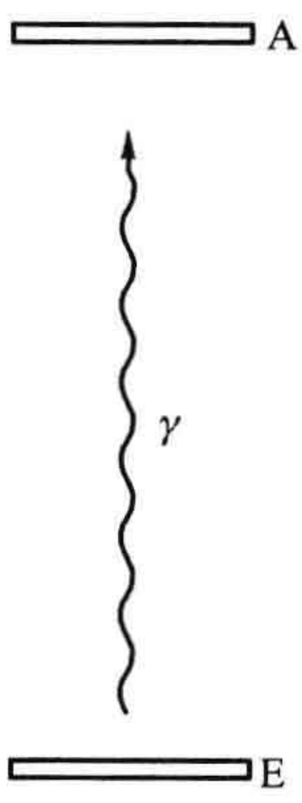


图 8-5　用穆斯堡尔效应测量地面附近的引力红移。

§8.4　水星近日点进动

按照牛顿力学,行星的轨道是以太阳为一个焦点的椭圆。然而观测结果与此略有歧离。以最靠近太阳的水星为例,虽然它在每一周期中的轨道很接近椭圆,但两个相邻周期的两个"椭圆"的长轴并不重合,表现在它的近日点(perihelion)的微小改变上(见图 8-6)。随着时间的推移,由于积累效应,"椭圆"的长

轴(因而近日点)绕太阳的缓慢转动变得可以观测。这现象叫做近日点的**进动**(precession)。从1697年至1848年的天文观测表明水星近日点的进动率是每世纪5 600″(″代表角秒)。如何解释?法国数学家**勒威耶**(Le Verrier)用牛顿力学对水星运动做过长期研究,并于1859年再次发表研究报告。根据这一研究,水星近日点由于某些原因(例如其他行星的影响)的确应该进动,但所有原因造成的进动率也只有每世纪5 557″,余下的每世纪43″无法解释,这就是著名的“43秒问题”。他猜想这个43″来自一个尚未认识的行星的影响,并将它命名为**祝融星**(Vulcan)。由于不久之前他在预言海王星上取得巨大成功,很快就在国际天文学界掀起了寻找祝融星的热潮,许多天文学家(包括业余的)先后报告他们发现了祝融星的踪迹,但是要么禁不起考验,要么互相矛盾。虽然勒威耶在1878年去世时仍然坚信他的猜想,但是绝大多数天文学家都不认可。直至1915年爱因斯坦的广义相对论文章[爱因斯坦等《相对论原理》(1980中译本)]发表后,人们才有足够理由相信祝融星的假设是多余的。爱因斯坦在文章末尾写道:

“如果我们用更高的近似程度计算引力场,……就会发现计算结果与开普勒-牛顿的行星运动定律有以下偏差:即行星的轨道椭圆在运动方向上做缓慢转动,……近似给出水星轨道的转动为每世纪43″,准确地符合天文学观测(勒威耶);因为天文学家已经发现,在扣除其他行星的摄动之后,这行星的近日点运动中仍留下这个无法解释的数值。”

爱因斯坦的这一结果大大加强了人们对广义相对论的信心。下面介绍用史瓦西解对近日点进动的推证。

设太阳系只有太阳和水星并略去水星的引力场,即只讨论水星在太阳引力场(外引力场)作用下的运动。先用牛顿引力论讨论。以 M,m 分别代表太阳和水星的质量,r 代表两者的距离,则水星的引力势能为(用几何单位制,其中 $G=1$)

$$U(r)=-\frac{Mm}{r}, \tag{8-4-1}$$

取球坐标系使水星轨道躺在赤道面上[即 $\theta=\pi/2$,这总可做到,证明见梁灿彬,周彬(2006)§9.1],则水星的速度 $\boldsymbol{u}$ 只有径向分量 $u_r=\mathrm{d}r/\mathrm{d}t$ 和切向分量 $u_\varphi=r\mathrm{d}\varphi/\mathrm{d}t$,故动能为

$$\frac{1}{2}m(u_r^2+u_\varphi^2)。$$

由机械能守恒律有

$$\frac{1}{2}m(u_r^2+u_\varphi^2)+U(r)=A, \tag{8-4-2}$$

其中常数 A 代表总机械能。设水星单位质量的角动量的大小为 $|L|$，则

$$L = ru_{\varphi} = r^2 \frac{\mathrm{d}\varphi}{\mathrm{d}t}, \tag{8-4-3}$$

由式(8-4-1)、式(8-4-2)、式(8-4-3)出发经计算得

$$\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2 + r^2 = \frac{2Mr^3}{L^2} + \frac{2Ar^4}{mL^2}。 \tag{8-4-4}$$

令 $\mu \equiv r^{-1}$，则 $\mu \neq 0$，故上式成为

$$\left(\frac{\mathrm{d}\mu}{\mathrm{d}\varphi}\right)^2 + \mu^2 = \frac{2A}{mL^2} + \frac{2M}{L^2}\mu。 \tag{8-4-5}$$

对 φ 求导得

$$\frac{\mathrm{d}\mu}{\mathrm{d}\varphi}\left(\frac{\mathrm{d}^2\mu}{\mathrm{d}\varphi^2} + \mu - \frac{M}{L^2}\right) = 0,$$

因此，要么 $\mathrm{d}\mu/\mathrm{d}\varphi = 0$（圆轨道）；要么

$$\frac{\mathrm{d}^2\mu}{\mathrm{d}\varphi^2} + \mu = \frac{M}{L^2}。 \tag{8-4-6}$$

上式的解为

$$\mu(\varphi) = \frac{M}{L^2}[1 + e\cos(\varphi - \varphi_0)], \tag{8-4-7}$$

其中 e 和 φ_0 为积分常数。不失一般性，取 $\varphi_0 = 0$，则

$$\mu(\varphi) = \frac{M}{L^2}(1 + e\cos\varphi)。 \tag{8-4-8}$$

这是圆锥截线方程，e 是偏心率，把式(8-4-8)及其导数代回式(8-4-5)便得

$$e^2 = 1 + \frac{2AL^2}{mM^2}。 \tag{8-4-9}$$

当 $0 \leqslant e < 1$ 时为椭圆，$\mathrm{d}\mu/\mathrm{d}\varphi = 0$（圆轨道）已作为 $e = 0$ 的特例含于其中。

然而广义相对论却给出略微不同的结果。前已说过（见 §6.5 末），由于太阳引力场很弱，对太阳系的绝大多数问题而言，使用牛顿引力论已经足够，无须出动广义相对论这把“牛刀”。然而近日点进动问题却不同，因为牛顿引力论的结果与观测数据并不吻合，一直遗留着一个“43 秒问题”，所以爱因斯坦期望广义相对论与牛顿引力论在这个问题上的微小差别恰好能解决这个难题。计算结果的确非常理想。从广义相对论看来，太阳外部的弯曲时空由史瓦西线元描述，而水星可看作是这个时空中的一个自由质点，其世界线是该时空的一条类时测地线。对测地线的计算给出

$$\frac{\mathrm{d}^2\mu}{\mathrm{d}\varphi^2} + \mu = \frac{M}{L^2} + 3M\mu^2。 \tag{8-4-10}$$

[注:本书并未介绍测地线的数学定义及计算公式,读者只能承认上述结果。欲知推导详情可参阅梁灿彬,周彬(2006)小节 9.2.2。] 与式(8-4-6)对比发现净多一项 $3M\mu^2$(广义相对论修正项),因而难于求解。幸好水星的 r 比太阳的 M 大得多,即 $M/r<<1$①,故修正项

$$3M\mu^2=\frac{3M}{r}\mu<<\mu, \tag{8-4-11}$$

(其中第一步用到 $\mu\equiv r^{-1}$)可设法求近似解。牛顿引力论的解[式(8-4-8)]可看作零级近似,为明确起见记作 $\mu_0(\varphi)$,即

$$\mu_0(\varphi)=\frac{M}{L^2}(1+e\cos\varphi)。 \tag{8-4-12}$$

把这零级近似代入式(8-4-10)右边第二项后所得方程可看作一级近似解 $\mu_1(\varphi)$ 所应满足的方程,即

$$\frac{\mathrm{d}^2\mu_1}{\mathrm{d}\varphi^2}+\mu_1=\frac{M}{L^2}+3M\mu_0^2=\frac{M}{L^2}+\frac{3M^3}{L^4}(1+2e\cos\varphi+e^2\cos^2\varphi)。 \tag{8-4-13}$$

不难验证其解为

$$\mu_1(\varphi)=\mu_0(\varphi)+\frac{3M^3}{L^4}\left[1+e\varphi\sin\varphi+e^2\left(\frac{1}{2}-\frac{1}{6}\cos2\varphi\right)\right]。 \tag{8-4-14}$$

令

$$\varepsilon\equiv 3M^2/L^2, \tag{8-4-15}$$

再与式(8-4-12)一同代入式(8-4-14),整理后得

$$\mu_1(\varphi)=\frac{M}{L^2}\left\{1+\varepsilon\left[1+e^2\left(\frac{1}{2}-\frac{1}{6}\cos2\varphi\right)\right]+e(\cos\varphi+\varepsilon\varphi\sin\varphi)\right\}。 \tag{8-4-16}$$

由式(8-4-12)可知 M/L^2 与 μ 有相同量级(记作 $M/L^2\sim\mu$),故 $\varepsilon\equiv 3M^2/L^2\sim 3M\mu=3M/r$,改为国际单位制并利用式(8-4-11)脚注的数据,得

$$\varepsilon\sim\frac{3GM}{c^2r}\approx\frac{3\times 1.5\ \mathrm{km}}{5\times10^7\ \mathrm{km}}\approx10^{-7}<<1, \tag{8-4-17}$$

于是式(8-4-16)右边花括号内第二项(含 ε 的项)与第一项(即 1)相较可以略去,再利用三角公式 $\cos(\alpha-\beta)=\cos\alpha\cos\beta+\sin\alpha\sin\beta$ 以及 $\cos\varepsilon\varphi\approx1$,$\sin\varepsilon\varphi\approx\varepsilon\varphi$ 改写第三项便得

$$\frac{1}{r(\varphi)}\approx\mu_1(\varphi)\approx\frac{M}{L^2}[1+e\cos(\varphi-\varepsilon\varphi)]。 \tag{8-4-18}$$

---

① 做数量计算时最好改回国际单位制,即补上物理常数 G 和 c。由附录 A 例 3 可知 M/r 实为 $(GM/c^2)/r$。太阳质量 M 对应的 $GM/c^2\approx1.5$ km,水星近日点与太阳的距离约为 5×10^7 km,故 $(GM/c^2)/r<<1$。

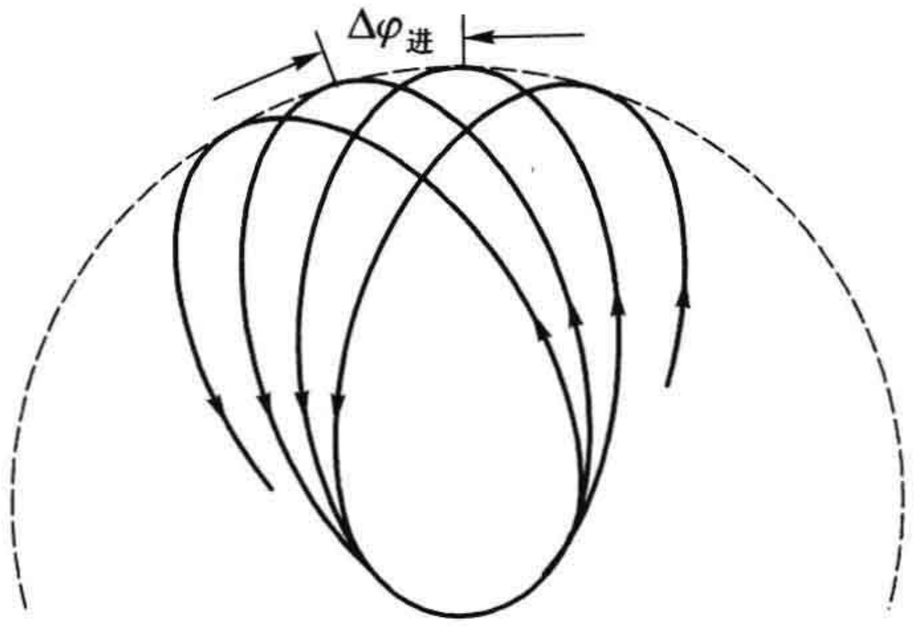


图 8 - 6　水星近(远)日点每周期进动角 $\Delta\varphi_{进}$(明显夸大)

这表明水星的轨道近似为椭圆。上式右边虽仍是周期函数,但周期已不像式(8 - 4 - 12)那样为 2π。近日点是 r 最小的点,即 $\cos(\varphi-\varepsilon\varphi)=1$ 的点。$\varphi=0$ 当然是近日点;但当 $\varphi=2\pi$ 时

$$\cos(\varphi-\varepsilon\varphi)=\cos(2\pi-2\pi\varepsilon)\neq 1。$$

设 $\hat{\varphi}$ 为满足 $\cos(\hat{\varphi}-\varepsilon\hat{\varphi})=1$ 而又最接近 2π 的 φ 值,则不难验证(略去高阶小量 $2\pi\varepsilon^2$)

$$\hat{\varphi}\approx 2\pi+2\pi\varepsilon。\tag{8-4-19}$$

可见水星近日点每周期内的进动角为(见图 8 - 6)

$$\Delta\varphi_{进}\approx 2\pi\varepsilon=6\pi M^2/L^2。\tag{8-4-20}$$

以上讨论其实对任何行星都适用。代入具体数据可得水星近日点的进动率为 43″/世纪。这可以看作广义相对论在实验验证方面的第一个成功,爱因斯坦对这一计算结果感到兴奋不已。至 2005 年为止,观测结果与上述理论结果之间的差别已达到只有千分之几的程度,见 Will(2005)。

§8.5　星 光 偏 转

8.5.1　理论预言值和日全食观测

远方恒星射到地球的光线经过太阳附近时受太阳引力场影响而弯曲(简称**星光偏转**),这是爱因斯坦根据广义相对论做出的一个重要预言。本节介绍这一预言的导出和验证。

因为光子除引力外不受任何力,所以来自恒星的光子在太阳周围的史瓦西时空中的世界线是该时空的类光测地线。仿照上节,取球坐标系 $\{t,r,\theta,\varphi\}$ 使光子的世界线躺在赤道面上(即线上的 θ 恒为 $\pi/2$),并令 $\mu\equiv r^{-1}$。对类光测地线

的计算给出

$$\frac{\mathrm{d}^2\mu}{\mathrm{d}\varphi^2}+\mu=3M\mu^2(M\text{为太阳质量})。\tag{8-5-1}$$

(与上节一样,此处同样不能给出测地线的计算过程。)当 $M=0$ 时(平直时空),方程(8-5-1)的通解为

$$\mu(\varphi)=\frac{1}{l}\sin(\varphi+\varphi_0),\tag{8-5-2}$$

其中 l 和 φ_0 为积分常数。约定当 $\varphi=0$ 时光子在无限远,即 $\mu(0)=1/r(0)=0$,则 $\varphi_0=0$,故

$$\mu(\varphi)=\frac{1}{l}\sin\varphi,\tag{8-5-3}$$

这是2维欧氏空间中用极坐标 r,φ 表述的直线方程。要看出这一点,取 $r=0$ 为直角坐标系 $\{x,y\}$ 的原点,则

$$x=r\cos\varphi,\tag{8-5-4}$$

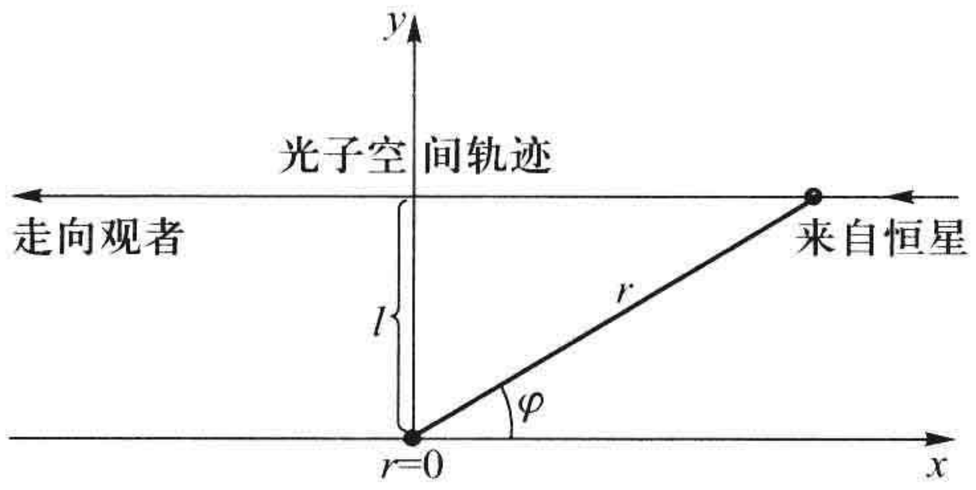


图8-7 平直时空中光子的空间轨迹

$$y=r\sin\varphi=\frac{1}{\mu}\sin\varphi=l=\text{常数},\tag{8-5-5}$$

其中第三步用到式(8-5-3)。所以光子的空间轨迹是平行于 x 轴、与原点距离为 l 的直线(见图8-7)。请注意 r 和 φ 沿该直线都在变化,不变的是 y。由于 r 的取值范围是 $0<r<\infty$,式(8-5-4)表明 x 的取值范围是 $-\infty<x<\infty$。

要讨论星光偏转当然不能取 $M=0$,但因 $M/r<<1$,求一级近似解 $\mu_1(\varphi)$ 便已足够。把式(8-5-3)的 $\mu(\varphi)$ 看作零级近似解 $\mu_0(\varphi)$,代入式(8-5-1)右边便得 $\mu_1(\varphi)$ 所应满足的微分方程:

$$\frac{\mathrm{d}^2\mu_1}{\mathrm{d}\varphi^2}+\mu_1(\varphi)=\frac{3M}{l^2}\sin^2\varphi。\tag{8-5-6}$$

不难验证下式是方程(8-5-6)的解:

$$\mu_1(\varphi)=\frac{1}{l}\sin\varphi+\frac{M}{l^2}(1-\cos\varphi)^2。\tag{8-5-7}$$

由上式知 $\mu_1(0)=0$，即 $r(0)=\infty$，说明光子的 φ 坐标为零时光子离太阳无限远（远方恒星的 r 可看作 ∞），但式(8-5-7)与式(8-5-3)的不同表明光子经过太阳附近又离太阳而去时的“去向”不同：由式(8-5-3)知 $\mu(\pi)=0$，说明光子远离“太阳”（这个“太阳”的 $M=0$）而去时 φ 坐标为 π；而由式(8-5-7)却有 $\mu_1(\pi)\neq 0$，我们预期它沿一个与 π 略有差别的方向 $\pi+\beta$ 远离太阳而去（见图8-8），即 $\mu_1(\pi+\beta)=0$。为求偏转角 β，把 $\varphi=\pi+\beta$ 代入式(8-5-7)并利用 $\mu_1(\pi+\beta)=0$ 得

$$0=\mu_1(\pi+\beta)=\frac{1}{l}\sin(\pi+\beta)+\frac{M}{l^2}[1-\cos(\pi+\beta)]^2。$$

β 很小导致 $\sin(\pi+\beta)\approx-\beta$，$\cos(\pi+\beta)\approx-1$，代入上式便得

$$\beta\approx\frac{4M}{l}。\tag{8-5-8}$$

上式是几何单位制的形式，改回国际单位制形式则为

$$\beta\approx\frac{4GM}{c^2l}。\tag{8-5-8'}$$

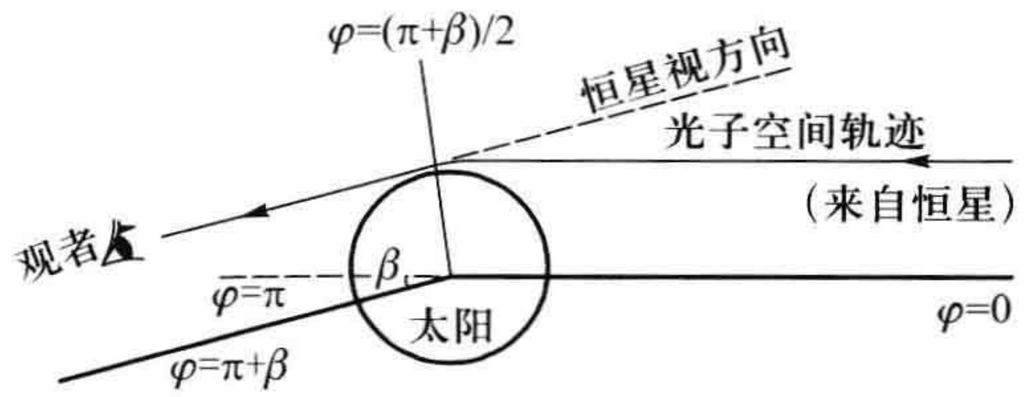


图8-8　太阳使时空弯曲，星光的空间轨迹受偏转。偏转角 β 明显夸大。

上式表明偏转角 β 随 l 的减小而反比地增大。l 的最小值等于太阳半径，相当于光子在太阳边缘掠过（“擦边光子”）。以此为 l 值代入式(8-5-8')便可求得擦边光子的偏转角 $\beta=1.75''$。

上式就是广义相对论对星光偏转的预言。为了用观测验证这一预言，可设法拍摄当星光被太阳偏转时恒星的视位置（角度），并与半年后（或前）当地球转到太阳另一侧时摄得的恒星真实位置做比较。然而观测恒星的视位置并非易事，因为太阳离地球比所要观测的恒星近得多，在太阳的光芒中根本无法看见星光。（“大白天看星星”！）于是想到利用日全食。阳光在日全食时被位于太阳与地球之间的月球所遮挡，但远方恒星的光却可以“绕过”太阳到达地球。只要将日全食时以及几个月后（当太阳离开星场时）对同样几个恒星拍摄的照片做对比，便可求得这些恒星所发光子的偏角，再用偏角与 l 的反比关系就能推得擦边光子的偏角。

由于对爱因斯坦的广义相对论欣赏有加,爱丁顿早在1917年写给英国皇家天文学会的报告中就强调了用星光偏转检验广义相对论的重要性。在第一次世界大战后不久,就有两支考察队从英国出发分别到巴西和西非对1919年3月29日的日全食进行观测(爱丁顿是西非队的领队)。两队观测的结果分别是1.98″±0.12″和1.61″±0.30″,亦即分别是理论预期值的1.13±0.07倍和0.92±0.17倍,被认为是对广义相对论的有力支持。这一消息由英国皇家学会和天文学会于1919年11月6日公布,立即引起了对战争厌倦了的西方各国公众的注意,并使爱因斯坦家喻户晓,名声显赫。(由于当时通信不便,中国人最早是在1920年2月10日发行的《东方杂志》上得知此消息的。)然而,据说爱因斯坦对这一观测结果反应平静。他对自己的理论是如此相信(基于其优雅性和内部自洽性),以致曾说过一句名言,大意是:如果实验竟然与理论矛盾,他将"对亲爱的主(Lord,意为上帝)感到遗憾"。

8.5.2 用星光偏转检验引力理论[选读]

星光偏转其实不是广义相对论的专利,牛顿引力论(以及其他多种引力理论)对此也有预言。牛顿在1704年出版了专著《光学》("Opticks"),他在书末提出了31个悬而未决、有待他人进一步研究的问题,其中就有下面一句:

"难道物体就不会对某个距离以外的光施加作用,从而使光线偏转……吗?"

他虽然倾向于肯定的答案,但似乎从未计算过偏转角,更没有表示过用观测进行验证的企图。一个世纪后的1804年,后来担任慕尼黑天文台台长的**索尔德纳**(Soldner)把这一想法具体化,用牛顿引力论求得星光在擦过太阳时的偏角为0.875″=1.75″/2,即恰为广义相对论预言之半①。(文章在1801年完成,1804年发表。)因此,日食观测的任务不仅是证实星光确有偏转,还要在广义相对论与牛顿引力论之间做一决断。糟糕的是,由于种种原因,英国人的1919年日食资料和归算处理存在诸多疑点。在后续的许多年中,天文学家几乎逢日全食必观测,可是结果起伏不定,难以定案。德国人在重新分析英国队的资料时发现,用不同方式处理这份资料会导致很不相同的结论。例如,只要删除成像特别差的个别恒星,结果就会大为改变。他们还发现英国人的归算方式总是使偏角变大,从而有利于广义相对论。直至20世纪60年代初,当多数天文学家逐渐相信偏转角更有利于广义相对论(而不是牛顿引力论)时又冒出了一种崭新的引力

① 开文迪许(Cavendish)早在1784年就曾做过计算(但计算方法已经无据可查),只是没有发表。他的结果与索尔德纳的结果只在一阶近似上相等,对太阳都是0.875″。

理论,即§6.3末段提及的布-迪理论,是广义相对论的最为强有力的竞争对手,它的挑战大大提高了对天文观测精度的要求。与广义相对论不同,布-迪理论中有一个可以调节的常参数 ω。“可以调节”是指可以人为选定任一实数作为 ω 值;“常参数”则是指一旦选定就不再变。所以可以说存在着无数种布-迪理论(一个 ω 值决定一种)。当 $\omega=\infty$ 时布-迪理论就回到广义相对论。可调参数 ω 的存在使布-迪理论比广义相对论灵活得多,因为可以根据天文观测数据确定 $|\omega|$ 的下限。事实上,这一下限随着观测的进步而不断提高,在1973年是 $|\omega|>5$,到1979年升至 $|\omega|>500$[见 Reasenberg et al.(1979)],而到了2003年竟达 $|\omega|>40\ 000$[见 Bertotti et al.(2003)]。这说明观测不断朝着有利于广义相对论的方向发展。不过这里的“观测”早已不是指单纯的日食观测。由于固有的系统误差以及诸多不利因素①,靠日全食测定偏角的误差一直未能越过0.1这道关卡。于是有人想到非可见光。**夏皮罗**(Shapiro)于1967年率先提出把观测对象从可见光源转移至射电天体的倡议。他还指出,传统的雷达接收法对观测如此小的偏角无能为力,建议采用由两个接收器组成的射电干涉仪,而且,因为角度测量误差与接收器之间的距离成反比,干涉仪的基线越长越好。从可见光观测发展为射电观测有许多好处:为了观测日食,天文学家不但要耐心等待日全食的时机,还必须带着笨重仪器匆忙赶到遥远的、人生地不熟的观测地点;而射电观测却可以在天文台里用最好的仪器按照周密的既定计划行事,占尽天时、地利、人和的优势。所以,这一建议一经提出就拉开了用非可见光(非日食法)观测星光偏角的序幕,精度不断提高的观测结果相继涌现。这里要特别强调的是观测精度,或说标准误差。如果观测结果竟然是 $1.75''\pm1.75''$,你不应在看到第一个数字1.75″时高兴过早,因为第二个数字 $\pm1.75''$ 是标准误差,在最坏的情况下(取-号时)意味着偏角为零($1.75''-1.75''=0$)! 对射电光线偏角观测的标准误差在1967年后创造了 10^{-2},10^{-3} 乃至 10^{-4} 的记录。除了用甚长基线干涉仪对射电源(如类星体)的观测外,又发展出一种对多年积累的大量资料进行统计分析以得出结论的巧方法[见 Shapiro et al.(2004)]:被利用的数据来自“大地测量甚长基线干涉仪资料库”(该测量的原始目的是对涉及地球运动的各个方面进行全天候监测),他们采用的数据包括1979—1999年间所做的近200万次观测的记录,涉及87个甚长基线干涉仪天文台(最长基线达1万km)和541个射电源。通过对这些数据的统计分析,他们得出的结论等于广义相对

① 首先,通过对比日食底片与数月后的比较底片得出偏角的做法就充满系统误差。其次,(透过大气层看星光的)光学畸变、温度影响(例如底板的冷缩热胀)以及气候干扰(例如沙尘弥漫会严重削弱星光透过率)都会带来误差。

论预言值的 0.999 83 ±0.000 45,标准误差已降至 4.5×10^{-4}。

8.5.3 牛顿引力论对星光偏角的推导[选读]

索尔德纳在 1801 年推导星光偏角公式时,光速的有限性已被证实。只要承认光的微粒说并默认光微粒(即现在的光子)在引力场中的表现与普通质点的唯一不同就是以光速运动,就不难用牛顿引力论推出偏角公式。此式的推导方法很多,现在介绍一种"借他山之石"的简便方法。

首先考虑质量为 m 的普通质点在太阳附近飞过时因受太阳引力而出现的偏角。这种情况与卢瑟福 1911 年研究的 α 粒子散射类似,因此可以借用其有关公式。α 粒子在原子核库仑力的作用下发生偏转,以 l 代表核与入射方向的垂直距离,则偏角 β 满足[推导见胡镜寰等(1989)P.7~8]

$$\tan\frac{\beta}{2}=\frac{D}{2l}, \tag{8-5-9}$$

式中的 D 由下式决定:

$$D=\frac{1}{4\pi\varepsilon_0}\frac{2Ze^2}{mv_0^2/2}, \tag{8-5-10}$$

其中 e 为基本电荷,Z 为核的原子序数,m 和 v_0 分别是 α 粒子的质量和初速。α 粒子是氦原子核,电荷为 $2e$,将上式的库仑斥力 $2Ze^2/4\pi\varepsilon_0r^2$ 改为太阳对质点 m 的万有引力 GMm/r^2(其中 M 为太阳质量),则式(8-5-10)成为

$$D=\frac{2GM}{v_0^2}, \tag{8-5-11}$$

由于星光偏角 β 很小,式(8-5-9)左边可近似写为

$$\tan\frac{\beta}{2}=\frac{\sin(\beta/2)}{\cos(\beta/2)}\approx\sin(\beta/2)\approx\frac{\beta}{2},$$

与式(8-5-11)及式(8-5-10)联立得

$$\beta\approx\frac{2GM}{lv_0^2}。 \tag{8-5-12}$$

由上式可得重要结论:偏角 β 与入射质点的质量 m 无关。于是,只要默认光子在引力场中的表现与普通质点相同,将上式的 v_0 改为光速 c 便得

$$\beta\approx\frac{2GM}{c^2l}。 \tag{8-5-13}$$

由此也可以猜测牛顿当年没有给出光线偏角的一个可能原因——当时的牛顿对光速的有限性尚不能肯定。假若光速无限,偏角势必为零。反之,索尔德纳所在时代早已测得有限的光速,所以不难推出偏角公式。

将式(8-5-13)与式(8-5-8′)对比可知,牛顿引力论给出的偏角正好是

广义相对论的预言值之半。

下面介绍一个有趣的历史故事。爱因斯坦在1905年发表狭义相对论后不久就致力于考虑引力问题,终于在1915年创立了广义相对论。在1911年发表的、题为“论引力对光传播的影响”的文章中[爱因斯坦等,《相对论原理》(1980中译本)],他用一种特别手法推出了与式(8-5-13)完全相同的偏角公式,由此求得的太阳边缘光线偏角与索尔德纳的结果相同,都是广义相对论预言值之半,即0.875″=1.75″/2。(限于当时的天文数据,他求得0.83″)。“特别手法”是指下列几步:① 把只对力学现象适用的等效原理假设性地推广为适用于一切物理现象的原理,即后来所称的爱因斯坦等效原理(EEP);② 用EEP证明光(辐射)也受引力作用;③ 在此基础上找出光在引力场中的坐标速度与引力势的关系,再同惠更斯原理结合便求得星光偏角。应该特别注意,该文发表于广义相对论诞生之前,他对引力的处理只能依据牛顿引力论,所得偏角不同于广义相对论是不足为奇的。

爱因斯坦在1911年的那篇文章结束时说:

“由于日全食时,太阳近旁这部分天空的恒星是看得见的,我们的理论结果就可以同实验比较。……假如天文学家愿意着手考察在此提出的问题,那是再好没有了。”

这一建议很快就受到响应,但是,由于观测地点的气候问题以及第一次世界大战的爆发(1914年德国对俄宣战,俄兵把在俄观测日食的德国天文学家扣为战俘),日食观测无果而终,爱因斯坦深感遗憾。不过这也许是件好事,因为这给爱因斯坦留出时间修正他1911年的结果。他在1915年关于广义相对论的文章[爱因斯坦等(1980)]中修正了这一错误,求得太阳边缘光线偏角为1.7″。

广义相对论的问世使情况发生根本性改观。虽然爱因斯坦是旷世奇才,但他在广义相对论尚未成型的1911年的文章中的某些讲法并不正确,一切应以广义相对论的讲法为准,而且最好用4维语言讲解。我们不必再去论证光子也受引力,因为广义相对论把光子必受引力暗含在“光子世界线是类光测地线”这一基本假设之中(走测地线就表明光子受且只受引力),由此便可按小节8.4.1顺理成章地推出星光偏角公式。何乐而不为也?

8.5.4　雷达回波的时间延缓[选读]

作为一种物理理论,广义相对论当然应该经受实验检验;作为一种关于引力的理论,广义相对论尤其需要更多更精确的实验检验,因为物理学中存在着许许多多的引力理论(本书涉及的就有牛顿引力论、广义相对论和布-迪理论三种),亟待实验决定取舍。引力理论分为度规理论和非度规理论两大类(见

§6.3)。讨论表明[见 Will(1993),Will(2001)],如果 EEP(爱因斯坦等效原理)成立,则只有度规理论才可能正确。由于 EEP 越来越取得实验的支持,非度规理论已经越来越没有立锥之地。牛顿引力论可以用作度规理论在弱场低速条件下的一级近似(最粗略的近似),它可以解释太阳系的大多数观测结果,但在涉及精度很高的实验(如水星近日点进动)时就无能为力。于是人们提出了"后牛顿近似"(post - Newtonian approximation)的想法(是比牛顿引力论更接近广义相对论的一种近似),并构造出一个可在弱场低速条件下比较各种度规理论以及将它们同实验做对比的理论框架。这个框架涉及十个参数,在不同的度规理论中可取不同数值,我们只谈其中之一,即参数 γ。对广义相对论,$\gamma=1$;对牛顿引力论,$\gamma=0$;对布 - 迪理论,γ 取决于其可调节参数 ω,两者关系为

$$\gamma=\frac{1+\omega}{2+\omega}。\tag{8-5-14}$$

牛顿引力论和各种度规理论对星光偏角的预言值之所以不同,就在于它们有不同的 γ 值。太阳系中各种观测的一个重要目的就是要测定参数 γ。可以证明,星光偏角在几何单位制中服从如下公式:

$$\beta\approx 2(1+\gamma)M/l,\tag{8-5-15}$$

其中 M 和 l 的含义与式(8-5-8)相同。对广义相对论,$\gamma=1$,上式回到式(8-5-8)。

1919 年的日全食观测可以看作测量 γ 值的第一个太阳系实验。为了从根本上提高测量精度,又从可见光线偏角观测发展为射电偏角观测。夏皮罗等的文章[Shapiro et al.(2004)]宣称用这种方法测得 $\gamma=0.999\,83\pm0.000\,45$,标准误差已降至 4.5×10^{-4}。

光线(包括非可见光)偏转现象还有一个一母同胞的孪生姐妹,那就是**雷达回波的时间延缓**效应。根据广义相对论("一母同胞"的"母"),太阳的存在使空间变得弯曲[见式(8-1-3)],光子在空间中要走较远的路径,因而要用较长的时间。假定从地球发出的雷达信号到达另一行星(或飞船)、被它反射并沿原路返回地球(图 8-9),则由广义相对论求得的信号来回时间 Δt 比用牛顿引力论(认为空间平直)求得的时间 $(\Delta t)_{牛}$ 要长一点点,即

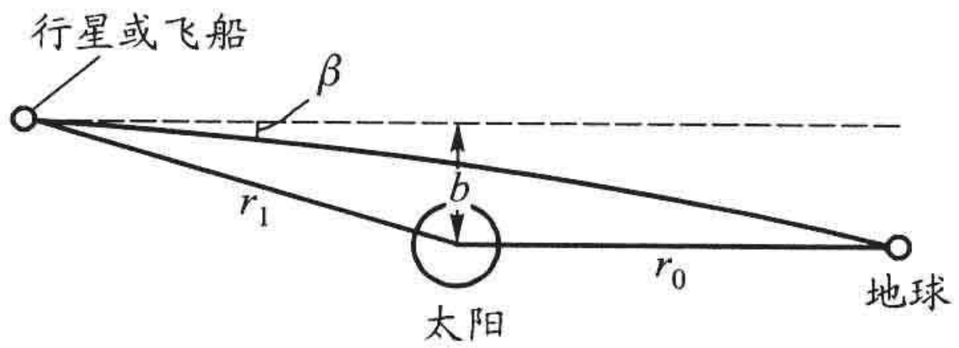


图 8-9 雷达回波的时间延缓

$$\Delta t = (\Delta t)_{牛} + \delta t,\ \delta t > 0,\ \delta t << (\Delta t)_{牛}。 \qquad (8-5-16)$$

这种回波时间延缓效应通常也被称为广义相对论的第四个经典验证。当地球、太阳和行星(或飞船)连成一条直线而且太阳在两者之间时,延缓时间 δt 最长,可以表为

$$\delta t = 2(1+\gamma) M \ln(4 r_0 r_1 / b^2), \qquad (8-5-17)$$

其中 r_0、r_1 及 $b(<<r_0, r_1)$ 如图 8-9 所示。虽然 $\delta t << (\Delta t)_{牛}$,但是 δt 对 b 的依赖是对数关系,对数函数的特有性质使得人们可以从测得的 Δt 中把极其微小的 δt 从很大的"背景"$(\Delta t)_{牛}$ 中识别出来。

夏皮罗于 1964 年首次从理论(广义相对论)上提出了上述时间延缓效应并建议用天文观测加以证实。后来的几次实验测得的 γ 值都非常有利于广义相对论,精度也在不断提高。例如,1979 年利用海盗号火星登陆飞船(Viking lander)测得从地球到飞船(火星)的回波延缓时间,由此求得 $\gamma = 1.000 \pm 0.002$,精度提高至 2×10^{-3},再用式(8-5-14)便可把布-迪参数 ω 限制在 $|\omega| > 500$。可惜这种方法的测量精度在以后的二十余年停步不前,直至 2003 年才因为技术上的改进而出现重大突破。这次突破的关键原因是实验者把测量延缓时间改为测量频移。虽然射电波在地球与飞船之间的往返时间很短,但地球和飞船在这段时间内的运动也会带来射电波的多普勒频移。事实上,太阳引力场的存在对光的传播就有引力频移效应,加上由地球和飞船运动造成的多普勒频移,合称引力多普勒频移(gravitational Doppler shift)。地球和飞船的运动通过改变式(8-5-17)中的 b 而改变 δt,由此可推出回波与发射波的相对频移 $(\nu' - \nu)/\nu$ 的公式:

$$\frac{\nu' - \nu}{\nu} = 4(1+\gamma)\frac{M}{b}\frac{\mathrm{d}b}{\mathrm{d}t}。 \qquad (8-5-18)$$

利用上式就可以通过测量频移而测定 γ。与测量时间延缓相比,这种用测量频移来测定 γ 的方法可把精度提高一个量级,但过去由于日冕造成的干扰,频移无法测准。2002 年发射的、探测土星的卡西尼飞船(Cassini spacecraft)由于装备有最先进的射电系统而可以克服这一困难,使准确测量频移成为可能。实验在 2002 年 6 月 6 日至 7 月 7 日进行,其中的 6 月 21 日最为特殊,那时地球、太阳和卡西尼飞船在一条直线上(地球与飞船分居太阳两侧)。根据实验者发表的文章[Bertotti et al. (2003)],实验测得

$$\gamma = 1 + (2.1 \pm 2.3) \times 10^{-5},$$

以 2×10^{-5} 的高精度验证了广义相对论。借式(8-5-14)可以推知布-迪理论的参数的下限高达 $|\omega| > 40\ 000$。

广义相对论从 1915 年诞生至今已是近百岁的"老人"。虽然存在着不可胜数的引力理论与它竞争,而且它又没有任何参数可供调节以适应实验数据(大

有“宁折不弯”的味道),但竟然能在近百年来精度越来越高的大量实验的考验下屡战屡捷,长寿不衰,实在难能可贵。(反之,许多引力理论已经先后“中箭落马”。)然而,越是如此,人们越是跃跃欲试地设计出更多的实验来考验它,并很想知道当精度提高到哪一量级时它会与实验有所偏离。例如,欧洲的大型计划项目 GAIA 在众多的探测任务中就有检验广义相对论的一项,并且有望将检验精度再提高一个量级(将标准误差降至10^{-6}),这个历时约十年的项目计划于2013年开始执行。(GAIA 是 Global Astrometric Interferometer For Astrophysics 的缩写,汉语译为“全天天体测量干涉仪”。)

§8.6 相对论在GPS(全球定位系统)中的应用

相对论在许多人的心目中有如下特点:① 新奇有趣;② 难度颇高;③ 毫不实用。下面谈谈特点③。狭义相对论在日常生活中毫不实用的关键原因是日常接触到的速率都太低。一旦涉足高能物理学领域,由于高能粒子通常都有很高的速率,狭义相对论就大有用武之地,而且变得不可或缺。然而,与物理学的其他分支相比,狭义相对论在国民经济和日常生活中的应用似乎还是非常罕见。至于广义相对论,虽然它在近代天体物理学(特别是黑洞和宇宙论)中必不可少,但过去一直认为它与国计民生以及日常生活毫无关联。有趣的是,有一件与日常生活息息相关的事情逐渐引起了公众的注意,这就是狭义和广义相对论在GPS中的不可替代的应用。无论你坐飞机还是开汽车,为了定时和定位,你都离不开 GPS,它是 Global Positioning System 的缩写,中译名为**全球定位系统**或**全球卫星定位系统**。这是美国最初出于独霸全球的军事目的而研制的、耗资300亿美元、历时20余载、于1994年建成的新一代卫星导航和定位系统,后来又不断改进,并迅速发展成为一个蓬勃兴旺的商业性产业。除了明显的军事目的(如导弹制导)之外,GPS 的应用实在是不可胜数,以下仅是部分例子:① 陆地应用,如车辆导航、大气观测、地球资源勘探、桥梁和隧道测量;② 海洋应用,如最佳航线测定、船只导航、海洋救援和探宝、海平面升降监测;③ 航空航天应用,如飞机导航、低轨卫星定轨、航空救援、载人航天器防护探测等等。总之,GPS 是一个全球性、高精度、全天候的定位系统,其应用领域早已遍及国民经济的各个部门,而且逐渐深入每个人的日常生活。你只要拥有一台用户接收机就可享受免费服务,为自己确定所在地的纬度、经度、高度以及当时的国际标准时刻。

我国也在独立自主地研制开发类似的导航系统,称为**北斗卫星导航系统**,它能提供高精度、高可靠的定位、导航和授时服务,具有导航和通信相结合的服务特色。通过19年的发展,这一系统在测绘、渔业、交通运输、电信、水利、森林防

火、减灾救灾和国家安全等诸多领域得到应用，产生了显著的经济效益和社会效益，特别是在四川汶川、青海玉树抗震救灾中发挥了非常重要的作用。北斗卫星导航系统将在2020年形成全球覆盖能力。

下面仅以GPS为例进行介绍。

作为一个系统，GPS包括空间设备(space segment)、控制设备(control segment)和用户设备(user segment)三个部分。

空间设备由携带原子钟的24个人造卫星组成，它们分布在6个轨道平面上以近圆形轨道绕地球公转(卫星高度 $h=20\times10^6$ m，轨道周期约12 h)，使得地面上几乎每一点的上空都有至少4个卫星。

控制设备位于地面，由1个主控站、 5个监测站和3个控制站组成。监测站负责把不断搜集到的卫星和大气数据发送给主控站，主控站由此算出卫星的轨道和时钟参数，并将结果发送给3个控制站，后者在每一卫星行至上空时将这些数据和指令注入卫星。

用户设备就是GPS接收机(导航仪)，又可分为车载式、船载式、机载式、星载式、弹载式和手持式等，都具有对GPS信号的接收、发射和解算功能。

以下侧重从相对论角度讲解GPS的工作原理。

每个卫星都不断发射无线电波(电磁信号)，其传播会受到地球引力场的影响。幸好这个引力场很弱，暂时忽略由它引起的时空弯曲，亦即默认时空平直，就可选惯性坐标系$\{t,x,y,z\}$把时空线元表为如下的简单形式：

$$\mathrm{d}s^2=-c^2\mathrm{d}t^2+\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2,\tag{8-6-1}$$

因为光子世界线是类光曲线，把上式用于该曲线的任一元段便得

$$0=\mathrm{d}s^2=-c^2\mathrm{d}t^2+\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2,$$

由此得

$$\frac{\sqrt{\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2}}{\mathrm{d}t}=c。$$

说明光在$\{t,x,y,z\}$系中的坐标速率等于c。

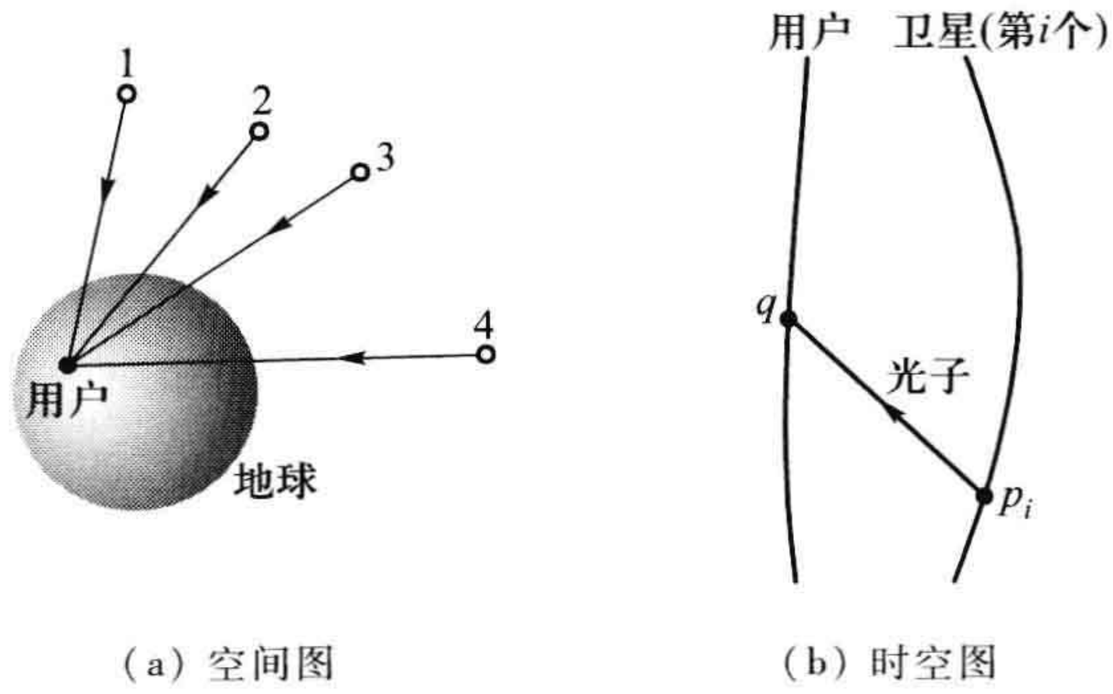
设地面用户在某一时刻同时接收到4个卫星的信号，图8-10(a)和(b)分别是相应的空间图和时空图。在图8-10(b)中，q和$p_i(i=1,2,3,4)$分别代表收信事件和发信事件，它们的惯性坐标分别为t,x,y,z和t_i,x_i,y_i,z_i，即

$$q=(t,x,y,z),\quad p_i=(t_i,x_i,y_i,z_i),\quad i=1,2,3,4,$$

便可列出以下4个方程[称为**传播延迟方程**(propagation delay equation)]：

$$c(t-t_i)=\sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2},\quad i=1,2,3,4。\tag{8-6-2}$$

卫星所发信号中含有发射时刻t_i以及该时刻的卫星位置x_i,y_i,z_i等信息，用户接收机的内置电脑根据这些已知数据(t_i,x_i,y_i,z_i)就可从上列4个方程中



(a) 空间图　　(b) 时空图

图 8 - 10　用户同时收到 4 个卫星的电磁信号

解出 4 个未知数 t,x,y,z(用户收信时的时空坐标),从而实现定时和定位。

稍有力学常识的人都能接受式(8 -6 -2),从而理解 GPS 的基本原理。然而,由于对 GPS 的精度要求很高,许多附加问题都不得不考虑。原子钟走时率的精度是足够高的,但其表观走时率却会因为相对论效应而受到影响,如果对此不加修正,GPS 所给出的位置误差可达每天 11 km 的程度,这种误差的积累效应将使 GPS 很快就变得毫无用处。GPS 的相对论效应主要表现为两个方面,分述如下。

效应 1　狭义相对论的动钟变慢效应

地球上的时间以国际原子时间(International Atomic Time)为标准,它由静止于海平面的原子钟给出。以下称此为**海面钟**。相对于地心而言,卫星钟以速率 $u_{卫}=3.84\times10^{3}$ m/s 做高速运动。海面钟虽然也随地球自转而运动,但即使是赤道上的海面钟的速率也不过 $u_{赤}=465$ m/s(下面就以赤道钟作为海面钟的代表)。由于狭义相对论的动钟变慢效应,相对于地心钟而言,卫星钟的变慢比赤道钟要严重得多。既然以赤道钟为时间标准,就应对卫星钟进行修正。地心钟的世界线可画成竖直线,赤道钟和卫星钟的世界线都是螺旋线,但在示意图中不妨画成斜率不同的斜直线(图 8 - 11)。以 $\sum_{t_1}$ 和 $\sum_{t_2}$ 代表两个相邻的同时面(等 t 面),令 $\Delta t\equiv t_2-t_1$,以 $\Delta\tau_{心}$、$\Delta\tau_{赤}$ 和 $\Delta\tau_{卫}$ 依次代表三个钟在这两张同时面之间经历的固有时间,则由狭义相对论的动钟变慢效应可知[见式(4 -2 -1)]

$$\Delta\tau_{卫}=\gamma_{卫}^{-1}\Delta t,\ \Delta\tau_{赤}=\gamma_{赤}^{-1}\Delta t,\ \Delta\tau_{心}=\Delta t,\qquad(8-6-3)$$

其中

$$\gamma_{卫}\equiv[1-(u_{卫}/c)^2]^{-1/2},\ \gamma_{赤}\equiv[1-(u_{赤}/c)^2]^{-1/2}。\qquad(8-6-4)$$

因为 $(u_{卫}/c)^2\ll1$,$(u_{赤}/c)^2\ll1$,上式可用牛顿二项式定理

$$(1-x)^{-1/2}\approx1+x/2\qquad(8-6-5)$$

简化为

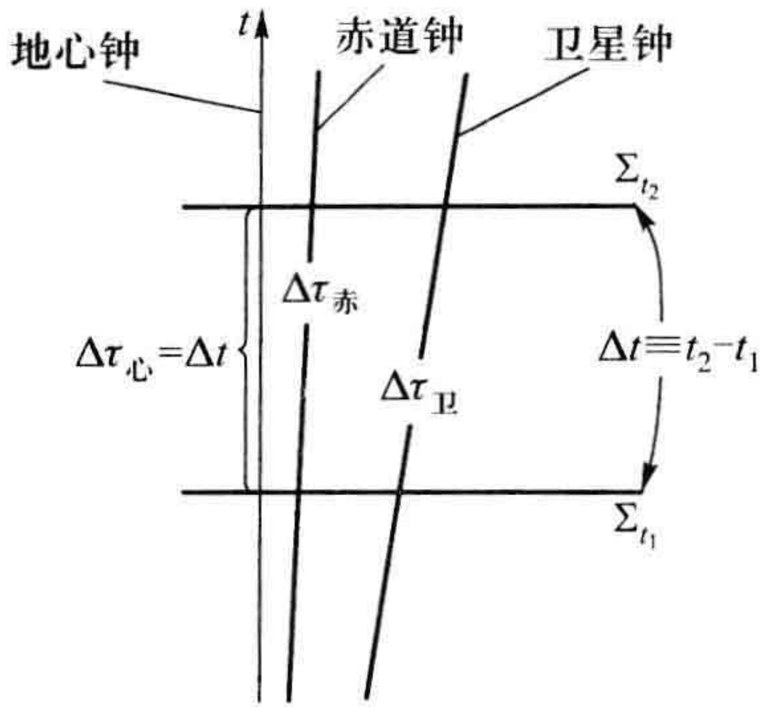


图 8－11　地心钟、赤道钟和卫星钟世界线示意

$$\gamma_{卫}=1+(u_{卫}/c)^2/2,\ \gamma_{赤}=1+(u_{赤}/c)^2/2,$$

而式(8－6－3)给出$\dfrac{\Delta\tau_{卫}}{\Delta\tau_{赤}}=\dfrac{\gamma_{赤}}{\gamma_{卫}}$,因而

$$\frac{\Delta\tau_{赤}-\Delta\tau_{卫}}{\Delta\tau_{赤}}=\frac{\gamma_{卫}-\gamma_{赤}}{\gamma_{卫}}\approx\frac{[(u_{卫}/c)^2-(u_{赤}/c)^2]/2}{1+(u_{卫}/c)^2/2}\approx\frac{1}{2}\left[\left(\frac{u_{卫}}{c}\right)^2-\left(\frac{u_{赤}}{c}\right)^2\right]=$$

$$\frac{1}{2}\left[\left(\frac{3.84\times10^3}{3\times10^8}\right)^2-\left(\frac{465}{3\times10^8}\right)^2\right]=\frac{1}{2}(1.64\times10^{-10}-0.024\times10^{-10})=0.81\times10^{-10}。$$

如果取

$$\Delta\tau_{赤}=1\ 天=24\times3\ 600\ \mathrm{s}=8.64\times10^{10}\ \mu\mathrm{s},\qquad(8-6-6)$$

则

$$\Delta\tau_{赤}-\Delta\tau_{卫}=(0.81\times10^{-10})\times8.64\times10^{10}\ \mu\mathrm{s}=7\mu\mathrm{s}。\qquad(8-6-7)$$

这就表明,由于速率较大,卫星钟比赤道上的海面参考钟每天要慢 7μs。

效应 2　广义相对论的引力钟慢效应

根据广义相对论,引力较强处的钟较慢。卫星钟位于高空,所受地球引力比赤道钟较弱,故赤道钟应较慢。下面用式(8－3－9′)计算两者的读数差。

令

$$\varepsilon_{卫}\equiv\frac{GM}{c^2r_{卫}},\ \varepsilon_{赤}\equiv\frac{GM}{c^2R},\qquad(8-6-8)$$

其中 G 和 M 分别代表引力常量和地球质量,$r_{卫}$ 代表卫星与地心的距离,R 代表海平面与地心的距离,亦即地球半径。由式(8－3－9′)得

$$\frac{\Delta\tau_{卫}}{\Delta\tau_{赤}}=\sqrt{\frac{1-2\varepsilon_{卫}}{1-2\varepsilon_{赤}}}=(1-2\varepsilon_{卫})^{1/2}(1-2\varepsilon_{赤})^{-1/2}\approx$$

$$(1-\varepsilon_{卫})(1+\varepsilon_{赤})\approx1+\varepsilon_{赤}-\varepsilon_{卫},\qquad(8-6-9)$$

[其中第三步用到式(8-6-5)及类似公式$(1-x)^{1/2}\approx 1-x/2$。] 于是

$$\frac{\Delta\tau_{卫}-\Delta\tau_{赤}}{\Delta\tau_{赤}}=\varepsilon_{赤}-\varepsilon_{卫}=\frac{GM}{c^2}\left(\frac{1}{R}-\frac{1}{r_{卫}}\right)=\frac{GM}{c^2}\frac{r_{卫}-R}{Rr_{卫}}。\quad (8-6-10)$$

利用地球数据和物理常数(均指各量在国际单位制中的数值)

$$M=5.977\times10^{24},\ G=6.67\times10^{-11},\ c=3\times10^{8},\quad (8-6-11)$$

求得

$$\frac{GM}{c^2}=\frac{(6.67\times10^{-11})\times(5.977\times10^{24})}{(3\times10^{8})^2}\approx4.43\times10^{-3}(单位为\ m)。\quad (8-6-12)$$

又已知

$$R\equiv地球半径=6.4\times10^{6}\ m,\ h\equiv卫星高度=20\times10^{6}\ m,$$

$$r_{卫}\equiv卫星与地心距离=R+h=26.4\times10^{6}\ m,$$

代入式(8-6-10)得

$$\frac{\Delta\tau_{卫}-\Delta\tau_{赤}}{\Delta\tau_{赤}}=(4.43\times10^{-3}\ m)\times\frac{20\times10^{6}\ m}{(6.4\times10^{6}\ m)\times(26.4\times10^{6}\ m)}=5.2\times10^{-10}。\quad (8-6-13)$$

仍取 $\Delta\tau_{赤}=1$ 天 $=8.64\times10^{10}\ \mu s$,便得

$$\Delta\tau_{卫}-\Delta\tau_{赤}=(5.2\times10^{-10})\times(8.64\times10^{10}\ \mu s)\approx45\mu s。\quad (8-6-14)$$

这就表明,由于处在较强的引力场中,海面参考钟比卫星钟每天要慢 45μs。刚才已求得狭义相对论效应使卫星钟比海面钟每天慢 7μs,所以相对论的总效应是卫星钟比海面参考钟每天快(45-7)μs=38μs。注意到光速 $c=3\times10^{8}$ m/s,不难相信这 38μs 导致的空间距离误差为

$$\delta l=(3\times10^{8}\ m/s)\times(38\times10^{-6}\ s)=11.4\ km。$$

就是说,即使 GPS 对用户距离的测定在今天中午 12 点时完全准确,那么明天中午 12 点时的误差(错报)竟达到 11.4km! 当今 GPS 定位的精度可达 10 m,这每天 11.4 km 的误差岂能容忍? 因此,GPS 内部必须包含对这种误差的修正机制。

以上简介了 GPS 中两个最主要的相对论修正,分别称为狭义相对论修正(效应 1)和广义相对论修正(效应 2)。效应 2 的根源是地球引力场,它相应的时空弯曲导致引力钟慢效应。但是地球引力场毕竟很弱,时空弯曲甚为轻微,所以在讨论效应 1 时就近似认为时空平直,并且直接借用狭义相对论的定量结论。这至少从物理直觉上是可以接受的。你当然还可以问:为什么对两种效应简单求代数和就能得到总效应(38 μs/天)? 下面的选读给出一个从广义相对论角度(从地球附近的弯曲线元出发)的完整讨论,并将推出一个集两种效应于一身的优雅公式(8-6-46)(由此可知总效应的确就是两种效应的代数和),此外还将讨论一些更为深刻的问题。

[选读8-2]

本选读准备就相对论在GPS中的应用问题进行更为深入和定量的讨论，特别强调如何从相对论的角度考虑问题。

地球附近的时空弯曲可近似用史瓦西线元描述：

$$ds^2 = -\left(1-\frac{2M}{\hat{r}}\right)d\hat{t}^2 + \left(1-\frac{2M}{\hat{r}}\right)^{-1} d\hat{r}^2 + \hat{r}^2(d\theta^2 + \sin^2\theta d\varphi^2), \tag{8-6-15}$$

其中M代表地球质量。对上式要做四点说明。

(1) 因为后面要涉及若干个坐标系，为使最有用的坐标系有最简单的坐标记号，即t, r, θ, φ，宁愿将此处的史瓦西坐标写成$\hat{t}$, $\hat{r}$, θ, φ。

(2) 由于地球有自转(非静态)，其外部时空几何与史瓦西线元略有偏离。幸好，量级估算表明，在GPS要求的精度内完全可用史瓦西线元讨论。

(3) 虽然地球自转对外部时空几何的影响可以忽略，但海面钟随地球自转的运动速率必须考虑。选择坐标系时要格外留意，上式的空间坐标系$\{\hat{r}, \theta, \varphi\}$必须是不随地球自转的。

(4) 上式是史瓦西线元在几何单位制的表达式。为便于定量计算，下面改用国际单位制表达式(其中G为引力常量，c为真空光速)

$$ds^2 = -\left(1-\frac{2GM}{c^2\hat{r}}\right)(cd\hat{t})^2 + \left(1-\frac{2GM}{c^2\hat{r}}\right)^{-1} d\hat{r}^2 + \hat{r}(d\theta^2 + \sin^2\theta d\varphi^2)。 \tag{8-6-15'}$$

用下式引入新的径向坐标r：

$$\hat{r} = \left(1+\frac{GM}{2c^2 r}\right)^2 r, \tag{8-6-16}$$

则线元(8-6-15′)在新坐标系$\{\hat{t}, r, \theta, \varphi\}$中取如下形式：

$$ds^2 = -\left(\frac{1-\dfrac{GM}{2c^2 r}}{1+\dfrac{GM}{2c^2 r}}\right)^2 (cd\hat{t})^2 + \left(1+\frac{GM}{2c^2 r}\right)^4 (dr^2 + r^2 d\theta^2 + r^2 \sin^2\theta d\varphi^2)。 \tag{8-6-17}$$

仿照牛顿引力势的概念，用下式定义**引力势** $\Phi(r)$：

$$\Phi(r) \equiv -\frac{GM}{r}, \tag{8-6-18}$$

则

$$ds^2 = -\left(\frac{1+\frac{\Phi}{2c^2}}{1-\frac{\Phi}{2c^2}}\right)^2 (c\mathrm{d}\hat{t})^2 + \left(1-\frac{\Phi}{2c^2}\right)^4 (\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\varphi^2)。$$

(8-6-19)

式(8-6-18)表明引力势$|\Phi(r)|$与 r 成反比,故地面的$|\Phi(R)|$大于空中的$|\Phi(r)|$。由数字计算可知[见式(8-3-23)]

$$\frac{|\Phi(R)|}{c^2} = \frac{GM}{c^2R} = \frac{(6.7\times10^{-11})\times(6\times10^{24})}{(3\times10^8)^2\times(6.4\times10^6)} \approx 7\times10^{-10} << 1,$$

(8-6-20)

故可用牛顿二项式定理

$$(1-x)^{-1} \approx 1+x \qquad (8-6-21)$$

求得

$$\frac{1+\frac{\Phi}{2c^2}}{1-\frac{\Phi}{2c^2}} \approx \left(1+\frac{\Phi}{2c^2}\right)\left(1+\frac{\Phi}{2c^2}\right) \approx 1+\frac{\Phi}{c^2}, \quad \left(\frac{1+\frac{\Phi}{2c^2}}{1-\frac{\Phi}{2c^2}}\right)^2 \approx \left(1+\frac{\Phi}{c^2}\right)^2 \approx 1+\frac{2\Phi}{c^2},$$

以及

$$\left(1-\frac{\Phi}{2c^2}\right)^4 = \left[\left(1-\frac{\Phi}{2c^2}\right)^2\right]^2 \approx \left(1-\frac{\Phi}{c^2}\right)^2 \approx 1-\frac{2\Phi}{c^2},$$

于是式(8-6-19)近似简化为

$$ds^2 = -\left(1+\frac{2\Phi}{c^2}\right)(c\mathrm{d}\hat{t})^2 + \left(1-\frac{2\Phi}{c^2}\right)(\mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\varphi^2)。$$

(8-6-22)

上式是史瓦西线元在坐标系$\{\hat{t}, r, \theta, \varphi\}$的近似表达式。设某钟静止于地面上某点 B,则其空间坐标为

$$r = R = 常数,\quad \theta = \theta_B = 常数,\quad \varphi = \varphi_B,$$

故 $\mathrm{d}r=0$, $\mathrm{d}\theta=0$ 而 $\mathrm{d}\varphi\neq0$(该钟因地球自转而做圆周运动),把式(8-6-22)用于该钟的世界线的任一元段得

$$ds^2 = -\left[1+\frac{2\Phi(R)}{c^2}\right](c\mathrm{d}\hat{t})^2 + \left[1-\frac{2\Phi(R)}{c^2}\right]R^2\sin^2\theta_B\mathrm{d}\varphi^2 =$$

$$-\left[1+\frac{2\Phi(R)}{c^2}-\left(1-\frac{2\Phi(R)}{c^2}\right)\frac{(R\sin\theta_B\mathrm{d}\varphi)^2}{(c\mathrm{d}\hat{t})^2}\right](c\mathrm{d}\hat{t})^2。$$

(8-6-23)

令

$$\hat{u} \equiv R\sin\theta_B\frac{\mathrm{d}\varphi}{\mathrm{d}\hat{t}}, \qquad (8-6-24)$$

因 $d\varphi/d\hat{t}$ 是该钟随地球自转而做圆周运动的角速率，故 $\hat{u}$ 就是它的线速率。把式(8-6-24)代入式(8-6-23)得

$$ds^2 = -\left[1+\frac{2\Phi(R)}{c^2}-\frac{\hat{u}^2}{c^2}+\frac{2\Phi(R)}{c^2}\frac{\hat{u}^2}{c^2}\right](cd\hat{t})^2 \approx -\left[1+\frac{2\Phi(R)}{c^2}-\frac{\hat{u}^2}{c^2}\right](cd\hat{t})^2,$$

于是该钟世界线任一元段的固有时间为

$$d\tau = \sqrt{\frac{-ds^2}{c^2}} = \left[1+\frac{2\Phi(R)}{c^2}-\frac{\hat{u}^2}{c^2}\right]^{1/2} d\hat{t} \approx \left[1+\frac{\Phi(R)}{c^2}-\frac{\hat{u}^2}{2c^2}\right] d\hat{t} \neq d\hat{t}。 \tag{8-6-25}$$

[其中第一步用到式(3-2-2′)，末步再次用到 $(1+x)^{1/2}\approx 1+x/2$。] 地球上的时间标准是由海平面上的静止原子钟的固有时间定义的，这称为国际原子时间(International Atomic Time)，而式(8-6-25)表明海平面上静止钟的坐标时间 $d\hat{t}$ 不等于其固有时间 $d\tau$，可见坐标时间 $\hat{t}$ 不符合国际时间标准。假定 GPS 用户从4个卫星信号求解而得自己收信的时刻竟然是坐标时间 $\hat{t}$，那就与国际时间对不上号，用 $\hat{t}$ 值的计算结果将使用户受骗上当。为了解决这一问题，有必要再做一次坐标变换，使新坐标包含地球自转的信息。为此，可以保持 $\hat{t}$, r, θ 不变而把 φ 变为新坐标 $\tilde{\varphi}$，定义为

$$\tilde{\varphi} \equiv \varphi - \Omega\hat{t}, \quad (\Omega \text{ 是地球自转角速率}) \tag{8-6-26}$$

于是线元(8-6-22)在新坐标系 $\{\hat{t}, r, \theta, \tilde{\varphi}\}$ 中取如下形式：

$$\begin{aligned} ds^2 = &-\left[\left(1+\frac{2\Phi}{c^2}\right)-\left(1-\frac{2\Phi}{c^2}\right)\left(\frac{\Omega r\sin\theta}{c}\right)^2\right](cd\hat{t})^2 + 2\left(1-\frac{2\Phi}{c^2}\right)\Omega r^2\sin^2\theta d\tilde{\varphi}d\hat{t} \\ &+\left(1-\frac{2\Phi}{c^2}\right)(dr^2+r^2d\theta^2+r^2\sin^2\theta d\tilde{\varphi}^2) \approx \\ &-\left[1+\frac{2\Phi}{c^2}-\left(\frac{\Omega r\sin\theta}{c}\right)^2\right](cd\hat{t})^2 + 2\left(1-\frac{2\Phi}{c^2}\right)\Omega r^2\sin^2\theta d\tilde{\varphi}d\hat{t} \\ &+\left(1-\frac{2\Phi}{c^2}\right)(dr^2+r^2d\theta^2+r^2\sin^2\theta d\tilde{\varphi}^2)。 \end{aligned} \tag{8-6-27}$$

令

$$\Psi \equiv \Phi - \frac{1}{2}(\Omega r\sin\theta)^2, \tag{8-6-28}$$

则

$$\begin{aligned} ds^2 = &-\left(1+\frac{2\Psi}{c^2}\right)(cd\hat{t})^2 + 2\left(1-\frac{2\Phi}{c^2}\right)\Omega r^2\sin^2\theta d\tilde{\varphi}d\hat{t} + \\ &\left(1-\frac{2\Phi}{c^2}\right)(dr^2+r^2d\theta^2+r^2\sin^2\theta d\tilde{\varphi}^2)。 \end{aligned} \tag{8-6-29}$$

这是史瓦西线元在坐标系 $\{\hat{t}, r, \theta, \tilde{\varphi}\}$ 的近似表达式，其中的 Ψ 由式(8-6-28)定义，由此式得

$$|\Psi| = \frac{GM}{r} + \frac{1}{2}(\Omega r\sin\theta)^2。\qquad (8-6-30)$$

有一件事情很有趣,也很有用。由于自转,地球略呈扁球状——赤道半径较大而两极半径较小,如图 8-12。把海平面的 Ψ 记作 Ψ_0。如果地球是球体,则式(8-6-30)的 $r=R$ 为常数,而 θ 在海平面上不是常数,故 Ψ_0 不是常数。然而扁球状的地球改变了这种情况。设 A, B 是海平面的两点,$\theta_A=\pi/2$(A 在赤道面上)而 $\theta_B\neq\pi/2$(见图 8-12),则 $r_A>r_B$。令 $l\equiv r\sin\theta$,则 $l_A>l_B$。而式(8-6-30)的 $|\Psi|$ 是两项之和,不妨写成

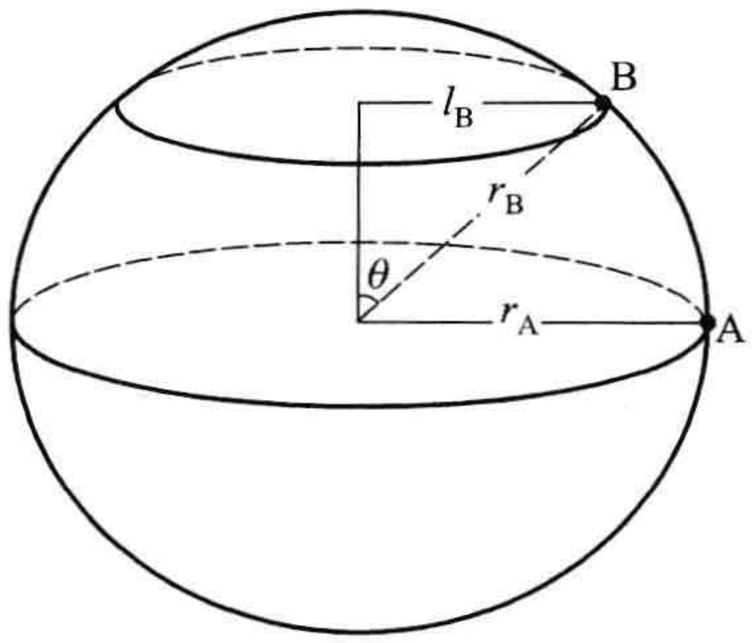


图 8-12　地球略呈扁球状

$$|\Psi| = \mu+\nu,\quad 其中\ \mu\equiv\frac{GM}{r},\quad \nu\equiv\frac{1}{2}(\Omega l)^2。$$

由 $r_A>r_B$ 和 $l_A>l_B$ 可知 $\mu_A<\mu_B$,　$\nu_A>\nu_B$,即 μ, ν 此消彼长。有趣的是,μ, ν 的消长情况竟然如此凑巧,使得在 GPS 所要求的精度内可以认为 $\Psi_A=\Psi_B$,因而可以认为 Ψ_0(海平面的 Ψ)是常数。利用这一常数就可以给我们关心的整个时空区域定义一个新的时间坐标 t:

$$t\equiv\left(1+\frac{\Psi_0}{c^2}\right)\hat{t},\quad 即\ \hat{t}=\left(1+\frac{\Psi_0}{c^2}\right)^{-1}t\approx\left(1-\frac{\Psi_0}{c^2}\right)t。\qquad (8-6-31)$$

(无论是否在海平面,其 t 都用上式定义。)由于 Ψ_0 是常数,上式无非是对时间尺度做一个常数性的伸缩,但很快就会看到这一简单的伸缩将带来巨大的好处。由式(8-6-31)得

$$d\hat{t}^2=\left(1-\frac{\Psi_0}{c^2}\right)^2dt^2=\left(1-\frac{2\Psi_0}{c^2}+\frac{\Psi_0^2}{c^4}\right)dt^2\approx\left(1-\frac{2\Psi_0}{c^2}\right)dt^2,\qquad (8-6-32)$$

故

$$式(8-6-29)右边第一项=-\left(1+\frac{2\Psi}{c^2}\right)c^2\left(1-\frac{2\Psi_0}{c^2}\right)dt^2\approx-\left[1+\frac{2(\Psi-\Psi_0)}{c^2}\right]c^2dt^2,$$

于是史瓦西线元从表达式(8-6-29)变为在 $\{t, r, \theta, \tilde{\varphi}\}$ 系的近似表达式

$$ds^2=-\left[1+\frac{2(\Psi-\Psi_0)}{c^2}\right](cdt)^2+2\left(1-\frac{2\Phi}{c^2}\right)\Omega r^2\sin^2\theta d\tilde{\varphi}d\hat{t}+$$

$$\left(1-\frac{2\Phi}{c^2}\right)(dr^2+r^2d\theta^2+r^2\sin^2\theta d\tilde{\varphi}^2)。\qquad (8-6-33)$$

再次考虑静止在海平面B点的钟，它在新坐标系$\{t, r, \theta, \tilde{\varphi}\}$的坐标自动满足$r$，$\theta$，$\tilde{\varphi}$=常数（因它随地球自转而动），故$\mathrm{d}r=0$，$\mathrm{d}\theta=0$，$\mathrm{d}\tilde{\varphi}=0$，把式(8-6-33)用于其世界线的任一元段得

$$\mathrm{d}\tau=\sqrt{\frac{-\mathrm{d}s^2}{c^2}}=\left[1+\frac{2(\Psi_B-\Psi_0)}{c^2}\right]^{1/2}\mathrm{d}t\approx\left[1+\frac{\Psi_B-\Psi_0}{c^2}\right]\mathrm{d}t=\mathrm{d}t。\tag{8-6-34}$$

（末步是因为B点在海平面保证$\Psi_B=\Psi_0$。）上式说明，海平面上静止钟在$\{t, r, \theta, \tilde{\varphi}\}$系的坐标时间$\mathrm{d}t$等于其固有时间$\mathrm{d}\tau$，而$\mathrm{d}\tau$正是国际时间标准，可见新定义的坐标时间$t$的重大优点：海平面上静止钟的$t$就是国际标准时。文献中称此$t$为**GPS时间**。进一步，为了找出任一个钟（特别是卫星钟）的坐标时间（GPS时间）$\mathrm{d}t$与该钟固有时间$\mathrm{d}\tau$的关系，可以利用式(8-6-31)和式(8-6-32)把史瓦西线元在$\{\hat{t}, r, \theta, \varphi\}$系的表达式(8-6-22)变换为在$\{t, r, \theta, \varphi\}$系的如下近似形式：

$$\mathrm{d}s^2=-\left[1+\frac{2(\Phi-\Psi_0)}{c^2}\right](c\mathrm{d}t)^2+\left(1-\frac{2\Phi}{c^2}\right)(\mathrm{d}r^2+r^2\mathrm{d}\theta^2+r^2\sin^2\theta\mathrm{d}\varphi^2)。\tag{8-6-35}$$

也可改写为

$$\mathrm{d}s^2=-\left[\left(1+\frac{2(\Phi-\Psi_0)}{c^2}\right)-\left(1-\frac{2\Phi}{c^2}\right)\frac{\mathrm{d}r^2+r^2\mathrm{d}\theta^2+r^2\sin^2\theta\mathrm{d}\varphi^2}{(c\mathrm{d}t)^2}\right](c\mathrm{d}t)^2。\tag{8-6-36}$$

令
$$u\equiv\frac{\sqrt{\mathrm{d}r^2+r^2\mathrm{d}\theta^2+r^2\sin^2\theta\mathrm{d}\varphi^2}}{\mathrm{d}t}。\tag{8-6-37}$$

则u是所论钟在$\{t, r, \theta, \varphi\}$系的坐标速率，注意到式(8-6-20)，便有

式(8-6-36)方括号内第二大项$\equiv-\left(1-\frac{2\Phi}{c^2}\right)\frac{\mathrm{d}r^2+r^2\mathrm{d}\theta^2+r^2\sin^2\theta\mathrm{d}\varphi^2}{(c\mathrm{d}t)^2}\approx-\frac{u^2}{c^2}$，

于是式(8-6-36)近似成为

$$\mathrm{d}s^2=-\left[1+\frac{2(\Phi-\Psi_0)}{c^2}-\frac{u^2}{c^2}\right](c\mathrm{d}t)^2,\tag{8-6-38}$$

因而

$$\mathrm{d}\tau=\sqrt{\frac{-\mathrm{d}s^2}{c^2}}=\left[1+\frac{2(\Phi-\Psi_0)}{c^2}-\frac{u^2}{c^2}\right]^{1/2}\mathrm{d}t\approx\left(1+\frac{\Phi-\Psi_0}{c^2}-\frac{u^2}{2c^2}\right)\mathrm{d}t。\tag{8-6-39}$$

也可反表为

$$\mathrm{d}t=\left(1+\frac{\boldsymbol{\Phi}-\boldsymbol{\Psi}_0}{c^2}-\frac{u^2}{2c^2}\right)^{-1}\mathrm{d}\tau\approx\left(1-\frac{\boldsymbol{\Phi}-\boldsymbol{\Psi}_0}{c^2}+\frac{u^2}{2c^2}\right)\mathrm{d}\tau。\qquad(8-6-39')$$

这就是任一钟的 GPS 时间 t 与固有时间 τ 的微分关系。用于卫星钟,以 μ 代表卫星钟世界线上我们关心的一段,由于卫星高度 $r_{卫}$ 及速率 $u_{卫}$ 近似为常数,将上式沿 μ 段积分得

$$\int_\mu \mathrm{d}t=\int_\mu \mathrm{d}\tau_{卫}-\left(\frac{\boldsymbol{\Phi}_{卫}-\boldsymbol{\Psi}_0}{c^2}-\frac{u_{卫}^2}{2c^2}\right)\int_\mu \mathrm{d}\tau_{卫}。\qquad(8-6-40)$$

令　$\Delta t\equiv\int_\mu \mathrm{d}t$，　$\Delta\tau_{卫}\equiv\int_\mu \mathrm{d}\tau_{卫}$，　$(\Delta t)_{修}\equiv-\left(\frac{\boldsymbol{\Phi}_{卫}-\boldsymbol{\Psi}_0}{c^2}-\frac{u_{卫}^2}{2c^2}\right)\Delta\tau_{卫}$，

$$(8-6-41)$$

便有

$$\Delta t=\Delta\tau_{卫}+(\Delta t)_{修},\qquad(8-6-42)$$

卫星经历的 GPS 时间 Δt 之所以不等于固有时间 $\Delta\tau_{卫}$,就是因为上式右边多了修正项 $(\Delta t)_{修}$。现在可以明确指出,传播延迟方程(8-6-2)中的 t 和 t_i 都是 $\{t,r,\theta,\varphi\}$ 系的坐标时(GPS 时),因此,从卫星钟读出固有时间 $\Delta\tau_{卫}$ 后,应先加上 $(\Delta t)_{修}$ 以得到 GPS 时间 Δt,然后发出信号。信号应包含发信时的 GPS 时刻 t_i 以及卫星在此时的空间坐标 r_i，θ_i，φ_i 等内容,以便用户当作已知数求解传播延迟方程(8-6-2),从而得到自己的时空坐标 t，r，θ，φ,达到定时定位的目的。这里的 t 就是用户收信时的 GPS 时刻,它与国际标准一致。

现在把式(8-6-41)的 $(\Delta t)_{修}$ 与选读前的两个修正项(效应 1 和 2)做一对比。由式(8-6-18)知

$$\frac{\boldsymbol{\Phi}_{卫}}{c^2}\equiv-\frac{GM}{c^2}\frac{1}{r_{卫}}。\qquad(8-6-43)$$

常数 $\boldsymbol{\Psi}_0$ 可借赤道上的 A 点求得。由式(8-6-28)和式(8-6-18)得

$$\frac{\boldsymbol{\Psi}_0}{c^2}=\frac{\boldsymbol{\Psi}|_{\mathrm{A}}}{c^2}=\frac{\boldsymbol{\Phi}|_{\mathrm{A}}}{c^2}-\frac{(\Omega R\sin\theta_{\mathrm{A}})^2}{2c^2}=-\frac{GM}{c^2}\frac{1}{R_{赤}}-\frac{(\Omega R)^2}{2c^2}=-\frac{GM}{c^2}\frac{1}{R_{赤}}-\frac{u_{赤}^2}{2c^2},$$

$$(8-6-44)$$

其中 $R_{赤}$ 是赤道半径,$u_{赤}$ 是海平面赤道处的静止钟随地球自转的线速率。把式(8-6-43)、式(8-6-44)代入式(8-6-41)得

$$(\Delta t)_{修}=\left[\frac{GM}{c^2}\left(\frac{1}{r_{卫}}-\frac{1}{R_{赤}}\right)+\frac{u_{卫}^2-u_{赤}^2}{2c^2}\right]\Delta\tau_{卫}。\qquad(8-6-45)$$

仍以 $\Delta\tau_{赤}$ 代表赤道钟在坐标时间 Δt 内经历的固有时间,从以上定量计算中不难看出 $\Delta\tau_{赤}$ 与 $\Delta\tau_{卫}$ 的相对差别为 10^{-10} 的量级,所以式(8-6-45)又可改写为

$$(\Delta t)_{修}=\left[\frac{GM}{c^2}\left(\frac{1}{r_{卫}}-\frac{1}{R_{赤}}\right)+\frac{u_{卫}^2-u_{赤}^2}{2c^2}\right]\Delta\tau_{赤}。\qquad(8-6-46)$$

上式右边方括号内第一、二大项分别对应于选读前的效应2和1，是“集两种效应于一身”的优雅表达式。此式不仅说明总效应等于两种效应的简单代数和，而且还给出一个有趣的“副产品”。同是海面钟，赤道钟与极地钟随地球自转有不同线速率（前者为465 m/s而后者为零），我们一直只用赤道钟计算，如果改用极地钟，结果如何？关键在于地球呈扁球状，两极半径$R_{极}<R_{赤}$，当式(8-6-46)的$u_{赤}$改为$u_{极}$时其$R_{赤}$也应改为$R_{极}$，两种改动的效果在GPS精度内抵消，$(\Delta t)_{修}$不变。这其实是Ψ_0的常数性保证的：u和R所在项无非是Ψ_0项裂变（一分为二）的产物[见式(8-6-44)]，无论如何裂变，万变不离其宗（指Ψ_0）。也可不用Ψ_0而用物理语言理解：从狭义相对论看，赤道钟因速率较大而慢于极地钟；从广义相对论看，它又因位于较高处（$R_{赤}>R_{极}$）而快于极地钟，总效果是两者快慢一样。

现在还可顺便回答选读8-1前提过的问题：美国国家标准局和英国格林尼治天文台的钟都因地球自转而不是静态钟，为何能用式(8-3-14)计算时间差？是的，由于地球自转，两钟都有非零线速率，式(8-3-14)看来应该修正。以u_1和u_2分别代表两钟的线速率，仿照式(8-6-46)的推导，不难得出式(8-3-14)的修正版本

$$\Delta\tau_1-\Delta\tau_2=\left[\frac{GM}{c^2}\left(\frac{1}{r_1}-\frac{1}{r_2}\right)+\frac{u_1^2-u_2^2}{2c^2}\right]\Delta\tau_2,\qquad(8-6-47)$$

其中r_1和r_2分别代表两钟与地心的距离[上式右边第一项实质就是式(8-3-14)的右边]。以Ω代表地球自转角速率，θ代表钟所在处的纬度，则$u\equiv\Omega r\sin\theta$。$u$与纬度有关使计算复杂，更讨厌的是可能导致式(8-6-47)右边第二项（含$u_1^2-u_2^2$）不能被忽略。幸好，扁球状的地球使问题大为简化，分析如下。$\Delta\tau_1$（$\Delta\tau_2$也一样）由两项组成，第一项来自引力势，第二项来自自转线速率。第二项当然依赖于纬度，第一项由于地球为扁球状也与纬度有关，而且纬度对两项的影响恰好抵消[见式(8-6-46)后的一段]。所以纬度变得不重要，不妨认为两钟所在纬度相同，加之两钟高度差Δh对u的影响又可被地球半径所忽略，于是计算$\Delta\tau_1-\Delta\tau_2$时两钟的第二项相消。但两钟第一项之差却与$\Delta h$有关，因为

$$GM[(1/r_1)-(1/r_2)]=GM(r_2-r_1)/r_1r_2\approx GM\Delta h/r_2^2=g\Delta h。$$

故式(8-6-47)变成式(8-3-14)，结论：式(8-3-14)仍然适用。

[选读8-2完]

[选读 8-3]

选读 8-2 虽然做过多次坐标变换,但万变不离其宗,所有线元式都代表同一线元——史瓦西线元(更准确地说是“近似史瓦西线元”,因为变换时经常用到近似处理)。这就又出现一个问题。GPS 定位定时的理论基础是传播延迟方程(8-6-2),这组方程的前提是“光以光速 c 传播”。对平直时空这没有问题,事实上,在写出式(8-6-2)之前我们的确写出了平直线元表达式(8-6-1),并指出这正是“光以光速 c 直进”的基础。不幸的是,GPS 涉及的是稍微弯曲的线元[例如式(8-6-22)],只当取近似 $1\pm2\Phi/c^2\approx1$ 时才是平直线元(8-6-1),而一旦回到平直线元就不再有引力钟慢效应,就只能给出 7 μs/天(而不是 -38 μs/天)的错误结果。这其实是笔者在阅读 GPS 有关文献时遇到的一个难题。在我们能读到的、关于相对论在 GPS 的应用的一些文献中[例如 Ashby (2002,2003)],一方面保留式(8-6-22)的 $2\Phi/c^2$ 项[或式(8-6-35)的 $(\Phi-\Psi_0)/c^2$ 项]以得出引力钟慢效应;另一方面,在写传播延迟方程(8-6-2)时又说这是“光速常数性原理”的结果,而且还给出 $c=299\ 792\ 458$ m/s 这样精确的数值。“光速不变(常数性)原理”诚然是狭义相对论的两个基本假定(出发点)之一,问题在于现在涉及的不是狭义相对论而是广义相对论,是稍微弯曲的时空而不是平直时空。解决问题的出路肯定要用近似,关键是要用得合理。文章在写出传播延迟方程后的讨论一直以式(8-6-22)的弯曲线元为基础,其中 $(cd\hat{t})^2$ 项的系数是 $-(1+2\Phi/c^2)$。$2\Phi/c^2$ 当然很小于 1,但为什么在写传播延迟方程时可以略去而在导出引力钟慢效应时又不能略去?总不能不给理由地呼之则来挥之则去吧?由于我们未能在有关文献中找到答案,只能自我思考,下面是我们杜撰的一种可能解决方案,未必正确,权且充当引玉之砖。

现在从稍微弯曲的线元(8-6-35)出发,看看在 GPS 要求的精度内,光的传播是否近似满足传播延迟方程(8-6-2)。我们只讨论光子沿径向(r 向)传播的简单情况(卫星就在用户头顶正上方),其世界线满足 $d\theta=0$, $d\varphi=0$。把式(8-6-35)用于该线的任一元段得

$$ds^2=-\left[1+\frac{2(\Phi-\Psi_0)}{c^2}\right](cdt)^2+\left(1-\frac{2\Phi}{c^2}\right)dr^2。\qquad(8-6-48)$$

由光子世界线的类光性可知 $ds^2=0$,故

$$(cdt)^2=\left(1-\frac{2\Phi}{c^2}\right)\left[1+\frac{2(\Phi-\Psi_0)}{c^2}\right]^{-1}dr^2\approx\left(1-\frac{2\Phi}{c^2}\right)\left[1-\frac{2(\Phi-\Psi_0)}{c^2}\right]dr^2,$$

$$cdt\approx\left(1-\frac{\Phi}{c^2}\right)\left(1-\frac{\Phi-\Psi_0}{c^2}\right)dr\approx\left(1-\frac{2\Phi-\Psi_0}{c^2}\right)dr=\left(1+\frac{\Psi_0}{c^2}\right)dr+\frac{2GM}{c^2}\frac{dr}{r}。$$

将上式沿图8-10(b)的光子世界线从 p_i 点积分到 q 点，由 $p_i=(t_i, r_i)$ 及 $q=(t, r)$ 可知变量 r 的积分下、上限是 r_i 和 r（因此宜将积分变量改记作 r'），得

$$c(t-t_i)=\int_{r_i}^{r}\left(1+\frac{\Psi_0}{c^2}\right)\mathrm{d}r'+\int_{r_i}^{r}\frac{2GM}{c^2}\frac{\mathrm{d}r'}{r'}=\left(1+\frac{\Psi_0}{c^2}\right)(r-r_i)+\frac{2GM}{c^2}\ln\frac{r}{r_i}。\tag{8-6-49}$$

上式中的 r 是用户所在空间点的待求径向坐标，取决于该点的海拔高度，但在量级估算时不妨取 $r=R=6.4\times10^6$ m。结合已知数据：

$$GM/c^2=4.43\times10^{-3}\ \mathrm{m}\quad 及\quad r_i(卫星径向坐标)=26.4\times10^6\ \mathrm{m},$$

可知

$$式(8-6-49)右边第二项\equiv\frac{2GM}{c^2}\ln\frac{r}{r_i}\approx2\times(4.43\times10^{-3}\ \mathrm{m})\times(-1.4)$$

$$\approx-1.2\ \mathrm{cm}。\tag{8-6-50}$$

另一方面，将式(8-6-28)用于赤道上的海平面得

$$\frac{\Psi_0}{c^2}=\frac{\Phi(R)}{c^2}-\frac{(\Omega R)^2}{2c^2}=-\frac{GM}{c^2}\frac{1}{R}-\frac{u_{赤}^2}{2c^2}=$$

$$-\frac{4.43\times10^{-3}}{6.4\times10^6}-\frac{1}{2}\left(\frac{465}{3\times10^8}\right)^2\approx-7\times10^{-10}。\tag{8-6-51}$$

故

$$式(8-6-49)右边第一项=(r-r_i)+(r-r_i)\Psi_0/c^2\approx$$

$$r-r_i+(-2\times10^7)(-7\times10^{-10})=r-r_i+1.4\times10^{-2}\mathrm{m}=r-r_i+1.4\ \mathrm{cm},\tag{8-6-52}$$

于是式(8-6-49)成为

$$c(t-t_i)=r-r_i+1.4\ \mathrm{cm}-1.2\ \mathrm{cm}=r-r_i+0.2\ \mathrm{cm}。\tag{8-6-53}$$

目前GPS测距的最高精度虽然可达10 m左右，但0.2 cm仍可忽略。所以就有

$$c(t-t_i)\approx r-r_i,\tag{8-6-54}$$

这正是传播延迟方程(8-6-2)用于径向光子世界线的结果。非径向光子世界线虽然更为复杂，但也有类似结果。就是说，传播延迟方程(8-6-2)在GPS的要求精度内是近似成立的。请注意我们只说近似成立，而我们读到的文献则认为（至少强烈暗示）它准确成立。　　**[选读8-3完]**

§8.7　史瓦西时空的最大延拓

8.7.1　奇点和奇性

§8.1 末讲过，通常天体的半径 R 远大于其史瓦西半径 $r_S \equiv 2M$，所以史瓦西线元的坐标 r 的取值范围是 $R < r < \infty$。恒星之所以处于静态，是因为其内部不断进行着烧氢变氦的核聚变反应（详见小节 7.2.1）。一旦内部核燃料消耗殆尽，恒星将在自己引力的作用下猛烈坍缩，而如果剩余质量仍超过中子星的质量上限（约为 $2M_\odot$），其半径会很快就缩至小于 r_S 的程度，使 r_S“暴露在星外的真空中”，这时史瓦西真空线元

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2) \qquad (8-7-1)$$

也应适用于 $r = r_s \equiv 2M$ 的点，而 $(1-2M/r_S)^{-1}$ 是无限大（准确地说是以 0 为分母的分数），导致线元在 $r=2M$ 的点上失去意义，人们称这些点为**奇点**，并说史瓦西线元在 $r=2M$ 处存在**奇性**（singularity）。进一步说，随着恒星的不断坍缩，r 可以取“要多小有多小”的值，但 r 不能为零，因为当 $r=0$ 时 $2M/r$ 没有意义，人们说 $r=0$ 是史瓦西线元的另一个奇点，或说史瓦西线元在 $r=0$ 处也存在奇性。史瓦西真空解发表后不久，物理学家们就注意到它存在这两个奇性，而且多年来认为它们在物理上是异乎寻常、不可思议的。后来才逐渐认识到这两个奇性在本质上截然不同：$r=0$ 处的奇性的确是奇怪得不可思议（计算表明时空曲率在 r 趋于零时竟然趋于无限大），因而称之为**物理奇性**、**时空奇性**或者**真奇性**；反之，$Sr=2M$ 处的奇性却只是“坐标奇性”：线元系数在 $r=2M$ 处的发散性只是因为 $r=2M$恰好超出了史瓦西坐标系 $\{t, r, \theta, \varphi\}$ 的适用范围（恰好“擦边在外”）。多数坐标系都存在一个适用范围，称为**坐标域**，以 2 维欧氏空间为例，直角坐标系 $\{x, y\}$ 的坐标域当然是全空间，但极坐标系 $\{r, \varphi\}$ 就不然：原点 $x=y=0$ 就不在极坐标系的坐标域内，因为该点的 φ 值无意义。在弯曲时空中使用坐标系时尤其要注意其坐标域。从一个好端端的线元出发做一个适当的坐标变换就能制造奇性。例如，2 维欧氏线元在直角坐标系的表达式

$$ds^2 = dx^2 + dy^2 \qquad (8-7-2)$$

既简单又正常，当然没有奇性。但只要定义新坐标 $x' \equiv x^3/3$，则线元在新坐标系 $\{x', y\}$ 的表达式成为

$$ds^2 = (3x')^{-4/3}dx'^2 + dy^2。$$

这个线元在 $x'=0$ 处存在奇性！这当然只是坐标奇性，通过坐标变换 $\{x', y\} \to$

$\{x, y\}$就能还原为线元(8-7-2),就能消除奇性。

发现线元存在奇性后,如何判断它是真奇性还是坐标奇性?这不是一个可以简单回答的问题,没有一种"旱涝保收"的判断方法。不过有一点可以肯定:若能找到另一坐标系(称为新系),使得在原坐标系表现奇异的点在新系的线元表达式中表现正常(不奇异),这种奇性就一定是坐标奇性。所以可以肯定地说,凡是能用坐标变换消除的奇性都是坐标奇性。我们之所以敢说史瓦西线元在$r=2M$处的奇性是坐标奇性,就是因为它的确可以用坐标变换消除。**列梅特**(Lemaitre)是历史上第一个明确地完成这一工作的人,他的文章发表在1933年。其实,爱丁顿于1924年的一篇文章中就已把史瓦西线元改写为另一形式,其中$r=2M$不再奇异,但他似乎并未注意到这一点,而且文中还存在一个把$r-2M$写成$r-M$的印刷错误。爱丁顿的这一变换直到1958年才被**芬克斯坦**(Finkelstein)重新发现。1960年,**克鲁斯科尔**(Kruskal)找到了一个更好的坐标变换①,它不但能消除$r=2M$处的奇性,而且史瓦西线元在这个新坐标系的表达式可以定义在一个延拓至最大的时空中,从而实现了史瓦西时空的最大延拓,详见下一小节。

8.7.2　克鲁斯科尔坐标系

虽然史瓦西线元是4维的,但导致它奇异的因素只出现在前2维,不妨先考虑式(8-7-1)右边前2项所代表的2维线元,并记作$d\hat{s}^2$,即

$$d\hat{s}^2 = -\left(1-\frac{2M}{r}\right)dt^2+\left(1-\frac{2M}{r}\right)^{-1}dr^2。\tag{8-7-3}$$

要使上式不奇异,可对坐标的取值范围做如下限制:

$$-\infty<t<\infty,\quad 2M<r<\infty。\tag{8-7-4}$$

也可说上式给出了2维史瓦西坐标系$\{t, r\}$的坐标域,$r=2M$的点恰恰不在坐标域内("擦边在外")。把式(8-7-3)改写为

$$d\hat{s}^2=\left(1-\frac{2M}{r}\right)\left[-dt^2+\left(1-\frac{2M}{r}\right)^{-2}dr^2\right],\tag{8-7-5}$$

上式方括号内第二项只含变量r,总可找到r的函数[记作$r_*(r)$]使得

$$dr_*=\left(1-\frac{2M}{r}\right)^{-1}dr,\tag{8-7-6}$$

从而把式(8-7-5)改写为

$$d\hat{s}^2=\left(1-\frac{2M}{r}\right)(-dt^2+dr_*{}^2)。\tag{8-7-7}$$

---

① 另一学者Szekeres也几乎同时地找到了这一坐标变换。

对式(8 -7 -6)积分便知函数 $r_*(r)$ 可取如下形式:

$$r_*(r)=r+2M\ln\left(\frac{r}{2M}-1\right)。\tag{8-7-8}$$

刚才无非是做了一次坐标变换——从 2 维史瓦西系 $\{t, r\}$ 变换到新系 $\{t, r_*\}$,其中的 r_* 称为**乌龟坐标**。式(8 -7 -7)就是 2 维史瓦西线元在新坐标系 $\{t, r_*\}$ 的表达式(请记住:看任何线元时都应注意它是在哪一个坐标系的表达式)。由式(8 -7 -8)和式(8 -7 -4)可知 t, r_* 的取值范围是

$$-\infty<t<\infty,\quad -\infty<r_*<\infty。\tag{8-7-9}$$

用下式定义两个新坐标 v 和 u:

$$v\equiv t+r_*,\quad u\equiv t-r_*,\tag{8-7-10a}$$

其逆变换为

$$t=(v+u)/2,\quad r_*=(v-u)/2,\tag{8-7-10b}$$

则

$$-\mathrm{d}t^2+\mathrm{d}r_*^2=-\mathrm{d}v\mathrm{d}u,\tag{8-7-11}$$

故线元(8 -7 -7)成为

$$\mathrm{d}\hat{s}^2=-\left(1-\frac{2M}{r}\right)\mathrm{d}v\mathrm{d}u。\tag{8-7-12}$$

读者对这一线元形式也许不熟悉。此处要补充一点有关线元的知识,仅以 2 维线元和坐标系(v, u)为例,其最一般的形式为

$$\mathrm{d}s^2=g_{vv}\mathrm{d}v^2+g_{vu}\mathrm{d}v\,\mathrm{d}u+g_{uv}\mathrm{d}u\mathrm{d}v+g_{uu}\mathrm{d}u^2,\tag{8-7-13}$$

其中各项的系数 g_{vv},g_{vu},g_{uv} 和 g_{uu} 都可以是坐标 v, u 的函数。线元必须满足两点要求(这是包含在线元定义中的,只不过我们一直未曾交代过):

(1) 对称性,即 $g_{vu}=g_{uv}$,因而式(8 -7 -13)可简化为三项:

$$\mathrm{d}s^2=g_{vv}\mathrm{d}v^2+2\,g_{vu}\mathrm{d}v\,\mathrm{d}u+g_{uu}\mathrm{d}u^2,\tag{8-7-14}$$

(2) 可逆性,是指由 4 个系数排成的方阵 $\begin{bmatrix}g_{vv} & g_{vu}\\ g_{uv} & g_{uu}\end{bmatrix}$ 必须有逆,等价于要求这个方阵的行列式 $\begin{vmatrix}g_{vv} & g_{vu}\\ g_{uv} & g_{uu}\end{vmatrix}$ 非零。行列式为零也是一种奇性,以 g 记此行列式,若线元在某处的 $g=0$,就说它在该处有奇性。

现在回到式(8 -7 -12),它无非是式(8 -7 -14)的特例,其系数行列式

$$g=\begin{vmatrix}0 & -(1-2M/r)/2\\ -(1-2M/r)/2 & 0\end{vmatrix}=-\frac{1}{4}\left(1-\frac{2M}{r}\right)^2$$

在 $r=2M$ 处为零,说明线元(8 -7 -12)在 $r=2M$ 处仍然奇异,所以还要继续努力。由式(8 -7 -10a)和式(8 -7 -9)可知 v, u 的取值范围是

$$-\infty < v,\quad u < \infty 。 \tag{8-7-15}$$

再用下式定义新坐标 V 和 U：

$$V \equiv e^{\beta v},\quad U \equiv -e^{-\beta u},\quad (\beta \text{ 为待定常数}) \tag{8-7-16}$$

则 V 和 U 的取值范围为

$$0 < V < \infty ,\quad -\infty < U < 0。 \tag{8-7-17}$$

而且

$$dv\,du = \beta^{-2} e^{\beta(u-v)}\,dV dU, \tag{8-7-18}$$

故式(8-7-12)在新系的表达式为

$$d\hat{s}^2 = -\left(1-\frac{2M}{r}\right)\beta^{-2} e^{\beta(u-v)}\,dV\,dU = -\left(\frac{r-2M}{r}\right)\beta^{-2} e^{-2\beta r_*}\,dV\,dU, \tag{8-7-19}$$

其中第二步用到式(8-7-10b)。上式的系数行列式含因子 $(r-2M)/r$，在 $r=2M$ 处为零，仍为奇性。消除奇性的关键一步是设法消去这个因子。因子 $e^{-2\beta r_*}$ 可借式(8-7-8)表为

$$\begin{aligned} e^{-2\beta r_*} &= e^{-2\beta r}\exp\left[-4\beta M\ln\left(\frac{r-2M}{2M}\right)\right] \\ &= e^{-2\beta r}\exp\left[\ln\left(\frac{r-2M}{2M}\right)^{-4\beta M}\right] = e^{-2\beta r}\left(\frac{2M}{r-2M}\right)^{4\beta M}, \end{aligned} \tag{8-7-20}$$

代入式(8-7-19)便得

$$d\hat{s}^2 = -\left(\frac{r-2M}{r}\right)\beta^{-2} e^{-2\beta r}\left(\frac{2M}{r-2M}\right)^{4\beta M} dV\,dU。 \tag{8-7-21}$$

只要把待定常数 β 选为

$$\beta = 1/4M, \tag{8-7-22}$$

式(8-7-21)便成为

$$d\hat{s}^2 = -\left(\frac{r-2M}{r}\right)\left(\frac{1}{4M}\right)^{-2} e^{-\frac{2r}{4M}}\left(\frac{2M}{r-2M}\right)dV\,dU = -\frac{32M^3}{r}e^{-\frac{r}{2M}}\,dV\,dU。 \tag{8-7-23}$$

上式就是2维史瓦西线元在新坐标系 (V, U) 的表达式。由于能导致线元在 $r=2M$ 处有奇性的因子 $r-2M$ 已被消去，上式在 $r=2M$ 处表现正常，不再奇异。可见 $r=2M$ 处的奇性的确可以通过坐标变换消除。你现在可以确定无疑地说史瓦西线元在 $r=2M$ 的奇性只是坐标奇性而已。不过，上式在 $r=0$ 处仍然是奇异的，事实上，计算表明，当你沿着一条曲线不断趋近 $r=0$ 时，你测得的时空曲率将趋于无限大(发散)，强烈暗示 $r=0$ 是不能消除的真奇点。

因为式(8-7-23)是线元在坐标系 $\{V, U\}$ 的表达式，所以式中的 r 不应再看作坐标而应看作坐标 V, U 的函数 $r(V, U)$。由 $V = e^{v/4M}$，$U = -e^{-u/4M}$ 及式(8-7-10b)、式(8-7-8)不难求得函数 $r(V, U)$ 的如下隐表达式：

$$\left(\frac{r}{2M}-1\right)\mathrm{e}^{\frac{r}{2M}}=-VU, \tag{8-7-24}$$

现在做一小结。我们为消除 $r=2M$ 的奇性做了一系列坐标变换，即

$$\{t,\ r\}\ \mapsto\{t,\ r_*\}\ \mapsto\{v,\ u\}\ \mapsto\{V,\ U\}。$$

确定每组坐标取值范围的原则是要使 $r=2M$“擦边地在坐标域之外”。然而，既然 $r=2M$ 已经不再是线元(8-7-23)的奇点，就没有必要坚守 V，U 的原定范围 $0<V<\infty$，$-\infty<U<0$，而是可以把范围适当延拓至 $V\leqslant 0$ 和 $U\geqslant 0$ 的区域。但是，由于 $r=0$ 仍然是奇点，r 的取值范围仍要受到 $r>0$ 的限制，所以 V 和 U 的范围也并非完全自由，它们的搭配必须保证 $r>0$ 的条件得到满足。由式(8-7-24)可知 $r>0$ 等价于

$$VU<1。$$

至此，我们已经成功地消除了坐标奇性 $r=2M$。作为结果，史瓦西线元的定义域被延拓至更大的时空范围(从 $V>0$，$U<0$ 延拓至 $VU<1$)。这一工作是克鲁斯科尔于 1960 年完成的，后人称之为史瓦西线元的**克鲁斯科尔延拓**(Kruskal extension)。虽然我们一直只涉及 2 维史瓦西线元，但只需简单地补上后两维便可得到 4 维的克鲁斯科尔线元：

$$\mathrm{d}s^2=-\frac{32M^3}{r}\mathrm{e}^{-\frac{r}{2M}}\,\mathrm{d}V\,\mathrm{d}U+r^2(\mathrm{d}\theta^2+\sin^2\theta\mathrm{d}\varphi^2)。 \tag{8-7-25}$$

8.7.3　克鲁斯科尔最大延拓

为便于讨论和画图，还可用最后一次坐标变换把式(8-7-25)的交叉乘积形式(指 $\mathrm{d}V\mathrm{d}U$)变为平方项形式。用下式定义新坐标 T，X：

$$T\equiv(V+U)/2,\quad X\equiv(V-U)/2, \tag{8-7-26}$$

则式(8-7-25)成为

$$\mathrm{d}s^2=\frac{32M^3}{r}\mathrm{e}^{-\frac{r}{2M}}(-\mathrm{d}T^2+\mathrm{d}X^2)+r^2(\mathrm{d}\theta^2+\sin^2\theta\mathrm{d}\varphi^2)。 \tag{8-7-27}$$

这时 r 应看作新坐标 T，X 的函数，即 $r(T,\ X)$。由式(8-7-24)和式(8-7-26)易得函数 $r(T,\ X)$ 的隐表达式：

$$\left(\frac{r}{2M}-1\right)\mathrm{e}^{\frac{r}{2M}}=X^2-T^2。 \tag{8-7-28}$$

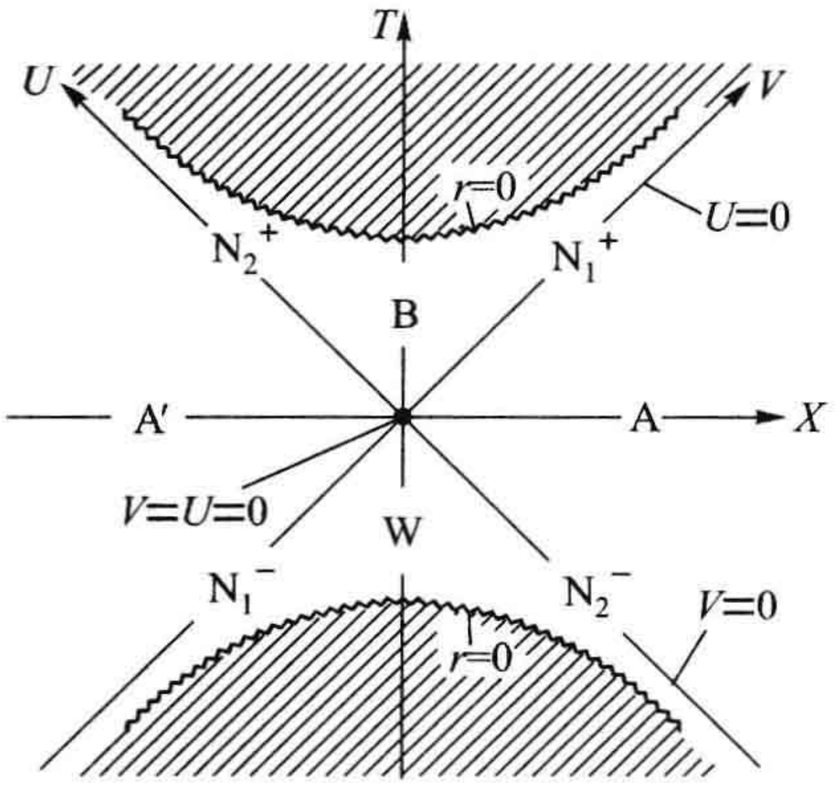
与 V，U 类似，坐标 T，X 的取值范围只受 $r>0$ 的限制，而式(8-7-28)表明 $r=0$ 对应于 $T^2-X^2=1$，所以 T，X 的取值范围只受下式限制：

$$T^2-X^2<1。 \tag{8-7-29}$$

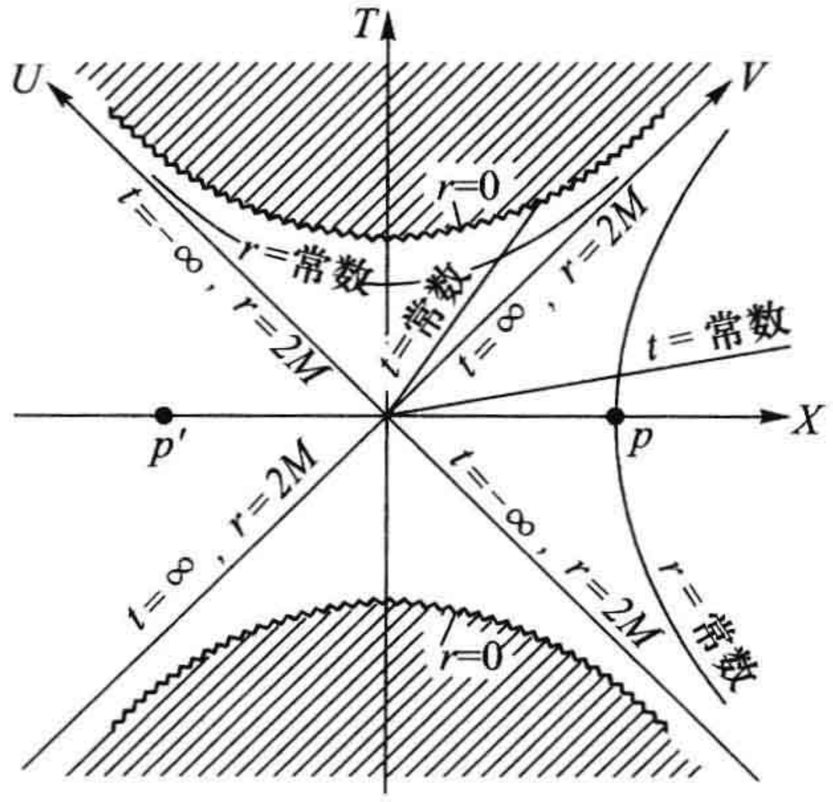
$r=0$ 是有曲率发散性的真奇点，不可能把 r 再延拓至 $r=0$ 之外[图 8-13

(a)的阴影部分],所以克鲁斯科尔延拓就是史瓦西时空的**最大延拓**(maximal extension)。图8-13(a)的阴影部分不属于延拓后的时空,两条锯齿状双曲线代表时空奇性所在处("奇点"在时空中并非一点)。

我们对这一最大延拓再做以下几点讨论:



(a) A和A′代表两个无因果联系的渐近平直区,B为黑洞区,W为白洞区。N_1^+和N_2^+是黑洞的视界。锯齿状曲线代表奇性所在处(不属于时空)。



(b) 过原点的±45°斜直线上$r=2M$, $t=\pm\infty$,以此为渐近线的双曲线上$r=$常数。

图8-13　史瓦西时空的克鲁斯科尔最大延拓

1. 图8-13虽然只是2维时空图,但由式(8-7-25)[或式(8-7-27)]可知补上后2维相应于把图8-13阴影区外的每个点看作一个2维球面,其半径就是该点的r值。就是说,只要你会看(只要你能把阴影区外的每点想象为一个2维球面),该图就可看作4维史瓦西时空最大延拓后的时空图。请特别注意阴影区外每点都代表一个(而不是半个)球面,两个不同的点代表两个不同的球面,尤其是,例如p和p'点,它们并非是同一球面上的两点,而是每点各代表一个不同的球面。

2. 设L是$T-X$面内的曲线,则由式(8-7-28)还可看出

$$r=\text{常数} \Leftrightarrow X^2-T^2=\text{常数} \Leftrightarrow \text{L是(以45°线为渐近线的)双曲线。}$$

3. 考虑图8-13中非阴影区内的任一条径向类光曲线("径向"是指θ和φ沿线为常数,亦即在$T-X$面内),注意到类光曲线的线元$ds^2=0$,将式(8-7-27)用于其任一元段得

$$0=-\,dT^2+dX^2=(-\,dT+dX)(\,dT+dX),$$

因而$-\,dT+dX=0$或$dT+dX=0$,积分便得如下两族径向类光曲线(其实还是

类光测地线)：

(a)$T=X+\alpha$；　(b)$T=-X+\alpha$。(α 为积分常数)　(8-7-30)

可见,在以$\{T, X, \theta, \varphi\}$为基准系的时空图中,径向类光测地线都是 ±45°斜直线①[见图 8-13(a)]。这将给今后的讨论提供很大方便。

4. 积分常数 $\alpha=0$ 的两条径向类光测地线是过原点的 ±45°的斜直线,即图 8-13(a)的 N_1 和 N_2。由式(8-7-28)可知 N_1 和 N_2 线的 r 值都是 $r=2M$。

5. 如果把 N_1 和 N_2 线的每点看作一个 2 维球面,则 N_1 和 N_2 各代表一个 3 维面(称为 3 维超曲面),以它们为边界就可在最大延拓的史瓦西时空中定义 4 个互不相交的子时空(也称为 4 个区),在图 8-13 中依次记作 A 区、B 区、A′区和 W 区,它们的范围由下式界定：

A 区：　$X>0$ 及 $X^2>T^2$,　(即 $V>0$, $U<0$),由式(8-7-28)知其 $r>2M$;

B 区：　$T>0$ 及 $X^2<T^2$,　(即 $V>0$, $U>0$),由式(8-7-28)知其 $r<2M$;

A′区：　$X<0$ 及 $X^2>T^2$,　(即 $V<0$, $U>0$),由式(8-7-28)知其 $r>2M$;

W 区：　$T<0$ 及 $X^2<T^2$,　(即 $V<0$, $U<0$),由式(8-7-28)知其 $r<2M$。

A 区的 r 值满足 $2M<r<\infty$,是原始坐标系$\{t, r, \theta, \varphi\}$的坐标域,也是克鲁斯科尔延拓的出发区("根据地")。其他三区 B, W, A′都是从 A 区出发做延拓的产物。

6. 对 A 区,由式(8-7-10b)及 $V=e^{v/4M}$,　$U=-e^{-u/4M}$得

$$t=(v+u)/2=2M\ln[V/(-U)]=2M\ln[V/|U|]。\quad (8-7-31)$$

X 轴有 $V=|U|$,故 $\ln[V/|U|]=0$,由式(8-7-31)可知 X 轴上有 $t=0$。为求得 N_1^+(N_1 线 $V>0$ 的部分)上的 t 值,请读者在图 8-13(a)的 A 区与$\{T, X\}$系的第一象限的交域中补画一条过原点的直线(记做 μ)。μ 线上 $V>|U|$导致 $\ln[V/|U|]>0$,由式(8-7-31)可知 $t>0$。令 μ 线越来越趋近直线 N_1^+,则因 $|U|$值趋于零而使 t 值趋于 $+\infty$,所以 N_1^+ 线上有 $t=+\infty$。同理还知 N_2^- 线上有 $t=-\infty$,如图 8-13(b)所示。这就说明坐标 t 在 N_1^+ 和 N_2^- 上没有定义(发散),这也正是史瓦西线元(8-7-1)在 $r=2M$ 处表现奇异的原因。(N_1^+ 和 N_2^- 都不属于坐标系$\{t, r, \theta, \varphi\}$的坐标域,恰恰是擦边在外。)

---

① 弯曲时空中"直线"及" ±45°"等词并无意义,此处只是借欧氏语言的称谓。

7. 我们从史瓦西坐标系$\{t, r, \theta, \varphi\}$出发,通过如下的一系列坐标变换

$$\{t, r, \theta, \varphi\} \mapsto \{t, r_*, \theta, \varphi\} \mapsto \{v, u, \theta, \varphi\} \mapsto \{V, U, \theta, \varphi\}$$

把线元表达式(8-7-25),发现它在$r=2M$处表现良好,于是从 A 区出发得到最大延拓(包含四个区 A, B, W, A′以及两张 3 维面 N_1, N_2)。因为延拓是借助于坐标 V, U 获得的,所以 V, U 在全时空都有明确定义。现在问:在除 A 区外的三个区中,其他坐标(只需重点讨论坐标 t, r)是如何定义的?式(8-7-23)中含有变量 r,前已讲过它应看作 V, U 的函数 $r(V, U)$,其隐表达式为式(8-7-24)。既然延拓是借式(8-7-23)获得的,式(8-7-24)就可用作变量 r 在其他三区的定义式。但是变量 t 在其他三区却不曾下过定义。注意到 A 区中坐标 V, U 与坐标 t, r_* 的关系:

$$V=\mathrm{e}^{\frac{r_*+t}{4M}},\quad U=-\mathrm{e}^{\frac{r_*-t}{4M}}, \tag{8-7-32}$$

不妨反过来用 V, U 依照以下关系定义其他三区的 t 坐标:

B 区　$V=\mathrm{e}^{\frac{r_*+t}{4M}},\quad U=\mathrm{e}^{\frac{r_*-t}{4M}};$

W 区　$V=-\mathrm{e}^{\frac{r_*+t}{4M}},\quad U=-\mathrm{e}^{\frac{r_*-t}{4M}};$　(8-7-32′)

A′区　$V=-\mathrm{e}^{\frac{r_*+t}{4M}},\quad U=\mathrm{e}^{\frac{r_*-t}{4M}},$

其中 r_* 由 r 依下式定义:

$$r_*\equiv r+2M\ln\left|\frac{r}{2M}-1\right|。 \tag{8-7-33}$$

把线元(8-7-25)用于 B, W, A′区并借式(8-7-32′)、(8-7-33)改写为以 t, r 表出的线元,结果仍为式(8-7-1),其中 r 的取值范围对 B, W 区为 $0<r<2M$,对 A′区为 $2M<r<\infty$(同 A 区)。四个区中坐标 T, X 与 t, r 的关系如下(证明留作习题,式中的 sh 和 ch 分别代表双曲正弦和余弦):

$$\text{A 区}\begin{cases}T=\sqrt{\dfrac{r}{2M}-1}\ \mathrm{e}^{\frac{r}{4M}}\mathrm{sh}\left(\dfrac{t}{4M}\right),\\[2ex] X=\sqrt{\dfrac{r}{2M}-1}\ \mathrm{e}^{\frac{r}{4M}}\mathrm{ch}\left(\dfrac{t}{4M}\right),\end{cases} \tag{8-7-34}$$

$$\text{B 区}\begin{cases}T=\sqrt{1-\dfrac{r}{2M}}\ \mathrm{e}^{\frac{r}{4M}}\mathrm{ch}\left(\dfrac{t}{4M}\right),\\[2ex] X=\sqrt{1-\dfrac{r}{2M}}\ \mathrm{e}^{\frac{r}{4M}}\mathrm{sh}\left(\dfrac{t}{4M}\right),\end{cases} \tag{8-7-35}$$

$$\text{W 区}\begin{cases}T=-\sqrt{1-\dfrac{r}{2M}}\ \mathrm{e}^{\frac{r}{4M}}\mathrm{ch}\left(\dfrac{t}{4M}\right),\\[2ex] X=-\sqrt{1-\dfrac{r}{2M}}\ \mathrm{e}^{\frac{r}{4M}}\mathrm{sh}\left(\dfrac{t}{4M}\right),\end{cases} \tag{8-7-36}$$

$$\text{A}'\text{区}\begin{cases} T=-\sqrt{\dfrac{r}{2M}-1}\ \mathrm{e}^{\frac{r}{4M}}\ \mathrm{sh}\left(\dfrac{t}{4M}\right), \\ X=-\sqrt{\dfrac{r}{2M}-1}\ \mathrm{e}^{\frac{r}{4M}}\mathrm{ch}\left(\dfrac{t}{4M}\right)。 \end{cases} \tag{8-7-37}$$

注记 2　史瓦西线元(8-7-1)在 4 个时空区内都与坐标 t 无关,但是不能简单地认为 t 一定是时间坐标。讨论表明(见选读 8-4),在 B 区和 W 区内的时间坐标是 r 而不是 t,与此密切相关的事实是:B 区和 W 区并非静态时空区。这就导致重要结论:史瓦西最大延拓时空就其整体而言不是静态时空。

[选读 8-4]

把图 8-13 看作 2 维时空图,则图中 A 区内的 $r=$常数 的曲线是类时曲线(这可由“径向类光曲线是 ±45°斜直线”直接看出),可充当某个观者的世界线,r 为常数表明他是静态观者,沿线改变的坐标只有 t,代表他的坐标时间在流逝。然而,B 区中“$r=$常数”的曲线是类空曲线,不是任何观者的世界线。反之,B 区中“$t=$常数”的线却是类时曲线,所以可以充当某个观者的世界线。由于 t 值沿线不变,t 在任何意义下都不能代表时间,但坐标 r 值却随着该线固有时 τ 的增大而减小(渐趋于零),所以 $-r$ 是该线固有时 τ 的常增函数,因而在 B 区内可以解释为坐标时间。(类似地,W 区内的坐标 r 可看作坐标时间。)其实这也可从线元表达式直接看出:

$$\mathrm{d}s^2=-\left(1-\frac{2M}{r}\right)\mathrm{d}t^2+\left(1-\frac{2M}{r}\right)^{-1}\mathrm{d}r^2+r^2(\mathrm{d}\theta^2+\sin^2\theta\mathrm{d}\varphi^2),$$

上式在 $r=0$ 和 $r=2M$ 处奇异,r 的取值范围只能在以下二者中任择其一:① $2M<r<\infty$(A 和 A′区);② $0<r<2M$(B 和 W 区)。对① 有 $1-2M/r>0$,此时 $\mathrm{d}t^2$ 的系数为负而 $\mathrm{d}r^2$ 的系数为正,所以 t 是时间坐标而 r 是空间坐标。对② 有 $1-2M/r<0$,此时 $\mathrm{d}t^2$ 的系数为正而 $\mathrm{d}r^2$ 的系数为负,所以 t 是空间坐标而 r 是时间坐标。可见,坐标 t 和 r 在从 A 区过渡到 B 区后出现“角色互换”的“怪事”——在 A 区,r 为空间坐标,t 为时间坐标;在 B 区,$-r$ 为时间坐标,t 为空间坐标。与此密切相关的事实是:B 区和 W 区都不是静态时空区。

然而这里有个微妙之处,下面以对话方式澄清,其中乙代表笔者。

甲　现在我明白 B 和 W 区不是静态区的理由了,这无非是因为线元系数随时间坐标 $-r$ 或 r 而变(随时间而变当然不是静态)。由此我悟到一个结论:只要线元系数随时间坐标而变,就一定不是静态线元。

乙　然而你悟到的结论是错误的,下面就是一个反例。某 2 维线元在某坐标系 $\{t, x\}$ 的表达式为

$$ds^2 = -t^{-4}dt^2 + dx^2, \quad (t>0),$$

dt^2 的系数为负表明 t 是时间坐标，而线元系数含有 t，如果你的结论正确，这个线元就非静态。然而，一个简单的坐标变换 $T = t^{-1}$，$X = x$ 就可把线元式变为

$$ds^2 = -dT^2 + dX^2,$$

从而发现所论的线元无非是2维闵氏线元，当然是静态的！

甲　那么，怎样从线元式判断非静态性？

乙　此处应该重提选读7.1关于用线元判断静态性的判据："对于静态时空，你总能找到这样的坐标系 $\{t, x^1, x^2, x^3\}$，线元在该系的表达式满足 ① 各项的系数与 t 无关（默认 dt^2 项的系数为负）；② 不含时空交叉项（即形如 $dtdx^i$ 的项）。"这意味着，如果存在一个坐标系，线元在该系的表达式不含时间坐标（而且没有时空交叉项），这时空一定是静态的；然而，如果存在一个坐标系，线元在该系的表达式含有时间坐标，却不能肯定时空一定非静态。

甲　这岂不是说明单从线元式无法判定时空的非静态性吗？具体到史瓦西最大延拓时空，您是怎么肯定B和W区不是静态区的？

乙　线元式是时空几何在坐标系的表现，用线元式讨论就是用坐标语言讨论。事实上，静态性是绝对性质，不因坐标系的改变而改变，所以最好用纯几何语言讨论。线元静态性的几何定义涉及本书没有讲到的微分几何知识[Killing矢量场，可参阅梁灿彬，周彬(2006)上册§8.1]，此处不能细述。以此定义衡量，B和W区的确不是静态区。　　**[选读8-4完]**

8.7.4　黑洞、白洞和视界

我们再来回顾一下克鲁斯科尔延拓的原始动机。任何人只要见到史瓦西真空线元(8-7-1)，都能看出它在 $r=0$ 和 $r=2M$ 处有奇性，因而 r 的取值范围只有两种可能：①$2M<r<\infty$；②$0<r<2M$（分别称此为①区和②区）。你也许会问："是否可以既取 $2M<r<\infty$ 又取 $0<r<2M$（只把 $r=2M$ 抠去）？"答案是否定的，因为这种"兼而取之"的做法将得到一个不连通的时空，而这是物理上不能接受的。你也许会接着问："$r=2M$ 不过是一张没有厚度的几何面（见图8-14），如果①区发生的物理过程有可能涉及②区（例如粒子从①区掉进②区并最终落

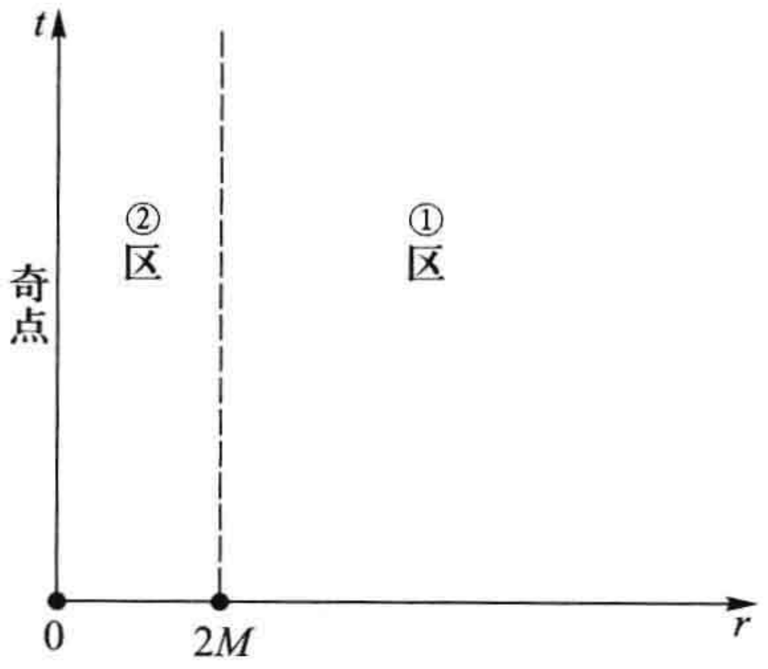


图8-14　用史瓦西坐标系为基准系画时空图

入奇点 $r=0$),它只要穿过 $r=2M$ 这张面不就可以了吗?”是的,作为事后诸葛亮,在克鲁斯科尔找到新坐标系从而消除了 $r=2M$ 的奇性之后,你的说法还说得过去,但在此之前,$r=2M$ 有可能是个真奇点,“从①区穿过 $r=2M$ 这张面进入②区”就完全说不清楚了。可见,我们需要一个克鲁斯科尔式的“诸葛亮”,就是说,必须找到一个坐标变换,使得 $r=2M$ 在新坐标系中真的就是一张“没有厚度的膜”。准确地说,这就是图 8-13(a)的 $\pm 45°$ 斜直线 N_1 和 N_2(实际上是两张 3 维面)。由于径向类光曲线在图 8-13 中是 $\pm 45°$ 斜直线,从 A 区任一点出发的“内向”(指 r 值不断减小)的、指向未来的类光或类时曲线将不可避免地穿越 N_1^+ 进入 B 区。反之,B 区任一点发出的、指向未来的类光或类时曲线都不可能穿越 N_1^+ 进入 A 区,它们的必然归宿是掉进奇点。可见 N_1^+ 是个有进无出的“单向膜”,A 区中的任何物体(连同光子)一旦穿过它而进入 B 区就永不可能回到 A 区(只能掉进奇点)。因此称 B 区为**黑洞**(black hole),称 N_1^+ 为**视界**(horizon)①。考虑到图 8-13 中每点代表一个球面,可知黑洞是个 4 维时空区域,而视界则是个 3 维面。另一方面,A′区由 $X<0$ 及 $X^2>T^2$ 表征,它也有 $r>2M$,事实上它与 A 区有完全一样的性质,它与黑洞 B 的关系也类似于 A 区与 B 区的关系,故 N_2^+ 也是视界。但 A′与 A 区之间没有任何因果联系:从 A 出发的任一类时或类光曲线都不能进入 A′区,反之亦然。在这个意义上也常把 A 和 A′区称为两个互不关联的“宇宙”。W 区由 $T<0$ 及 $X^2<T^2$ 表征,它也有 $r<2M$。W 区与 A(及 A′)区也只有“一膜之隔”,这张“膜”就是 3 维面 N_2^- 或 N_1^-,两者都是有出无进的单向膜,W 区中任一指向未来的类时或类光曲线都将穿越 N_2^- 或 N_1^- 进入 A 或 A′区。既然 B 区叫黑洞,自然就把 W 区称为**白洞**(white hole)。

以上是在全时空为真空的前提下得到的最大延拓。虽然这一延拓包含了黑洞、白洞、视界以及两个全同“宇宙”等诱人术语,其物理存在性(真实性)却是另一个问题。从初值理论的角度考虑,整个时空存在的可能性微乎其微,但其中的一部分(包括 A, B 区以及其间的视界的一部分)却很有意义,详见 §8.9。

§8.8 伯克霍夫定理

正如 §6.5 所云,史瓦西之所以能够很快就求得真空爱因斯坦方程的第一个精确解,是因为他巧妙地利用了问题本身的对称性——恒星既是球对称的,又是静态的。然而,任何恒星都不是永恒不变的,它们也像人一样要经历生、老、病、死等各个演化阶段,恒星在许多阶段中并不处于静态之中,这时恒星外部的

① 相对论中存在许多不同种类的视界,这里的“视界”准确地应称为**事件视界**(event horizon)。

时空几何还可用史瓦西真空线元(8-7-1)描述吗？如果有人向史瓦西提出这个问题,恐怕他也难以回答。直到1923年,伯克霍夫(Birkhoff)才对此问题给出了肯定的答案,因为他证明了如下定理:

伯克霍夫定理(1923)　真空爱因斯坦方程的球对称解必静态。

根据这一定理,即使恒星处于急剧的收缩、膨胀、径向震荡甚至爆炸的过程中(已经远离静态),只要它的物质分布仍然保持球对称性,它外面(真空中)的时空线元必然仍是静态球对称线元,因而仍可由史瓦西真空解(8-7-1)描述。这就为星体演化的研究提供了很大的方便(见§8.9)。

然而后来发现伯克霍夫定理的提法不够准确[**彼得罗夫**(Petrov)于1963年最先提出质疑],应该修正。修正后的表述为

修正后的伯克霍夫定理　真空爱因斯坦方程的球对称解必为史瓦西线元。

表面看来修正前后的两个定理似乎一样,其实却有重要不同。关键在于式(8-7-37)下面注记2的结论,即史瓦西最大延拓时空就其整体而言不是静态时空。所以说伯克霍夫定理的原始提法必须修正。修正后的伯克霍夫定理也非常有用,因为恒星在各个演化阶段中通常都能保持球对称性,于是星外(真空中)的时空几何仍然由史瓦西线元描述。

§8.9　恒星的引力坍缩和史瓦西黑洞

如小节7.2.2所讲,大质量球对称恒星如果不能在演化中通过抛出足够质量成为稳定的白矮星或中子星,就根本没有稳定状态可言,只能不断坍缩,最终成为黑洞。根据伯克霍夫定理(见上节),爱因斯坦方程的真空球对称解必为史瓦西线元,因此可用图8-15表示。图中无阴影部分与图8-13的相应部分全同,但阴影部分则由星内线元(爱因斯坦方程的非真空解)描述,所以球对称坍缩星的时空根本没有白洞区W,也没有A′区,但黑洞区B以及A区的一部分在此情况下有重要意义。无论构成星体的物质是如何坚实,只要星体表面越过视界,就只能不断收缩,直至整个星体被压缩为奇点$r=0$。理由很简单:星体表面任一点的世界线都必须位于该点的光锥之内(以保证类时性),因而与T轴的夹角必须小于45°(请注意图8-13和图8-15中$\pm45°$斜直线代表径向类光曲线)。史瓦西坐标系$\{t, r\}$只能覆盖$r>2M$(或$0<r<2M$)的时空区域,因而不能表现恒星晚期坍缩为史瓦西黑洞的全过程,特别是不能表现全过程中最关键的一步——星体表面缩进视界以内的那一步。如果一定要用史瓦西坐标描述恒星的晚期坍缩,就只能画成图8-16。由于史瓦西坐标t在$r=2M$处没有定义(t在该处为“无限大”,就是没定义),此图实际上是把两个图(分别描述$r>2M$和

$0<r<2M$)生拼在一起的结果。图的右边($r>2M$)容易让人误以为坍缩星表面永远处于视界 $r=2M$ 以外,这种误解来自把 $t=\infty$ 混同于“永远”(知道芝诺佯谬的读者应注意到坐标时间 t 同“阿基里斯时间”的类似性)。由图 8-13(b)可知星体表面与 $r=2M$ 的交点(见图 8-15 的 p)对应于 $t=\infty$,但星体表面的观者在 p 点的固有时 τ 却是有限值,他从此进入黑洞并在很短的固有时间 $\Delta\tau$ 内掉入奇点(对 $M=3M_\odot$ 的黑洞,$\Delta\tau$ 约为 2×10^{-5} s)。

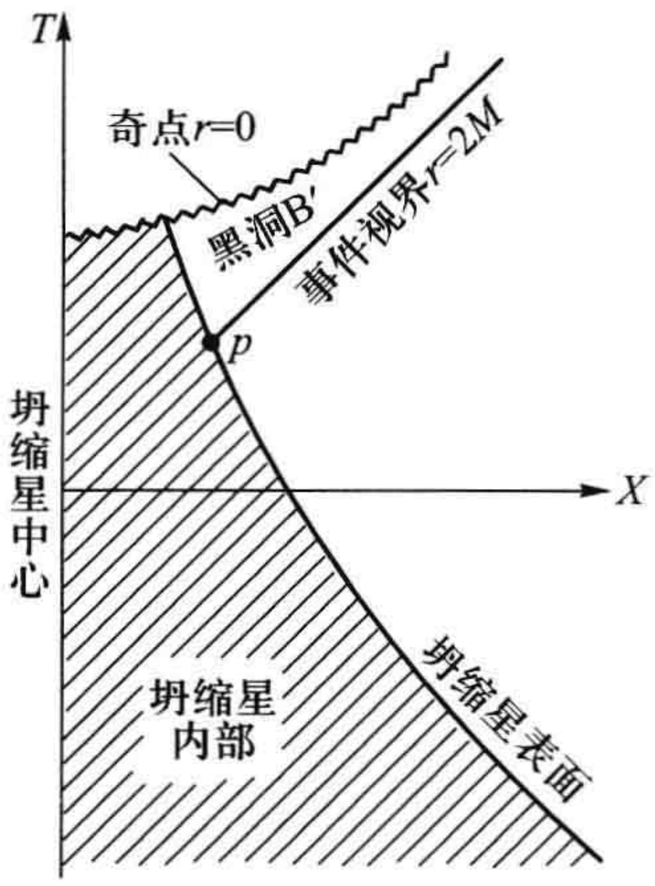


图 8-15　用克鲁斯科尔坐标描述恒星晚期坍缩。史瓦西真空解只适用于星外,B′区(不含阴影部分)代表坍缩造成的黑洞。

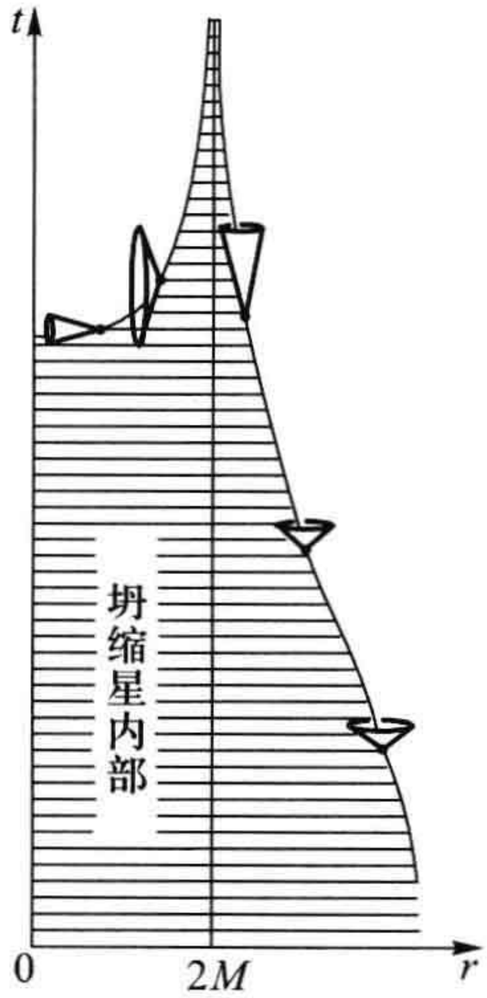


图 8-16　用史瓦西坐标描述恒星晚期坍缩。虽然星体表面要到 $t=\infty$ 时才缩至 $r=2M$,却不表明它永远在视界以外,因为坐标时 t 趋于无穷不代表“永远”。

恒星坍缩成黑洞的过程还可更形象地用另一坐标系描述。作为预备,我们先介绍内向爱丁顿坐标系$\{v, r, \theta, \varphi\}$,它虽然不像克鲁斯科尔系那样能覆盖最大延拓的史瓦西时空,但能覆盖 A 区和 B 区(而史瓦西坐标系则只能覆盖 4 个区中的一个)。就是说,它能消除 A, B 区之间的 $r=2M$ 面的奇性。该系的 r, θ, φ 与史瓦西系的对应坐标相同,而

$$v\equiv t+r_*。\tag{8-9-1}$$

史瓦西系的前两维线元

$$\mathrm{d}\hat{s}^2=-\left(1-\frac{2M}{r}\right)\mathrm{d}t^2+\left(1-\frac{2M}{r}\right)^{-1}\mathrm{d}r^2\tag{8-9-2}$$

在内向爱丁顿系成为

$$\mathrm{d}\hat{s}^2=-\left(1-\frac{2M}{r}\right)\mathrm{d}v^2+2\mathrm{d}v\mathrm{d}r。\tag{8-9-3}$$

上式的线元系数 $g_{vv}=-(1-2M/r)$，$g_{vr}=1$，$g_{rr}=0$ 在 $r=2M$ 处都表现良好，而且系数行列式 $g=-1\neq 0$，所以 $r=2M$ 不再是奇点。注意到 $-\infty<v<\infty$ 对应于 $0<V<\infty$，可知 $\{v, r\}$ 系的坐标域能覆盖图 8-13 的 A，B 区，它消除了这两个区的交界面 N_1^+ 的坐标奇性。

为了使讨论更为形象，我们借助时空图。画时空图前要明确选定一个基准坐标系，因为以不同基准系画出的时空图在外观上可以很不相同（虽然实质一样），图 8-15 与图 8-16 就是例子，它们表现的都是恒星坍缩为黑洞的过程，特别是恒星表面一点的世界线，但这两条世界线在两个图中的形状甚为悬殊。现在又有了内向爱丁顿系，我们想看看以它为基准系画出的时空图是否更有帮助。最好先画出各点的光锥（因为任何质点的世界线都必须躺在光锥以内）。为此应先求得内向和外向的类光曲线。把式（8-9-3）用于类光曲线的任一元段，利用类光性 $d\hat{s}^2=0$ 便得

$$0=dv\left[-\left(1-\frac{2M}{r}\right)dv+2dr\right],$$

它有两个解（分别代表内向和外向类光曲线）：

（1）$dv=0$；　　（8-9-4）

（2）$-\left(1-\frac{2M}{r}\right)dv+2dr=0$。　　（8-9-5）

式（8-9-4）沿类光曲线积分给出 $v=$常数（无论 r 值如何），说明这族类光曲线在以 $\{v, r\}$ 为基准系的时空图中是一组水平直线，这并无任何不对，但毕竟不利于直观地考虑问题，所以最好再做一次坐标变换。令

$$\tilde{t}\equiv v-r,　　(8-9-6)$$

得坐标系 $\{\tilde{t}, r\}$。线元（8-9-3）在 $\{\tilde{t}, r\}$ 系的形式为

$$d\hat{s}^2=-\left(1-\frac{2M}{r}\right)d\tilde{t}^2+\frac{4M}{r}d\tilde{t}dr+\left(1+\frac{2M}{r}\right)dr^2。　　(8-9-7)$$

用于类光曲线，全式除以 dr^2 得

$$0=-\left(1-\frac{2M}{r}\right)\left(\frac{d\tilde{t}}{dr}\right)^2+\frac{4M}{r}\frac{d\tilde{t}}{dr}+\left(1+\frac{2M}{r}\right)。$$

上式是关于未知量 $d\tilde{t}/dr$ 的 2 次代数方程，有两个解：

（1）$\frac{d\tilde{t}}{dr}=-1$；　　（8-9-8）

（2）$\frac{d\tilde{t}}{dr}=\frac{r+2M}{r-2M}$。　　（8-9-9）

相应的两族类光曲线示于图 8-17 中。第一族类光曲线都是斜率为 -45°的斜

直线,代表内向类光曲线族,看似一般;但第二族类光曲线却非常特别:它有一条竖直线(在 $r=2M$ 处,该处有 $\mathrm{d}\tilde{t}/\mathrm{d}r=\infty$),而且其他各线在这条竖直线两边"左右开花"(因为 $\mathrm{d}\tilde{t}/\mathrm{d}r$ 右正左负)——竖直线以右的线随 $\tilde{t}$ 的增大而增大(真正外向);但竖直线以左的线的 r 值却随 $\tilde{t}$ 的增大而减小(实际向内,但仍属外向族)。这一怪事反映了黑洞的重要特征——$r=2M$ 是视界,在视界以内($r<2M$)的任何光子都不能穿越视界而到达洞外($r>2M$),它们的 r 值只能不断减小至零(掉入奇点)。由这两族类光曲线可方便地画出各点的光锥,它们对于正确画出质点的世界线大有帮助(因为质点的世界线必须在光锥以内)。由此可知视界外的质点可穿越视界进入黑洞,而一旦进入就无法退出,只能落入奇点。将图 8-17 以 $\tilde{t}$ 轴为对称轴旋转便得 3 维时空图(图 8-18),再补上坍缩星表面的世界面(图中的炮弹形),就可形象地表现坍缩为黑洞的星体的外部时空几何。为便于理解,讨论以下假想实验("假想"包括忽略潮汐力)。设某人自愿充当黑洞探险者,他坐着燃料充足的飞船从远方实验室出发做黑洞探险。如果他不开动发动机,飞船将在黑洞外面不断自由下落,必然穿过视界进入黑洞并葬身奇点。但如果他在到达视界前"悬崖勒马",掉转船头,开足马力(即在 r 尚未小到 $2M$ 时就让它重新变大),是可以安全返航并提交探险报告的。然而,如果他贪婪一点,多走一步到达视界(须知当他的世界线与视界的世界面相交时他并无特殊感觉),就将"一失足成千古恨",因为由视界上的光锥可知一到视界就无法逆转,就连向远方实验室的朋友用手机打个无线电话也传不出去,因为在视界上发

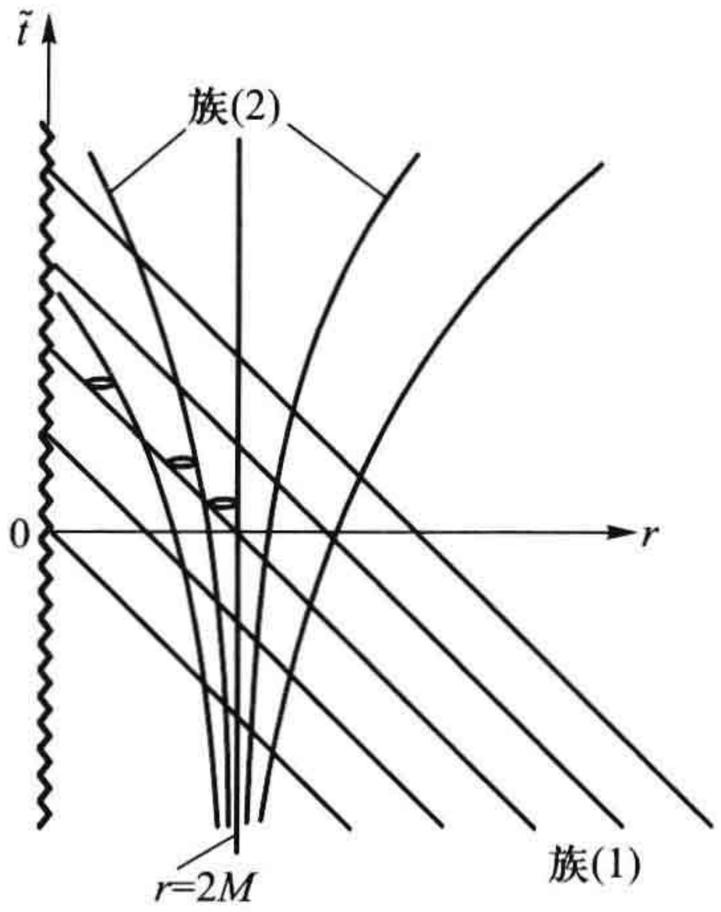


图 8-17 2 维史瓦西时空的两族类光曲线在 $\{\tilde{t}, r\}$ 系的表现

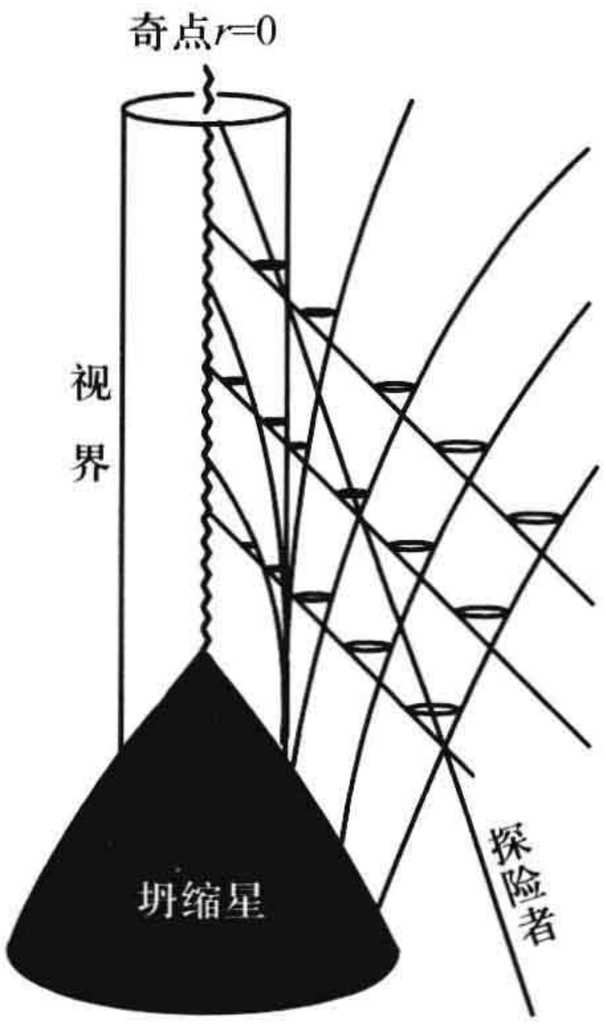


图 8-18 恒星坍缩为黑洞的时空图[以 $\{\tilde{t}, r\}$ 系为基准系]。黑洞探险者如不在到达视界前掉头就必落入奇点。

出的“外向”光子只能沿视界竖直向上(r 值永远保持为 $2M$)。以上就是黑洞探险者的遭遇,下面再讨论黑洞探险者的外观。假定探险者身上装有永能发光的灯泡,远方实验室的静态观者用最理想的观测仪器密切观测(“最理想”是指可以观测到任何波长、非常微弱的电磁波)。图 8-18 示出探险者世界线所发出的外向光子到达外部静态观者的情况。由于探险者世界线在视界以内(黑洞区内)的一段发出的光子不能到达视界以外,外部观者似乎会看到探险者在走到某处时忽然消失。然而细观图 8-18 可知情况并非如此。由于视界以外的外向光子世界线越靠近视界越陡峭,在视界上完全竖直(躺在视界面上永不离开),外部观者将永远(无论其固有时为多大)收到探险者所发出的光(只不过是他到达视界之前所发的光)。他会感到探险者越走越慢(但永不停止),不断逼近某个地方(视界),并且越来越像是“冻结”在那里。这就是黑洞探险者的外观。真实遭遇和外观的如此不同正是奇点周围引力奇强的表现。如果外部观者观察坍缩星的表面,也会看到类似现象——星体越来越趋于某一大小,其半径以越来越小的速率趋于 $2M$ 并将“冻结”在这一大小上(所以在“黑洞”一词提出前曾有过“冻结星”的称谓)。这种现象也可看作引力场中的钟慢效应(星表面的钟慢于远方静态观者的钟),但不全同于小节 8.3.1 的引力钟慢效应(图 8-4)。两者的相同点是被比较的两钟位于引力势不同的两地;区别在于,图 8-4 的两钟都是静态钟,可以认为两钟没有相对速度,钟慢效应完全由引力势的不同所导致,所以可称为纯粹的引力钟慢效应;而图 8-18 中有一个钟(坍缩星表面的钟)并非静态,可解释为两钟有相对速度,除引力钟慢效应之外还应叠加一个“动钟变慢”效应,与 GPS 情况类似。

有人问:图 8-17 和 8-18 中的外向光子的速率是多少?还会等于 c 吗?如果听到“不等于 c”的答案,就会说:岂不是违反光速不变原理了吗?本书 §6.2 对这一问题早已做过详细讨论,完全不必大惊小怪。

我们在小节 8.3.2 曾经推出过史瓦西时空的引力红移公式[式(8-3-33)]:

$$\frac{\lambda_1}{\lambda_2}=\frac{\Delta\tau_1}{\Delta\tau_2}=\sqrt{\frac{1-2M/r_1}{1-2M/r_2}}>1 \quad (\because r_2<r_1)。 \tag{8-9-10}$$

当时默认这个时空是某个静态恒星的外部时空(见图 8-1),所以式(8-3-33)的径向坐标 r_1 和 r_2 都要大于恒星的半径 R,自然也就大于史瓦西半径 $r_S\equiv 2M$。现在讨论的对象是黑洞。既然史瓦西黑洞在视界外面也是史瓦西时空,引力红移公式(8-9-10)对它同样成立,这时光源的径向坐标 r_2 只要大于 $2M$ 就可以[不再受到恒星半径 R 的限制(根本没有恒星)]。用下式定义相对红移量(简称**红移**,记作 z):

$$z\equiv\frac{\lambda_1-\lambda_2}{\lambda_2}=\frac{\lambda_1}{\lambda_2}-1。\tag{8-9-11}$$

如果接收者的径向坐标 r_1 保持不变，z 就是光源的径向坐标 r_2 的函数，记作 $z(r_2)$，由式(8-9-10)易见 $z(r_2)$ 是常减函数[$\mathrm{d}z(r_2)/\mathrm{d}r_2<0$]，而且 $\lim\limits_{r_2\to 2M}z(r_2)=\infty$。这说明：①光源的位置越是靠近视界，红移越大；②当光源无限逼近视界时红移趋向无穷。[但每次实验时光源位置总是固定而不是变化的(光源不动)，因为式(8-9-10)只适用于静态光源和接收者。] 因此，史瓦西黑洞的视界又起着**无限红移面**的作用。有些初学者爱问：当光源位于视界上时，红移是否就是无限大？这个问题含有两个错误。①如果光源真的位于视界上，它所发出的外向光子世界线将永远停在视界的世界面上，永远到不了视界外的接收者，也就无红移可言；②红移 z 是接收者测出的一个确定数值，而“无限大”不是一个确定数值，“红移是无限大”的提法根本没有意义。

第9章　虫洞和时间机器

美国康奈尔大学的天体物理学家**萨根**(Sagen)在1985年出版的著名小说《接触》(Contact)中描写了一个通过虫洞作快速星际旅行的故事。小说初稿写的是穿越黑洞的旅行,出版前,作者曾将书稿寄给美国加州理工学院的相对论专家**索恩**(Thorne),请他协助把小说中涉及引力物理的部分尽可能加以准确化。索恩对黑洞非常熟悉,深知穿越黑洞的旅行几乎无望,便想试着把黑洞改为虫洞(wormhole)。此事促使索恩开始认真研究并探索有关虫洞的问题。这一研究硕果累累:他不但找到可供星际旅行的虫洞解,而且发现虫洞可以被转化为时间机器。时间机器又称时间隧道,本是纯科幻小说的话题,说的是人类可以利用时间机器实现回到过去的时间旅行,出现"今日出门昨夜归"的咄咄怪事,由此自然诱发出一系列的因果疑难。严肃的物理学家过去从来不曾相信过时间机器能够存在。直到1988年,索恩才使这一问题正式登上严肃科学的殿堂,成为广义相对论的一个研究课题。本章介绍虫洞和时间机器,为此要先从嵌入图讲起。

§9.1　嵌　入　图

广义相对论经常讨论弯曲空间,嵌入图是直观地反映空间弯曲性的一种工具,会经常被用到。本节通过两个例子讲解嵌入图的绘制原则、方法和结果。

例1　静态球对称恒星内外空间的嵌入图

我们早已知道静态球对称恒星外部时空的几何由史瓦西真空解

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1}dr^2 + r^2(d\theta^2+\sin^2\theta d\varphi^2), \quad R\leqslant r<\infty \tag{9-1-1}$$

描述,其中 M 和 R 分别是恒星的质量和半径。星内时空的几何则要复杂得多,此处只介绍结论:星内线元可以表为

$$ds^2 = -e^{2A(r)}dt^2 + \left[1-\frac{2m(r)}{r}\right]^{-1}dr^2 + r^2(d\theta^2+\sin^2\theta d\varphi^2), \quad 0<r\leqslant R, \tag{9-1-2}$$

其中 $m(r)$ 和 $A(r)$ 代表 r 的两个函数,分别满足

$$m(R)=M, \qquad e^{2A(R)}=1-2M/R, \tag{9-1-3}$$

上式保证星内外线元 ds^2 在恒星表面的连续性。物理地说，$m(r)$ 代表半径为 $r(\leqslant R)$ 的任一同心球内的质量（包括静能、内能和引力势能）。函数 $m(r)$ 和 $A(r)$ 本来只在星内（即 $0<r\leqslant R$）有意义，但是，只要在星外补充定义：

$$m(r)\equiv M,\quad e^{2A(r)}\equiv 1-2M/r,\quad \text{对 } R\leqslant r<\infty, \tag{9-1-4}$$

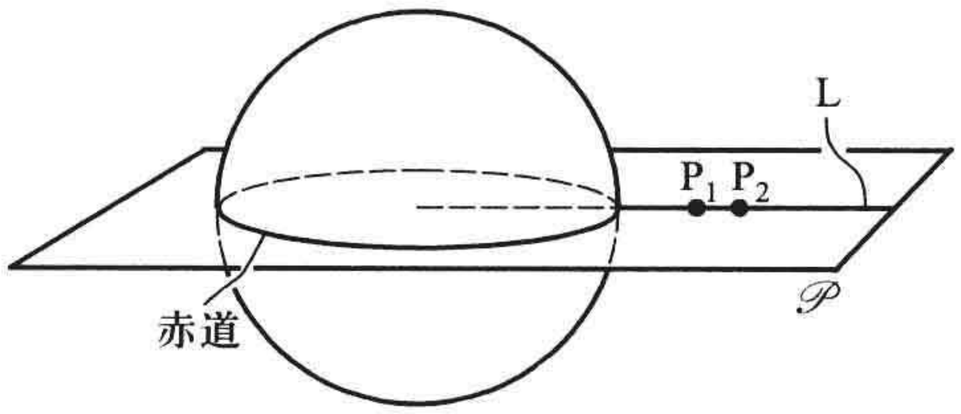


图 9-1 星球的赤道面 $\mathscr{P}$

则式（9-1-2）就可看作适用于恒星内外的线元表达式。式中 dt^2 的系数显然为负，故 t 代表时间坐标；讨论表明（从略）总有 $2m(r)<r$，故 dr^2 的系数为正，因而 r 代表空间坐标。

仿照 §8.1，我们用史瓦西坐标系 $\{t,r,\theta,\varphi\}$ 对恒星内外时空做“3+1 分解”，每个等 t 面 Σ_t 代表 t 时刻的“全 3 维空间”。式（9-1-2）在 Σ_t 上诱导出如下的 3 维线元：

$$d\hat{s}^2=\left[1-\frac{2m(r)}{r}\right]^{-1}dr^2+r^2(d\theta^2+\sin^2\theta d\varphi^2), \tag{9-1-5}$$

它描述 Σ_t 的空间几何，因子 $1-2m(r)/r$ 的存在表明这一几何是弯曲几何（3 维黎曼张量非零）。由于有球对称性，不妨只关心赤道所在平面 $\mathscr{P}$ 上的情况（图 9-1）。$\mathscr{P}$ 上各点有 $\theta=\pi/2$，故面上任一曲线的任一元段有

$$d\hat{\hat{s}}^2=\left[1-\frac{2m(r)}{r}\right]^{-1}dr^2+r^2d\varphi^2, \tag{9-1-6}$$

设 L 是赤道面 $\mathscr{P}$ 上的任一径向（$\varphi=$ 常数）线段，P_1，P_2 是 L 上任意两个邻点，径向坐标分别为 r_1，r_2。令 $dr\equiv r_2-r_1$，以 $dl|_{P_1P_2}$ 代表 P_1，P_2 之间的距离，则［也可参见式（8-1-6）］

$$dl|_{P_1P_2}=\sqrt{d\hat{\hat{s}}^2}=\left[1-\frac{2m(r)}{r}\right]^{-1/2}dr>dr。 \tag{9-1-7}$$

这个 $dl|_{P_1P_2}>dr$ 正是 Σ_t 的内禀弯曲性的关键体现。为了直观地把它表现出来，可以借助于 3 维欧氏空间。3 维欧氏线元在柱坐标系 $\{z,r,\varphi\}$ 的表达式为

$$ds^2(\text{欧})=dz^2+dr^2+r^2d\varphi^2, \tag{9-1-8}$$

我们想在 3 维欧氏空间的 $z-r$ 面上画出一条曲线，使得与 r 轴上的 r_1r_2 段对应的曲线段 Q_1Q_2 的欧氏线长等于 $dl|_{P_1P_2}$（图 9-2），如此就可从曲线的外形看出

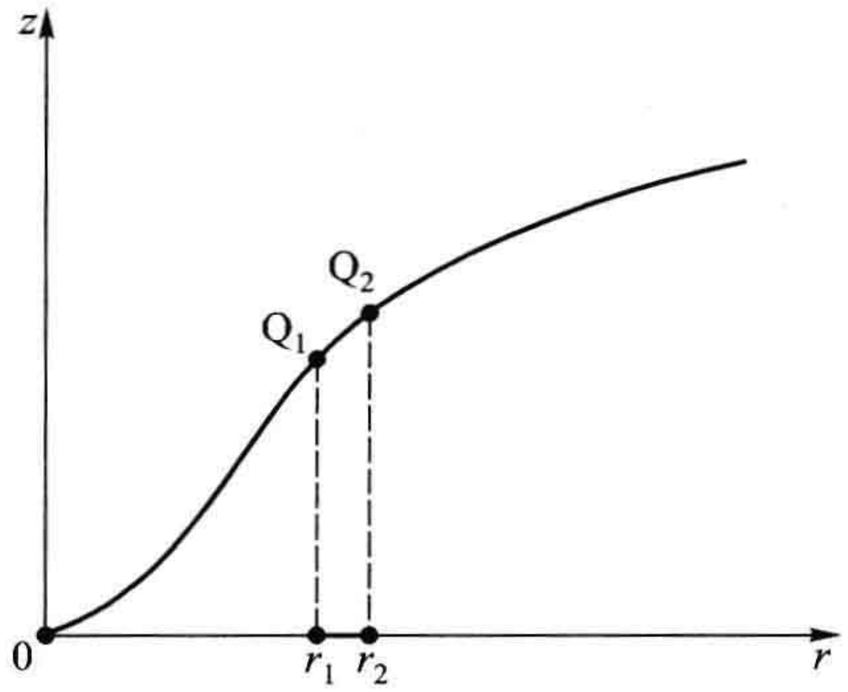


图 9－2　待求曲线的 Q_1Q_2 段的欧氏线长应等于 $\mathrm{d}l\big|_{P_1P_2}$

L 线的内禀弯曲情况。再以 z 轴为对称轴将曲线旋转一周便扫出一个曲面，其外部弯曲情况就反映赤道面 $\mathscr{P}$ 的内禀弯曲情况。这样得到的曲面叫做静态球对称恒星的嵌入图。刚才所谈的是嵌入图的绘制原则，下面介绍绘制方法。

设欧氏空间中 $z-r$ 面上待求曲线的函数表达式为 $z(r)$，则由式(9－1－8)可知该曲线的任一元段有

$$\mathrm{d}s^2(\text{欧}) = \left[\frac{\mathrm{d}z(r)}{\mathrm{d}r}\mathrm{d}r\right]^2 + \mathrm{d}r^2 + r^2\mathrm{d}\varphi^2 = \left[\left(\frac{\mathrm{d}z}{\mathrm{d}r}\right)^2 + 1\right]\mathrm{d}r^2 + r^2\mathrm{d}\varphi^2, \tag{9-1-9}$$

将上式与式(9－1－6)对比可知，为使 Q_1Q_2 段的欧氏线长等于 $\mathrm{d}l\big|_{P_1P_2}$，只需

$$\left[\left(\frac{\mathrm{d}z}{\mathrm{d}r}\right)^2 + 1\right]\mathrm{d}r^2 + r^2\mathrm{d}\varphi^2 = \left[1 - \frac{2m(r)}{r}\right]^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\varphi^2, \tag{9-1-10}$$

为此又只需

$$\left(\frac{\mathrm{d}z}{\mathrm{d}r}\right)^2 + 1 = \left[1 - \frac{2m(r)}{r}\right]^{-1}, \quad \text{即} \quad \left(\frac{\mathrm{d}z}{\mathrm{d}r}\right)^2 = \frac{2m(r)}{r-2m(r)}, \tag{9-1-11}$$

因而

$$\frac{\mathrm{d}z(r)}{\mathrm{d}r} = \sqrt{\frac{2m(r)}{r-2m(r)}}。 \tag{9-1-12}$$

约定 $z(0)=0$，则

$$z(r) = \int_0^r \sqrt{\frac{2m(r')}{r'-2m(r')}}\mathrm{d}r', \qquad (\text{对 } 0 < r < \infty)。 \tag{9-1-13}$$

因为 $r \geqslant R$ 时 $m(r)=M$[见式(9－1－4)]，所以对 $r \geqslant R$ 有

$$z(r) = \int_0^R \sqrt{\frac{2m(r)}{r-2m(r)}}\mathrm{d}r + \int_R^r \sqrt{\frac{2M}{r'-2M}}\mathrm{d}r' =$$

$$\int_0^R \sqrt{\frac{2m(r)}{r-2m(r)}}\mathrm{d}r + \sqrt{8M(r-2M)} - \sqrt{8M(R-2M)} = \sqrt{8M(r-2M)} + \alpha, \tag{9-1-14}$$

其中常数

$$\alpha \equiv \int_0^R \sqrt{\frac{2m(r)}{r-2m(r)}}\mathrm{d}r - \sqrt{8M(R-2M)}。 \tag{9-1-15}$$

令 $x \equiv r-2M$,则式(9-1-14)成为

$$z-\alpha = \sqrt{8Mx}, \tag{9-1-16}$$

可见,若把 $z-\alpha$ 看作 x 的函数,则函数关系是抛物线(的上半)。注意到上式只当 $r \geqslant R$ 时成立,便知函数 $z(r)$(对 $r \geqslant R$)的曲线如图 9-3 的实线所示。

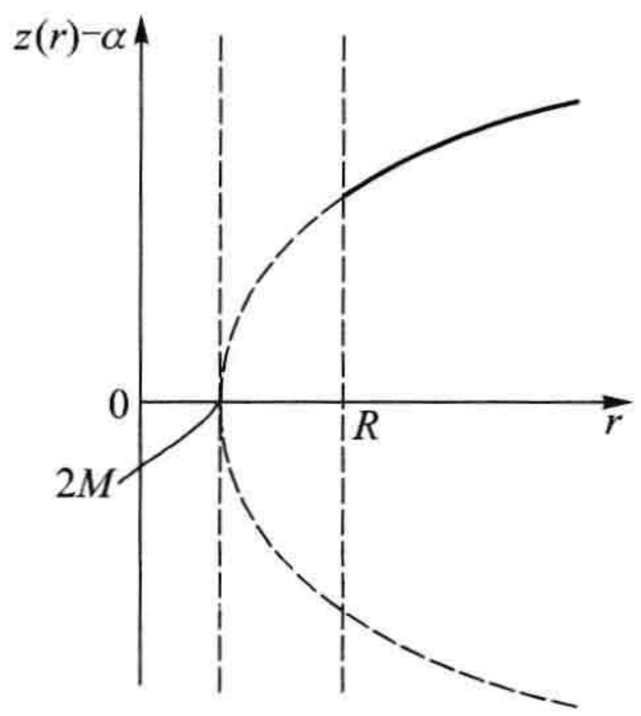


图 9-3 抛物线的实线段代表 $r \geqslant R$ 的函数 $z(r)$

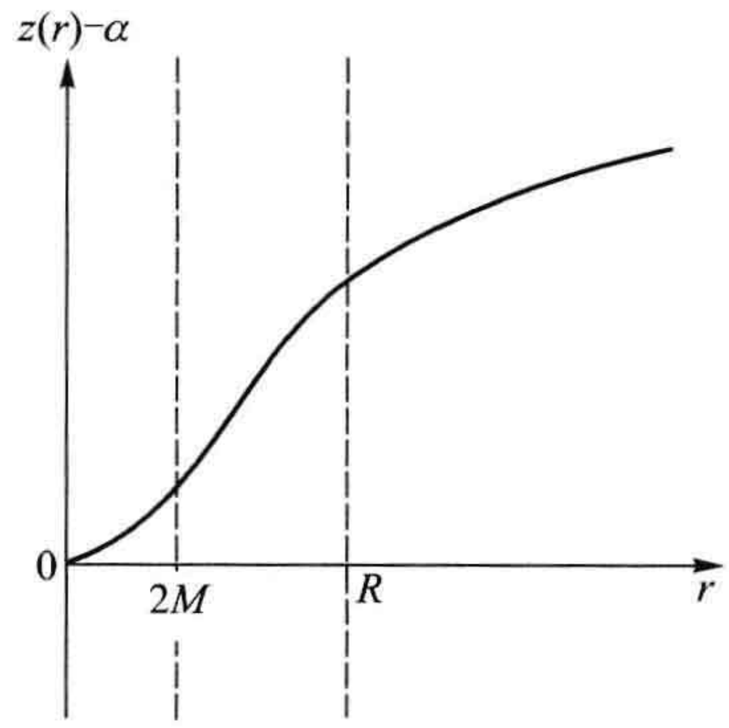


图 9-4 $z(r)$的函数曲线,其中 $r<R$ 一段只是定性画出

至于 $z(r)-\alpha$ 在 $r<R$ 时的曲线,虽然 $m(r)$的函数曲线因恒星的内部结构而异,但式(9-1-13)表明 $z(r)$是单调常增函数,不妨定性地画成图 9-4 的样子(图中曲线在 $0<r<R$ 的一段)。将该曲线以 z 轴为对称轴转一圈便得到图 9-5的“草帽状”曲面,它的外在弯曲形状代表着图 9-1 的赤道面的内禀弯曲情况,称为静态球对称恒星(内外空间)的嵌入图。请注意,嵌入图的背景欧氏空间只是为画出曲面而人为引入的,不要以为“草帽”里可以装进什么东西,真正有物理意义的点只是曲面(“有草”部分)上的点。这个曲面之所以能代表 3 维空间,是因为从一开始选赤道面时就压缩了一维。只要把图 9-5 中每个圆周看作一个球面[面积为 $4\pi r^2$,r 是该圆周的径向坐标值(半径)]。图 9-5 还形象地告诉你,时空是渐近平直的(越远越平)。想象一只昆虫从“草帽”边缘(r 很大处)沿径向朝着 r 值减小的方向爬行,它必将觉得空间弯曲得越来越甚。(这里用“昆虫爬行”来讲解是为了适应于嵌入图。在嵌入图 9-5 中,3 维空间

被压缩成2维面，所以3维运动在图中就成了贴着2维面的爬行。）

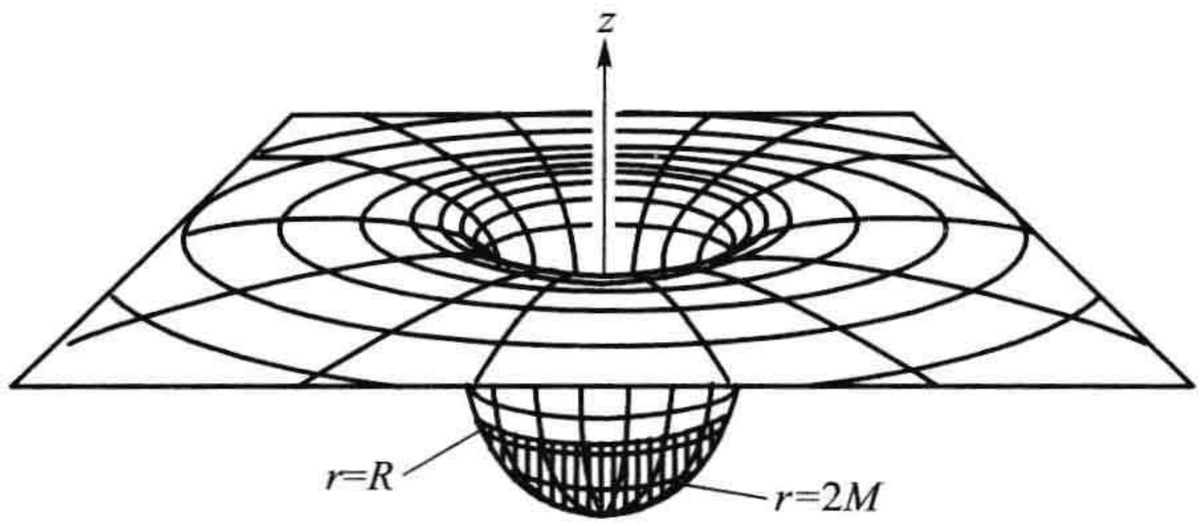


图9-5　静态球对称恒星内外空间（赤道面上）的嵌入图（压缩一维）

例2　史瓦西最大延拓时空（图8-13）中 $T=0$ 时刻的全空间的嵌入图[①]

画图前应注意图8-13的以下三点。

（1）X 轴上除原点外的 r 值都大于 $2M$（r 的范围是 $2M<r<\infty$），原点的 r 值等于 $2M$。

（2）X 轴的正负半轴各自代表一个物理空间，两者关于原点 $X=0$ 对称。正负半轴的每一点各代表一个球面，面积为 $4\pi r^2$，其中 r 是该点的径向坐标。

（3）原点 $X=0$（其 r 值为 $2M$）也代表一个球面，面积为 $4\pi(2M)^2$，是 X 轴的正负半轴所代表的两个3维空间的连接处，称为**喉**（throat）。

仿照例1的做法，以 M 代替式（9-1-11）中的 $m(r)$，便得到欧氏空间中待求曲线的函数 $z(r)$ 所应满足的方程

$$\left[\frac{\mathrm{d}z(r)}{\mathrm{d}r}\right]^2=\frac{2M}{r-M}。\tag{9-1-17}$$

与例1不同的是，现在要在一张图中表现 X 轴的正负半轴所代表的两个空间，而且它们要在 $r=2M$（喉）处连接，自然就想到用 $z>0$ 的半条曲线和 $z<0$ 的半条曲线的合成曲线作为待求曲线，所以在对上式开方时要取 ± 号：

$$\frac{\mathrm{d}z(r)}{\mathrm{d}r}=\pm\sqrt{\frac{2M}{r-M}}。\tag{9-1-18}$$

约定 $z(2M)=0$，便得

$$z(r)=\pm\int_{2M}^{r}\sqrt{\frac{2M}{r'-2M}}\mathrm{d}r'=\pm\sqrt{8M(r-2M)}。\tag{9-1-19}$$

所以 $z(r)$ 的函数曲线是图9-3的整条抛物线（连实带虚都要），绕 z 轴转一周便

---

① 由 $t=2M[\ln V-\ln(-U)]$［式（8-7-31）］及 $V=T+X$，$U=T-X$ 可知 $T=0$ 对应于史瓦西时间 $t=0$，所以 Σ_0 也就是 $t=0$ 时刻的全空间。

得到图 9 -6 的“线轴状”嵌入图。与图 9 -5 类似,①只有“线轴”上的点才有物理意义;②设该点的径向坐标值为 r,则该点所在的圆周代表一个面积为 $4\pi r^2$ 的球面。

观察图 9 -6 时应特别注意以下三点。

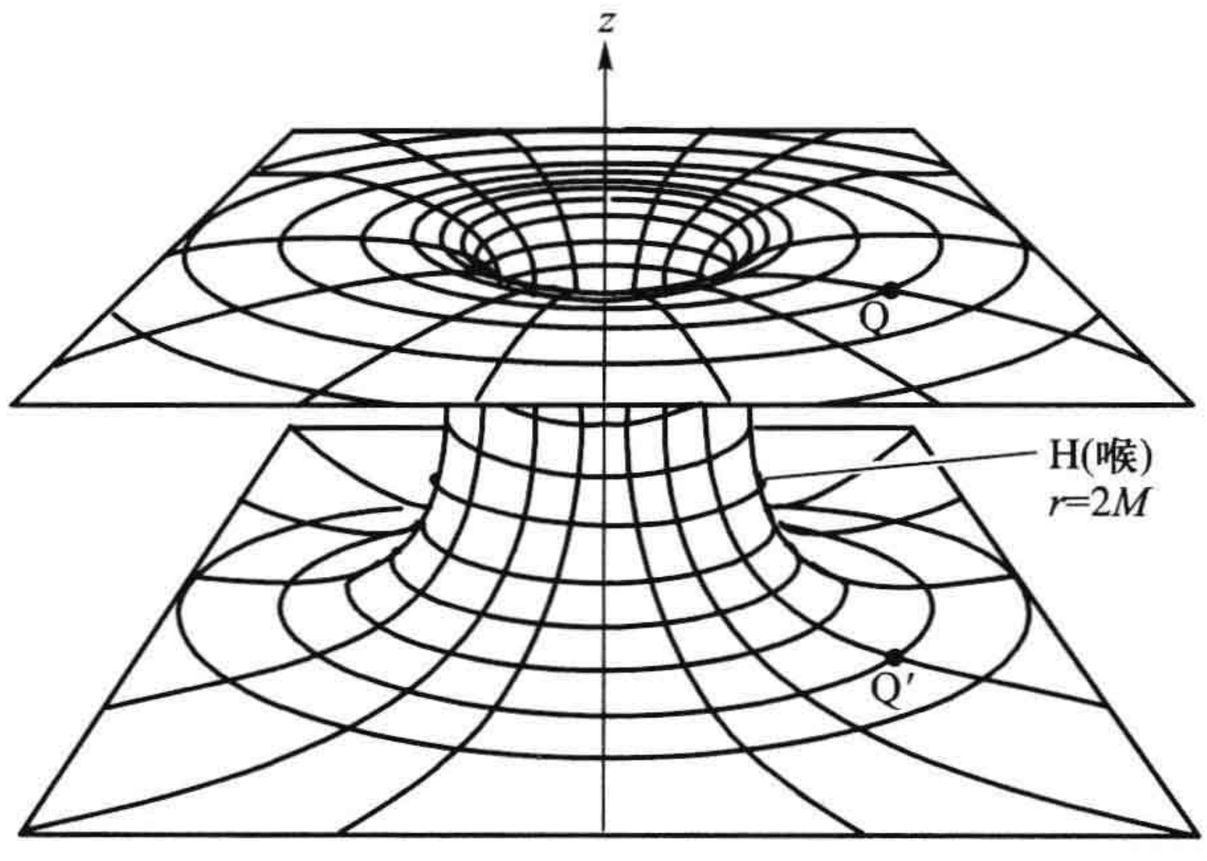


图 9 -6　史瓦西最大延拓时空 $T=0$ 时刻的嵌入图
(压缩一维)

(1) $r=2M$ 的圆周 H 代表喉[实际上是个面积为 $4\pi(2M)^2$ 的球面]。“线轴”上除 H 外的点的 r 值都大于 $2M$,故 r 在“线轴”上的取值范围是 $2M \leqslant r < \infty$。

(2) 喉以上和以下的两半个“线轴”分别就是 X 轴的正、负半轴所代表的互无因果联系的空间,由图易见这两个空间都渐近平直(越远越平)。

(3) Q 点(类似地, Q′点)所在圆周代表一个面积为 $4\pi r_Q^2$ 的球面,相当于图 8 -13(b)中 p 点(类似地, p'点)所代表的球面。

在此基础上,以下两节就可介绍虫洞和时间机器这一有趣话题,主要参考资料是 Thorne(1994),有中译本,详见书末参考文献。

§9.2　虫　　洞

9.2.1　史瓦西虫洞

再次观察图 9 -6。在“线轴”上半部分选定某个圆周(其实是球面,其 r 值比 $2M$ 略大),例如 Q 点所在的圆周,则“线轴”的下半部分必有一个 r 值相同的圆周,介于这两个圆周之间的“线轴”部分就称为一个**虫洞**(wormhole),这两个球面则称为两个**洞口**(mouth)。洞口的选法有一定的任意性[你也可选 r 值比

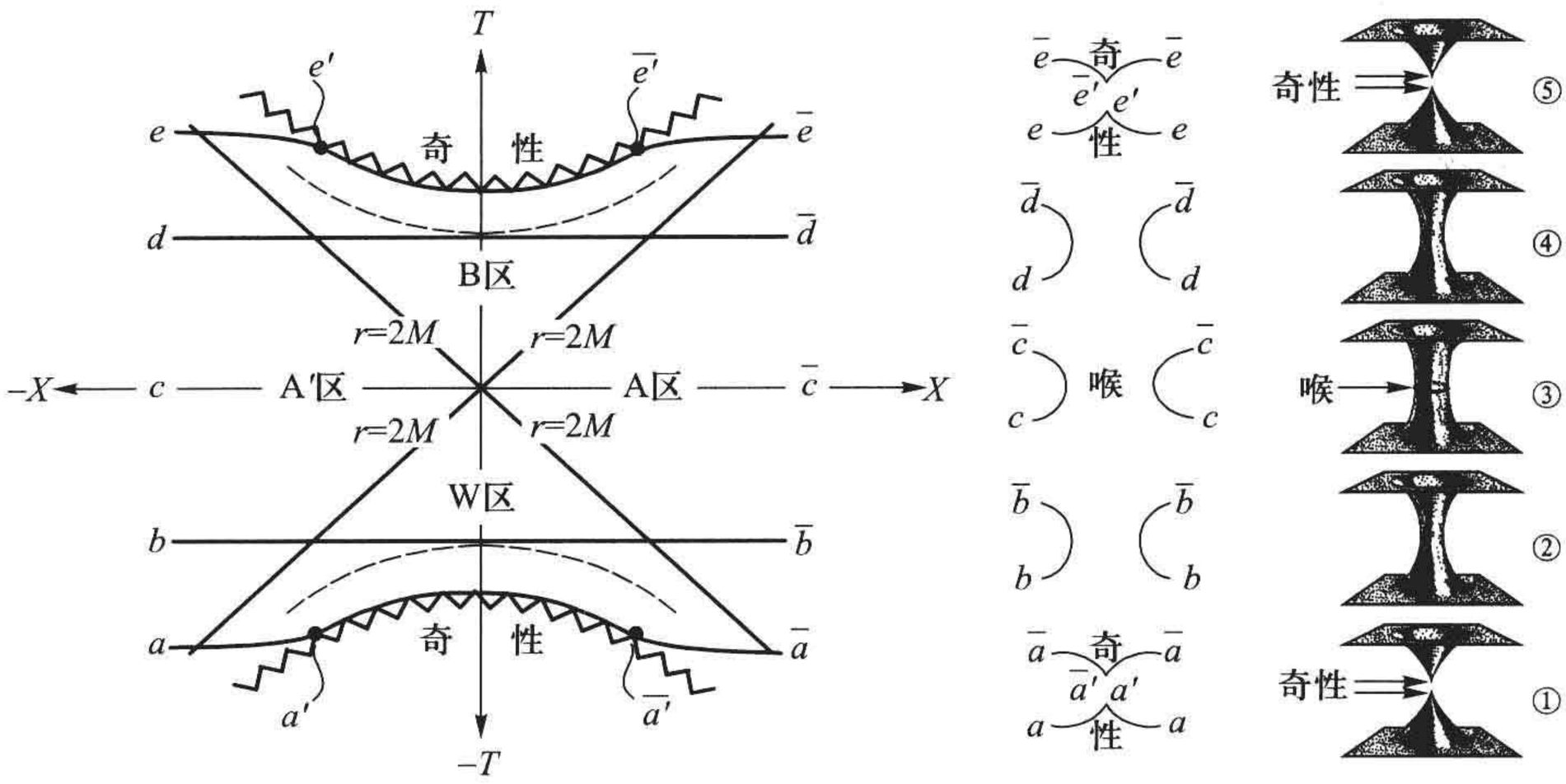
r_Q 略大(或小)的圆周为洞口],所以"洞口"的概念有一定的含糊性,不过它们一定对称地分居于喉(其 $r=2M$)的两侧,而且 r 值不能比 $2M$ 大很多。

史瓦西最大延拓时空的 A 区和 A′区(见图 8-13)是两个渐近平直区,越远(r 越大)越平直。然而这一延拓不能排除如下可能性:A 区和 A′区在 r 很大处连在一起,如图 9-7 的嵌入图所示。物理上(包括使用爱因斯坦方程)无法排除两者中之任一,所以不妨就图 9-7 进行讨论,并暂时把它想象成我们的宇宙。根据天文测量,地球与织女星的距离是 26 光年,假定你坐飞船从地球飞到织女星,就算飞船的速率高达 $0.9c$,也要花费将近 29 年。但是,如果宇宙中有这么一个虫洞,它的两个洞口分别在地球和织女星附近(图 9-7),你就有望从地球出发经过洞内空间(沿着"线轴"的内表面)到达织女星。由于洞内距离可以很小(例如只有 1 km),虫洞似乎能为星际旅行提供一条捷径! 然而,这只是无法实现的奢望,关键在于史瓦西最大延拓时空的 B 区和 W 区是动态时空区[见式(8-7-37)后的注记 2],虽然虫洞在 $T=0$ 时刻是张开的(图 9-6),但在其他时刻不尽如此。设想一只昆虫从洞口 M 的 Q 点出发沿着该点所在的抛物线向下爬行至喉 H,其 r 值将从 $r_Q>2M$ 开始渐减至 $2M$,越过喉后 r 值又复增加,到另一洞口的 Q′点时增至 $r_{Q'}=r_Q$。然而爬行(运动)是需要时间的(不可能在同一时刻 $T=0$ 完成这个运动),所以还必须考察虫洞在其他时刻的表现。

图 9-7　两个渐近平直区在很远处连在一起,虫洞提供一条从地球到织女星的捷径。

图 9-8(a)示出五张类空面,即 $a\bar{a}$, $b\bar{b}$, $c\bar{c}$, $d\bar{d}$和 $e\bar{e}$,每张面都可看作某一时刻的全空间。$c\bar{c}$面就是 $T=0$ 的全空间,其嵌入图早已示于图 9-6,现在又重画成图 9-8(c)之③。与其他各张类空面相较,虫洞在 $c\bar{c}$面有最大张开,不妨与 $d\bar{d}$面做一对比。图 9-8(a)中直线 $d\bar{d}$与 B 区的一条双曲线(虚线)相切,虚线上各点有相同 r 值,介于 $2M$ 与 0 之间,例如 $r=1.5M$,而 B 区中位于此双曲线以下的点的 r 值都大于 $1.5M$,可见 r 在直线 $d\bar{d}$上的取值范围是 $1.5M\leqslant r<\infty$,所以虫洞在这一时刻的张开程度比 $c\bar{c}$时要小,如图 9-8(c)之④。类似地可以理解 $b\bar{b}$面,如图 9-8(c)之②。$a\bar{a}$和 $e\bar{e}$代表两张最为特殊的类空面,从 a

到 a'(以及从 $\bar{a}$到 $\bar{a}'$)会遇到奇性,虫洞根本就不张开,上、下两个区域[即图9-8(a)的左、右两个区域]互相分离。



(a) 五张类空面 $a\,\bar{a}, b\,\bar{b}, c\,\bar{c}, d\,\bar{d}$和 $e\,\bar{e}$　　(b) 嵌入图截面　　(c) 嵌入图

图 9-8　史瓦西最大延拓时空的动态研究

小结　按时间顺序而言,虫洞先是不张开(有奇性),然后逐渐张开,到 $T=0$ 时张至最大,继而又渐渐收窄,最终又完全不张开(有奇性)。人或昆虫是否能够趁着虫洞张开的短暂时间迅速从 A 区"爬到" A′区? 不可能,因为由图 9-8(a)可知,从 A 区出发的任一类时(甚至类光)曲线都不可能进入 A′区(只能进入 B 区),而人和昆虫的世界线只能是类时曲线。直观地说,这意味着人和昆虫的运动都慢于虫洞的开合,所以说史瓦西虫洞是不可穿越的虫洞。不可能利用这种虫洞进行星际旅行。

9.2.2　可穿越虫洞

那么,宇宙中是否可能存在可穿越的虫洞? 为了简化问题,不妨先讨论球对称的情况。根据伯克霍夫定理(见 §8.8),真空爱因斯坦方程的球对称解必为史瓦西线元,而史瓦西虫洞是不可穿越的,所以要获得可穿越虫洞就得放弃真空条件,就是说,可穿越虫洞所在的时空一定存在物质场。物理地说,我们正是要利用物质的引力来撑开虫洞。更有甚者,计算表明能撑开虫洞的物质是极其特别的,是经典物理学认为不可能存在的物质,因此被称为**奇异物**(exotic matter)。研究还证明,无论球对称与否,任何可穿越虫洞都要靠奇异物撑开。既然经典物理学不容许存在奇异物,自然就想到量子物理学。根据量子场论,经典物理学认

为空无一物的真空其实存在着量子涨落，直观地说，就是某些地方（甲处）会自发地在短期内出现正的能量密度，这些能量是从别处（乙处）暂时借来的，于是乙处就有了负的能量密度。很快地，乙处的能量密度由负变正，这份能量又是从他处借来的。总之，虽然全空间的能量密度平均为零，但空间各处的能量密度都在快速地时涨时落，所以称为**真空涨落**。研究表明，在适当的弯曲时空区域中（例如视界附近），真空涨落可能被扭曲，从而相当于奇异物。由于至今尚未建成完整的量子引力论，物理学家对这些问题尚未取得完全肯定的结论，但至少可以说，量子物理学未必像经典物理学那样禁绝奇异物。在暂时默认量子物理学容许奇异物存在的前提下，一些物理学家就开展了对于可穿越虫洞的研究。索恩及其学生找到了爱因斯坦方程的一批简单的（静态、球对称的）虫洞解。下面介绍其中之一[Morris and Thorne(1988)]。该解在坐标系$\{t,l,\theta,\varphi\}$的线元为

$$ds^2 = -dt^2 + dl^2 + (b^2 + l^2)(d\theta^2 + \sin^2\theta d\varphi^2)，\text{其中 } b \text{ 为常数，} \tag{9-2-1}$$

坐标的取值范围是$-\infty < t,l < \infty$，$0 \leqslant \theta \leqslant \pi$，$0 \leqslant \varphi \leqslant 2\pi$。定义函数$r^2(l) \equiv b^2 + l^2$，则$dl = rdr/l$，代入式(9-2-1)得

$$ds^2 = -dt^2 + \frac{r^2}{l^2}dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2)。 \tag{9-2-2}$$

由此便可（仿照例1的做法）画出$t=0$时刻的嵌入图，为此只需将式(9-1-11)改为下式：

$$\left[\frac{dz(r)}{dr}\right]^2 + 1 = \frac{r^2}{l^2}。 \tag{9-2-3}$$

于是

$$\left[\frac{dz(r)}{dr}\right]^2 = \frac{r^2 - l^2}{l^2} = \frac{1}{(r/b)^2 - 1}，$$

故

$$\frac{dz(r)}{dr} = \pm\frac{1}{\sqrt{(r/b)^2 - 1}}。 \tag{9-2-4}$$

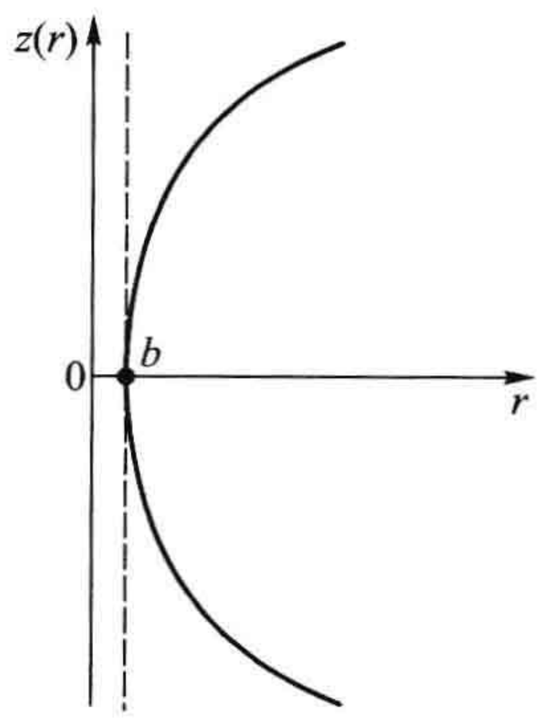


图9-9　粗实线代表式(9-2-5)的函数$z(r)$

上式表明r不能小于b，而且，当r从大于b出发趋于b时$dz(r)/dr \to \pm\infty$，可见曲线$z(r)$在$r=b$处的切向是竖直的，暗示$r=b$处很可能是喉的所在处。不失一般性，选$z(b)=0$，对式(9-2-4)积分给出

$$z(r) = \pm\int_b^r \frac{dr'}{\sqrt{(r'/b)^2 - 1}} = \pm b\ln\left[\frac{r}{b} + \sqrt{\left(\frac{r}{b}\right)^2 - 1}\right]。 \tag{9-2-5}$$

函数 $z(r)$ 的曲线如图 9-9 的粗实线所示，绕 z 轴旋转一周便得到线元(9-2-1)所在时空的嵌入图。由于静态，虫洞每时每刻都像图 9-6 那样张开，因而是可以穿越的。

前已说过，可穿越虫洞的存在必然涉及奇异物。事实上，如果从线元(9-2-1)出发计算该时空的曲率，再代入爱因斯坦方程便会发现该时空充满物质场，而且它的确是经典物理学所不容许的奇异物。

§9.3 时 间 机 器

9.3.1 把虫洞转化为时间机器

嵌入图表现的是空间弯曲情况，但时间机器必然涉及时间，所以还要画时空图。先介绍静态虫洞的时空图。虫洞有两个洞口，分别记作 M 和 N(参见图 9-7)。先看洞口 M，它本是个 2 维球面，压缩一维后是个圆周。由于静止，其世界面可画成竖直圆柱面(图 9-10)。喉的世界面是位于洞口 M 世界面之内的同轴圆柱面。其实，图 9-7 中洞口 M 附近的每个圆周在图 9-10 中都表现为一个同轴圆柱面，只是没有画出。为了简单醒目，不妨用一根粗的竖直线代表喉的圆柱面(即把喉看成一个点)，并在竖直线上画上许多圆点，圆点旁的数字代表喉的固有时，如图 9-11。还可把图 9-11 再压缩一维，成为 2 维时空图，洞口的世界面表现为两条竖直线(图 9-12)。现在遇到一个问题：洞口 M 的世界面

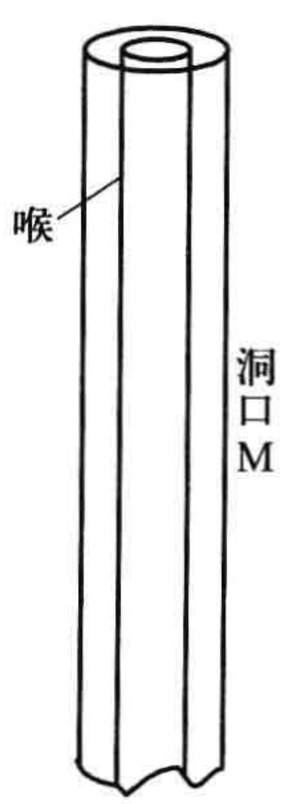


图 9-10 洞口 M 的世界面和喉的世界面(压缩一维)

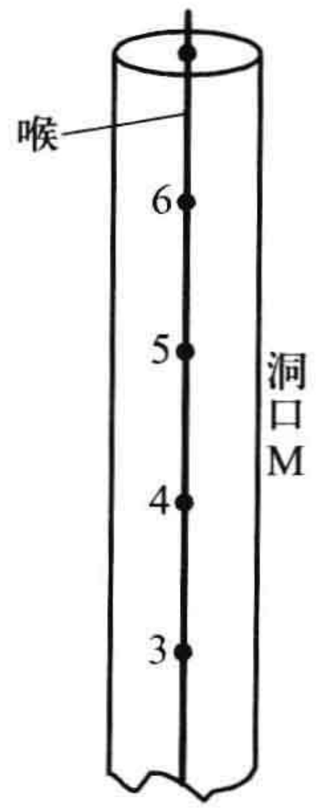


图 9-11 喉的世界面简画成世界线，圆点旁数字为固有时。

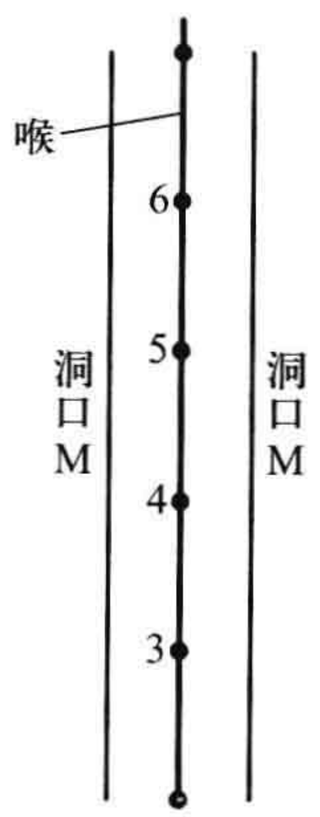


图 9-12 把图 9-11 再压缩一维，成为 2 维时空图。洞口世界面表现为两条竖直线。

既然在竖直方向上画成无限长，洞口 N 的世界面又能怎么画？办法是在图 9－12 的右边再画上类似的 3 条竖直线，粗直线仍代表喉，旁边的两条代表洞口 N，成为图 9－13。这岂非有两个喉？不是的，因为我们还约定要将两条粗直线“认同”。准确地说，就是要把两条线上固有时相等的一对点认为是同一时空点。

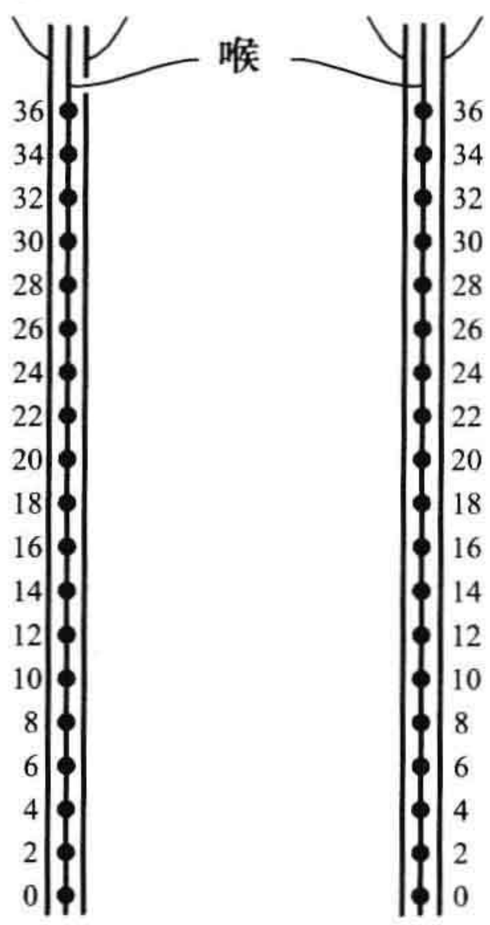


图 9－13　两条粗实线认同后代表喉，圆点旁的数字是喉的固有时。固有时相等的两点代表同一时空点。

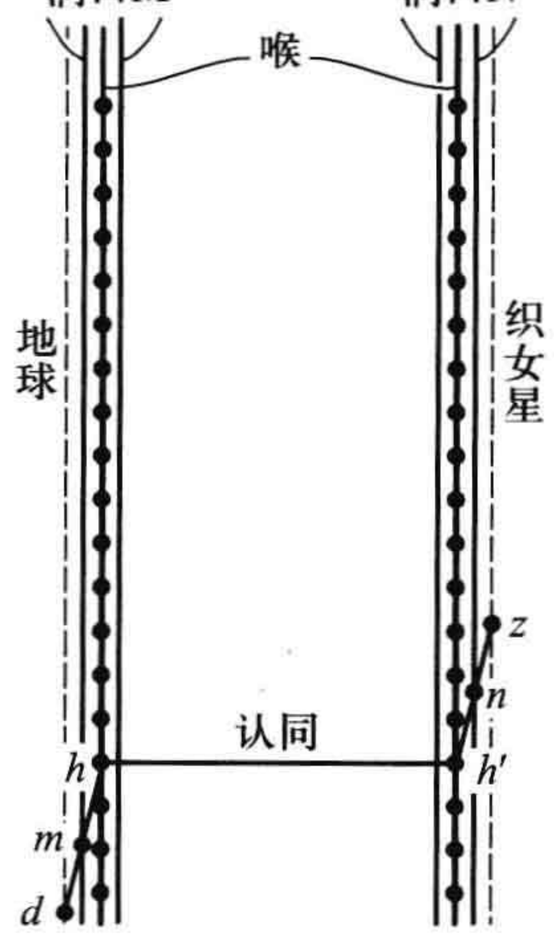


图 9－14　王五从地球经虫洞到织女星的世界线

为帮助读者进一步理解虫洞，我们把图 9－7 看作含静态虫洞的空间，并讨论如下过程。某人（王五）从图 9－7 的地球出发经洞口 M 进入虫洞，先“爬”到喉处，再“爬”到洞口 N，出洞后很快就“爬”到织女星。这一“捷径式星际旅行”也可借助于图 9－14 做 4 维描述：王五从地球出发（事件 d），很快到达洞口 M（事件 m），接着走到喉（事件 h）。由于两粗线对应点认同，也可说他到了右边粗线的 h' 点（h 与 h' 是同一时空点），然后再到洞口 N（事件 n），由此出洞并很快到达织女星（事件 z）。

讨论虫洞时务必注意洞外和洞内有非常不同的几何。例如，不妨认为图 9－7 的洞外几何是闵氏几何（事实上，我们的 3 维宇宙可以认为是平直的），但洞内几何由于存在物质（而且是奇异物）而非常弯曲。所谓静态虫洞，是指两个洞口相对于洞外时空的某个惯性系 $\mathrm{K}\equiv\{T,X,Y,Z\}$ 为静止，即两洞口的世界面都是竖直的（图 9－13）。现在，假定洞口 N 从洞外时刻 $T=0$ 开始相对于 K 系做加速运动，加速至接近光速后掉头返航并做减速运动，直至 $T=31$ 回到原位后

再度静止,相应的时空图是图 9 – 15,它非常类似于双子效应的时空图[图 4 – 16(b)]。仿照双子问题中外出兄弟固有时的计算,不难接受图 9 – 15 中洞口 N 内喉边所标的固有时数字。切勿忘记左右两条喉线(一直一弯)数字相同的点要认同。借助于这样的“加速虫洞”就可以实现回到过去的时间旅行。图 9 – 16 示出这一激动人心的旅行过程。旅行者王五原先一直静止于洞口 N 附近,所以他的固有时 τ 一直等于洞外惯性坐标时 T。设他在洞外时刻 $T=31$ 时从洞口 N 进入虫洞,则他进洞时(事件 p)的固有时自然是 $\tau_p=31$(不妨说他进洞时为 31 岁),与洞外日历显示的惯性坐标时 T_p 一样,即 $\tau_p=T_p=31$。假定洞口与喉非常靠近,他几乎不用时间就从洞口 N 到达喉,到达事件记作 h,由图可知喉在 h 点的固有时为 20。以 h'代表洞口 M 内竖直喉线上固有时为 20 的点(见图 9 – 16),则 h 和 h'代表同一时空点,所以王五到达 h 点也就是到达 h'点。接着,他很快又到达洞口 M 并经 M 出洞,出洞事件记作 q。由于洞内运动几乎不占时间,王五在事件 q 的固有时 τ_q 仍为 31,但由图可知 q 点(看作洞外时空的一点)的坐标时 $T_q=20$。就是说,王五出洞时发现当时当地的日历显示的是 11 年前的数字,周围全是 11 年前的事物。于是他实现了回到过去的旅行!更有甚者,

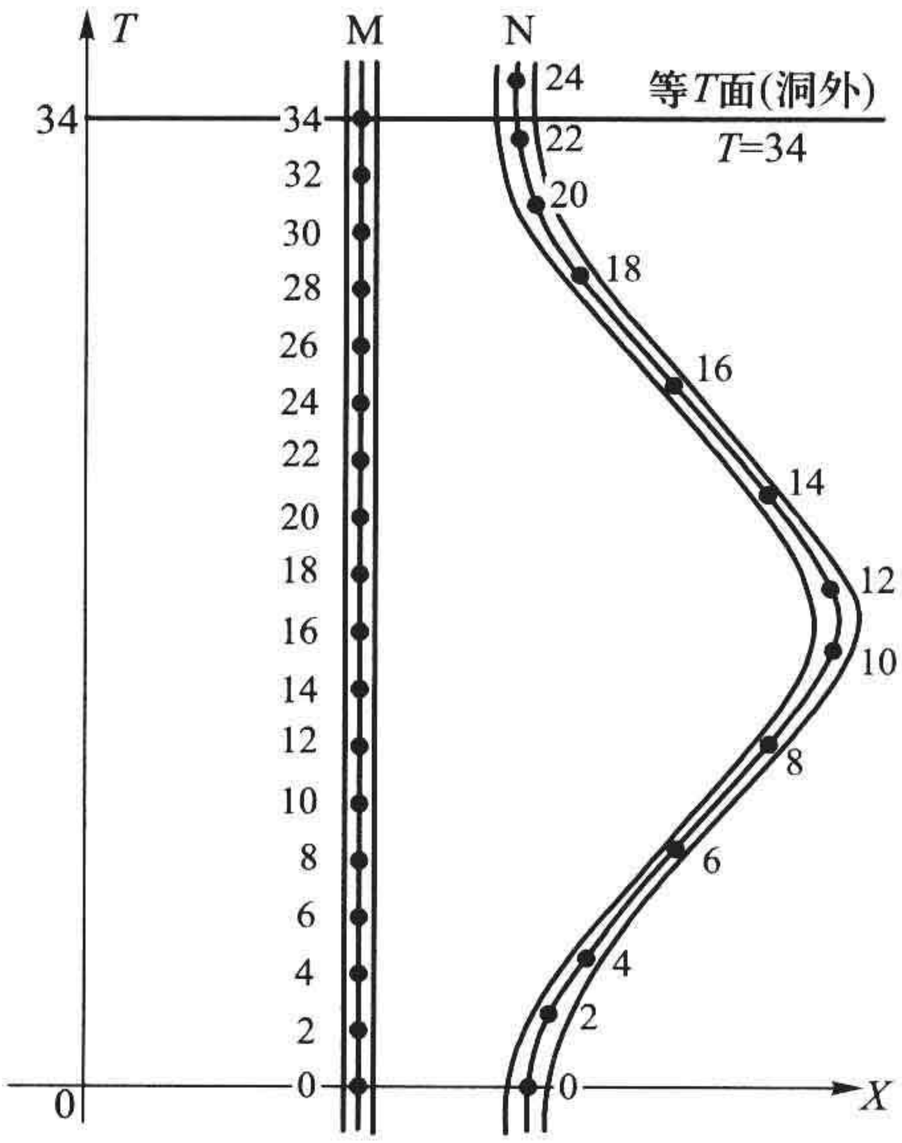


图 9 – 15 “加速虫洞”:洞口 N 从洞外时刻 $T=0$ 开始相对于 K 系做加速—调头—减速运动,直至 $T=31$ 时复位并静止。圆点旁的数字是喉的固有时。

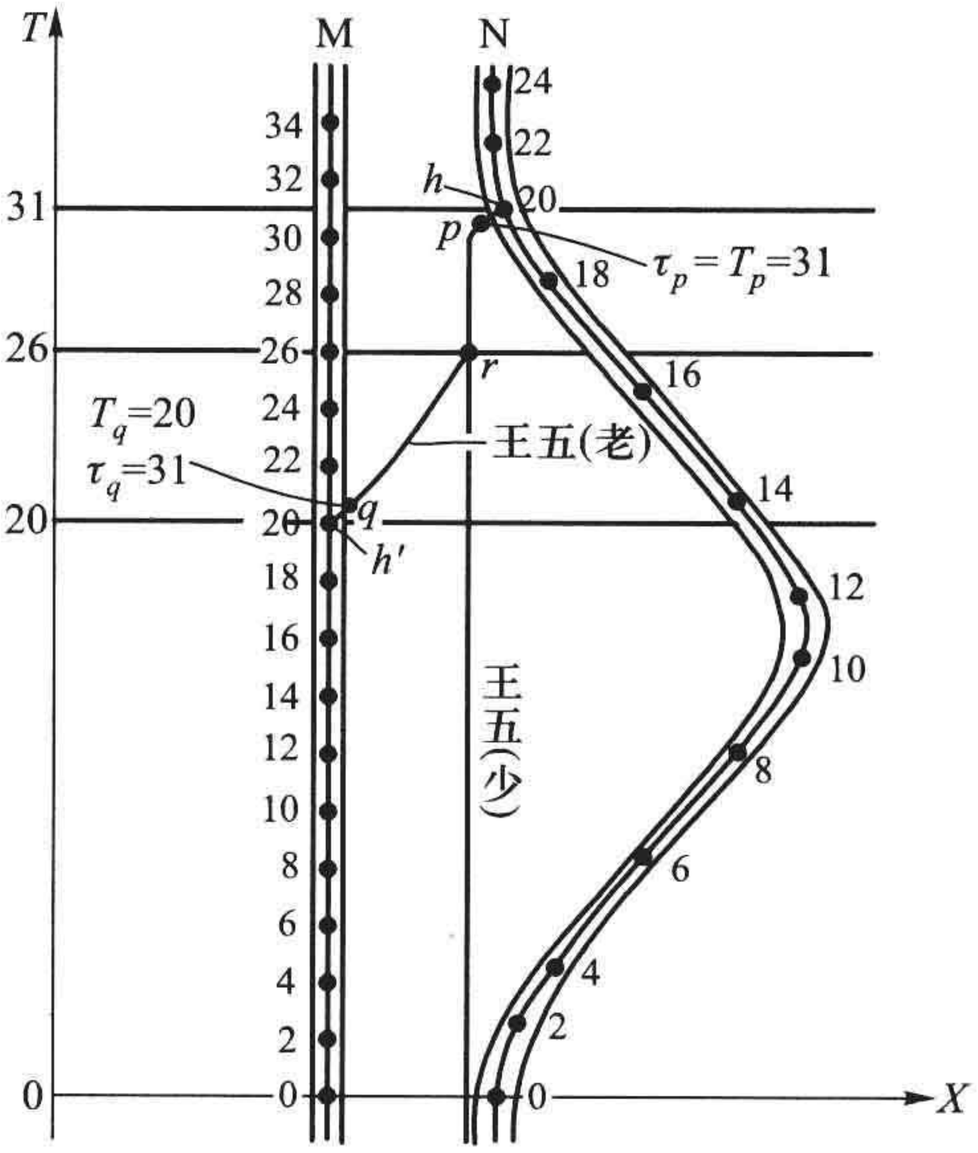


图 9 – 16 王五的时间旅行:洞内距离很短,从进洞到出洞几乎不花固有时间 τ,但进出洞时的洞外时间 T 却从 31 降至 20,所以他年轻了 11 岁!

如果他从 q 起高速向右运动(由斜置世界线代表),就能与过去的(较年轻的)自己相遇。以 r 代表相遇事件,由图可知 $T_r = 26$,因而较年轻的王五(竖直世界线)在事件 r 的固有时(年龄)τ_r(少) $= 26$。另一方面,从洞口 M 出来的王五(斜线)在事件 q 时为 31 岁(因为 $\tau_q = 31$),故到达 r 时他的固有时大于 31,即 τ_r(老) > 31,这意味着较老的王五竟然与较年轻的自己(26 岁)相遇甚至可以互相握手并互道"你好"。虫洞成了一部时间机器!过去只有纯科幻小说才敢于描述的情景现在看来真有可能实现!

9.3.2　弒母悖论

时间机器(又称时间隧道)一旦存在,就会导致一系列因果疑难。常见的一个质疑是:"如果我通过时间隧道回到出生之前并杀死我娘,那么还有我吗?既然没有我,我又怎能回去杀死我娘?"这可称为**弒母悖论**。由于这一悖论还涉及人的"自由意志"以及道德伦理等问题,讨论相当麻烦。美国德克萨斯大学的物理教授**珀尔钦斯基**(Polchinski)巧妙地提出了一个本质一样但大为简化的问题(图 9 - 17):一只台球(白色)以适当的位置和速度出发经洞口 N 进入虫洞并从洞口 M 复出(图中画成灰色),出洞时洞外时刻早于它的进洞时刻(例如早了半小时),它恰好击中进洞前的(较年轻的)自己并改变其运动方向,使之不能到达洞口 N,因而不会有台球从洞口 M 飞出并击中自己。这同弒母悖论有相同物理实质,都是回到过去并改变历史,从而造成因果疑难。

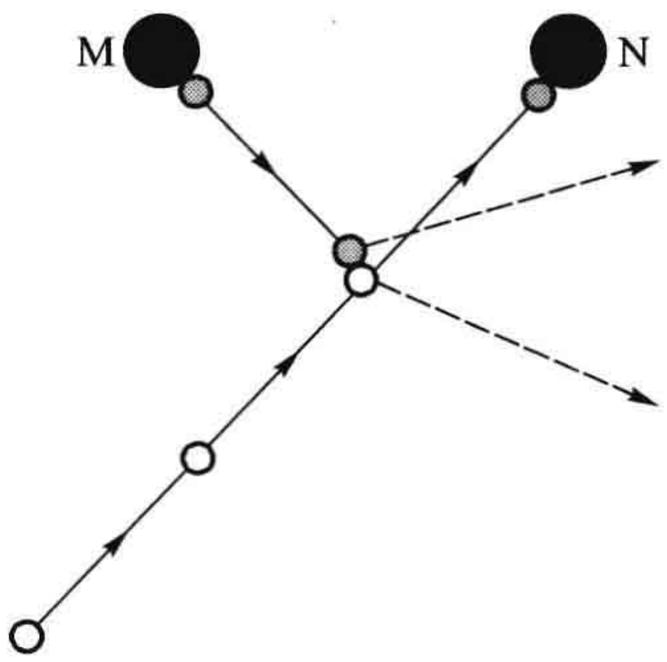
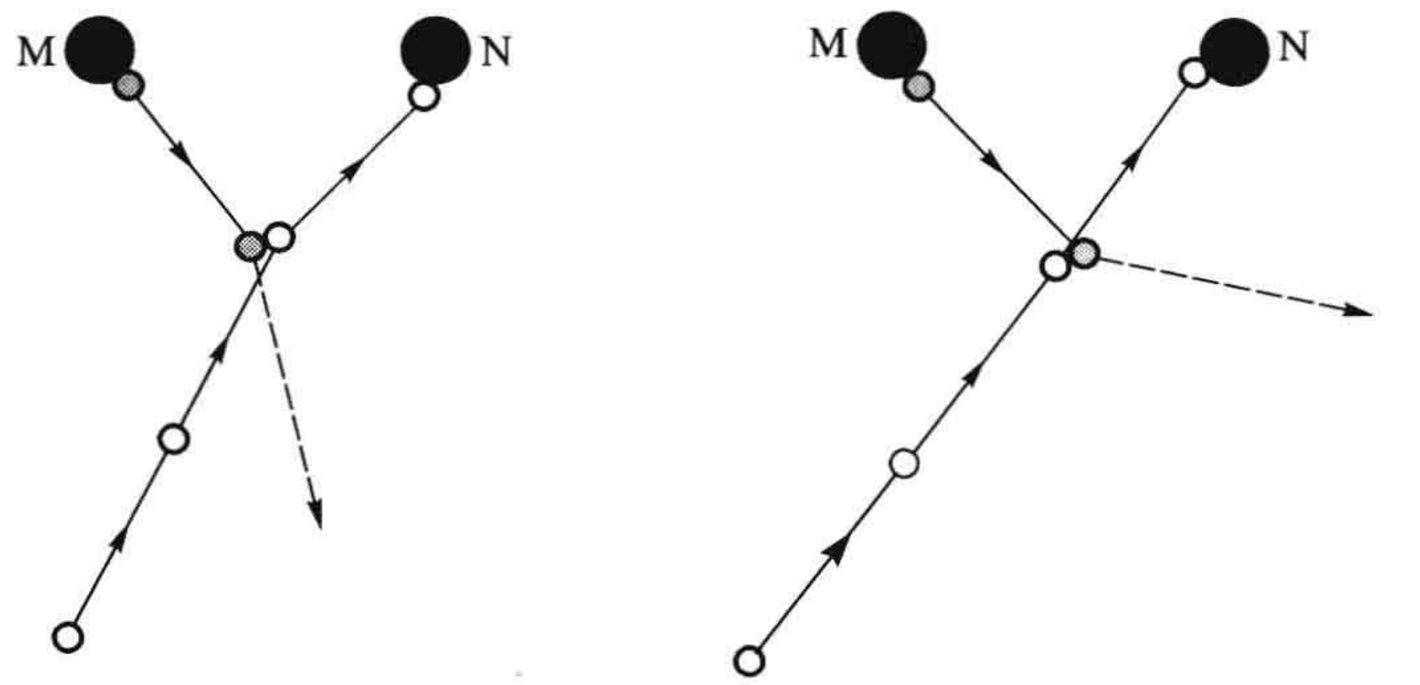


图 9 - 17　台球悖论:台球从洞口 N 进虫洞,经洞口 M 出洞并回到过去。它恰好击中进洞前的自己并使之不能进洞。

从 4 维观点看来,时间旅行者的世界线是一条闭合的类时曲线,即图 9 - 16 的 $rphh'qr$。通常时空中的类时线当然不是闭合的,它总是从过去走向未来。图 9 - 16 是索恩挖空心思地构造的时空(但原则上可能存在),其中竟然存在闭合类时线。弒母悖论的提出反映这样一种信念:闭合类时线的存在使人们可以改变过去(改变历史),从而带来因果性的严重疑难。然而许多物理学家认为并非如此。弒母(台球)悖论只是人们想象的一个非自洽过程,它的成因是没有考虑到"未来"对"过去"同样会产生影响。(事实上,闭合类时线上任何两点中之一点都可以认为另一点在自己的未来,因而可以影响对方。)许多物理学家相信任何物理过程在逻辑上一定是自洽的,相信所有过程都遵守如下原理:

自洽性原理(Principle of self-consistency) 闭合类时线上任意两个事件都以自洽的方式互相影响(一定能自动调整到这种方式):每个事件只发生一次,而且不会被改变。

只要接受这一原理,就可断定:①图 9 - 17 的非自洽台球运动不可能发生;②在相同的台球初值(初位置和初速度)下应该存在自洽解。果然,约定台球做低速运动,利用牛顿力学(非相对论力学)的能量和动量守恒定律,索恩的两个博士生通过计算的确找到了初值与图 9 - 17 相同的自洽解,而且共找到两个,分别示于图 9 - 18 的(a)和(b)。此二图与图 9 - 17 的关键区别在于,图 9 - 17 的两球相碰是强烈的正面碰撞,它明显改变较年轻台球的轨道并使之不能进洞;而图 9 - 18 的碰撞对较年轻台球的轨道只起到"微扰"作用。例如,在图 9 - 18(a)中,白色台球在行进过程中忽然遇到从洞口 M 飞出的灰色台球,灰球只是从白球的左后边缘轻轻擦过,白球仍可(沿着稍微改变了的轨道、以稍微不同的速率)到达洞口 N,当它穿过虫洞并从洞口 M 复出(成为灰球)时,位置和方向与图 9 - 17 相较稍有不同,正是这种稍微的不同保证它只是从较年轻的自己(白球)的左后边缘轻轻地擦过[详细计算见 Echeverria et al. (1991)]。这显然是个自洽的运动过程。



(a) 灰球在白球左后边缘轻轻擦过　　(b) 灰球在白球右前边缘轻轻擦过

图 9 - 18　台球运动的两个自洽解(初值与图 9 - 17 相同)

物理学家**诺维科夫**讨论了某些更为复杂的非自洽过程[Novikov(1992)],下面是其中一例。一根 Y 形管以图 9 - 19(a)的方式与虫洞的两个洞口相连,管的内壁理想光滑,使管内的活塞得以无摩擦地运动。活塞从某一初值出发沿路径 α 和 α' 到达洞口 N,穿过虫洞后从洞口 M 复出并回到过去,继续沿路径 β 运动并赶在较年轻的自己到达前占领接头 J,挡住较年轻自己的去路,使之不能到达洞口 N。这显然也是个非自洽过程。但诺维科夫用计算证明的确存在初值相同的自洽解,物理图像见图 9 - 19(b):活塞沿路径 α 走到接头 J 时恰与从洞口

M 出来的较老的自己的前端相接触，由于受到较老活塞前端的摩擦而略有减速，以这个较小的速率走过 α'，从 N 进洞再从 M 复出。由于速率略小，它不能抢先挡住较年轻自己的去路，只能用前端与它的侧面接触。这显然是个自洽的运动过程。

(a) 非自洽运动情况　　(b) 自洽运动情况

图 9 - 19　管内活塞穿过虫洞的两种运动情况(初值相同，都满足牛顿运动方程)

再来回顾台球悖论。刚才讲过，在相同的台球初值下不但存在非自洽解，而且存在两个自洽解。我们当然摒弃非自洽解，但两个自洽解的存在又提出了新的问题。台球运动速率很低，牛顿力学当然适用。在牛顿力学中存在着初值决定论①，即牛顿运动方程满足初值的解是唯一的。然而现在却有两个满足相同初值的自洽解。自然要问：给定初值后，台球到底按哪个解运动？这是在承认存在时间机器之后所遇到的崭新问题。更糟的是，索恩等后来发现台球的同一初值竟允许无限多个自洽解，图 9 - 20 就是简单例子。该图的五个分图代表台球在同一初值下的五种自洽运动过程，其中图(a)不必解释，对图(b)说明如下：台球在向上走到 M、N 连线时被从洞口 M 飞出的、较老的自己碰撞并到达洞口 N，穿越虫洞后从洞口 M 飞出并把较年轻的自己撞到洞口 N。对图(c)、(d)则可用如下方式简述。

图(c)：从最下端向上——碰(与年老)——向右上——洞口 N(进)——洞内(图中不显示)——洞口 M(出)——洞外(向右下)——碰(与年轻)——向下(永远)。

图(d)：从最下端向上——碰(与年老)——向右上——洞口 N(进)——洞内——洞口 M(出)——洞外(向右)——洞口 N(再进)——洞内——洞口 M(再出)——洞外(向右下)——碰(与年轻)——向下(永远)。

图(d)无非是在图(c)的基础上添加了用重点号标出的三步而已。由此就

① 这里不拟涉及有关混沌的问题。

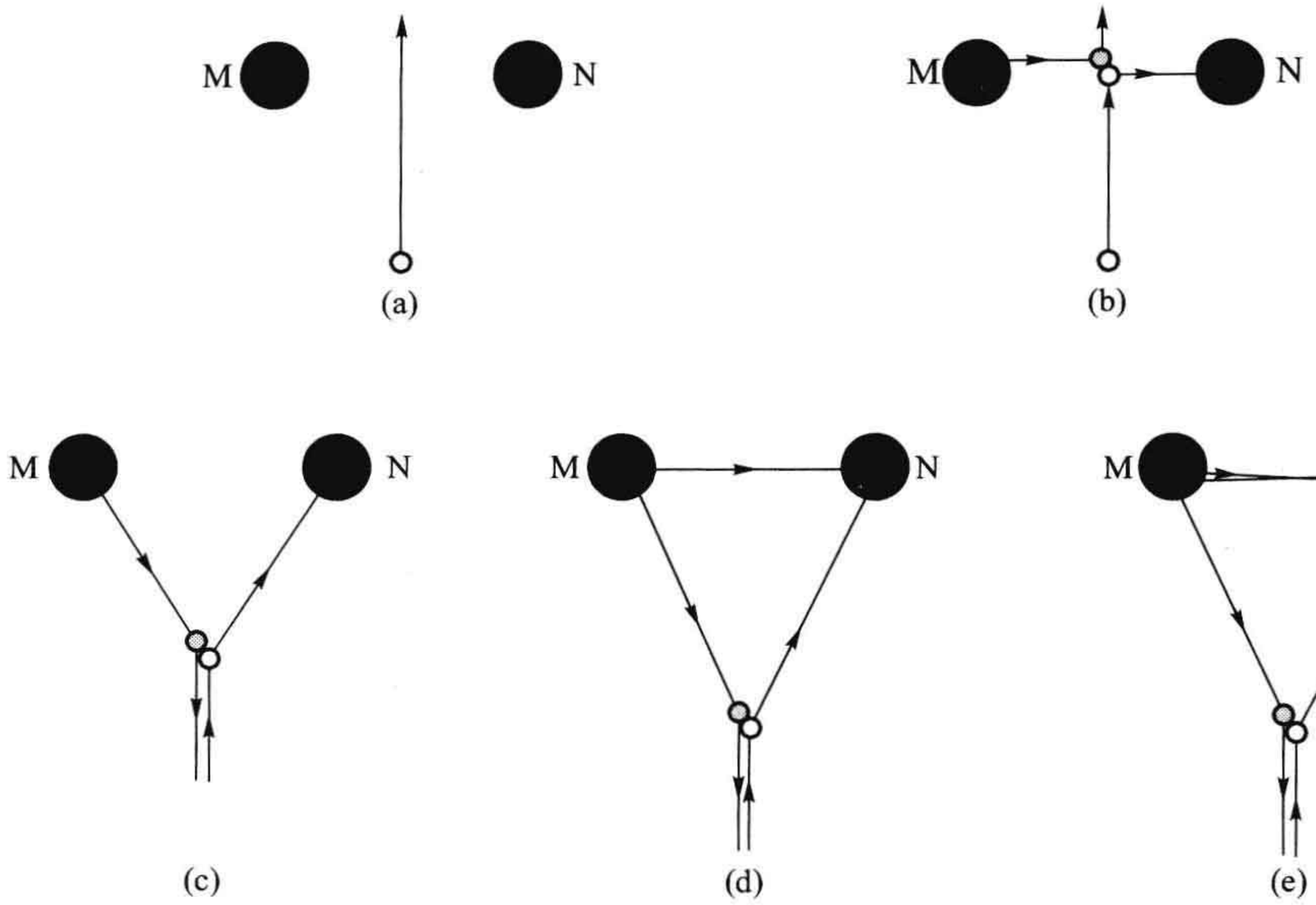


图 9 - 20　台球在同一初值下的 5 种可能运动

不难理解图(e),它不过是在图(d)的基础上又一次添加三步。(就是说,在第三次从洞口 M 出来后才去碰撞较年轻的自己。)由于这三步可以不断添加(可以不断增加穿越虫洞的次数),不难相信这种初值下存在着无限多种自洽运动情况。类似地,满足图 9 - 18 的初值的自洽情况也有无限多种。这是在没有闭合类时线的时空中绝不可能出现的。只要承认存在时间机器,物理学家就必须回答:在给定初值后,物理学能够告诉我们台球到底按哪一种方式运动吗?经典物理学对此完全无能为力,但量子物理学却能给出答案。与经典物理学不同,量子物理学对实验结果只做概率性的预言。索恩及其学生根据量子力学发现,在各种可能路径(自洽解)中,台球走哪条路径都有可能(都有非零的概率)。以图 9 - 18为例,两种运动情况[图(a)和图(b)]发生的概率各是 48%,而其他各种可能路径(未在图中画出)的概率则要小得多。

9.3.3　物理学允许时间机器吗?

在索恩、诺维科夫等致力于研究各种自洽解的同时,另一些物理学家(霍金是突出代表)则从根本上质疑时间机器存在的可能性。霍金认为,即使量子物理学允许奇异物存在,时间机器也不可能形成。关键在于真空中不断存在着电磁场的涨落,简称**电磁真空涨落**(小节 9.2.2 简介过真空涨落并说过它使奇异物的存在成为可能)。假如,像王五那样,微弱的真空涨落也从洞口 N 进入虫

洞、从洞口 M 出洞并与“较年轻”的自己相遇，就会自我加强，如此反复地进洞——出洞——加强，最后就可能强大到足以破坏（摧毁）虫洞的程度。“成也萧何败也萧何”：真空涨落相当于奇异物，它使得可穿越虫洞的存在成为可能；但也正是真空涨落反复进出虫洞而很可能毁掉时间机器！我们不免要问：“真空涨落是否一定毁灭时间机器？”然而，要回答这一问题，现在似乎还为时尚早，因为只有量子引力论才有可能给出最终结论，而这一理论至今尚未创建成功。

与索恩、诺维科夫的研究路线相反，霍金早在 1992 年就在大量论证的基础上提出了如下的猜想：

时序保护猜想（chronology protection conjecture）　物理定律不允许闭合类时线的存在（因而不会有时间机器）。

索恩后来也越来越觉得霍金可能是对的。但是，正如索恩[Thorne（1994）]所说的：“量子引力将虫洞能否成为时间机器的答案藏起来了”……“在物理学家深刻认识量子引力定律之前，我们谁也不能肯定。”

附录 A　单位制、量纲和公式转换

A.1　单位制基本知识

A.1.1　基本单位和导出单位

为了定量研究物理现象，必须对物理量进行测量。所谓测量，就是用一个与被测的量同类的量与该量作比较。这个用以比较的量叫做**单位**。用某一单位测某一量的结果是一个数，例如以 m 为单位测百米跑道的长度得数为 100。用不同单位测同一量所得的数不同。若以 cm 为单位测百米跑道的长度，得数为 10 000。一般地说，以 A 代表被测的量，$\hat{A}_1$ 代表测量单位，a_1 代表测得的数，则

$$A = a_1\hat{A}_1。$$

若以 a_2 代表用单位 $\hat{A}_2$ 测 A 所得的数，又有

$$A = a_2\hat{A}_2,$$

故

$$a_1\hat{A}_1 = a_2\hat{A}_2, \tag{A-1}$$

上式也可改写为

$$\frac{a_1}{a_2} = \frac{\hat{A}_2}{\hat{A}_1}。 \tag{A-1'}$$

如果 $a_1 < a_2$，就说单位 $\hat{A}_1$ 比 $\hat{A}_2$ 大（都作为量）。于是，用不同单位测同一量时，大单位得数小，小单位得数大。

物理规律是物理量之间关系的反映。既然选定单位后每个量可用一个数表示，物理规律也就可用**数的等式**表示。（最狭义地说，数学就是研究数的关系的学问。）反映物理规律的数的等式称为物理规律的**数值表达式**（numerical - value equation），式中每一字母代表用某一单位测该量所得的数。例如功的公式

$$w = f\,l \tag{A-2}$$

中的 w、f 及 l 可理解为用 J、N 及 m 为单位测量功、力和距离所得的数。

如果把上式的 w、f 及 l 分别理解为功、力及位移等物理量本身，则上式成为

量的等式(quantity equation),它直接描述这 3 个量的关系。为了明确区分量和数以及区分量的等式和数的等式,本小节分别以大、小写字母代表量和数,于是数的等式 $w = f\,l$ 相应的量的等式为

$$W = FL。 \tag{A-2'}$$

既然用不同单位测同一量得数不同,反映同一物理规律的数的等式就可能由于单位的改变而改变。例如,如果力的单位由 N 改为 dyn 而功和距离的单位不变,式(A-2)就要改变。以 f_{N} 及 f_{dyn} 分别代表以 N 及 dyn 为单位测同一力 F(是个量)所得的数,由于

$$1\ \mathrm{N} = 10^5\ \mathrm{dyn},\quad (\text{这是个量的等式})$$

故由式(A-1')得

$$\frac{f_{\mathrm{N}}}{f_{\mathrm{dyn}}} = \frac{\mathrm{dyn}}{\mathrm{N}} = 10^{-5},$$

即

$$f_{\mathrm{N}} = 10^{-5} f_{\mathrm{dyn}}。$$

因为式(A-2)中的 f 是以 N 为单位测力所得的数,所以可更明确地表为

$$w = f_{\mathrm{N}} l。$$

把 $f_{\mathrm{N}} = 10^{-5} f_{\mathrm{dyn}}$ 代入上式得

$$w = 10^{-5} f_{\mathrm{dyn}} l。$$

就是说,若分别以 J、dyn 及 m 为单位测量功、力及距离等物理量,则它们之间的数的关系应写成

$$w = 10^{-5} f\,l。 \tag{A-3}$$

可见同一物理规律在不同单位选择下的数值表达式不同。但也不难相信,同一规律的各个数值表达式之间的差别仅体现为一个附加因子①。若在表达式等号右边补一个依赖于各量单位选择的比例系数 k,则所得式子对任何单位都成立。

单位的选择本来是任意的,但若对每类量都任意选择一个单位("一盘散沙"),则物理规律数值表达式的比例系数可能复杂得难以记住。为使数值表达式尽量简单,应该采用单位制。一个**单位制**由以下两个要素构成:

(1) 选出几个最基本的量并独立地指定它们各自的单位。这些量叫做**基本量**,这些单位叫做**基本单位**。

(2) 不是基本量的量叫做**导出量**,它们的单位叫做**导出单位**。

导出单位是由基本单位出发按下述方法逐一规定的:设欲规定导出量 A 的单位,先找出一个把量 A 与基本量联系起来的物理规律,写出该规律的数值表

① 在等式含有不止一项的情况下,各项的附加因子可以不同,故等号右边各项所补的 k 也可不同。

达式,指定表达式中的比例系数 $k=1$,便规定了量 A 的导出单位。例如,国际单位制的力学部分指定长度、质量和时间为基本量,指定 m、kg 和 s 为基本单位。为规定导出量速度的导出单位,可利用速度与基本量(长度及时间)的关系的数值表达式:

$$v=k\frac{l}{t}。\tag{A-4}$$

请特别注意其中各字母都代表数。上式对长度、时间和速度取任何单位的情况都成立,不同单位选择的区别只体现在 k 值的不同。指定上式的 $k=1$ 便得

$$v=\frac{l}{t}。\tag{A-5}$$

这一指定就意味着规定了速度的国际单位制单位:设一个做匀速直线运动的质点在某一时间内走了某一距离,式中的 t 和 l 分别代表以 s 和 m 为单位测量该时间和距离所得的数。只有用适当单位测此速度,得数 v 才恰好等于 l/t。可见式(A-5)的确规定了速度的单位。为了看出这个单位是怎样的一种速度,只需考虑一个特例:设质点在 1 s 内走了 1 m,分别以 s 和 m 为单位测量质点运动的时间及距离,得值便为 $t=1$ 及 $l=1$,代入式(A-5)得 $v=1$,说明这个质点的速度正是由式(A-5)定义的速度单位。可见,速度的国际单位制单位就是每 s 走 1 m这样一种速度。

在许多情况下,导出量 A 与基本量的关系中还涉及另一些导出量,例如,设导出量 A 与量 J 及 B 由某一物理规律联系起来,其中 J 是基本量而 B 是导出量,这时应设法先通过 B 与基本量的联系规定 B 的单位,再在 A、J、B 的关系的数值表达式中指定系数 $k=1$ 来规定导出量 A 的导出单位。

用以定义导出单位的数值等式叫做该**导出单位的定义式**,例如式(A-5)是速度的国际单位制单位(m/s)的定义式。请注意定义式是数的等式。定义式中涉及的各量中除待定义单位的那个量外,其他各量的单位必须已经指定。

如果由某物理规律所联系的几个量的单位已经事先选定,这个规律的数值表达式中的系数就不能再任意指定,而应由实验测得。例如在 CGS 制中,长度、质量和时间是基本量,力的导出单位已由牛顿定律

$$f=ma$$

规定(上式是力的 CGS 制单位的定义式),于是万有引力定律

$$f=g\frac{m_1m_2}{r^2}$$

中的系数 g 就不能再任意指定而只能由实验测得。

虽然物理规律的数学表达式既可看作数的等式又可看作量的等式,但有些

公式(不是指物理规律的数学表达式)不能看作数的等式,例如

$$\text{跑道长度} = 100\ \text{m} = 10\ 000\ \text{cm}$$

就只能理解为量的等式,它表明 100 m 和 10 000 cm 这两个量相等。我们不能写成

$$\text{跑道长度} = 100 = 10\ 000,$$

因为 100 无论如何不等于 10 000,而跑道长度是一个量,它也不能与一个数(100 或 10 000)相等。

笔者一贯偏爱和习惯于把物理规律的数学表达式看作数的等式,这是在五十余年前受当时苏联作者赛纳的书[赛纳(1959)]的影响而逐渐形成的。该书认为物理规律的数学表达式都是数的等式而不是量的等式。例如,书中第 3 页明确写道:"**通常在公式中的一些符号并非是量的本身,而是一些数值,这些数值是表示这些量用任一单位来量度时的数值。**"笔者认为,数的等式最为清晰,而量的等式则只有在刻意讲清之后才有明确含义。例如,数的等式 $3 \times 4 = 12$ 中乘号的含义是 3 的 4 倍,等式的含义是 3 的 4 倍等于 12;而量的等式 $q = I \times t$ 或"库 = 安 × 秒"中的乘号是什么意思?该式又代表什么意义?总不能说代表"安的秒倍等于库"这样一种毫无意义的"意义"吧!?因此,要使量的等式有明确意义就先要对量的乘法等运算下定义。笔者尚未在文献中查到这类定义,但自己已对此下了定义。由于该定义及其有关问题过于抽象和冗长,此处不拟介绍(将另写长文发表),只想列出几个有用的结论:① 物理规律的数值表达式不但清晰,而且可以导出量纲分析的全套理论。赛纳(1959)就是这样做的。这是最为清晰简洁的做法。② 对量的运算刻意地下了定义后,物理规律的数学表达式也可看做量的等式。③ 导出单位的定义式都是数的等式。当然,正如上面所讲,一旦用它对导出单位下了定义,也可把它看作量的等式。

A.1.2　量纲式和量纲

为便于讲解,本小节采用对话形式,其中乙代表笔者。

甲　量纲到底是什么?是个量?是个数?还是别的什么东西?

乙　这是个很好的问题,问与不问、懂与不懂大有区别。由于导出单位由基本单位通过定义式来定义,基本单位的改变自然会引起导出单位的改变。量纲就是为描述导出单位随基本单位改变而改变的依从关系才引入的。

甲　但是,如果两个单位制连基本量都选得不同,"导出单位随基本单位的改变而改变"又从何谈起!?

乙　你问得太好了!所以谈到量纲时必须相对于一个单位制族而言。

甲　什么叫做单位制族?

乙　两个单位制如果满足以下条件就称为同族的：①基本量选得相同；②各导出单位的定义式在两制中相同。例如力学范畴内的国际单位制和 CGS 制是同族的，因为它们都以长度、质量和时间为基本量，而且所有导出单位的定义式都相同。但各量的单位在两制中都可以不同。

甲　我明白了。同族的两个单位制的最根本差别就是基本单位可以选得不同，从而导出单位也就不同。

乙　所以就应特别关心导出单位如何随着基本单位的改变而改变。描述导出单位随基本单位改变而改变的依从关系的等式叫做该导出量的**量纲式**。既然国际单位制的力学部分以长度 L、质量 M 和时间 T 为基本量，任一力学量 A 在国际单位制所在族的量纲式就可表为

$$\dim A = \mathrm{L}^{\lambda}\mathrm{M}^{\mu}\mathrm{T}^{\tau}, \tag{A-6}$$

其中正体大写字母 L、M 和 T 依次代表长度、质量和时间的新旧单位的比值（都是数！），λ、μ、τ 称为量 A 在国际单位制所在族的**量纲指数**，$\dim A$ 称为量 A 在国际单位制所在族的量纲（dimension）。量纲指数全部为零的量称为**量纲为一的量**（或**无量纲量**）。索性把式（A-6）的 L、M 和 T 也称为三个基本量的量纲，该式便可改写为

$$\dim A = (\dim L)^{\lambda}\ (\dim M)^{\mu}\ (\dim T)^{\tau}。 \tag{A-6′}$$

甲　我终于明白了：量纲是个数，代表该量的新旧单位的比值。

乙　很对。知道这一点固然重要，但更重要的是还要知道量纲式的根本作用在于给出导出单位随基本单位的改变而改变的函数关系（依从关系），基本量的量纲 $\dim J_i$ 是自变数，导出量的量纲 $\dim C$ 是因变数（函数），函数关系是非常简单的幂单项式。当基本量选得不同时，同一导出量的量纲式自然不同。例如，设有两族单位制，第一族以长度、质量和时间为基本量，第二族以长度、力和时间为基本量，则导出量在第一族单位制中的量纲式为 $\dim L$、$\dim M$、$\dim T$ 的幂单项式，在第二族单位制中的量纲式为 $\dim L$、$\dim F$、$\dim T$ 的幂单项式。另一方面，在基本量一样的前提下，如果导出单位的定义式不同（指实质上的不同），则同一导出量的量纲式也会不同。可见，脱离（不明确）单位制族而侃侃而谈某量的量纲是没有意义的，正如脱离参考系而谈速度没有意义那样。

甲　上述知识有什么用处？

乙　在此基础上就可以学好**量纲分析**理论，这对于定性和半定量地讨论物理问题有非常重要的意义。关于量纲分析的理论及应用，我们将另写长文发表。下面只给出两个非常简单实用的结论：

（1）等式中各项所代表的量必有相同的量纲。

根据这一性质，如果演算中出现一个等式，其等号两边量纲不等或式中某项

与其他项量纲不等,则此式必错。这个方法有助于迅速发现演算中的某些错误。

(2) 超越函数符号(如 sin、lg、ln、e、…)下的变量一定是无量纲量。

此外,小节 A.2.2(公式转换)也是量纲分析的应用实例之一。

A.2　几何单位制

A.2.1　几何单位制

几何制规定 $c=G=1$,其中 c 是用速度的几何制单位测量(真空)光速所得的数,此数为 1 表明几何制以光速为速度单位。类似地,$G=1$ 表明几何制以(我们宇宙的)引力常量为引力常量的单位。一个常见的问题是:几何制有几个基本量?基本单位是什么?答案有一定灵活性,可根据需要采用不同的看法。

第一种看法认为几何制有三个基本量,即速度、引力常量和时间,基本单位是:速度以(真空)光速为单位,引力常量以(我们宇宙的)引力常量为单位,时间以 s 为单位。由此就可求得其他各量(视为导出量)的单位。

第二种看法认为几何制只有一个基本量,这是因为 $c=g=1$ 表明时间、长度和质量三者的单位只有一个可任意选择。这时可把这三个原本不同的量认同为同一类量,认同的"钥匙"是把 1 s、3×10^8 m 和 4×10^{35} kg 这三个量视为相等(理由见稍后),即

$$1\ \mathrm{s}=3\times10^8\ \mathrm{m}=4\times10^{35}\ \mathrm{kg}。\tag{A-7}$$

从而使所有量要么没有单位(例如速度就没有单位),要么以 s(或 s 的若干次方)为单位(例如加速度的单位是 s^{-1})。

根据上述两种看法,几何单位制与国际单位制是不同族的单位制,不利于两制之间的公式转换,于是还可采用第三种看法:① 认为几何制有三个基本量,即时间 T、长度 L 和质量 M(与国际单位制相同),基本单位则是 $\hat{T}_{几}=1$ s、$\hat{L}_{几}=3\times10^8$ m、$\hat{M}_{几}=4\times10^{35}$ kg(理由见稍后)。② 导出单位的定义式与国际单位制全同。各量的单位在第一、三种看法中完全相同,但第三种看法保证几何单位制与国际单位制是同族单位制。

现在说明把基本单位选为 $\hat{T}_{几}=1$ s、$\hat{L}_{几}=3\times10^8$ m、$\hat{M}_{几}=4\times10^{35}$ kg 的理由。几何单位制受到 $c=g=1$ 的约束,基本单位 $\hat{T}_{几}$、$\hat{L}_{几}$ 和 $\hat{M}_{几}$ 只有一个可以任取。为了简单,也为了与国际单位制尽量一致,当然取 $\hat{T}_{几}=1$ s。至于 $\hat{L}_{几}$ 和 $\hat{M}_{几}$,则可由下面的计算求得。

先求长度的几何制单位 $\hat{L}_{几}$。由速度定义易得量纲关系：

$$\dim L = (\dim T)(\dim V), \tag{A-8}$$

其中 $\dim L$ 代表长度的国际单位制单位与几何单位制单位的比值，就是说，若以 $\hat{L}_{国}$ 和 $\hat{L}_{几}$ 分别代表长度的国际单位制和几何单位制单位，则

$$\dim L \equiv \hat{L}_{国}/\hat{L}_{几}, \tag{A-9}$$

注意到式(A-8)，得

$$\hat{L}_{几} = \hat{L}_{国}/(\dim L) = \hat{L}_{国}(\dim T)^{-1}(\dim V)^{-1}。 \tag{A-10}$$

而

$$\dim T = \frac{\hat{T}_{国}}{\hat{T}_{几}} = \frac{1\text{s}}{1\text{s}} = 1, \quad \dim V = \frac{\hat{V}_{国}}{\hat{V}_{几}} = \frac{1\text{m/s}}{3\times10^{8}\text{m/s}} = \frac{1}{c_{国}}, \tag{A-11}$$

其中 $c_{国} = 3\times10^{8}$ 代表用国际单位制单位测光速所得的数。式(A-11)代入式(A-10)得出

$$\text{几何单位制长度单位 } \hat{L}_{几} = c_{国}\ \hat{L}_{国} = (3\times10^{8})\times1\ \text{m} = 3\times10^{8}\ \text{m}。 \tag{A-12}$$

再求质量的几何制单位 $\hat{M}_{几}$。由 $f = ma$ 及 $f = gm_1m_2/r^2$ 可知质量的量纲式为

$$\dim M = (\dim T)(\dim V)^{3}(\dim G)^{-1}。 \tag{A-13}$$

再由 $\dim M = \hat{M}_{国}/\hat{M}_{几}$ 得

$$\hat{M}_{几} = \hat{M}_{国}/(\dim M) = \hat{M}_{国}(\dim T)^{-1}(\dim V)^{-3}(\dim G)。 \tag{A-14}$$

其中的 $\dim G$ 可求之如下：

$$\dim G = \frac{\hat{G}_{国}}{\hat{G}_{几}} = \frac{g_{几}}{g_{国}} = \frac{1}{6.67\times10^{-11}}。 \tag{A-15}$$

将上式及式(A-11)代入式(A-14)便得

几何单位制质量单位

$$\hat{M}_{几} = \frac{c_{国}^{3}}{6.67\times10^{-11}}\hat{M}_{国} = \frac{(3\times10^{8})^{3}}{6.67\times10^{-11}}\times1\ \text{kg} = 4\times10^{35}\ \text{kg}。 \tag{A-16}$$

A.2.2　几何单位制公式转换为国际单位制公式

本书由于使用几何单位制而使公式大为简化，但如果需要代入具体数值，最好先转换为国际单位制形式。利用量纲分析可以完成这一任务。为此，对几何单位制应该采用第三种看法，这时几何单位制与国际单位制就是同族单位制，区

别只来自基本单位的不尽相同：

$$\hat{T}_{国}=1\ \mathrm{s},\quad \hat{L}_{国}=1\ \mathrm{m},\quad \hat{M}_{国}=1\ \mathrm{kg},\tag{A-17a}$$

$$\hat{T}_{几}=1\ \mathrm{s},\quad \hat{L}_{几}=3\times10^{8}\ \mathrm{m}=c_{国}\ \mathrm{m},\quad \hat{M}_{几}=4\times10^{35}\ \mathrm{kg}。\tag{A-17b}$$

（请特别注意上式中的 $c_{国}=3\times10^{8}$ 是以国际单位制单位测光速所得的数）。下面用三个例子介绍一种简便实用的转换方法[另一种方法见梁灿彬、周彬(2006)附录 A]。

例 1　已知洛伦兹变换(第一式)的几何单位制形式为 $t'=\gamma(t-vx)$，求其国际单位制形式。

解　为区分起见，给字母补注下标以表明它是用哪个单位制的单位测得的数。例如，$t_{几}$ 代表用几何单位制单位测时间所得的数。于是已知公式 $t'=\gamma(t-vx)$ 可以更明确地写成

$$t'_{几}=\gamma(t_{几}-v_{几}\,x_{几})。\tag{A-18}$$

注意到式(A-17)，有 $x_{几}/x_{国}=\hat{x}_{国}/\hat{x}_{几}=1/c_{国}$，故

$$x_{几}=x_{国}/c_{国},\tag{A-19}$$

再由

$$v_{几}/v_{国}=\hat{V}_{国}/\hat{V}_{几}=1\ \mathrm{ms}^{-1}/(3\times10^{8}\ \mathrm{ms}^{-1})=1/c_{国}\tag{A-20}$$

又得

$$v_{几}=v_{国}/c_{国},\tag{A-21}$$

所以

$$v_{几}\,x_{几}=v_{国}\,x_{国}/c^{2}_{国}。\tag{A-22}$$

另一方面，由 $\hat{T}_{几}=1\mathrm{s}=\hat{T}_{国}$ 又知 $t_{几}=t_{国}$ 及 $t'_{几}=t'_{国}$，与式(A-22)一同代入式(A-18)得 $t'_{国}=\gamma(t_{国}-v_{国}\,x_{国}/c^{2}_{国})$，去掉下标便得洛伦兹变换(第一式)的国际单位制形式：

$$t'=\gamma(t-vx/c^{2})。\qquad\text{[解毕]}$$

例 2　已知 2 维闵氏线元的几何单位制形式为 $\mathrm{d}s^{2}=-\mathrm{d}t^{2}+\mathrm{d}x^{2}$，求其国际单位制形式。

解　把上式明确写成

$$\mathrm{d}s^{2}_{几}=-\mathrm{d}t^{2}_{几}+\mathrm{d}x^{2}_{几}。\tag{A-23}$$

由 $t_{几}=t_{国}$ 及 $x_{几}=x_{国}/c_{国}$[式(A-19)]得

$$\mathrm{d}t^{2}_{几}=\mathrm{d}t^{2}_{国},\quad \mathrm{d}x^{2}_{几}=\mathrm{d}x^{2}_{国}/c^{2}_{国}。\tag{A-24}$$

由式(2-1-11)可知 $\mathrm{d}s^{2}$ 与 $\mathrm{d}x^{2}$ 单位相同，故仿照式(A-24)又有 $\mathrm{d}s^{2}_{几}=\mathrm{d}s^{2}_{国}/c^{2}_{国}$，与式(A-24)一同代入式(A-23)得 $\mathrm{d}s^{2}_{国}/c^{2}_{国}=-\mathrm{d}t^{2}_{国}+\mathrm{d}x^{2}_{国}/c^{2}_{国}$，去掉下

标便得闵氏线元的国际单位制形式：

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + \mathrm{d}x^2。$$ [解毕]

例 3　已知 2 维史瓦西线元的几何单位制形式为

$$\mathrm{d}s^2 = -\left(1-\frac{2M}{r}\right)\mathrm{d}t^2 + \left(1-\frac{2M}{r}\right)^{-1}\mathrm{d}r^2,$$

求其国际单位制形式。

解　把上式明确写成

$$\mathrm{d}s_{\text{几}}^2 = -\left(1-\frac{2M_{\text{几}}}{r_{\text{几}}}\right)\mathrm{d}t_{\text{几}}^2 + \left(1-\frac{2M_{\text{几}}}{r_{\text{几}}}\right)^{-1}\mathrm{d}r_{\text{几}}^2。 \tag{A-25}$$

关键是要将 $M_{\text{几}}/r_{\text{几}}$ 用 $M_{\text{国}}/r_{\text{国}}$ 表出。由式($A-19$)知 $r_{\text{几}} = r_{\text{国}}/c_{\text{国}}$，而

$$M_{\text{几}}/M_{\text{国}} = \hat{M}_{\text{国}}/\hat{M}_{\text{几}} = \dim M = (\dim T)(\dim V)^3(\dim G)^{-1} = G_{\text{国}}/c_{\text{国}}^3,$$ (A-26)

[其中第三步用到式(A-13)，第四步用到式(A-20)及式(A-15)。] 故

$$M_{\text{几}} = (G_{\text{国}}/c_{\text{国}}^3)M_{\text{国}}。 \tag{A-27}$$

上式与 $r_{\text{几}} = r_{\text{国}}/c_{\text{国}}$ 结合给出

$$\frac{M_{\text{几}}}{r_{\text{几}}} = \frac{(G_{\text{国}}/c_{\text{国}}^3)M_{\text{国}}}{r_{\text{国}}/c_{\text{国}}} = \frac{G_{\text{国}}\,M_{\text{国}}}{c_{\text{国}}^2\,r_{\text{国}}}。 \tag{A-28}$$

去掉下标并仿照例 2 的做法，便得史瓦西线元的国际单位制形式：

$$\mathrm{d}s^2 = -\left(1-\frac{2GM}{c^2r}\right)(c\mathrm{d}t)^2 + \left(1-\frac{2GM}{c^2r}\right)^{-1}\mathrm{d}r^2。$$ [解毕]

习题答案

第1章

题3　K′系相对于K系的速率为0.5；K″系相对于K系的速率为0.8。

题5　$u'_x = c/5$，$u'_y = u'_z = 0$。

题6　(a) 对牛顿力学 $u_x = (1-n)v$；对狭义相对论 $u_x = \dfrac{1-n}{1-nv^2}v$。

(b) 对牛顿力学 $u_x = \dfrac{1}{4}c$；对狭义相对论 $u_x = \dfrac{2}{7}c$。

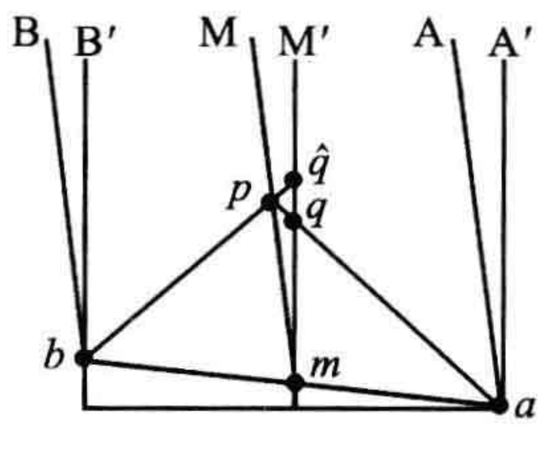
第2章

题2　$\mathrm{d}s^2 = -\mathrm{d}t^2 + \mathrm{d}\rho^2 + \rho^2\mathrm{d}\varphi^2 + \mathrm{d}z^2$。

第3章

题1　$D = (\tau_2 - \tau_1)/2$。

题2　见附图，由图易见观者M′同样是先收到车头来光后收到车尾来光。



3章题2解用图

题5　待找的坐标变换为 $T = -t^{-1}$，$X = x$，$Y = y$，$Z = z$。

第4章

题1　$\hat{t} = 1.67 \times 10^{-7}\ \mathrm{s}$，$\hat{\tau}_1 = 10^{-7}\ \mathrm{s}$，$\hat{\tau}_2 = 1.34 \times 10^{-7}\ \mathrm{s}$。

题2　$7.5 \times 10^{-9}\ \mathrm{s}$。

题3　6 s。

题4　(b) 甲钟3 s，乙钟3 s，丙钟3.5 s。

题5　(a) A钟4 s，B钟5 s；(b) A钟看见B钟指3 s。

题6　(a) 5s；(b) A钟看见B钟指 −20 s。

题7　比B大(由直线段组成的折线不是测地线)。

题8　5/3昼夜。

题10　(a) $\dfrac{2}{3} \times 10^{-8}\ \mathrm{s}$；(b) $\hat{l} = 2.5\ \mathrm{m}$；(c) $\dfrac{\hat{l}}{l} = \sqrt{\dfrac{1-u}{1+u}}$。

题11　(a) 4 μs；(b) 2.5 μs。

题12　(a) $\dfrac{\tau'}{\tau} = \sqrt{\dfrac{1+u}{1-u}}$；　(b) $u = 0.6c$ 时 $\dfrac{\tau'}{\tau} = 2$；　(c) $u = 0.8c$ 时 $\dfrac{\tau'}{\tau} = 3$。

题 13　$\Delta t = 8 \times 10^{-7}$ s 。

题 14　2.5×10^{4} s^{-1}。

题 15　(a)$v = 0.6c$; (b)9×10^{8} m 。

第 5 章

题 1　(a)3.7×10^{14} kg; (b)6.8×10^{-12}。

题 2　(a)4.7×10^{-14} kg; (b)4.7×10^{-11}。

题 4　$u = \sqrt{3}c/2$。

题 5　质子静能 $= 1.5 \times 10^{-10}$ J，电子静能 $= 8.2 \times 10^{-14}$ J。

题 6　$u = 2.4$ m/s 。

题 7　能量 $E = 4.7 \times 10^{-10}$ J；动量 $p = 1.5 \times 10^{-18}$ kg · m/s；速率 $u = 0.95c$。

题 8　(a)$u = 0.858$; (b)$u = 0.999\ 999\ 87$。

题 11　(a)$E = 5.3 \times 10^{-18}$ J，$p = 1.8 \times 10^{-26}$ kg · m/s; (b)1.7×10^{33}个。

参考文献

爱因斯坦,等.相对论原理(狭义相对论和广义相对论经典论文集)[M].赵志田,刘一贯,译.北京:科学出版社,1980.

胡镜寰,王忠烈,刘玉华.原子物理学[M].北京:北京师范大学出版社,1989.

朗道,栗弗席兹.场论[M].任朗,袁炳南,译.北京:人民教育出版社,1959.

梁灿彬,周彬.微分几何入门与广义相对论(上册)[M].北京:科学出版社,2006.

梁灿彬,周彬.微分几何入门与广义相对论(中册)[M].北京:科学出版社,2009.

梁灿彬.电磁学(第三版).北京:高等教育出版社,2012.

孟广达,王润华,霍瑞云,等.狭义相对论解决双生子问题之不可能[J].大学物理,1997,16(4):22.

钱尚武.关于时钟佯谬[J].大学物理,1982,1(7):4 .

赛纳.物理学单位[M].嵇储凤,等,译.上海:上海科学技术出版社,1959.

虞福春,郑春开.电动力学[M].北京:北京大学出版社,1992.

张元仲.狭义相对论实验基础.北京:科学出版社 ,1994.

赵凯华,罗蔚茵.新概念物理教程力学[M].北京:高等教育出版社,1995.

Adler C G. Does Mass Really Depend on Velocity, Dad? [J] Am J Phys, 1987, 55 (8).

Ashby N. Relativity and the global positioning system[J]. Phys Today, 2002, 13 (9):41 -47.

Ashby N. Relativity in the global positioning system[OL]. 2003. http://www.Relativity. Living Reviews. org/Articles/Irr 2003 -1.

Bertotti B, Iess L, and Tortora P. A test of general relativity using radio links with the Cassini spacedraft[J]. Nature ,2003 ,425 :374.

Brian D. Einstein: A Life[M]. New York: John Wiley and Sons, 1996.

Clark R. Einstein: The Life and Times[M]. New York: World Publishing, 1971.

Cropper W H. Great Physicists[M]. New York: Oxford University Press, 2001.

Dingle H. A problem in relativity theory[J]. Proc Phys Soc, 1956, A69:925.

Echeverria F, Klinkhammer G, and Thorne K S. Billiard balls in wormhole spacetimes with closed timelike curves: Classical theory[J]. Physical Review D, 1991, 44: 1077.

Frisch D H, and Smith J H. Measurement of the relativistic time dilation using μ - mesons[J]. Am J Phys, 1963, 5: 342.

Geroch R. General relativity from A to B[M]. Chicago and London: The University of Chicago Press, 1978.

Hafele J C, and Keating R E. Around - the - world atomic clocks: Predicted relativistic time gains[J]. Science, 1972, 177: 166 - 167; Around - the - world atomic clocks: Observed relativistic time gains[J]. Science, 1972, 177: 168 - 170.

Thorne K S, Black holes and time warps[M]. New York, London: Norton and Company, 1994. 黑洞与时间弯曲[M]. 李泳, 译. 长沙: 湖南科学技术出版社, 2000.

Novikov I D. Time machine and self - consistent evolution in problems with self - interaction[J], Physical Review D, 1992, 45: 1989 - 1994.

Morris M S, and Thorne K S. Wormholes in spacetime and their use for interstellar travel: A tool for teaching general relativity[J]. Am J Phys, 1988, 56(5): 395.

McCrea W H. The clock paradox in relativity theory[J]. Nature, 1951, 167: 681.

Misner C, Thorne K, and Wheeler J. Gravitation[M]. San Francisco: W H Freeman and Company, 1973.

Muller R A. The twin paradox in special relativity[J]. Am J Phys, 1972, 40: 966.

Penrose R. The apparent shape of a relativistically moving sphere[J]. Proc Cambridge Philos Soc, 1959, 55: 137.

Reasenberg R D, et al. Viking relativity experiment: verification of signal retardation by solar gravity[J]. Astrophys. J. 1979, 234: L219 - L221.

Resnick R. Introduction to special relativity[M]. New York: John Wiley & Sons Inc, 1968.

Rindler W. Essential relativity[M]. New York: Springer - Verlag, 1977.

Rossi B, and Hall D B. Variation of the rate of decay of mesotrons with momentum[J]. Phys Rev, 1941, 59: 223.

Shapiro S S, Davis J L, Lebach D E, and Gregory J S. Measurement of the solar Gravitational deflection of radio waves using geodetic very - long - baseline interferometry data, 1979 - 1999[J]. Phys Rev Lett, 2004, 92: 121101.

Synge J L. Relativity: The general theory[M]. Amsterdam: North Holland Pub-

lishing Company, 1960.

Terrell J. Invisibility of the Lorentz contraction[J]. Phys Rev, 1959, 116: 1041.

Ugarov V A. Special theory of relativity[M]. (Revised from the 1977 Russian edition, English translation) Moscow: Mir Publishers, 1979.

Wald R M. General Relativity. Chicago: The University of Chicago Press, 1984.

Weisskopf V F. The visual appearance of rapidly moving objects[J]. Phys Today, 1960, 13(9): 24//二十世纪的物理学. 北京: 科学出版社, 1979.

Will C M. Theory and experiment in gravitational physics[M]. 2nd edition. London: Cambridge University Press, 1993.

Will C M. The confrontation between general relativity and experiment [J/OL]. Living Rev Relativity, 2001, 4: 4. http://www.livingreviews.org/lrr-2001-4.

Will C M. Was Einstein right? Testing relativity at the centenary[M]//Ashtekar A. 100 years of relativity: space-time structure: Einstein and beyond. Biswas Hopr Press, 2005.

索　引

B

白矮星　7.2.2

白洞　8.7.4

保时序　§3.6

北斗卫星导航系统　§8.6 开头

标准钟　§3.2 开头

伯克霍夫定理　§8.8

不变量　§5.4 末

布兰斯－迪克(布－迪)理论　§6.3 末段，8.5.2

C

参考系　§3.2

测地偏离　§6.4

测地线　§2.5

测量形象　§4.4 开头

超大质量黑洞　§7.5

超新星　7.2.4

潮汐加速度　§6.4

潮汐力　§6.4

车库佯谬　4.1.2

尺缩效应　4.1.1

虫洞　第9章

初始设定　§3.2 注记5

传播延迟方程　式(8－6－2)

D

等效原理　§6.3 开头

　爱因斯坦等效原理　§6.3 开头

　弱等效原理　§6.3 开头

强等效原理　§6.3 末段

动量(守恒)　§5.1

动能　§5.3

多普勒频移　4.2.4

F

负能轨道　§7.4

G

GPS　§8.6

GPS 时间　式(8－6－37)后

伽利略变换　式(1－1－4)

改时序　§3.6

固有距离　§8.1

固有时　§3.2 开头，§3.4

观者　§3.2 开头

　惯性观者(自由质点)　§3.2 注记5 后

　非惯性观者　§3.2 注记5 后

惯性质量　§6.1 开头

惯性坐标时　§3.4

惯性坐标系　式(1－1－2)后，§3.2

光行差　[选读4－6]偏后

光锥面　§3.1 例2 后

光子　§3.1 开头，5.8

光子的能量和动量　式(5－8－1)

国际原子时间　式(8－6－28)后

H

哈勃空间望远镜　§7.5

核聚变　7.2.1

黑洞　§7.3，8.7.4

黑洞无毛定理 §7.4
恒星晚期坍缩 §8.9
恒星演化 §7.2
红巨星 7.2.1 偏后
红移因子 8.3.2
喉 §9.1 例 2

J

基准坐标系 §3.1 开头
简并压 7.2.2
渐近平直(区) 图 8-13(a), §9.1 例 1 末,图 9-7
渐近平直线元 §8.1
结合能 §5.6
静界 §7.4
静能 §5.4
静态观者 §8.1
静态球对称线元 §8.1
静质量 §5.1
绝对的 §3.7 开头

K

可穿越虫洞 9.2.2
克鲁斯科尔延拓 式(8-7-25)前,8.7.3
空间几何 §8.1

L

雷达回波的时间延缓 8.5.4
类空(时、光)曲线 式(2-2-3)后
类空(时、光)间隔 式(3-6-4)后
类星体 §7.5 偏后
力 §5.2
粒子 §3.1 开头, §5.8
量子引力论 §7.3,9.2.2
洛伦兹变换 式(1-3-1)

M

脉冲星 7.2.5
闵氏时空 §2.2 末
闵氏线元(几何) §2.2

N

能层 §7.4

P

彭罗斯过程 §7.4
平直空间(时空) §2.4

Q

奇性(奇点) §7.3,8.7.1 开头
奇异物 9.2.2 开头
钱德拉 X 射线天文卫星 §7.5
钱氏质量上限 7.2.2
嵌入图 §9.1
球面线元(几何) §2.3 末
曲率(黎曼张量) §2.4
全球卫星定位系统(GPS) §8.6

R

日全食 §8.5

S

铯原子钟环球飞行 4.3.4
时间机器 第 9 章
时空 §1.1 开头
时空(元)间隔 式(2-1-8)
时空点 §1.1 开头, §2.2 末
时空奇性(真奇性) 8.7.1 开头
时序保护猜想 9.3.3
史瓦西半径 §8.1 偏后
史瓦西时空 §8.1
史瓦西时空的无限红移面 §8.9 末
史瓦西时空的引力红移 式(8-9-10)
史瓦西真空解 式(6-5-4), §8.1

世界面　§3.1 末段
世界线　§3.1 开头
事件　§1.1 开头
视觉形象　§4.4 开头
视界(事件视界)　§7.3,8.7.4
弒母悖论　9.3.2
守恒量　§5.4 末
双子效应(佯谬)　§4.3
水星近日点进动　§8.4
斯皮策空间望远镜　§7.5
四种基本相互作用　§7.1

T

特列尔转动　§4.4
天鹅 X-1　§7.5
天狼 A(B)星　7.2.2
天球　[选读 4-6],§7.1 开头
同时面(线)　§3.1 注记 2

W

弯曲空间(时空)　§2.4
伪欧几何　§2.3 开头
乌龟坐标　式(8-7-8)后

X

X 射线星　§7.5
吸积盘　§7.5
线元　§2.1
相对的　§3.7 开头
相对论速度变换式　式(1-3-3)
相对性原理　§1.1 开头,§3.2 注记 5 后,
　　§6.1 开头
校准曲线　§3.5 末
星光偏转　§8.5
星内时空线元　式(9-1-2)
星团　§7.1 开头
星系　§7.1 开头
　银河系　§7.1 开头
　河外星系　§7.1 开头
星云　§7.1 开头
星座　§7.1 开头

Y

引力红移　8.3.2
引力势　式(8-3-17),(8-6-21)
引力收缩　7.2.1 开头
引力坍缩　7.2.3 开头
引力质量　§6.1 开头
引力钟慢效应　8.3.1
诱导线元　§2.3 末,§8.1
运动质量(相对论质量)　§5.1,§5.5

Z

真空爱因斯坦方程　§6.5
真空涨落　9.2.2 开头
正电子　§5.4 偏后
织女星　§9.2 开头
直接观测(当时当地观测)　§3.2
直线　§2.5 开头
质点　§3.1 开头
质量亏损　式(5-6-4)
质能动关系式　式(5-7-4)
质能相当性　§5.5 开头
中心球(星核)　7.2.1 开头
中子星 7.2.3
钟慢效应
　闵氏时空的钟慢效应　§4.2
　弯曲时空的钟慢效应　8.3.1
钟同步操作　[选读 3-2]前
祝融星　§8.4
自洽过程(非自洽过程)　9.3.2
自洽性原理　9.3.2
总能　§5.4
走时率　§3.2 注记 5

最大延拓 式(8-7-29)后
最简关联 式(1-1-3)后
最一般的洛伦兹变换 式(3-6-9)
最一般的伪转动 [选读3-3]
坐标距离 §8.1
坐标奇性 8.7.1开头
坐标时间 §8.1偏后
坐标域 8.7.1开头
坐标钟 §8.1偏后

从零学相对论

ISBN 978-7-04-038121-4

定价 25.70元