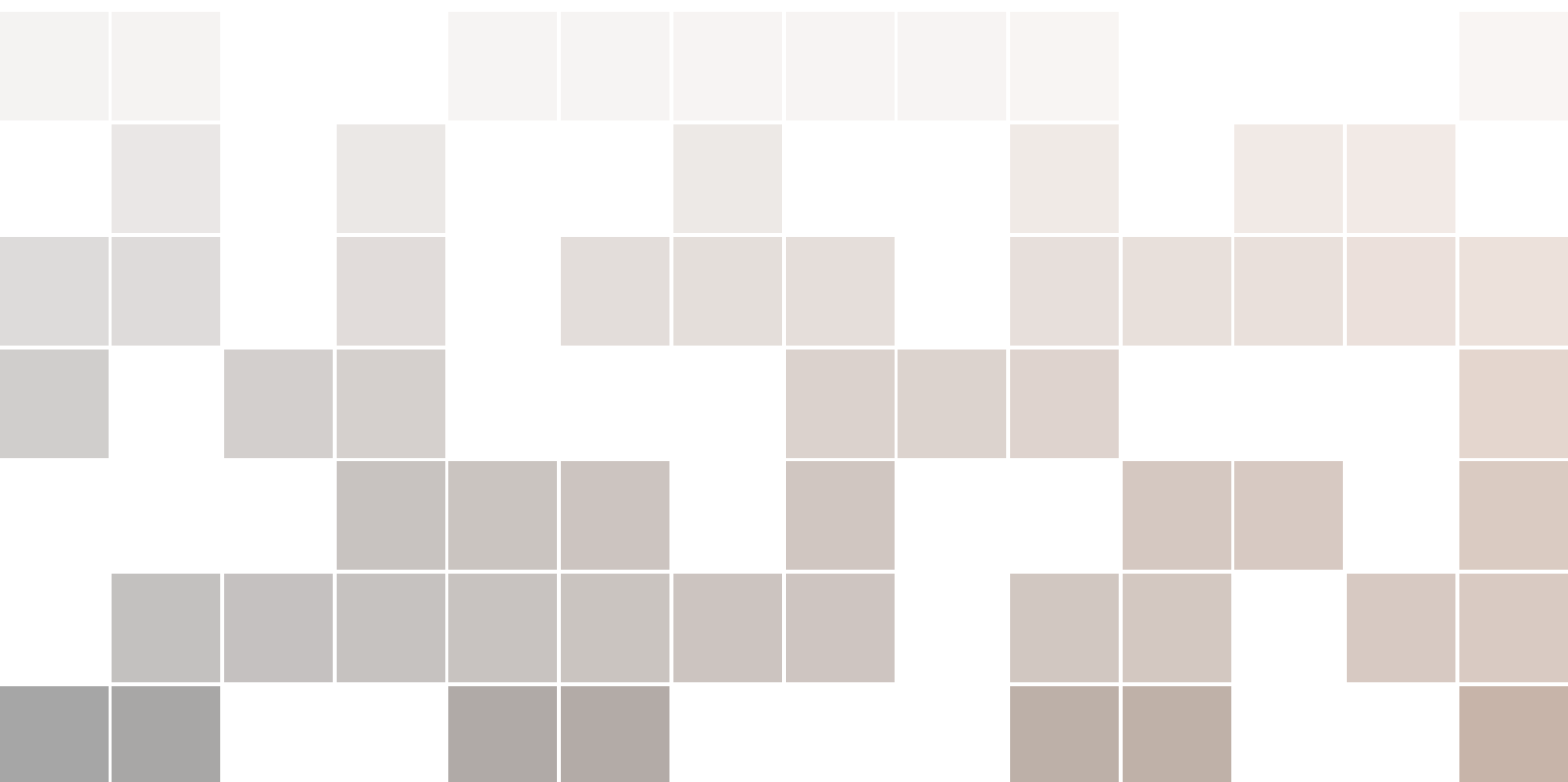


引力论讲义

现代导引

陈童



Copyright © 2026 陈童

PUBLISHED BY PUBLISHER

下载网址: <http://newquanta.com/>

作者本人 (即陈童) 对本讲义保有所有版权, 目前读者可在上面的网站上自由下载, 也可以私自传播, 但仅限于学习或教学等非商业用途, 禁止私自以本讲义的部分或全部来获利, 一经发现, 作者将保留诉诸法律的权利。如果你有任何反馈, 或者发现书中的任何错误, 请给作者发邮件, tongchen@ecut.edu.cn

2026 年 1 月

Contents

1	引力的普适性以及非欧几何	13
1.1	万有引力的特性	13
1.1.1	引力的普适性	13
1.1.2	万有引力是一种吸引力	15
1.2	非欧几何	16
1.2.1	内禀几何与高斯坐标	16
1.2.2	非欧双曲几何的发现	17
1.2.3	局部平坦与黎曼曲率	20
2	狭义相对性原理以及能量-动量张量	23
2.1	相对性原理	23
2.1.1	间隔不变性与闵可夫斯基几何	23
2.1.2	庞加莱群	28
2.1.3	四维闵可夫斯基时空的矢量和张量	29
2.2	狭义相对性原理与经典物理	32
2.2.1	狭义相对性原理与经典场论	32
2.2.2	狭义相对性原理与经典力学	35
2.3	时空对称性与能量动量张量	38
2.3.1	场论系统的能量-动量张量	38
2.3.2	多粒子系统的能量动量张量	43

3	从等效原理到时空弯曲	45
3.1	等效原理	45
3.1.1	自由下落的小房间	45
3.1.2	潮汐力	46
3.1.3	局部惯性系与等效原理	47
3.2	引力场中的自由粒子	50
3.2.1	统一自由落体定律	50
3.2.2	测地线方程	50
3.2.3	克里斯托夫联络与等效原理	53
3.3	等效原理的预言	55
3.3.1	引力红移	55
3.3.2	光线偏折	57
4	广义协变原理与微分同胚不变性	59
4.1	广义协变原理	59
4.2	张量场和微分同胚映射	60
4.2.1	矢量场和张量场	60
4.2.2	微分同胚映射和李导数	62
4.3	协变微分	66
4.4	等效原理和微分同胚不变性	69
4.4.1	最小耦合原理	69
4.4.2	微分同胚不变性与能量动量张量	71
4.5	弯曲时空中的高斯定理	77
5	黎曼曲率张量和爱因斯坦场方程	79
5.1	黎曼曲率张量	79
5.1.1	寻找黎曼曲率张量	80
5.1.2	黎曼曲率张量的代数性质	82
5.2	爱因斯坦-希尔伯特作用量和爱因斯坦场方程	83
5.2.1	爱因斯坦-希尔伯特作用量	83
5.2.2	爱因斯坦场方程	84
5.2.3	微分同胚不变性与比安基恒等式	86
5.2.4	广义相对论作为一个有效理论	88
5.3	附录：曲率张量的唯一性	91


6	坐标无关的数学语言初步	93
6.1	切向量场, 余切向量场, 张量场	93
6.1.1	切向量场	93
6.1.2	余切向量场	97
6.1.3	张量场	98
6.2	李导数与微分同胚群	100
6.3	协变导数, 黎曼曲率张量	103
6.4	* 附录: 一个并非重点的证明	107
7	曲率的意义	109
7.1	测地线偏离方程与牛顿近似	109
7.1.1	测地线偏离方程	109
7.1.2	测地线偏离方程与潮汐力	114
7.1.3	爱因斯坦场方程的牛顿近似	115
7.2	* 粒子固有时的一阶和二阶变分公式	118
8	施瓦西解以及广义相对论经典检验	127
8.1	球对称性与度规	127
8.2	联络与曲率的计算	128
8.2.1	如何计算克里斯托夫联络	128
8.2.2	计算球对称时空的联络和曲率	130
8.3	施瓦西解	132
8.3.1	真空爱因斯坦方程	132
8.3.2	施瓦西解	132
8.4	粒子在施瓦西时空中的运动	134
8.4.1	粒子在球对称时空中的运动	135
8.4.2	光线偏折再回顾	137
8.4.3	水星近日点进动	138
9	黑洞基础	143
9.1	黑洞视界	144
9.1.1	视界的定义	144
9.1.2	近视界几何	147
9.1.3	基林 (Killing) 视界	149
9.1.4	表面引力与霍金温度	152

9.2	克鲁斯卡延拓	153
9.2.1	克鲁斯卡延拓的一般原理	153
9.2.2	施瓦西黑洞的克鲁斯卡延拓	157
9.2.3	Reissner-Nordstrom 黑洞的克鲁斯卡延拓	162
9.2.4	Eddington-Finkelstein 坐标	165
9.3	彭罗斯图	168
9.3.1	外尔变换	168
9.3.2	闵可夫斯基时空的彭罗斯图	170
9.3.3	施瓦西时空的彭罗斯图	172
9.3.4	RN 解的彭罗斯图	176
10	带电黑洞与极端黑洞等等	181
10.1	RN 带电黑洞	181
10.1.1	Einstein-Maxwell 方程	181
10.1.2	RN 解	182
10.2	极端黑洞	185
10.3	带宇宙学常数的黑洞	189
11	时空对称性	191
11.1	基林向量场以及最大对称空间	191
11.1.1	基林向量场	191
11.1.2	最大对称空间	194
11.2	对称性与守恒定律	201
11.3	共形基林向量场	204
11.4	静态时空与引力红移	207
11.4.1	稳态时空与静态时空	207
11.4.2	静态时空的引力红移	208
11.5	基林视界	209
12	德西特空间与反德西特空间	213
12.1	德西特和反德西特空间的定义	213
12.2	德西特空间的一些常用坐标以及彭罗斯图	216
12.2.1	整体坐标与彭罗斯图	216
12.2.2	静态坐标以及其它	218
12.3	反德西特空间的一些常用坐标	221

13	克尔黑洞, 以及黑洞热力学定律	225
13.1	克尔黑洞的度规与视界	225
13.1.1	克尔黑洞的度规	225
13.1.2	克尔黑洞的视界和无限红移面	227
13.1.3	稳态观察者	229
13.2	彭罗斯过程	230
13.3	黑洞热力学定律	233
13.3.1	克尔黑洞的表面引力	233
13.3.2	再谈霍金辐射	235
13.3.3	黑洞热力学定律	237
13.4	克尔黑洞的彭罗斯图和时空奇性	241
13.5	超辐射以及黑洞无毛定理	243
13.5.1	超辐射	243
13.5.2	黑洞无毛定理	245
14	再谈能量动量张量	247
14.1	能量条件	247
14.1.1	能动张量的本征值分解	247
14.1.2	弱能量条件和类光能量条件	249
14.1.3	强能量条件和主能量条件等等	251
14.2	理想流体	252
14.2.1	能作为理想流体的两类物理系统	253
14.2.2	理想流体的运动方程	255
15	球对称恒星及其演化	259
15.1	星体结构微分方程	259
15.2	牛顿近似以及均匀密度星	263
15.3	恒星演化	265
16	极简宇宙学	271
16.1	从热大爆炸到宇宙学原理	272
16.1.1	热大爆炸的概念	272
16.1.2	早期宇宙简史	274
16.1.3	原初核合成	276
16.1.4	宇宙微波背景 (CMB) 和宇宙学原理	278

16.2	宇宙膨胀动力学	280
16.2.1	FLRW 度规	280
16.2.2	联络和曲率	283
16.2.3	弗里德曼方程	285
16.2.4	宇宙学红移	286
16.2.5	宇宙学参数及其确定	288
16.2.6	宇宙学视界	293
16.3	暗物质与暗能量	296
16.4	宇宙暴胀	299
16.4.1	暴胀的基本概念	299
16.4.2	暴胀场	300
16.5	参考文献	302
17	引力波	303
17.1	线性化的场方程	304
17.2	真空中的引力波	308
17.2.1	求解波动方程	308
17.2.2	随波起伏	310
17.2.3	精确解	312
17.3	引力波的产生	313
17.3.1	求解线性化爱因斯坦场方程	314
17.3.2	示例：双星系统	317
17.4	诺特技巧推导引力波能动张量	318
17.5	辐射功率	322
17.5.1	引力波能动张量的具体形式	322
17.5.2	再谈 TT 规范	324
17.5.3	辐射功率	326
17.5.4	估算辐射功率的量级	327
17.6	参考文献	328
18	哈密顿表述与 ADM 能量	329
18.1	时空中的超曲面	329
18.1.1	超曲面与 ADM 分解	329
18.1.2	外曲率	332
18.1.3	超曲面上的协变导数	333
18.1.4	Gauss-Codazzi 方程	334

18.2	引力作用量的边界项	337
18.3	哈密顿表述与 ADM 能量	340
18.3.1	构造类时边界	341
18.3.2	引力作用量的 ADM 分解	343
18.3.3	引力场的哈密顿量	345
18.3.4	ADM 能量	347
.1	诺特技巧与能动张量	352
.1.1	时空平移与正则能动张量	352
.1.2	洛伦兹对称性与对称能动张量	355
.2	诺特技巧推导对称能动张量	356
.2.1	任意场	356
.2.2	诺特技巧与对称能动张量	357



前言

本讲义可能要求读者熟悉狭义相对论，以及用拉格朗日力学方法处理的经典场论的一些基础知识。为此，读者可以预先翻看一下我之前的《经典场论新讲》。

本讲义将强调从等效原理和广义协变原理（或者说微分同胚不变性）两条基本原理的角度阐述广义相对论，真正讲清楚为什么要用黎曼几何，而不是简单假定时空是一个伪黎曼流形。

本讲义涵盖广义相对论从入门到提高的基本内容。除了基本的入门内容之外，还会系统讨论黑洞物理，包括彭罗斯图和黑洞热力学等内容。也包含有一个对宇宙学以及引力波的基本介绍。

1. 引力的普适性以及非欧几何

1.1 万有引力的特性

1.1.1 引力的普适性

引力是自然界中最为特殊的一种力，特殊就特殊在它的普适性，一切物质之间，无论是太空中的还是地面上的，无论是星辰还是大海，一切物质之间都有引力相互作用，这就是牛顿所谓的万有引力，万有二字就是强调它的普适性。实际上，我们可以利用引力的普适性给物质下一个定义：问，什么是物质？答，能够产生引力相互作用的东西就是物质！

而衡量一个物体产生引力相互作用能力的一个量即是质量，更准确地说是引力质量，记作 m_G 。根据物质的定义， m_G 也可以作为物质的量的一个度量。

除了引力之外，自然界中另一种普适性的存在即是物质的惯性 (Inertia)，也就是物质保持其运动状态的能力，一切物质都有惯性。我们也可以利用惯性的普适性给物质下一个不同的定义：问，什么是物质？答，具有惯性的东西就是物质。

而衡量一个物体惯性大小的量就是惯性质量，记作 m_I 。根据物质的这第二个定义， m_I 同样可以作为物质的量的一个度量。

由于起源看起来完全不同，物质的量的这两种度量， m_G 和 m_I ，原则上应该不同。但是，伽利略发现，对任何物质，都有引力质量和惯性质量刚好相等，

$$m_G = m_I. \tag{1.1}$$

这两者为何相等？这在爱因斯坦之前可以说是物理学最大的未解之谜。

伽利略是如何发现引力质量和惯性质量正好相等的呢？回答是，伽利略将引力的普适性和惯性的普适性联系起来，因为他发现，一切物质在重力作用下自由下落的规律是普适的！换言之，伽利略发现，一切物质（无论其组成成分为何）在相同的重力作用下都有相同的加速度，即（记 g 为重力强度）

$$a = (m_G/m_I)g \quad (1.2)$$

是一个不依赖于具体物质的普适量。很显然，这除非 m_G/m_I 是一个普适常数，进而只要适当选取单位，就可以使得 $m_G = m_I$ 。

爱因斯坦是如何解释引力质量为什么等于惯性质量的呢？简单地说，爱因斯坦发展了伽利略的思路，进一步提出，引力的普适性和惯性的普适性本质上是一回事。具体来说就是爱因斯坦的等效原理。好了，让我们暂且打住，关于等效原理我们留到后面的章节再详加讨论。暂时我们只需要知道，惯性质量和引力质量相等是将爱因斯坦导向广义相对论的关键线索，有时候也称之为弱等效原理。

$m_G = m_I$ 的弱等效原理可谓久经实验检验。伽利略本人就利用了单摆，通过选择不同物体作为摆锤，并分析单摆周期是否有变化来检验弱等效原理，精度可以达到约 10^{-3} 。19 世纪末，厄缶 (Eötvös) 提出利用扭秤平衡来精巧地检验弱等效原理，这可以使得精度达到 10^{-9} 。2017 年底，MICROSCOPE 卫星实验把精度提高到了 10^{-14} ，即是在 10^{-14} 的精度范围内，引力质量都等于惯性质量。而基于卫星试验的 SETP (Satellite Test of Equivalence Principle) 更是计划把精度提高到 10^{-17} 。

正因为引力质量等于惯性质量，所以之后再提到物体质量时，我们就不再区分是引力质量还是惯性质量了，而是统一称为质量，记为 m 。

前面说过，质量是衡量物质产生引力相互作用能力的一个量。根据牛顿的万有引力定律，质量为 M 和 m 的两个粒子之间的万有引力势能为

$$V = -G \frac{Mm}{r}, \quad (1.3)$$

式中 r 是两粒子之间的距离， G 为牛顿万有引力常数。而在场论的观点中，物体之间之所以有引力相互作用，是因为一个物体产生了一个引力场，这个引力场作用在另一个物体上。因此是 M 产生了一个引力场，记为引力势 Φ ， Φ 作用在 m 上，得到万有引力势能 $V = m\Phi$ 。也即是说，这里的 $\Phi = -G \frac{M}{r}$ ，很显然它满足方程

$$\nabla^2 \Phi = 4\pi G M \delta^3(\mathbf{x}). \quad (1.4)$$

这个方程可以从点粒子质量密度 $M\delta^3(\mathbf{x})$ 推广到任意的质量密度 $\rho(\mathbf{x}, t)$ ，它产生的引力势满足

$$\nabla^2 \Phi = 4\pi G \rho. \quad (1.5)$$

即是说，质量密度是引力场的源，质量密度产生引力场。而根据狭义相对论中爱因斯坦的质能关系，质量密度其实是一种能量密度。因此即是说，物质的能量密度是引力场的源！

然而能量密度不过是能量-动量张量的 00 分量，记为 T_{00} ，因此方程 (1.5) 就成为

$$\nabla^2\Phi = 4\pi GT_{00}. \quad (1.6)$$

但这个方程不是洛伦兹协变的，为了让它洛伦兹协变，我们显然应该将引力场的源推广为整个能量-动量张量 $T_{\mu\nu}$ (这是一个二阶对称张量)，那这时候相应的，引力势 Φ 也要推广成某个张量场 $h_{\mu\nu}$ 。即是说，在相对论中，引力场应该用某个二阶对称张量场 $h_{\mu\nu}$ 来刻画。出人意料的是，爱因斯坦将 $h_{\mu\nu}$ 与时空的几何联系起来，并确定了引力其实是时空弯曲的一种显现。

引力的普适性就反映在，一切物质都有能量-动量张量，而能量-动量张量正是引力场的源，所以一切物质都能产生引力。

接受引力是时空弯曲这一结论并不难。真正难的有两点：**第一点是搞清楚为什么？为什么时空可以弯曲？为什么引力是时空弯曲？第二点是，在数学上如何描述与刻画时空弯曲？**要搞懂第一个难点就需要爱因斯坦关于等效原理的洞察，我们留给后面的章节讨论。而我们下一节将要简单论述的非欧几何就涉及第二个难点。

另一方面，我们可以稍微深入地探讨一下弱等效原理。为此假设有一个引力场，其引力势为 $\Phi(\mathbf{x})$ ，即是说， \mathbf{x} 处质量为 m 的粒子具有引力势能 $m\Phi(\mathbf{x})$ 。进而粒子在引力场中自由下落的作用量就可以写成

$$S = m \int dt \left[\frac{1}{2} \dot{\mathbf{x}}^2 - \Phi(\mathbf{x}) \right]. \quad (1.7)$$

从这个作用量中，我们很容易看清伽利略发现的自由落体的普适性，也就是弱等效原理，那就是，粒子质量只作为一个整体的常数因子而出现！（所以粒子的运动微分方程和质量无关。）进一步，为了保证弱等效原理在相对论情形中依然成立，我们必须要求，粒子质量即使在相对论性的自由落体作用量中，也只能作为一个整体常数因子而出现！至于具体如何将自由落体作用量推广到相对论情形，我们也留到后面的章节。

1.1.2 万有引力是一种吸引力

万有引力的第二个特性是，它总是一种吸引力。这个特性看起来平凡，实际上却有很深远的影响。不妨设想一个质量非常大的物体，其质量大到这种程度以致于其自身无法支撑自身的吸引力，那它就会坍缩！

实际上，我们知道恒星是靠热核反应而发光发热的，当热核反应耗光能量而终止时，按照牛顿的万有引力理论，恒星就会坍缩，一直坍缩下去。当它坍缩到一定程度，引力场就会变得很强，这时候牛顿的理论就不适用了。但是，即使在爱因斯坦的广义相对论中，引力总是吸引这个特性依然存在。所以引力坍缩现象即使在广义相对论中也依然存在。

印度人钱德拉塞卡发现，当恒星的质量小于一个数值时，它最终就会坍缩成白矮星。白矮星的半径很小，但密度很大，并且会发出白光。如果恒星质量小于某个数值，那它坍缩成白矮星以后，其内部的电子简并压就能抗住自身引力的吸引，它就不再坍缩了。但是，如果恒星质量比这个数值要大，那它就还会继续坍缩。朗道认为，只要恒星的质量不超过 2 个太阳，它就会坍缩成一种更小的东西，叫中子星。这以后其内部的中子简并压就能抗住自身的引力，它就不再坍缩了。

但是，如果恒星的质量超过 2 个太阳会怎么样呢？这时候，中子简并压也抗不住自身引力了，它就会一直坍缩下去，最终形成黑洞！什么是黑洞呢？就是时空中的一个区域，它的吸引力强到这样的地步，最后甚至连光都无法从这个区域逃逸出去，所以它看起来是黑的，叫黑洞。

万有引力总是一种吸引力，然而天文观测发现，我们的宇宙在膨胀，不可思议的是，这还是一种加速膨胀，即是说，即使万有引力的相互吸引也没有使得宇宙膨胀的速度减慢。宇宙为什么会加速膨胀呢？这个问题我们留到学完广义相对论的基本方程以后再探讨。暂时只提一句，这和神秘的暗能量有关！

1.2 非欧几何

在搞清楚引力如何与时空弯曲产生联系之前，这一节先让我们简单回顾一下数学家是如何想到研究弯曲空间的，以及他们是如何研究弯曲空间的。

1.2.1 内禀几何与高斯坐标

因为研究大地测量学的需要，高斯系统地研究了三维空间的两维曲面。高斯的研究有一个深刻的发现，根据这个发现，三维空间中的球面和无穷长的圆柱面有本质的区别！你可能会说，两者不都是曲面吗？不都有弯曲吗？有什么本质区别呢？回答是，高斯发现，的确，球面是弯曲的，但是，圆柱面在某种意义上却是平坦的！

对于一个半径为 R 的两维球面，过球面上任意一点，我们总可以作两个正交的大圆，每个大圆的曲率均为 $1/R$ ，球面的弯曲情况可以很好地由这两个大圆的弯曲情况决定，因此可以将这两个大圆曲率的乘积 $1/R^2$ 定义成球面的曲率，称作球面的高斯曲率，记作 $K = 1/R^2$ 。类似的，对于竖直放置的圆柱面，过其上任意一点，我们都可以作一条高，它是直线，因此曲率为零，同时还可以作一个水平的圆，其曲率不为零。但是，这两条曲线曲率的乘积为零，因此，圆柱面的高斯曲率为零，记作 $K = 0$ 。

实际上，对于任何两维曲面，高斯都类似地定义了一个高斯曲率 K ，它一般来说是曲面上不同点的函数，因为对于非球面，它上面不同点的弯曲情况可以不同。高斯曲率的最初定义依赖于表面上的曲线在三维空间中的弯曲情况，也即是说，依赖于两维曲面是如何嵌入到三维空间中的。然而，高斯发现，这个曲率其实是曲面内禀的性质，与它们怎么嵌入在三维空间中没有关系！特别的，高斯曲率为零的曲面就可以在不拉伸也不挤压的情况下在平面上展平，而高斯曲率非零的曲面就做不到这一点。比方说，圆柱面可以沿着高的方向剪开，并摊平（因此本质是平坦的），但是球面就无法摊平。

高斯关于内禀曲率的发现可谓影响深远，实际上，高斯本人已经意识到这个发现的深刻含义，所以他称之为绝妙定理，现在通常称作 **高斯绝妙定理 (Gauss theorem egregium)**！

高斯如何研究两维曲面呢？其方法首先是通过在曲面上引入任意的曲线坐标，称作高斯坐标。比如，对于三维空间中由方程 $x^2 + y^2 + z^2 = R^2$ 给出的两维球面 S^2 ，可以引入球坐标 (θ, ϕ) ，即定义 $x = R \sin \theta \cos \phi, y = R \sin \theta \sin \phi, z = R \cos \theta$ 。很显然，球坐标并非在整个球面上都定义良好，比方对于 $\theta = 0$ 和 $\theta = \pi$ 的北极点和南极点，球坐标就没有定义。

也即是说, 高斯坐标通常不能覆盖整个曲面, 而只能覆盖曲面的某个局部区域。但是, 高斯的新想法是, 曲面上的高斯坐标可以是任意两个与曲面上的点一一对应的**独立变量**, 一个高斯坐标无法覆盖整个曲面, 那就可以用多个相互有交叠的高斯坐标来覆盖。

其次, 高斯引入了度规的概念 (虽然当时不叫度规), 也就是研究曲面上邻近两点之间的距离 ds 。比方说, 对于球面的球坐标而言, 由于球面嵌在三维欧几里得空间中, 所以可以代 $ds^2 = dx^2 + dy^2 + dz^2$, 代入球坐标, 即可得球面上相邻两点间的距离

$$ds^2 = R^2(d\theta^2 + \sin^2\theta d\phi^2). \quad (1.8)$$

这个 ds^2 就称作球面上的度规。

但是, 高斯说了, 可以使用任意局部坐标 (x^1, x^2) , 很显然, 对于任何曲面, 其上两邻近点的距离平方总可以写成

$$ds^2 = g_{ij}dx^i dx^j, \quad \text{这里 } i, j = 1, 2. \quad (1.9)$$

由于 $dx^i dx^j$ 关于 i, j 指标对称, 所以这里 g_{ij} 也关于 i, j 指标对称。很多时候, 我们也把 g_{ij} 同样称作度规或者度规系数。比方说, 对于球面的球坐标, 就有 $g_{\theta\theta} = R^2$, $g_{\phi\phi} = R^2 \sin^2\theta$, $g_{\theta\phi} = g_{\phi\theta} = 0$ 。

假设我们作一个局部坐标变换

$$(x^1, x^2) \rightarrow (x'^1, x'^2) \quad (1.10)$$

则由 $ds^2 = g_{ij}(x)dx^i dx^j = g'_{ij}(x')dx'^i dx'^j$, 就可以得到变换以后的度规系数 $g'_{ij}(x')$ 为

$$g'_{ij}(x') = g_{kl}(x) \frac{\partial x^k}{\partial x'^i} \frac{\partial x^l}{\partial x'^j}. \quad (1.11)$$

高斯绝妙定理说的就是, 曲面的高斯曲率 K 完全由曲面内禀的度规 g_{ij} 决定, 不依赖于曲面是如何嵌入三维空间中的。

1.2.2 非欧双曲几何的发现

从古希腊开始的两千年来, 数学家和哲学家一直认为欧几里得几何是我们的世界不言而喻的真理。其中, 空间邻近的两点 (x, y, z) 和 $(x + dx, y + dy, z + dz)$ 之间的距离 ds 满足毕达哥拉斯定理

$$ds^2 = dx^2 + dy^2 + dz^2. \quad (1.12)$$

然而, 在 19 世纪 20 年代, 由于对欧几里得几何平行公设的研究, 三个数学家, 高斯、鲍耶、以及罗巴切夫斯基, 相互独立地发现了非欧双曲几何。

非欧双曲几何是一种不同于欧几里得平面几何的两维几何, 庞加莱曾经用一个单位圆盘世界来描绘它。庞加莱圆盘世界就是 (x, y) 平面内半径为 $2R$ 的圆盘 $x^2 + y^2 \leq 4R^2$ 。在生活于这个圆盘世界的人看来, 两个邻近的点 (x, y) 和 $(x + dx, y + dy)$ 之间的距离由下式给出

$$ds^2 = \frac{dx^2 + dy^2}{\left(1 - \frac{x^2 + y^2}{4R^2}\right)^2}. \quad (1.13)$$

也就是说，同样的坐标差 (dx, dy) ，在圆盘中心的人看来是一个很小的距离，而在圆盘边缘的人看来却是一个很大的空间距离，(1.13) 包含了两点间距离的所有信息，称为庞加莱圆盘的度规。

奇妙的是，庞加莱圆盘世界的人眼中的直线（两点之间的最短线）不同于我们眼中的直线。比方说，连接圆盘边缘附近两点 A, B 的直线（最短线）就不是我们眼中的直线，为了使得走过的距离尽可能短，一个从 A 点出发的人不会直接走向 B ，而是会将路径向着圆盘中心的方向弯，因为同样的坐标差在圆盘中心处对应更短的距离，因此让路径弯向圆盘中心可以节省距离。

比方说，如果我们考虑的这两个点 A, B 就在圆盘边缘上，那么从 (1.13) 可以看出，直接沿着边缘从 A 到 B 将是一个无穷大的距离，因为度规在边缘是发散的。从 A 出发，为了最快地走向 B ，你应该先沿着圆盘的径向走，以使得度规尽快地降下来，然后你再弯向 B 。实际上，庞加莱圆盘世界的直线是所有和圆盘边缘垂直相交的圆弧，如图 (1.1) 所示。

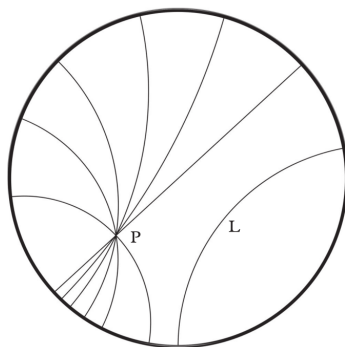


Figure 1.1: 在庞加莱圆盘世界中，过直线外一点，可以作无数多条直线与已知直线永不相交。

值得注意的是，在我们眼中，单位圆盘内的直线长度都是有限的，然而，在庞加莱世界的人看来，其直线的长度是无穷的，可以向两端无限延伸。为了看清楚这一点，我们不妨在单位圆盘上取极坐标 (r, θ) ，并且定义坐标变换 $\frac{r}{R} = \tanh(\rho/2)$ ， ρ 取值范围是 $[0, +\infty)$ ，在变换以后的坐标中，度规的表达式 (1.13) 就变成了，

$$ds^2 = R^2(d\rho^2 + \sinh^2 \rho d\theta^2). \quad (1.14)$$

庞加莱圆盘的边缘现在变到了无穷远处，因此，在庞加莱圆盘世界的人看来，圆盘的边缘实际上意味着无穷远。我们也注意到，只要 ρ 大到一定程度，(1.14) 中度规的系数就一定大于 R^2 ，因此 ρ 坐标值的无穷远实际上也意味着度规距离的无穷远，因此，在庞加莱圆盘世界的人看来，其直线可以向着无穷远无限延伸。

那么，在庞加莱世界中，何为两条直线相互平行呢？通常的定义是，两条永远不相交的直线就是相互平行的，这当然正是我们在欧几里德平面几何中对平行线的自然定义。然而，稍微画画图你就能够看出来，如果这样定义平行，那么在庞加莱圆盘世界中，过直线外一点可以作无数条直线和已知直线平行，如图 (1.1) 所示。并且，庞加莱圆盘世界的两

条平行线之间的距离并非固定不变。如果你从圆盘边缘出发，引两条在边缘处坐标值靠得很近的直线，假如这两条直线不相交，那么你可以很容易地看到，这样的两条平行线之间的距离会以一个恒定的速率增大。两条平行线之间的距离会越来越大实际上说明了庞加莱圆盘世界是一个空间弯曲的世界，而且高斯曲率是负的（如果高斯曲率为正，那么邻近的平行线只会越来越接近，比如球面上的两个大圆）。同时，平行线距离增加的速率为定值还说明庞加莱世界的高斯曲率是一个常数，实际上 $K = -1/R^2$ 。之所以庞加莱圆盘看起来在一个平坦的两维平面上，是因为我们通过局部的拉伸和挤压把这个弯曲世界画在了两维平面上。

现在，在庞加莱圆盘内取三条直线（三条和边缘垂直相交的圆弧），它们将会交出一个三角形，很显然，这个三角形的内角和小于 180 度。假如我们记这个三角形的内角和为 Δ ，那么高斯博内特公式告诉我们，

$$\Delta = \pi - S/R^2 \quad (1.15)$$

式中 S 表示这个三角形的面积。为了更好地理解这个公式，你可以在三维空间的半径为 R 的球面上取一个由三个大圆组成的三角形，你会发现，对于球面 S^2 上的这些三角形而言，其内角和 Δ 必定大于 180 度，并且符合公式 $\Delta = \pi + S/R^2$ 。这是因为球面的高斯曲率为 $+1/R^2$ ，所以面积对三角形内角和的修正是正的，而庞加莱圆盘的高斯曲率为 $-1/R^2$ ，所以面积对三角形内角和的修正应该是负的。

这就带来两个有趣的结论，第一，如果庞加莱圆盘世界的人只在很小的范围内活动，考虑的总是很小的三角形（边长 $\ll R$ ），那么这些三角形的内角和就近似地等于 π ，欧几里德几何近似地成立，这就是所谓的**局部平坦**。第二，庞加莱圆盘世界的三角形面积有上限，最大值为 πR^2 。

庞加莱圆盘世界的哪些三角形有最大面积 πR^2 呢？答案非常简单，所有三个角都落在圆盘边缘上的三角形面积都是 πR^2 。这是因为三角形必须由三条直线构成，而庞加莱圆盘世界的直线在圆盘边缘一定和边缘相垂直，这样一来，落在边缘的角的两条边只能相切，夹角为 0。也就是说，三个角都落在圆盘边缘的任意三角形内角和必定为零，由 (1.15)，这就意味着这些三角形的面积为 πR^2 。

由于非欧双曲几何的发现，以及对三维空间中曲面的研究，高斯成为了第一个思考我们所居住的三维空间可能不是人们一直假设的平坦的欧几里得空间，而可能是某种弯曲空间的人。由于知道在弯曲空间中，三角形的内角和不是 180 度，所以高斯在德国中部哈尔兹山脉的三座山峰上安装了测量设备，以连接这些山峰的最短线为三角形的边（高斯假设从一座山峰到另一座山峰的光线走这条最短路径）。高斯测量了在每座山峰相遇的两条边之间的夹角，以确定三角形内角和是否为 π 。他发现结果确实为 π ，并得出结论，空间确实是欧几里得的。不过，高斯的思想深刻地影响了他的学生黎曼，使得黎曼成为第一个研究高维（高于两维）弯曲空间的人，这就是所谓的黎曼几何。

要把高斯的内禀几何推广到任意维空间，关键是要推广高斯曲率的概念！黎曼提出黎曼几何的时候正是完成了这个推广工作。值得强调的是，这样的推广是高度非平凡的，这是因为，刻画两维空间的曲率只需要一个数，但是高维空间就有许多不同的弯曲方向，因

此就不能简单地用一个数来刻画其曲率，如何刻画高维空间的弯曲也就成了一个难题。最终黎曼通过定义黎曼曲率张量解决了这个问题。

1.2.3 局部平坦与黎曼曲率

为了将曲率的概念推向任意维弯曲空间，我们需要进一步研究两维内禀几何的一个重要性质，就是前面提及的局部平坦！下面让我们首先搞清楚局部平坦的准确含义。

古时候的人不知道地球是球体，他们一直以为大地是平的。这反映了两维球面的局部平坦性，即，在一个足够小的局部区域看来，球面是平坦的。为了看清楚局部平坦的准确含义，让我们考察球面的一种特殊局部坐标。具体来说，对于三维空间中由方程 $x^2 + y^2 + z^2 = R^2$ 给出的两维球面 S^2 ，它在南极点 $(0, 0, -R)$ 附近近似有 $z = -\sqrt{R^2 - x^2 - y^2} \approx -R + \frac{x^2 + y^2}{2R}$ ，代入 $ds^2 = (dx^2 + dy^2 + dz^2)|_{S^2}$ ，即可得

$$\begin{aligned} ds^2 &\approx dx^2 + dy^2 + \frac{(xdx + ydy)^2}{R^2} \\ &= dx^2 + dy^2 + O\left(\left(\frac{x}{R}\right)^2, \left(\frac{y}{R}\right)^2, \left(\frac{x}{R}\right)\left(\frac{y}{R}\right)\right) \\ &= \left(1 + \frac{x^2}{R^2}\right)dx^2 + \left(1 + \frac{y^2}{R^2}\right)dy^2 + 2\frac{xy}{R^2}dxdy. \end{aligned} \quad (1.16)$$

式中 O 表示二阶小量。也即是说，在南极点附近，球面 S^2 上局部坐标 (x, y) 的度规直到一阶小量为止都是平直的欧氏平面度规，只在二阶小量上有修正，这就是局部平坦。换言之，由于局部平坦，在南极点附近我们总可以取局部欧氏坐标系 (x, y) ，其相应度规系数在南极点处的一阶偏导都是零，而二阶偏导正比于球面的曲率 $1/R^2$ ！

但是，对于两维球面来说，南极点并没有什么特殊的。南极点附近局部平坦则球面任意点附近都局部平坦，而局部平坦的准确含义就是：**在任意给定点附近我们总可以取这样的局部坐标系，使得其度规在这点上等同于平坦的欧氏度规，并且这个度规在这点处的一阶偏导为零，而二阶偏导正好反映球面内禀的高斯曲率！**

黎曼推广高斯内禀几何的关键就在于注意到其本质，即任何弯曲的曲面都可以看成是将无限多个平坦的无穷小局部拼接起来的！也就是将无限多个局部平坦的坐标系在度规的二阶偏导上拼接起来，使得高斯曲率的定义不依赖于特定的坐标系。

黎曼首先将高斯坐标以及度规的概念推广到了任意 D 维空间，即对于 D 维空间的任意局部坐标 $x^\mu, \mu = 1, 2, \dots, D$ ，黎曼引入了两相邻点之间的距离平方（也称作线元）

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu, \quad (1.17)$$

同样的，度规系数 $g_{\mu\nu}$ 关于两个指标对称。局部坐标当然可以任意选，只要和 D 维空间的点一一对应即可，但刻画两点间距离的线元必须在局部坐标变换下保持不变。即假设将局部坐标从 x^μ 变换到 x'^μ ，必定有

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu = g'_{\rho\sigma}(x')dx'^\rho dx'^\sigma. \quad (1.18)$$

即在坐标变换下，度规系数按如下方式变换

$$g_{\rho\sigma}(x) \rightarrow g'_{\rho\sigma}(x') = g_{\mu\nu}(x) \frac{\partial x^\mu}{\partial x'^\rho} \frac{\partial x^\nu}{\partial x'^\sigma}. \quad (1.19)$$

特别的, 对于平坦的 D 维欧几里得空间, 我们总可以选择笛卡尔直角坐标系, 从而使得

$$ds^2 = \delta_{\mu\nu} dx^\mu dx^\nu. \quad (1.20)$$

当然, 笛卡尔直角坐标系不是唯一的, 因为可以作一个坐标系的正交旋转。由于任意两个不同的直角坐标方向都可以构成一个独立的转动平面, 所以独立的坐标系旋转共有 $\frac{1}{2}D(D-1)$ 个。

利用方程 (1.19), 黎曼可以证明, 高斯内禀几何局部平坦的性质在高维中依然可以保持。具体证明如下: 首先, 任选一个点 P , 不妨假设 P 点的局部坐标为 $x^\mu = 0$ 。将度规在 P 点附近展开到二阶

$$g_{\mu\nu}(x) = g_{\mu\nu}(0) + A_{\mu\nu,\rho} x^\rho + B_{\mu\nu,\rho\sigma} x^\rho x^\sigma + \dots \quad (1.21)$$

式中 $A_{\mu\nu,\rho}$ 和 $B_{\mu\nu,\rho\sigma}$ 均为关于 $\mu\nu$ 指标对称的常数, 而且由于 $x^\rho x^\sigma$ 关于指标 $\rho\sigma$ 对称, 所以 $B_{\mu\nu,\rho\sigma}$ 也关于 $\rho\sigma$ 指标对称。注意到在 D 维空间, 两个对称指标共有 $\frac{1}{2}D(D+1)$ 个独立的可能性, 所以 $A_{\mu\nu,\rho}$ 共有 $\frac{1}{2}D(D+1)D = \frac{1}{2}D^2(D+1)$ 个独立分量, 而 $B_{\mu\nu,\rho\sigma}$ 有 $(\frac{1}{2}D(D+1))^2$ 个独立分量。

其次, 作如下坐标变换

$$x^\mu = K^\mu{}_\nu x'^\nu + L^\mu{}_{\nu\rho} x'^\nu x'^\rho + M^\mu{}_{\nu\rho\sigma} x'^\nu x'^\rho x'^\sigma + \dots \quad (1.22)$$

式中 $L^\mu{}_{\nu\rho}$ 关于 $\nu\rho$ 指标对称, 而 $M^\mu{}_{\nu\rho\sigma}$ 关于 3 个指标 $\nu\rho\sigma$ 全对称。证明高维内禀几何依然局部平坦的关键在于: 证明总可以通过合适地选择系数 $K^\mu{}_\nu$ 、 $L^\mu{}_{\nu\rho}$ 使得在新坐标 x' 中的度规系数 $g'_{\mu\nu}(x')$ 满足两个条件, 其一, $g'_{\mu\nu}(P)$ 为欧氏度规系数 $\delta_{\mu\nu}$, 其二, $g'_{\mu\nu}(x')$ 在 P 点处的一阶偏导等于零, 也就是总可以完全消去度规的一阶展开系数 $A_{\mu\nu,\rho}$ 。

证明并不难, 我们可以按照上面级数展开逐阶进行证明。首先, 在级数展开的零阶, 由方程 (1.19), 显然有

$$g'_{\rho\sigma}(0) = g_{\mu\nu}(0) K^\mu{}_\rho K^\nu{}_\sigma. \quad (1.23)$$

假设将 $g_{\mu\nu}(0)$ 记作对称矩阵 g , $g'_{\rho\sigma}(0)$ 记作对称矩阵 g' , $K^\mu{}_\rho$ 记作矩阵 K , 那么上式即是

$$g' = K^T g K, \quad (1.24)$$

这里 T 表示矩阵转置。线性代数的定理告诉我们, 对于正定的对称矩阵 g , 总是存在矩阵 K , 使得它被对角化成单位矩阵。即总是存在 $K^\mu{}_\rho$, 使得变换之后的 $g'_{\mu\nu}(0) = \delta_{\mu\nu}$ 。

值得提及的是: $K^\mu{}_\rho$ 共有 D^2 个独立分量, 而对称的 $g_{\mu\nu}(0)$ 共有 $\frac{1}{2}D(D+1)$ 个独立分量。所以为了将 $g_{\mu\nu}(0)$ 变换为 $\delta_{\mu\nu}$ 需要用去 $K^\mu{}_\rho$ 的 $\frac{1}{2}D(D+1)$ 个独立分量, 还剩下 $D^2 - \frac{1}{2}D(D+1) = \frac{1}{2}D(D-1)$ 个自由分量。正好对应笛卡尔直角坐标系的 $\frac{1}{2}D(D-1)$ 个独立旋转!

现在, 假设 $g_{\mu\nu}(0)$ 已经变换成 $g_{\mu\nu}(0) = \delta_{\mu\nu}$ 了。也就是说假设度规可以展开为

$$g_{\mu\nu}(x) = \delta_{\mu\nu} + A_{\mu\nu,\rho} x^\rho + \dots \quad (1.25)$$

我们将继续证明可以通过坐标变换的下一阶完全消去 $A_{\mu\nu,\rho}$ 。为此只需考虑坐标变换

$$x^\mu = x'^\mu + L^\mu{}_{\nu\rho} x'^\nu x'^\rho \dots \quad (1.26)$$

结果几乎是显然的, 因为我们可以自由调节的系数 $L^\mu{}_{\nu\rho}$ 的个数和我们要消去的系数 $A_{\mu\nu,\rho}$ 一样多, 都是 $\frac{1}{2}D(D+1)D = \frac{1}{2}D^2(D+1)$ 个! 所以我们的确可以通过合适选取 $L^\mu{}_{\nu\rho}$ 来完全消去 $A_{\mu\nu,\rho}$ 。至此就已经证明了高维内禀几何也有局部平坦的性质!

现在, 我们可以假设

$$g_{\mu\nu}(x) = \delta_{\mu\nu} + B_{\mu\nu,\rho\sigma} x^\rho x^\sigma + \dots, \quad (1.27)$$

相应的坐标变换可以假设为

$$x^\mu = x'^\mu + M^\mu{}_{\nu\rho\sigma} x'^\nu x'^\rho x'^\sigma + \dots \quad (1.28)$$

根据前面关于局部平坦的讨论可以知道, 系数 $B_{\mu\nu,\rho\sigma}$ 反映的正是空间在 P 点的弯曲情况。但是空间在 P 点的曲率应该不依赖于局部坐标系, 所以并非 $B_{\mu\nu,\rho\sigma}$ 的所有分量都可以用来刻画曲率, 我们还要讨论能够通过调节坐标变换系数 $M^\mu{}_{\nu\rho\sigma}$ 消去多少个与曲率无关的分量。

为此我们需要考察有多少个可调节的变量 $M^\mu{}_{\nu\rho\sigma}$ 。首先将注意力集中在全对称指标 $\nu\rho\sigma$ 上, 它有多少种独立可能性呢? 不妨记有 $f(D)$ 种可能性, 由于是三重指标, 所以 $f(D)$ 必然是 D 的三次多项式, 并且当然 $f(0) = 0$, 所以不妨将 $f(D)$ 写成 $f(D) = (aD^2 + bD + c)D$, 其中 a, b, c 为 3 个待定常数。下面我们可以用数据拟合法找出 $f(D)$ 的具体函数形式。首先, $f(1) = 1$ (即只有三个指标全取 1 这一种可能性)。其次, 为了找出 $f(2)$, 我们不妨利用全对称性, 假设三个指标 ν, ρ, σ 间满足 $\nu \geq \rho \geq \sigma$, 所以只有 222, 221, 211, 111 这四种独立可能性, 即 $f(2) = 4$ 。类似的, 可以得到 $f(3) = 10$ 。根据 $f(1)$ 、 $f(2)$ 以及 $f(3)$ 的值, 我们就能定出待定常数 a, b, c , 进而得到 $f(D) = \frac{1}{6}D(D+1)(D+2)$ 。所以 $M^\mu{}_{\nu\rho\sigma}$ 的独立分量个数为

$$Df(D) = \frac{1}{6}D^2(D+1)(D+2). \quad (1.29)$$

因此, 在 $B_{\mu\nu,\rho\sigma}$ 的 $(\frac{1}{2}D(D+1))^2$ 个独立分量中, 我们可以通过坐标变换消去 $\frac{1}{6}D^2(D+1)(D+2)$ 个分量。所以 $B_{\mu\nu,\rho\sigma}$ 中真正能用以刻画 P 点曲率的只有

$$\left(\frac{1}{2}D(D+1)\right)^2 - \frac{1}{6}D^2(D+1)(D+2) = \frac{1}{12}D^2(D^2-1) \quad (1.30)$$

个。比方说, 对于 $D = 2$, 我们得到刻画曲率的分量个数为 1, 正好是高斯曲率。但是, 对于 $D > 2$ 的高维空间, 曲率必须用一个具有 $\frac{1}{12}D^2(D^2-1)$ 个分量的量才能刻画。

找到这样一个量的具体公式相当具有挑战性, 目前我们能利用的信息只有: 首先, 这个公式必须包含度规的二阶导数 (因为它来自于度规级数展开的二阶项 $B_{\mu\nu,\rho\sigma}$), 其次, 这个公式必须将无限多个局部平坦的坐标系拼接起来, 使得结果不依赖于特定的坐标系。黎曼完成了这个困难的挑战性任务, 得到的就是我们称之为黎曼曲率张量的一个量。

2. 狭义相对性原理以及能量-动量张量

本章主要是帮助不太熟悉狭义相对论和与之相关的经典场论的读者快速熟悉这两方面的内容。对这两方面很熟悉的读者可以跳过本章的大部分内容。但建议不要跳过最后引入能量动量张量的那一节，因为它对于本书后续章节的讲解有很关键的作用，即使你已经了解了相关内容，也请务必浏览一下最后这一节。

2.1 相对性原理

狭义相对论的基本原理是相对性原理，有时候也称之为狭义相对性原理，它说的是：在所有惯性系中，一切物理规律——包括相互作用的传播规律——都是相同的。

特别的，爱因斯坦提出，相互作用的传播速度不是无穷大，而是有限的。的确，按照今天对经典物理的理解，一个物体要对另一个物体施加作用，就要向它发出一个信号，而受作用的物体只是对这个信号进行响应，在场论中，这个传播相互作用的信号就是场的波动。总之，物体间的相互作用需要信息的传递，但是信息传递的速度不可能像超距作用说的那样是无穷大，而必定是有限的。不妨记信息传播的最大速度为 c (当然， c 就是真空中的光速)，按照相对性原理， c 必定不依赖于所选的惯性系，而是一个不变的常数，因此我们当然可以合适地选择时间的单位，使得 $c = 1$ 。

2.1.1 间隔不变性与闵可夫斯基几何

考察 S 系和 S' 系这两个惯性系，假定同一事件在这两个参考系中的时空坐标分别是 (t, x, y, z) 和 (t', x', y', z') 。则，由于时空的均匀性，这两个参考系之间的坐标变换一定是一个线性变换。

现在, 设想在 S 系中, 从 (t, x, y, z) 点发出一束光到达 $(t + dt, x + dx, y + dy, z + dz)$ 点, 由于 $c = 1$, 显然我们有

$$-dt^2 + dx^2 + dy^2 + dz^2 = 0. \quad (2.1)$$

根据 c 的不变性, 同样的两个事件在 S' 系看来也得满足

$$-dt'^2 + dx'^2 + dy'^2 + dz'^2 = 0. \quad (2.2)$$

换言之, 两个邻近事件在两不同参考系中的时空坐标必得满足一个约束关系, 即当 (2.1) 成立时必有 (2.2) 成立, 反之亦然。又由于两参考系之间的坐标变换是线性变换, 因此, 对任意的两个邻近事件, 我们必有

$$-dt'^2 + dx'^2 + dy'^2 + dz'^2 = D(\mathbf{v})(-dt^2 + dx^2 + dy^2 + dz^2), \quad (2.3)$$

式中 \mathbf{v} 是 S' 系相对于 S 系的速度。同样的, 由于 S 与 S' 地位平等, 如果从 S' 变换到 S , 就有

$$-dt^2 + dx^2 + dy^2 + dz^2 = D(-\mathbf{v})(-dt'^2 + dx'^2 + dy'^2 + dz'^2). \quad (2.4)$$

换言之, 我们必有 $D(-\mathbf{v})D(\mathbf{v}) = 1$ 。又由于空间的各向同性可知, D 对相对速度 \mathbf{v} 的依赖只能是依赖于其大小 v , 而必定和其方向无关, 因此我们必定有 $D(-\mathbf{v}) = D(\mathbf{v}) = D(v)$ 。因此, $D(-\mathbf{v})D(\mathbf{v}) = (D(v))^2 = 1$, 即 $D(v) = \pm 1$ 。又由于 $D(v)$ 是 v 的连续函数, 而且 $D(0) = 1$ (对应 S 和 S' 为同一个参考系的情形), 因此必有 $D(v) = 1$ 。因此, 对于任意两个邻近事件我们必有

$$-dt'^2 + dx'^2 + dy'^2 + dz'^2 = -dt^2 + dx^2 + dy^2 + dz^2. \quad (2.5)$$

通常将 $-dt^2 + dx^2 + dy^2 + dz^2$ 称为两个邻近事件的时空间隔的平方, 简称间隔平方, 并记为 ds^2 , 即

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 = -dt^2 + d\mathbf{x}^2. \quad (2.6)$$

用这个记号, 方程 (2.5) 就可以简记成

$$ds'^2 = ds^2, \quad (2.7)$$

称为两个事件的间隔不变性。

很显然, 狭义相对论中的间隔平方完全类似于平坦的欧几里得几何中的线元 ds^2 , 只不过现在处理的是时空而不是单纯的空间, 而且在 dt^2 前面多了一个负号。基于这个观察, 闵可夫斯基建议将狭义相对论看成是一种平坦时空的几何, 这种几何类似于描述平坦空间的欧几里得几何, 通常称作闵可夫斯基几何。根据这个观点, 惯性参考系之间的坐标变换就类似于欧几里得几何中的坐标系旋转, (只不过现在是时空坐标的“旋转”), 所以保持“两点距离”的平方, 即间隔平方 (现在也同样称作闵氏几何的线元)。

人们通常约定 $t = ct = x^0 (c = 1), x = x^1, y = x^2, z = x^3$, 这样就把四个时空坐标统一地记成了 $x^\mu, \mu = 0, 1, 2, 3$ 。利用这个记号, 我们就可以将闵氏几何的线元重写为

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu, \quad (2.8)$$

式中 $\eta_{\mu\nu}$ 为, $-\eta_{00} = \eta_{11} = \eta_{22} = \eta_{33} = 1$, 其它指标分量都等于零。类似于欧几里得空间的度规张量 $\delta_{\mu\nu}$, 我们称 $\eta_{\mu\nu}$ 为四维闵可夫斯基时空的度规张量。值得注意的是, $\eta_{\mu\nu}$ 关于它的两个指标是对称的, 即满足 $\eta_{\mu\nu} = \eta_{\nu\mu}$ 。

也可以将 $\eta_{\mu\nu}$ 看成是一个 4×4 矩阵的分量形式 (记这个矩阵为 η)

$$\eta_{\mu\nu} = \text{diag}(-1, +1, +1, +1), \quad (2.9)$$

进而引入这个矩阵的逆矩阵 η^{-1} , 其分量形式记为 $\eta^{\mu\nu}$,

$$\eta_{\alpha\beta} \eta^{\beta\gamma} = \delta_{\alpha}^{\gamma}, \quad (2.10)$$

式中 δ_{α}^{γ} 为 4×4 单位矩阵的分量形式。很容易看出, $\eta^{\mu\nu}$ 也为, $-\eta^{00} = \eta^{11} = \eta^{22} = \eta^{33} = 1$, 其它指标分量都等于零。

假设我们考察的不是两个邻近时空点, 而是两个有限间隔的时空点 x_1^μ 和 x_2^μ , 记 $\Delta x^\mu = x_2^\mu - x_1^\mu$, 记这两个事件的时空间隔为 Δs , 则有

$$(\Delta s)^2 = \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu = -(\Delta t)^2 + (\Delta \mathbf{x})^2. \quad (2.11)$$

同样, 无论在哪个参考系中计算, 间隔 Δs 都是不变的。1. 如果 $(\Delta s)^2 < 0$, 我们就称 x_1^μ 和 x_2^μ 这两个事件**类时相间** (timelike separated)。由于这时候 $(\frac{\Delta \mathbf{x}}{\Delta t})^2 < 1$, 所以我们总可以用信号将这两个事件联系起来, 所以这两个事件就存在因果关系, 而且正如马上就会看到的, 在不同的惯性系中, 事件的因果关系将会保持不变, 先发生的事件在任何惯性系中都会先发生。2. 如果 $(\Delta s)^2 = 0$, 我们就称这两个事件**类光相间** (lightlike separated)。这时候可以用光信号将两个事件联系起来。3. 如果 $(\Delta s)^2 > 0$, 我们就称这两个事件**类空相间** (spacelike separated)。这时候两事件没有任何因果关系, 不可能用任何信号将它们联系起来, 同时它们的先后顺序也是相对的, 在不同参考系中对哪个事件先发生会有不同的看法。

为了将上述两事件间的关系看得更清楚, 我们取其中一个事件为 $x_1^\mu = 0$, 即位于时空图的坐标原点, 另一个事件 $x_2^\mu = x^\mu$, 记事件 x^μ 与原点事件的间隔为 s , 则

$$s^2 = -t^2 + \mathbf{x}^2. \quad (2.12)$$

对于 $s^2 = 0$ 的类光相间情形, 方程 $-t^2 + \mathbf{x}^2 = 0$ 给出的是时空图上以原点为顶点的圆锥面, 称之为光锥 (lightcone), 如图 (2.1) 所示。

对于 $s^2 < 0$ 情形, 这时候方程 $t^2 - \mathbf{x}^2 = -s^2 > 0$ 给出的是具有两支的双曲面, 一支位于上半光锥所包围的内部区域, 一支位于下半光锥所包围的内部区域, 由于不同惯性系中 $t^2 - \mathbf{x}^2 = -s^2$ 保持不变, 因此在不同惯性系中事件 2 的坐标 x^μ 只能在这个双曲面上变动。不过由于双曲面被光锥分隔成了两支, 所以在不同惯性系中上半支的点只能在上半支上变

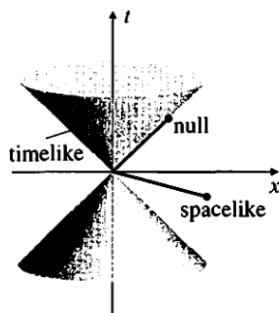


Figure 2.1: 时空图上原点处的光锥。

动，即恒有 $t > 0$ ，而下半支的点则恒有 $t < 0$ ，即是说，处于原点未来的事件在任何惯性系中都保持在未来，而处于原点过去的事件在任何惯性系中都在过去，即，不同参考系不会改变事件间的因果关系。要让双曲面上半支的点变到下半支只能是经过一个 $t \rightarrow -t$ 的时间反演。通常称上半光锥为未来光锥，称下半光锥为过去光锥。

最后，对于 $s^2 > 0$ 情形，方程 $\mathbf{x}^2 - t^2 = s^2 > 0$ 给出的是一个连通的双曲面，原则上，这个双曲面上的任何点可以在一个合适的参考系中变到双曲面上的任何其它点，特别的， $t > 0$ 的点可以变到 $t < 0$ ，反之亦然。即是说，与原点类空相间的事件是先于原点发生还是后发生并没有绝对的意义！

固有时

假如原来有一个参考系，有一个粒子从参考系的 (t, x, y, z) 点运动到 $(t + dt, x + dx, y + dy, z + dz)$ 点，这两点（它们当然类时相间）间的间隔当然满足 $ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$ 。现在，假设有一个钟固定在这个粒子上，记此钟走过的时间为 $d\tau$ ，并且我们依托这个钟建立一个固定在粒子上的参考系，那么在这个参考系中，粒子的空间位移当然是零，从而从这个固定在粒子的参考系看来，间隔 ds 应该满足， $ds^2 = -d\tau^2$ ，即

$$\eta_{\mu\nu} dx^\mu dx^\nu = -d\tau^2. \quad (2.13)$$

上式中的 τ 就称之为粒子走过的固有时，而 $x^0 = t$ 则称之为坐标时。很显然，固有时就是固连在粒子上的钟所走过的时间。

利用固有时，我们可以定义粒子的四维速度 u^μ

$$u^\mu = \frac{dx^\mu}{d\tau}. \quad (2.14)$$

根据 (2.13) 式，显然有

$$\eta_{\mu\nu} u^\mu u^\nu = -1. \quad (2.15)$$

我们称 u^μ 为一个类时矢量。即是说，假如把 u^μ 的起点画在光锥原点的话， u^μ 将完全躺在光锥之内。这就说明，一切从光锥原点经过的粒子，其在时空中的运动轨迹都将在这一点的光锥之内。当然，光子例外，光子的运动轨迹躺在光锥面上，因为光子的固有时总是零！

假设粒子的速度为 \mathbf{v} ，即 $d\mathbf{x} = \mathbf{v}dt$ ，则根据 (2.13) 式即有

$$d\tau^2 = dt^2 - d\mathbf{x}^2 = dt^2(1 - \mathbf{v}^2) \Rightarrow dt = \frac{d\tau}{\sqrt{1 - \mathbf{v}^2}}. \quad (2.16)$$

即是说，坐标时总是比固有时长的。测固有时长的钟当然是相对于原参考系运动的钟，所以这个结果常常也被人们说成是，运动的钟会变慢，因为对同一个参考系中的过程，它测出来的时间更短。

但是，这并非运动的钟本身有什么问题，而是从原参考系的静止观察者来看，运动的一切事物都变慢了，运动的人的生命过程也变慢了。不过，从运动的人自己来看，他自己的一切都是正常的，在他看来，反而是原参考系中的观察者在运动（运动是相对的），反而是这观察者的生命过程变慢了。

光锥与世界线

在狭义相对论中，由于类空相间事件的先后顺序是相对的，因此两个不同地的事件是否同时发生也是相对的，依赖于参考系，这就是所谓的的同时的相对性。在牛顿力学中，我们可以绝对地将时空用等时的空间超曲面分割，因为同时的相对性，在狭义相对论中却无法做到这一点。但这并不说明狭义相对论的时空完全没有结构。相反，在任何一个时空点，我们都可以定义一个光锥，如图 (2.2) 所示，它就是经过这点的光线在时空中的运动轨迹之集合。由于在狭义相对论中，一切粒子的运动速度都不能超过光速，所以经过任何时空点的粒子，其在四维时空中的运动方向都只能在这点的光锥之内，最多是在光锥面上（对于光子），而决定不能在光锥之外。

我们把粒子在时空中的运动轨迹称作粒子的世界线，记为 $x^\mu(\sigma)$ ，很多时候人们会选择粒子的固有时 τ 作为世界线的参数，即取 $\sigma = \tau$ 。根据上一段所说可知，粒子的世界线只能从其上各点的光锥内部穿过，如图 (2.2) 所示。

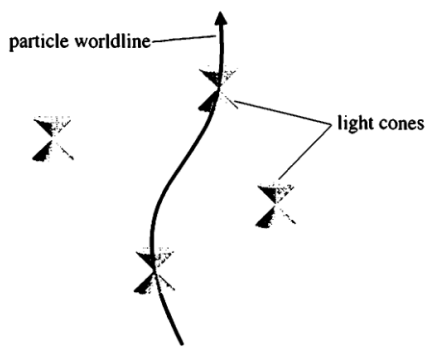


Figure 2.2: 时空图上的光锥和世界线。

前面我们说过运动的钟会变慢，同时我们也说过运动是相对的。那么，假设有两只钟，一只为 Bob 所持有，而 Bob 静止不动，另一只为 Alice 持有，它跟着 Alice 沿着闭合路径运动一圈再回到起点与 Bob 的钟比较，那到底哪只钟慢了呢？回答是，运动的钟绝对地慢了。但是运动不是相对的吗？从运动钟来看，不是静止的钟在运动吗？但是，从这以运

动的钟为参考系的后一种观点导不出静止的钟变慢的结论，因为这时候这个参考系不是一个惯性系，这是由于这个运动的钟是沿着闭合路径运动一圈而不是作匀速直线运动。

其实，这两只钟分别测量的是 Bob 和 Alice 的固有时，而 Bob 与 Alice 在时空中的世界线如图 (2.3) 所示。对于有限长世界线，固有时 T 是 $d\tau$ 沿着世界线的积分

$$T = \int d\tau = \int \sqrt{dt^2 - d\mathbf{x}^2}. \quad (2.17)$$

注意这个积分中的时间微元贡献和空间微元贡献是相减的关系，Bob 世界线的空间完全固定，而 Alice 的世界线在空间上扫过，由于空间的贡献是要减去的，所以很显然 Bob 的固有时比较长，而 Alice 的固有时比较短，所以 Alice 的钟慢！

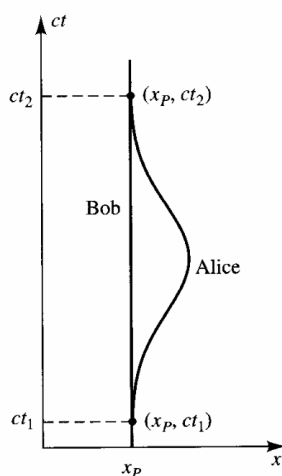


Figure 2.3: Bob 与 Alice 在时空中的世界线。

这就是闵可夫斯基几何与欧几里得几何不同的地方，在欧几里得几何中，两点之间直线总是最短的。但是在闵可夫斯基几何中，对于运动粒子的世界线，两点之间直线的固有时反而是最长的，而曲线由于要减去更多的空间贡献，固有时反而会比较短。

2.1.2 庞加莱群

狭义相对论中，两个惯性参考系之间的时空坐标变换称作洛伦兹变换，根据前面所说，它是一个保持间隔不变性的线性变换，通常写成

$$dx'^{\mu} = \Lambda^{\mu}_{\nu} dx^{\nu}, \quad (2.18)$$

式中 Λ^{μ}_{ν} 构成变换矩阵 Λ 的分量形式。根据间隔不变性，我们有

$$\eta_{\mu\nu} dx'^{\mu} dx'^{\nu} = \eta_{\mu\nu} \Lambda^{\mu}_{\alpha} \Lambda^{\nu}_{\beta} dx^{\alpha} dx^{\beta} = \eta_{\alpha\beta} dx^{\alpha} dx^{\beta}, \quad (2.19)$$

从而即有

$$\eta_{\mu\nu} \Lambda^{\mu}_{\alpha} \Lambda^{\nu}_{\beta} = \eta_{\alpha\beta}. \quad (2.20)$$

我们当然可以将 (2.20) 式写成矩阵形式, 即

$$\Lambda^T \eta \Lambda = \eta. \quad (2.21)$$

很容易验证, 如果洛伦兹变换 Λ_1 满足上面式子, Λ_2 也满足上面式子, 则 $\Lambda_1 \Lambda_2$ 必定也满足上面式子, 从而也是洛伦兹变换。即是说, 所有洛伦兹变换的集合在矩阵乘法下封闭。另外, 对 (2.21) 式两边求行列式, 并注意到 $\det(\eta) = -1$, 从而即可得

$$[\det(\Lambda)]^2 = 1 \Rightarrow \det(\Lambda) = \pm 1. \quad (2.22)$$

由此可知矩阵 Λ 必定存在逆矩阵 Λ^{-1} , 并且很明显 Λ^{-1} 也是洛伦兹变换, 即任何洛伦兹变换都有逆变换。满足乘法封闭性, 并且存在逆元素的元素集合就是数学上所谓的群, 所以, 所有洛伦兹变换的集合构成一个群, 称作洛伦兹群, 常常记作 $O(1,3)$ 。很明显, 所有 $\det(\Lambda) = 1$ 的洛伦兹变换也构成一个群, 它是 $O(1,3)$ 的子群, 通常记作 $SO(1,3)$, 实际上, 人们在谈到洛伦兹群的时候更多都是指的这个 $SO(1,3)$ 群。

进一步, 在 (2.20) 式中取 $\alpha = \beta = 0$, 即可得

$$(\Lambda^0_0)^2 = 1 + \sum_{i=1,2,3} (\Lambda^i_0)^2 \geq 1 \Rightarrow \Lambda^0_0 \geq 1 \text{ or } \Lambda^0_0 \leq -1. \quad (2.23)$$

从而根据 $\det(\Lambda)$ 的正负以及 Λ^0_0 的正负, 我们可以将洛伦兹变换的集合分成四个子集。其中所谓的正洛伦兹变换要求满足下面条件

$$\det(\Lambda) = 1, \quad \Lambda^0_0 \geq 1. \quad (2.24)$$

容易验证, 正洛伦兹变换的集合也构成一个群, 称作正洛伦兹群, 它是洛伦兹群的子群, 通常记为 $SO^+(1,3)$ 。

除了惯性参考系之间的洛伦兹变换之外, 很显然, 时空坐标的平移

$$x^\mu \rightarrow x'^\mu = x^\mu + a^\mu, \quad (2.25)$$

这也同样保持事件间的间隔不变。洛伦兹变换再加上时空平移就构成一个比洛伦兹群更大的群, 称作庞加莱群。由于四维时空闵可夫斯基几何的 ds^2 在庞加莱群变换下保持不变, 所以它是闵可夫斯基时空的对称群。

2.1.3 四维闵可夫斯基时空的矢量和张量

上一小节说过, 在不同的惯性参考系中, 时空坐标按照下面的洛伦兹变换而变换,

$$dx'^\mu = \Lambda^\mu_\nu dx^\nu, \quad (2.26)$$

式中变换矩阵 Λ 满足

$$\Lambda^T \eta \Lambda = \eta \Leftrightarrow \Lambda^{-1} = \eta^{-1} \Lambda^T \eta. \quad (2.27)$$

类比于四分量的 dx^μ ，假设一个任意的四分量量 $A^\mu = (A^0, \mathbf{A})$ 在参考系的变换下与 dx^μ 的变换规则相同，即满足

$$A'^\mu = \Lambda^\mu_\nu A^\nu, \quad (2.28)$$

则我们称 A^μ 为一个四维闵可夫斯基时空的矢量，简称四矢量，当然严格来讲 A^μ 是四矢量的分量形式。与时空间隔类似，我们可以定义四矢量的平方 A^2 为

$$A^2 = \eta_{\mu\nu} A^\mu A^\nu = -(A^0)^2 + \mathbf{A}^2 = -(A^0)^2 + (A^1)^2 + (A^2)^2 + (A^3)^2. \quad (2.29)$$

很明显， A^2 在洛伦兹变换下是不变的。假如 $A^2 < 0$ ，我们就称 A^μ 为一个类时矢量，假如 $A^2 = 0$ ，我们就称之为一个类光矢量或者零性矢量，假如 $A^2 > 0$ ，就称之为一个类空矢量。

我们也可以定义下指标的四分量量 A_μ 为，

$$A_\mu = \eta_{\mu\nu} A^\nu, \quad (2.30)$$

写得更清楚一点就是

$$A_\mu = (A_0, A_1, A_2, A_3) = (-A^0, A^1, A^2, A^3) = (A_0, \mathbf{A}). \quad (2.31)$$

则 A^2 就可以写成 $A^2 = A_\mu A^\mu$ ，而 A^2 在洛伦兹变换下的不变性则意味着

$$A'_\mu A'^\mu = A_\mu A^\mu. \quad (2.32)$$

注意到 A^μ 在洛伦兹变换下按照 (2.28) 式变换，因此上式就意味着 A_μ 必然按照下式变换

$$A'_\mu = (\Lambda^{-1})^\nu_\mu A_\nu. \quad (2.33)$$

在洛伦兹变换下按照这样变换的量同样叫做四矢量。不过为了区分上指标的四矢量和下指标的四矢量，有时候人们称 A^μ 为四矢量的逆变分量，而称 A_μ 为四矢量的协变分量。利用 $\eta_{\mu\nu}$ 我们可以把上指标降下来，进而将逆变分量转化为协变分量，反过来，我们也可以利用 $\eta^{\mu\nu}$ 将下指标升上去，即

$$A^\mu = \eta^{\mu\nu} A_\nu. \quad (2.34)$$

假设记 $\partial_\mu = \frac{\partial}{\partial x^\mu}$ ，则不难明白全微分 $d = dx^\mu \partial_\mu$ 是不依赖于坐标系的，由此即可以看出，偏导运算 ∂_μ 在洛伦兹变换下和协变四矢量的变换规则相同，即按下式变换

$$\partial'_\mu = (\Lambda^{-1})^\nu_\mu \partial_\nu. \quad (2.35)$$

归纳一下即是，在洛伦兹变换下，四矢量的逆变分量和 dx^μ 的变换规则相同，而协变分量则和 ∂_μ 的变换规则相同。

很显然，任意一个逆变四矢量 A^μ 和任意一个协变四矢量 B_μ 都可以构成一个在洛伦兹变换下保持不变的量，这个量即是 $A^\mu B_\mu$ ，有时候也记作 $A \cdot B$ ，称作两个四矢量 A 和 B

的内积, 有时候也称作矢量 A 和 B 的缩并。两个四矢量的内积 (缩并) 是洛伦兹不变的, 称作一个四维标量, 四维标量即是在洛伦兹变换下保持不变的量。

四维矢量是只有一个指标的量, 我们当然可以进一步考察多个指标的量, 比如 $B^{\mu\nu}$, 如果这个量的每一个指标在洛伦兹变换下都按逆变矢量那样变, 即是说

$$B'^{\mu\nu} = \Lambda^\mu_\alpha \Lambda^\nu_\beta B^{\alpha\beta}, \quad (2.36)$$

我们就称 $B^{\mu\nu}$ 为一个 2 阶逆变张量, 或者记作 (2,0) 张量, (2,0) 代表它有 2 个上指标 0 个下指标。类似的, 我们也可以考察 (0,2) 张量, 它即是两个下指标, 且在洛伦兹变换下按照下式变换的量,

$$B'_{\mu\nu} = (\Lambda^{-1})^\alpha_\mu (\Lambda^{-1})^\beta_\nu B_{\alpha\beta}. \quad (2.37)$$

进一步, 也可以考察混合型张量, 比如 (1,1) 张量, 它即是一个上指标一个下指标, 且在洛伦兹变换下按照下式变换的量,

$$B'^\mu_\nu = \Lambda^\mu_\alpha (\Lambda^{-1})^\beta_\nu B^\alpha_\beta. \quad (2.38)$$

类似的概念可以很容易推广到有 p 个上指标 q 个下指标的 (p,q) 张量。特别的, (0,0) 张量就是四维标量, (1,0) 张量就是四维逆变矢量, 而 (0,1) 张量则是四维协变矢量。

当然, 完全类似于四矢量情形, 我们同样可以用 $\eta_{\mu\nu}$ 来将张量的上指标降下来, 也可以用 $\eta^{\mu\nu}$ 来将张量的下指标升上去。而且, 对于一个 (p,q) 张量, 我们可以让它的某个上指标和某个下指标相同, 从而默认对这个指标求和, 结果就是一个 $(p-1, q-1)$ 张量, 这同样也叫做张量的缩并。比如说, 对于 (1,1) 张量 B^μ_ν , 我们可以考察 $B^\mu_\mu = B^0_0 + B^1_1 + B^2_2 + B^3_3$, 注意它的上指标和下指标已经求和掉了, 从而人们很容易验证它是洛伦兹不变的, 即是一个 (0,0) 张量, 或者说是一个四维标量。

另外, 比如说对于 (2,0) 张量 $B^{\mu\nu}$, 我们可以进一步要求它的两个指标对称, 即满足 $B^{\mu\nu} = B^{\nu\mu}$, 这就叫二阶对称张量。而如果我们要求两个指标反对称, 即满足 $B^{\mu\nu} = -B^{\nu\mu}$, 那就叫二阶反对称张量。对于反对称张量 $B^{\mu\nu}$, 我们有 $B^{00} = B^{11} = B^{22} = B^{33} = 0$, 这是因为比如说 $B^{00} = -B^{00}$, 从而必有 $B^{00} = 0$ 。对于 (0, p) 张量 $C_{\mu_1\mu_2\dots\mu_p}$, 如果它的任意两个指标均反对称, 我们就称之为 p 阶反对称张量。但是在四维时空中, 必定有 $p \leq 4$ 。这是因为, 在四维时空中, 任何指标都只能取 0,1,2,3, 从而对于 $p > 4$ 的情形, $C_{\mu_1\mu_2\dots\mu_p}$ 的任意 p 个下指标中必有两个取相同值, 考虑到反对称这就意味着 $C_{\mu_1\mu_2\dots\mu_p} = 0$, 即高于 4 阶的反对称张量必定为零。进一步, 由于 p 阶反对称张量场 p 个指标必须全不相同, 所以在四维时空中, 它的独立分量个数就是 $\frac{4!}{p!(4-p)!}$ 。

最后, 四维张量的概念很容易推广到场, 如果一个量既是一个四维张量, 同时还是一个场, 那就叫做张量场, 比如一个 (0,2) 型二阶张量场可以写成 $B_{\mu\nu}(x)$, 式中 x 表示时空点。 $B_{\mu\nu}(x)$ 在洛伦兹变换下按照下式变

$$B_{\mu\nu}(x) \rightarrow B'_{\mu\nu}(x') = (\Lambda^{-1})^\alpha_\mu (\Lambda^{-1})^\beta_\nu B_{\alpha\beta}(x). \quad (2.39)$$

其它四维张量场的变换规则可以类似地推广。特别的, 对于标量场 $\Phi(x)$, 我们有

$$\Phi(x) \rightarrow \Phi'(x') = \Phi(x). \quad (2.40)$$

2.2 狭义相对性原理与经典物理

2.2.1 狭义相对性原理与经典场论

现在我们可以将狭义相对性原理重新表述为，任何物理规律都应该在洛伦兹变换下保持不变。我们知道，经典场论的规律（也就是场方程）可以由最小作用量原理导出，因此这就意味着经典场论的作用量泛函必须在洛伦兹变换下保持不变！这就意味着经典场论的作用量泛函必须是洛伦兹标量。

另外，对于局域场论，作用量 S 总可以写成拉格朗日密度 \mathcal{L} 的积分，即 $S = \int d^4x \mathcal{L}$ ，注意到由于 $\det(\Lambda) = 1$ ，因此体积元 d^4x 显然是洛伦兹不变的，因此 S 要是洛伦兹标量当且仅当拉格朗日密度 \mathcal{L} 为洛伦兹标量！

标量场

进一步，假设我们考虑的是一个标量场论，场变量记为 ϕ ，则拉氏密度实际上是 ϕ 和 $\partial_\mu \phi$ 的函数（这里 $\partial_\mu = \frac{\partial}{\partial x^\mu}$ ），记为 $\mathcal{L} = \mathcal{L}(\phi, \partial_\mu \phi)$ 。根据最小作用量原理，我们可以进行如下推导

$$\begin{aligned} \delta S[\phi(x)] &= \int d^4x \delta \mathcal{L}(\phi, \partial_\mu \phi) \\ &= \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi} \delta \phi + \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \delta (\partial_\mu \phi) \right] = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi} \delta \phi + \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \partial_\mu (\delta \phi) \right] \\ &= \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) \right] \delta \phi + \int d^4x \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \delta \phi \right). \end{aligned} \quad (2.41)$$

很显然，最后一行的第二项是全微分项，因此积分结果完全取决于边界项，为

$$\int d^4x \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \delta \phi \right) = \int dS_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \delta \phi \right) |_\infty. \quad (2.42)$$

式中 dS_μ 表示四维时空无穷远三维边界的体积元。总之，结果仅仅在四维时空的无穷远边界上有贡献。但是，最小作用量原理要求在时空的边界上场位形是固定的，从而 $\delta \phi = 0$ ，因此这一项的最终结果其实等于零。

从而即有

$$\delta S[\phi(x)] = \int d^4x \left[\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) \right] \delta \phi, \quad (2.43)$$

进一步应用最小作用量原理 $\delta S[\phi(x)] = 0$ ，即可得到场方程，

$$\frac{\delta S}{\delta \phi(x)} = \frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \right) = 0. \quad (2.44)$$

而 $\partial_\mu \phi$ 能构造出来的最简单洛伦兹标量就是

$$\partial_\mu \phi \partial^\mu \phi = \eta^{\mu\nu} \partial_\mu \phi \partial_\nu \phi = -(\partial_t \phi)^2 + (\nabla \phi)^2. \quad (2.45)$$

要求动能项为正，并进一步通过将合适的常数吸进场 ϕ 的定义之中，我们总能把 $\partial_\mu \phi$ 对拉氏密度最简单的贡献写作

$$-\frac{1}{2} \partial_\mu \phi \partial^\mu \phi. \quad (2.46)$$

另外，很显然， ϕ 的任意函数 $-\mathcal{U}(\phi)$ 都是洛伦兹标量，因此可以加到拉氏密度中去，进而就得到如下最简单的洛伦兹不变的拉氏密度

$$\mathcal{L} = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \mathcal{U}(\phi). \quad (2.47)$$

当然，洛伦兹不变性并不能完全决定拉氏密度，比如，读者很容易发现下面的拉氏密度同样洛伦兹不变，

$$\mathcal{L} = -\frac{1}{2}g(\phi)\partial_\mu\phi\partial^\mu\phi - \mathcal{U}(\phi), \quad (2.48)$$

式中 $g(\phi)$ 为 ϕ 的任意函数。这也是一种很常见的标量场模型，虽然人们对它的研究可能比上面那个更简单的模型略少。在这个模型中取 $g(\phi)$ 为常数 g_0 ，然后再将 $\sqrt{g_0}$ 吸收到 ϕ 的定义中去，就回到了上面那个更简单的模型。

读者可能会想为什么只用 ϕ 和 $\partial_\mu\phi$ 构造拉氏密度呢？为什么不考虑二阶导数 ($\partial_\mu\partial_\nu\phi$)，甚至更高阶导数呢？的确，考虑二阶导数也能轻易构造出洛伦兹不变的拉氏密度，比如

$$\mathcal{L} = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi + f(\phi)(\partial_\mu\partial_\nu\phi)(\partial^\mu\partial^\nu\phi). \quad (2.49)$$

但实际上，人们几乎不会研究这种场论模型，原因有两个：第一，这种场论模型用最小作用量原理导出的场方程是四阶微分方程，而我们通常要求物理系统的运动微分方程为二阶微分方程。第二，可以证明，这样含高阶导数的模型导出来的哈密顿量（也就是能量）没有下界，即没有最低能量，从而物理上是不允许的，这就是所谓的 Ostrogradsky 不稳定性。

前面的标量场模型很容易推广，比如说，我们可以同时考察 n 个标量场，记为 $\phi^a, a = 1, 2, \dots, n$ ，这时候很容易构造出如下拉氏密度，

$$\mathcal{L} = -g_{ab}(\phi)\partial_\mu\phi^a\partial^\mu\phi^b. \quad (2.50)$$

式中 $g_{ab}(\phi)$ 是 ϕ^a 的函数，实际上人们通常让它是场空间的黎曼度规。这样的场论模型就是所谓的非线性 sigma 模型。之所以没有在非线性 sigma 模型的拉氏密度中加上 $-\mathcal{U}(\phi)$ 这样的项，是因为我们还要求了场空间的微分同胚不变性， $\mathcal{U}(\phi)$ 这样的项会破坏这种不变性。

前面考察的标量场 ϕ 都是实数值的，我们当然也可以考察复数值的标量场，不过由于作用量和拉氏密度必须是实数值的，所以这时候需要同时考虑 ϕ 以及它的复共轭场 $\bar{\phi}$ 。很显然，这时候最简单的拉氏密度可以取下面的形式

$$\mathcal{L} = -\partial_\mu\bar{\phi}\partial^\mu\phi - \mathcal{U}(\bar{\phi}\phi). \quad (2.51)$$

很容易看出，除了洛伦兹不变性之外，这个拉氏密度还在下面变换下保持不变，

$$\phi \rightarrow e^{i\theta}\phi, \quad \bar{\phi} \rightarrow e^{-i\theta}\bar{\phi}. \quad (2.52)$$

式中 θ 为一个任意常数。

电磁场

以上只考虑了四维闵可夫斯基时空中的标量场论。四维矢量场甚至高阶张量场当然也能构造相应的拉氏密度，进而得到相应的经典场论。但这时候为了得到真正有用的经典场论，往往需要在洛伦兹不变性之外进一步对系统加上更多的限制。比方说，对于电磁场四矢量 $A^\mu = (\phi, \mathbf{A})$ ，我们还要加上规范对称性，即要求物理可观测量和作用量在如下规范变换下保持不变

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \varepsilon(x). \quad (2.53)$$

式中 $\varepsilon(x)$ 为任意函数。这时候系统的作用量 S_g 可以写成，

$$S_g = - \int d^4x \frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \quad (2.54)$$

式中 $F_{\mu\nu}$ 称作规范场强，它是一个二阶反对称张量场，定义为

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (2.55)$$

很容易验证， $F_{\mu\nu}$ 在规范变换下保持不变。

将上面这个作用量对 A_μ 变分，可得

$$\begin{aligned} \delta S_g &= - \int d^4x \frac{1}{2} F^{\mu\nu} \delta F_{\mu\nu} = - \int d^4x F^{\mu\nu} \partial_\mu \delta A_\nu \\ &= \int d^4x \partial_\mu F^{\mu\nu} \delta A_\nu. \end{aligned} \quad (2.56)$$

为了得出上式最后一行的结果，我们需要分部积分，并丢弃边界项（因为假设无穷远边界上场的变分等于零）。

人们也可以将电磁场与电流四矢量 $J^\mu = (\rho, \mathbf{J})$ 相耦合，即给电磁场作用量加上一项 S_I ，它满足

$$\delta S_I = \int J^\mu \delta A_\mu. \quad (2.57)$$

相应的，根据最小作用量原理 $0 = \delta S = \delta S_g + \delta S_I$ ，即可得规范场 A_μ 满足的运动微分方程，

$$-\partial_\mu F^{\mu\nu} = J^\nu. \quad (2.58)$$

另外，根据 $F_{\mu\nu}$ 的定义 (2.55)，很容易验证它还必然满足如下比安奇恒等式

$$\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0. \quad (2.59)$$

(2.58) 式和 (2.59) 式一起就构成麦克斯韦方程组的四维协变形式，更多的讨论可以参看我的《经典场论新讲》。

2.2.2 狭义相对性原理与经典力学

如果我们要考察的不是一个场论系统，而是一个粒子，那狭义相对性原理告诉我们，相对论粒子的作用量也得是洛伦兹变换不变的！唯一的和粒子坐标有关的这种不变量就是粒子的固有时 $d\tau$ 。所以，粒子的作用量必定正比于 $d\tau$ 沿着粒子运动路径的积分。固有时具有时间量纲，而作用量的量纲为能量量纲乘以时间量纲，刚好粒子质量 m 是能量量纲（由于 $E = mc^2$ ，而 $c = 1$ ），从而我们知道，相对论粒子的作用量必定可以写成

$$S[x(\sigma)] = -m \int d\tau = -m \int d\sigma \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}}. \quad (2.60)$$

式中 σ 为粒子世界线的参数，第一个式子的负号是为了使得作用量有极小值（因为固有时有极大值，直线的固有时最长）。

值得指出的是，上述作用量 (2.60) 显然具有重参数不变性，即在 $\sigma \rightarrow \tilde{\sigma}(\sigma)$ 的参数变换下保持不变。这意味着我们可以在一定意义上任意选择世界线参数 σ 。最常见的选择有两种，第一种是，取 $\sigma = \tau$ ，即取固有时本身为路径的参数。第二种选择是取 $\sigma = x^0 = t$ ，即取通常的时间坐标为参数，这时候注意到

$$d\tau^2 = dt^2 - d\mathbf{x}^2 = dt^2(1 - \mathbf{v}^2). \quad (2.61)$$

式中 $\mathbf{v} = \frac{d\mathbf{x}}{dt}$ 是粒子的速度。从而就可以将作用量 (2.60) 写成

$$S[\mathbf{x}(t)] = -m \int dt \sqrt{1 - \mathbf{v}^2}. \quad (2.62)$$

假设粒子的运动速度远低于光速，即 $\mathbf{v} \ll 1$ ，那这时候就可以利用关于 \mathbf{v}^2 的泰勒展开将相对论的作用量 (2.62) 近似成

$$S = -m \int dt + \int dt \frac{1}{2} m \mathbf{v}^2 + \dots \quad (2.63)$$

省略号表示 \mathbf{v}^2 的高阶项。很显然，除了相差一个对变分没有影响的常数项 $-m \int dt$ 之外，这个近似作用量正是非相对论的自由粒子的作用量！

为了导出相对论粒子在闵可夫斯基时空中的运动方程，我们记

$$L = \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}}, \quad (2.64)$$

因此当取固有时本身为世界线参数时，即 $\sigma = \tau$ 时，我们有 $L = 1$ ，即

$$-\eta_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 1. \quad (2.65)$$

现在，我们可以把相对论粒子的作用量写成 $S[x(\sigma)] = -m \int d\sigma L$ ，为了导出粒子的运动方程，我们需要计算变分 $\delta S[x(\sigma)] = -m \int d\sigma \delta L$ 。利用

$$\begin{aligned} L\delta L &= \frac{1}{2} \delta(L^2) = -\eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \delta\left(\frac{dx^\nu}{d\sigma}\right) \\ &= -\eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{d(\delta x^\nu)}{d\sigma}. \end{aligned} \quad (2.66)$$

可以得到

$$\begin{aligned}\delta S[x(\sigma)] &= -m \int d\sigma \delta L = m \int d\sigma \frac{1}{L} \eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{d(\delta x^\nu)}{d\sigma} \\ &= m \int d\sigma \frac{d}{d\sigma} \left(\frac{1}{L} \eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \delta x^\nu \right) - m \int d\sigma \frac{d}{d\sigma} \left(\frac{1}{L} \eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \right) \delta x^\nu \\ &= -m \int d\sigma \frac{d}{d\sigma} \left(\frac{1}{L} \eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \right) \delta x^\nu.\end{aligned}\quad (2.67)$$

式中最后一个等于是利用了路径两端是固定的 (即在两端 $\delta x^\nu = 0$), 从而全微分项的积分结果为零。进而根据最小作用量原理 $\delta S[x(\sigma)] = 0$, 就可以得到运动微分方程

$$\frac{\delta S}{\delta x^\nu(\sigma)} = -m \frac{d}{d\sigma} \left(\frac{1}{L} \eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \right) = 0. \quad (2.68)$$

我们可以通过取 $\sigma = \tau$ 来简化这个运动方程, 这时候 $L = 1$, 从而即有

$$\frac{d^2 x^\mu}{d\tau^2} = 0. \quad (2.69)$$

(2.69) 式就是自由的相对论性粒子在时空中的运动微分方程, 这个方程的解显然是

$$x^\mu(\tau) = u^\mu \tau + a^\mu, \quad (2.70)$$

式中 u^μ 和 a^μ 均是常矢量, u^μ 当然就是粒子的四维速度, 而且 (2.65) 式告诉我们

$$\eta_{\mu\nu} u^\mu u^\nu = -1. \quad (2.71)$$

也即是说, 相对论性自由粒子在闵可夫斯基时空中是作匀速直线运动! 也可以定义粒子的四维动量 p^μ , 为

$$p^\mu = m u^\mu = m \frac{dx^\mu}{d\tau}. \quad (2.72)$$

如果允许引入一个辅助性的力学变量 $e(\sigma)$, 那我们还可以将作用量 (2.60) 写成一个更加顺眼的形式,

$$S[e(\sigma), x(\sigma)] = \frac{1}{2} \int d\sigma \left(e^{-1} \eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} - e m^2 \right). \quad (2.73)$$

为了证明这个作用量与前面的 (2.60) 相等价, 人们只要先对辅助力学变量 $e(\sigma)$ 使用最小作用量原理即可, 即先利用下式求出 $e(\sigma)$, 再代入上面作用量中消去 $e(\sigma)$

$$\frac{\delta S}{\delta e(\sigma)} = 0. \quad (2.74)$$

特别的, (2.73) 允许我们取 $m = 0$ 的极限, 从而就得到零质量粒子 (比如光子) 的经典作用量, 为

$$S[e(\sigma), x(\sigma)] = \frac{1}{2} \int d\sigma \left(e^{-1} \eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} \right). \quad (2.75)$$

这是引入辅助变量 $e(\sigma)$ 之后的一个意外好处, 即允许我们用作用量原理统一描述有质量粒子和零质量的粒子。

上面所讨论的只不过是一个自由的相对论粒子。而实际的相对论粒子会和诸如引力场或者电磁场这样的基本力场发生相互作用，如何写出一个包含了相互作用的作用量呢？我们以电荷为 q 的带电粒子与电磁场的相互作用为例来说明这一点。要点依然是通过考虑对称性，不过，这时候我们要考虑的是电磁场的规范对称性。

仔细思考以后我们可能会发现下面这一项

$$+q \int_a^b A_\mu dx^\mu. \quad (2.76)$$

它就是电磁势沿着带电粒子世界线的积分，式中我们假设这条世界线起于 a 点终止于 b 点。这一项显然是洛伦兹不变的，那它规范不变吗？很明显不是，因为在规范变换 (2.53) 的作用下，它会变为

$$q \int_a^b A_\mu dx^\mu \rightarrow q \int_a^b A_\mu dx^\mu + q \int_a^b d\varepsilon \quad (2.77)$$

但是，我们发现多出来的部分是一个全微分，它完全取决于路径的两个端点，即

$$q \int_a^b d\varepsilon = q(\varepsilon(b) - \varepsilon(a)). \quad (2.78)$$

而我们早就知道，在使用最小作用量原理时，路径的两个端点是固定不变的，因此规范变换下多出来的这一项对变分完全没有贡献。即是说，虽然给作用量加上的这一项 (2.76) 不是规范不变的，但是，由它导出来的运动微分方程却是规范不变的！所以，(2.76) 这一项实际上符合要求！

因此，我们可以写出相对论性带电粒子完整的作用量，为

$$S[x(\sigma)] = -m \int d\sigma \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}} + q \int d\sigma A_\mu \frac{dx^\mu}{d\sigma}. \quad (2.79)$$

如果将参数 σ 取成坐标时 t ，并考虑 $\mathbf{v} \ll 1$ 的非相对论极限，那上式就可以近似成

$$S[\mathbf{x}(t)] = -m \int dt + \int dt \left[\frac{1}{2} m \mathbf{v}^2 - q\phi + q\mathbf{A} \cdot \mathbf{v} \right] + \dots \quad (2.80)$$

将作用量 (2.79) 对规范势 A_μ 变分 (为了避免混淆，下面将带电粒子在时空中的位置坐标改记为 $x_e^\mu(\sigma)$)，即可得到

$$\begin{aligned} \delta S &= q \int d\sigma \frac{dx_e^\mu}{d\sigma} \delta A_\mu \\ &= \int d^4x \left[q \int d\sigma \frac{dx_e^\mu}{d\sigma} \delta^4(x - x_e(\sigma)) \right] \delta A_\mu. \end{aligned} \quad (2.81)$$

将这个式子与标准的电流四矢量 J^μ 与规范势的耦合 $\delta S = \int d^4x J^\mu \delta A_\mu$ 进行比较，即可得到此带电粒子系统的电流四矢量，为

$$J^\mu(x) = q \int d\sigma \frac{dx_e^\mu}{d\sigma} \delta^4(x - x_e(\sigma)). \quad (2.82)$$

为了看清楚 (2.82) 式的物理意义，我们取世界线参数 $\sigma = x_e^0$ ，并做出 (2.82) 式中的积分，则有

$$\begin{aligned} \rho &= q \delta^3(\mathbf{x} - \mathbf{x}_e(t)) \\ \mathbf{J} &= q \mathbf{v}_e \delta^3(\mathbf{x} - \mathbf{x}_e(t)). \end{aligned} \quad (2.83)$$

很显然，结果正符合我们对电荷密度以及电流密度表达式的预期。

2.3 时空对称性与能量动量张量

这一节我们将引入物质系统的一个特征性物理量，即所谓的能量-动量张量。根据著名的诺特定理，任何一个连续的对称性都必然有一个相应的守恒流四矢量，能量-动量张量就是与时空平移对称性相对应的守恒流，由于时空平移本身是一个四维矢量，所以它对应的守恒流就是一个二阶张量。

2.3.1 场论系统的能量-动量张量

标量场

为了简单起见，我们首先讲述标量场系统的能量-动量张量。

假设有一个标量场 $\phi(x)$ ，其拉格朗日密度可以写成 $\mathcal{L}(\phi, \partial_\mu \phi)$ ，特别的，这个拉格朗日密度不显含时空坐标 x 。现在，假定将整个场论系统进行一个时空平移，使得 x 点的场平移到 x' 点，

$$x^\mu \rightarrow x'^\mu = x^\mu + a^\mu, \quad (2.84)$$

其中 a^μ 为某常数四矢量。记平移之后的场为 $\phi'(x)$ ，很显然 ϕ' 在 x' 点的场值来自于平移之前 ϕ 在 x 点的场值，即

$$\phi'(x') = \phi(x). \quad (2.85)$$

很容易验证作用量 $S[\phi] = \int d^4x \mathcal{L}(\phi, \partial_\mu \phi)$ 在此平移之下保持不变，具体验证过程如下

$$\begin{aligned} S[\phi'] &= \int d^4x \mathcal{L}(\phi'(x), \partial_\mu \phi'(x)) \\ &= \int d^4x' \mathcal{L}(\phi'(x'), \partial'_\mu \phi'(x')) = \int d^4x' \mathcal{L}(\phi(x), \partial'_\mu \phi(x)) \\ &= \int d^4x \mathcal{L}(\phi(x), \partial_\mu \phi(x)) = S[\phi] \end{aligned} \quad (2.86)$$

式中第二个等号只是将 x 变量重写成了 x' ，倒数第二个等号是用了 $d^4x' = d^4x$ 以及 $\partial'_\mu = \partial_\mu$ 。

为了通过诺特定理引入相应的守恒流，下面考察无穷小时空平移，即将 a^μ 取成无穷小量 ε^μ 。进一步，我们使用一个**关键技巧**，即将 ε^μ 变成依赖于时空坐标 x 的无穷小量 $\varepsilon^\mu(x)$ ，即考察如下无穷小时空平移

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x). \quad (2.87)$$

当然我们依然有

$$\phi'(x') = \phi(x). \quad (2.88)$$

但是，在这种依赖于时空点 x 的局域平移之下，作用量当然无法保持不变，因为这种局域平移根本不是系统的对称性。

很显然，在一阶近似上有

$$\frac{\partial x'^\mu}{\partial x^\nu} = \delta_\nu^\mu + \partial_\nu \varepsilon^\mu, \quad \frac{\partial x^\nu}{\partial x'^\mu} = \delta_\mu^\nu - \partial_\mu \varepsilon^\nu. \quad (2.89)$$

另外, 记坐标变换的雅可比行列式为 $|\frac{\partial x'}{\partial x}|$, 利用矩阵恒等式 $\det(A) = e^{\text{Tr}(\ln A)}$, 易得在一阶近似上有

$$|\frac{\partial x'}{\partial x}| = 1 + \partial_\mu \epsilon^\mu(x). \quad (2.90)$$

下面我们来计算在局域平移 (145) 之前和之后, 系统作用量的改变量。平移以后的作用量 $S[\phi'(x)]$ 为

$$\begin{aligned} S[\phi'(x)] &= \int d^4x \mathcal{L}(\phi'(x), \partial_\nu \phi'(x)) \\ &= \int d^4x' \mathcal{L}(\phi'(x'), \partial'_\mu \phi'(x')) = \int d^4x' \mathcal{L}(\phi(x), \partial'_\mu \phi(x)) \\ &= \int d^4x |\frac{\partial x'}{\partial x}| \mathcal{L}(\phi(x), \frac{\partial x^\nu}{\partial x'^\mu} \partial_\nu \phi(x)) \\ &= \int d^4x (1 + \partial_\mu \epsilon^\mu(x)) \mathcal{L}(\phi(x), \partial_\mu \phi(x) - \partial_\mu \epsilon^\nu \partial_\nu \phi(x)) \\ &= \int d^4x \left[-\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi(x) + \delta_\nu^\mu \mathcal{L}(\phi(x), \partial_\rho \phi(x)) \right] \partial_\mu \epsilon^\nu \\ &\quad + \int d^4x \mathcal{L}(\phi(x), \partial_\nu \phi(x)). \end{aligned} \quad (2.91)$$

由上面的推导易知, 变换前后作用量的改变量为

$$\begin{aligned} \delta S &= S[\phi'(x)] - S[\phi(x)] \\ &= - \int d^4x \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi(x) - \delta_\nu^\mu \mathcal{L} \right] \partial_\mu \epsilon^\nu \\ &= \int d^4x T^\mu{}_\nu \partial_\mu \epsilon^\nu = \int d^4x T_{\mu\nu} \partial^\mu \epsilon^\nu. \end{aligned} \quad (2.92)$$

式中

$$T^{\mu\nu} = - \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial^\nu \phi(x) - \eta^{\mu\nu} \mathcal{L} \right]. \quad (2.93)$$

注意 (150) 式, 当 ϵ^ν 是一个不依赖于时空点的整体无穷小平移时, 即有 $\delta S = 0$, 这当然是因为这种整体时空平移是系统的对称性。正因为如此, 不仅对于标量场系统可以找到这样一个 $T^{\mu\nu}$, 实际上对闵可夫斯基时空的任何物质系统都能找到一个类似的 $T^{\mu\nu}$, 因为任何物质系统的作用量在局域平移 (145) 前后的改变量都必定能写成如下形式

$$\delta S = \int d^4x T_{\mu\nu} \partial^\mu \epsilon^\nu. \quad (2.94)$$

而这又是因为, 当将 ϵ^ν 取成常矢量 (即考察无穷小整体时空平移时) 时, 由于时空平移对称性, 必有 $\delta S = 0$, 所以在局域化平移情形, δS 只能依赖于 ϵ^ν 的偏导 $\partial^\mu \epsilon^\nu$, 从而 δS 必定能写成 (152) 式的形式!

下面进一步假定 $\epsilon^\nu(x)$ 在时空的无穷远边界上都趋于零。另外, 上面的 (152) 式对于任意场位形都成立, 下面我们考察真实的满足场运动微分方程的场位形, 那这时候由于这些场位形满足最小作用量原理, 当然就有 $\delta S = 0$ 对于任意的在时空无穷远处趋于零的场

变分都成立，那当然也对局域时空平移 (145) 引起的场改变成立。即是说，对于真实场位形，必有

$$0 = \delta S = \int d^4x T_{\mu\nu} \partial^\mu \varepsilon^\nu. \quad (2.95)$$

将上式分部积分，即有

$$0 = \int d^4x T_{\mu\nu} \partial^\mu \varepsilon^\nu = - \int d^4x \partial^\mu T_{\mu\nu} \varepsilon^\nu(x). \quad (2.96)$$

由于 $\varepsilon^\nu(x)$ 为任意无穷小函数，从而即有

$$\partial^\mu T_{\mu\nu} = 0. \quad (2.97)$$

即是说， $T_{\mu\nu}$ 就是与时空平移对称性相应的守恒流！特别的，如果取 $\varepsilon^\mu = \varepsilon \delta^{\mu 0}$ ，即假设考察的是时间平移，那相应的守恒流 $T^{\mu 0}$ 当然就是能量流四矢量，换言之， T^{00} 必定为能量密度， $T^{i0} (i=1,2,3)$ 为能量流密度。而如果取 $\varepsilon^\mu = \varepsilon \delta^{\mu j}$ 为空间平移，那相应的守恒流 $T^{\mu j}$ 当然就是动量流四矢量，即是说 T^{0j} 必为动量密度，而 T^{ij} 为动量流密度。将所有这些合起来，我们就称 $T_{\mu\nu}$ 为能量动量张量！上面的讨论也说明，**闵可夫斯基时空的任何物质系统都存在这么一个守恒的能量动量张量。**

比如，假设我们考虑 $\mathcal{L} = -\frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \mathcal{U}(\phi)$ 的场论模型，则容易算得，

$$\begin{aligned} T^{\mu\nu} &= \partial^\mu \phi \partial^\nu \phi + \eta^{\mu\nu} \mathcal{L} \\ &= \partial^\mu \phi \partial^\nu \phi - \frac{1}{2} \eta^{\mu\nu} \partial_\rho \phi \partial^\rho \phi - \eta^{\mu\nu} \mathcal{U}(\phi). \end{aligned} \quad (2.98)$$

很明显，这个 $T^{\mu\nu}$ 的两个指标是对称的。实际上，可以证明，对于任何洛伦兹不变的标量场论，其 $T^{\mu\nu}$ 都必定是对称张量。但是当我们的考察范围超出标量场论时，其按照上面类似的办法求出来的 $T^{\mu\nu}$ 就不一定为对称张量了，比方说对于矢量场，它的 $T^{\mu\nu}$ 就不是对称张量。不过， $T^{\mu\nu}$ 的流守恒方程 $\partial_\mu T^{\mu\nu} = 0$ 告诉我们， $T^{\mu\nu}$ 的定义不是唯一的，实际上人们很容易看出，对于任何 $X^{\rho\mu\nu}$ ，只要 $X^{\rho\mu\nu} = -X^{\mu\rho\nu}$ ，则 $T^{\mu\nu} + \partial_\rho X^{\rho\mu\nu}$ 同样满足流守恒方程，因此可以定义为新的能动量张量。从而只要我们合适地选取 $X^{\rho\mu\nu}$ ，我们总可以重新定义以后的能动量张量为一个对称张量。实际上，马上我们就会给出一个找到这样一个对称能量动量张量的巧妙办法。

洛伦兹对称性

我们考察的经典场论都是具有洛伦兹不变性的场论，即拉格朗日密度为洛伦兹标量的理论，具体来说即是拉氏密度在如下坐标变换下保持不变的理论，

$$x^\mu \rightarrow x'^\mu = \Lambda^\mu_{\nu} x^\nu, \quad (2.99)$$

Λ^μ_{ν} 就是所谓的洛伦兹变换。那么，洛伦兹变换对应的守恒流是什么呢？

为此我们需要考察无穷小洛伦兹变换，即取

$$\Lambda^\mu_{\nu} = \delta^\mu_{\nu} + \varepsilon^\mu_{\nu}, \quad (2.100)$$

式中 ε_ν^μ 为无穷小量。很显然，无穷小洛伦兹变换由于可以和恒等变换连续过渡，从而必定是正洛伦兹变换，即满足¹

$$\det(\Lambda) = 1 \Rightarrow \varepsilon_\mu^\mu = 0. \quad (2.101)$$

进一步，利用洛伦兹变换的定义 $\eta_{\mu\nu}\Lambda_\alpha^\mu\Lambda_\beta^\nu = \eta_{\alpha\beta}$ ，易得

$$\eta_{\mu\nu}(\delta_\alpha^\mu + \varepsilon_\alpha^\mu)(\delta_\beta^\nu + \varepsilon_\beta^\nu) = \eta_{\alpha\beta} \Rightarrow \varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} = 0. \quad (2.102)$$

即 $\varepsilon^{\mu\nu}$ 是一个二阶反对称张量。

注意到无穷小洛伦兹变换为 $x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon_\nu^\mu x^\nu$ 。为了利用诺特定理，我们使用关键技巧，即将无穷小参数 ε_ν^μ 变成依赖于时空坐标的 $\varepsilon_\nu^\mu(x)$ ，进而考察如下无穷小局域时空坐标变换，

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x), \quad \text{这里 } \varepsilon^\mu(x) = \varepsilon_\nu^\mu(x)x^\nu. \quad (2.103)$$

则完全类似于前面对 (152) 式的论证，可知在此时空变换之下作用量的改变量必定可以写成

$$\begin{aligned} \delta S &= \int d^4x T^\mu{}_\nu \partial_\mu \varepsilon^\nu(x) \\ &= \int d^4x T^\mu{}_\nu (\varepsilon_\mu^\nu + x^\rho \partial_\mu \varepsilon_\rho^\nu). \end{aligned} \quad (2.104)$$

(注意，对于比方说矢量场，它的各指标分量也要按照无穷小局域洛伦兹变换 $\Lambda_\nu^\mu(x) = \delta_\nu^\mu + \varepsilon_\nu^\mu(x)$ 的规则变，这是不同于前面时空平移情形的。在时空平移情形，矢量场的分量指标并不变换。因此上式中的 $T_{\mu\nu}$ 一般不同于前面按照时空平移直接求出来的 $T_{\mu\nu}$ 。) 注意到 $\varepsilon_{\mu\nu}$ 关于指标反对称，从而有

$$\begin{aligned} \delta S &= \frac{1}{2} \int d^4x \left[(T^{\mu\nu} - T^{\nu\mu}) \varepsilon_{\nu\mu} + (x^\rho T^{\mu\nu} - x^\nu T^{\mu\rho}) \partial_\mu \varepsilon_{\nu\rho} \right] \\ &= -\frac{1}{2} \int d^4x \left[(T^{\mu\nu} - T^{\nu\mu}) \varepsilon_{\mu\nu} \right] - \frac{1}{2} \int d^4x \left[(x^\mu T^{\rho\nu} - x^\nu T^{\rho\mu}) \partial_\rho \varepsilon_{\mu\nu} \right]. \end{aligned} \quad (2.105)$$

如果我们将 $\varepsilon_{\mu\nu}$ 取回常数，则 $\partial_\rho \varepsilon_{\mu\nu} = 0$ ，从而上式最后一行只剩下前面那项，但是洛伦兹不变性告诉我们，当 $\varepsilon_{\mu\nu}$ 为常数时作用量应该不变，即这时候必有 $\delta S = 0$ ，由此可知

$$T^{\mu\nu} - T^{\nu\mu} = 0, \quad (2.106)$$

即这样求出来的能动量张量必定是对称张量！这就是我们给出的如何寻找对称能量动量张量的办法。

在 (163) 式中代入 $T^{\mu\nu} = T^{\nu\mu}$ ，即有

$$\delta S = -\frac{1}{2} \int d^4x \left[(x^\mu T^{\rho\nu} - x^\nu T^{\rho\mu}) \partial_\rho \varepsilon_{\mu\nu} \right]. \quad (2.107)$$

¹利用矩阵恒等式 $\det(A) = e^{\text{Tr}(\ln A)}$.

完全类似于前面导出能动量张量守恒的讨论, 这个结果意味着 $M^{\rho\mu\nu} = x^\mu T^{\rho\nu} - x^\nu T^{\rho\mu}$ 为守恒流, 满足守恒方程

$$\partial_\rho M^{\rho\mu\nu} = 0. \quad (2.108)$$

综合关于无穷小局域时空平移和无穷小局域洛伦兹变换的结果, 可知, 在如下无穷小局域时空变换之下

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x), \quad (2.109)$$

任何物质系统作用量的改变量必定可以写成

$$\delta S = \int d^4x T^{\mu\nu} \partial_\mu \varepsilon_\nu(x) = \frac{1}{2} \int d^4x T^{\mu\nu} (\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu). \quad (2.110)$$

其中我们利用了 $T^{\mu\nu}$ 是一个对称张量。这个式子是一个很关键的式子, 在后面的章节中有大用。

电磁场

下面我们来考察一个自由电磁场的能量-动量张量。所谓的自由电磁场就是不与其它任何东西耦合, 单独只由拉氏密度 $\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu}$ 描述的无源电磁场。当然, 我们可以按照前面讨论无穷小局域洛伦兹变换中给出的办法直接找出对称的能量-动量张量。但更快的方法实际上是下面的办法。

我们已经知道标量场能动量张量的公式, 为了得到电磁场的能动量张量, 我们将 $A_\mu (\mu = 0, 1, 2, 3)$ 的每一个分量看成一个实标量场, 从而容易给出电磁场能动量张量的公式

$$T^{\mu\nu} = - \left[\frac{\partial \mathcal{L}}{\partial (\partial_\mu A^\rho)} \partial^\nu A^\rho - \eta^{\mu\nu} \mathcal{L} \right]. \quad (2.111)$$

代入自由电磁场的拉氏密度可以得到,

$$T^{\mu\nu} = F^\mu{}_\rho \partial^\nu A^\rho - \frac{1}{4} \eta^{\mu\nu} F_{\rho\sigma} F^{\rho\sigma}. \quad (2.112)$$

很显然, 这个能动量张量不是一个对称张量, 而更为严重的问题是, 它不是规范不变的! 而作为物理可观测量, 能动量张量必须规范不变! 好在, 正如前面所说的, 我们可以通过给它加上一个合适的 $\partial_\rho X^{\rho\mu\nu}$ 来重新定义能动量张量, 使它成为一个对称张量。具体来说, 我们可以给上述能动量张量加上 $\partial^\rho (-F^\mu{}_\rho A^\nu) = -F^\mu{}_\rho \partial^\rho A^\nu$ (注意到对于自由电磁场, 电流四矢量等于零, 从而方程 (2.58) 变成 $\partial^\rho F^\mu{}_\rho = 0$), 很显然, 加上这一项修正以后, 电磁场的能动量张量就变成了如下对称张量

$$T^{\mu\nu} = F^\mu{}_\rho F^{\nu\rho} - \frac{1}{4} \eta^{\mu\nu} F_{\rho\sigma} F^{\rho\sigma}. \quad (2.113)$$

这才是一个真正规范不变的量。这个能动量张量有一个重要的特征, 即

$$T^\mu{}_\mu = 0. \quad (2.114)$$

2.3.2 多粒子系统的能量动量张量

考虑多个相对论性自由粒子所构成的系统，我们以 $n = 1, 2, \dots, N$ 来标记不同的粒子。根据本章前面的知识可以知道，这个系统的作用量可以写成

$$S[x(\sigma)] = -\sum_n m_n \int d\tau_n = -\sum_n m_n \int d\sigma_n \sqrt{-\eta_{\mu\nu} \frac{dx_n^\mu}{d\sigma_n} \frac{dx_n^\nu}{d\sigma_n}}. \quad (2.115)$$

式中 x_n^μ 为粒子 n 的时空坐标， m_n 为它的质量， τ_n 为它的固有时， σ_n 为它的世界线参数。由于上式在 $\sigma_n \rightarrow \tilde{\sigma}_n(\sigma_n)$ 的重参数化之下保持不变，所以 σ_n 的选择有很大的任意性，特别的，我们可以将 σ_n 选作固有时 τ_n 。

不妨以固有时参数（即 $\sigma_n = \tau_n$ ）下的粒子作用量来进一步讨论。作用量 (2.115) 显然具有 $x_n^\mu \rightarrow x_n'^\mu = x_n^\mu + a^\mu$ (a^μ 为常矢量) 的时空坐标平移不变性。为了考察这一时空平移对称性所对应的能动量张量，我们考虑如下局域化的无穷小时空坐标变换

$$x_n^\mu \rightarrow x_n'^\mu = x_n^\mu + \varepsilon^\mu(x_n) \Leftrightarrow \delta x_n^\mu = \varepsilon^\mu(x_n). \quad (2.116)$$

(从而 $\delta(dx_n^\mu) = d(\delta x_n^\mu) = \frac{\partial \varepsilon^\mu}{\partial x_n^\nu} dx_n^\nu$) 注意到

$$\begin{aligned} \delta S &= -\sum_n m_n \int \delta(d\tau_n) = -\sum_n m_n \int \frac{1}{2} \frac{\delta(d\tau_n^2)}{d\tau_n} \\ &= \sum_n m_n \int \frac{1}{2} \delta(\eta_{\mu\nu} dx_n^\mu dx_n^\nu) / d\tau_n \\ &= \sum_n m_n \int \frac{dx_n^\mu}{d\tau_n} \eta_{\mu\nu} \delta(dx_n^\nu) \\ &= \sum_n m_n \int \frac{dx_n^\mu}{d\tau_n} \eta_{\mu\nu} d(\delta x_n^\nu). \end{aligned} \quad (2.117)$$

可得在上面的局域时空坐标变换下，作用量的改变量为

$$\begin{aligned} \delta S &= \sum_n m_n \int \frac{dx_n^\mu}{d\tau_n} \eta_{\mu\nu} d(\delta x_n^\nu) \\ &= \sum_n m_n \int \left(\frac{dx_n^\mu}{d\tau_n} \right) \frac{\partial \varepsilon_\mu}{\partial x_n^\nu} dx_n^\nu \\ &= \sum_n m_n \int d\tau_n \left(\frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n} \right) \frac{\partial \varepsilon_\mu}{\partial x_n^\nu}. \end{aligned} \quad (2.118)$$

通过引入四维时空 δ 函数 $\delta^4(x - x_n(\tau_n)) = \prod_{\mu=0}^3 \delta(x^\mu - x_n^\mu(\tau_n))$ ，可以进一步将上面结果写成

$$\begin{aligned} \delta S &= \sum_n m_n \int d\tau_n \left(\frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n} \right) \frac{\partial \varepsilon_\mu}{\partial x_n^\nu} \\ &= \int d^4x \left[\sum_n m_n \int d\tau_n \left(\frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n} \right) \delta^4(x - x_n(\tau_n)) \right] \partial_\nu \varepsilon_\mu(x). \end{aligned} \quad (2.119)$$

完全类似于前面关于标量场能量-动量张量的讨论，可知，与时空平移对称性对应的能动量张量为

$$T^{\mu\nu}(x) = \sum_n m_n \int d\tau_n \left(\frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n} \right) \delta^4(x - x_n(\tau_n)). \quad (2.120)$$

显然，这个能动量张量是一个对称张量。

为了看清楚上述能动量张量的意义，我们将它重写成，

$$T^{\mu\nu} = \sum_n m_n \int d\sigma_n \left(\frac{dx_n^\mu}{d\sigma_n} \frac{dx_n^\nu}{d\tau_n} \right) \delta^4(x - x_n(\sigma_n)). \quad (2.121)$$

然后取世界线参数 $\sigma_n = x_n^0$ ，则当我们做完对 σ_n 的积分后，即有能动量密度

$$T^{0\mu} = \sum_n m_n \left(\frac{dx_n^\mu}{d\tau_n} \right) \delta^3(\mathbf{x} - \mathbf{x}_n(t)) = \sum_n p_n^\mu \delta^3(\mathbf{x} - \mathbf{x}_n(t)). \quad (2.122)$$

类似的，在取 $\sigma_n = x_n^0$ 并做完对 σ_n 的积分以后，也有流密度

$$\begin{aligned} T^{i\mu} &= \sum_n m_n \left(\frac{dx_n^\mu}{d\tau_n} \frac{dx_n^i}{dt} \right) \delta^3(\mathbf{x} - \mathbf{x}_n(t)) \\ &= \sum_n p_n^\mu v_n^i \delta^3(\mathbf{x} - \mathbf{x}_n(t)). \end{aligned} \quad (2.123)$$

综合 (2.122) 式和 (2.123) 式，可知粒子系统的能动量张量可以表达成

$$T^{\mu\nu} = \left[\sum_n \frac{p_n^\mu p_n^\nu}{p_n^0} \delta^3(\mathbf{x} - \mathbf{x}_n(t)) \right]. \quad (2.124)$$

3. 从等效原理到时空弯曲

3.1 等效原理

3.1.1 自由下落的小房间

1907年十一月的一天，坐在瑞士伯尔尼专利局里的爱因斯坦突然想到一个他称之为这辈子最幸福的想法。爱因斯坦想到，一个从屋顶上自由下落的人将感觉不到引力存在(失重)，或言之，对自由下落者来说，引力不存在！这个看似平凡其实深刻的想法最终把爱因斯坦导向了正确的引力理论，即广义相对论。

为了看清爱因斯坦的想法有何深刻之处，设想某一天你睡醒来，发现自己在一个密闭的小房间里，你看不到外面，房间布置得就和一个太空舱一样，但其实房间是在地面上方从高空自由下落，不过这一点你并不知道，请问你能否发现这个事实，你能否判断出自己不是被送上了外太空自由漂浮？

爱因斯坦告诉你，你不能！因为这两种情况在物理学上是等价的。外太空远离所有物质的自由漂浮当然是一个惯性系，因此爱因斯坦断言，自由下落的一个小房间也是惯性系！也即是说，自由下落的你在小房间里观察到的一切物理规律都和惯性系中的规律一样，都是狭义相对论中的那些物理规律，而对你来说，引力场是不存在的！

如果用牛顿的语言来分析爱因斯坦的想法，那就是，自由下落的小房间是一个加速参考系，其中的你会感受到一个向上的惯性力，这个惯性力正好将你的重力抵消了，因此你就失重了。但是爱因斯坦说，惯性力之所以总是能抵消重力，是因为它们本质是一回事，都是引力(或者也可以说都是惯性力)！只不过，对于地面上静止不动的人来说，引力存在，但是对于自由下落的你来说，引力不存在，所以随着你自由下落的小房间实质是一个惯性

系，就和漂浮在外太空的太空舱一样，在这个惯性系中，不含引力的狭义相对论规律总是成立。

反过来，如果你真在外太空的太空舱里，那只要将太空舱加速，那你也无法区分自己感受到的是惯性力还是引力，因为它们两者是一回事！换言之，按照爱因斯坦的观点，引力就是相对于牛顿力学的惯性系加速的观察者所感受到的力。

当然，正如爱因斯坦自己所强调的，以上所述只在一个足够小的小房间里成立。自由下落的小房间只在一个足够小的局部上抵消掉了地球引力，但在美国自由下落的小房间和在中国自由下落的小房间，是两个不同的小房间！不存在一个参考系能整体地抵消掉全部地球引力。

3.1.2 潮汐力

自由下落的房间在多大的局部上抵消掉了引力呢？这个问题的答案很容易分析，因为地球就受月球的引力作用，地球就是一个向着月球自由下落的参考系，只不过地球不够小，所以月球的影响不能完全抵消掉，这个抵消不掉的影响就表现为潮汐力！同样，在地球表面自由下落的参考系里也有地球施加的潮汐力，潮汐力就是不能被自由下落的局域参考系完全抵消的引力。图 (3.1) 中 (b) 的小箭头表示的就是潮汐力。

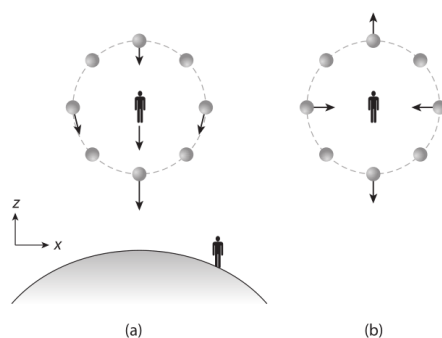


Figure 3.1: 一圈正在自由下落的小球：(a) 从地面上观察者的角度看，(b) 从跟随小球一起自由下落的观察者角度看。图片来自 A Zee, Einstein Gravity in a Nutshell

为了进一步分析潮汐力，假设有一个在引力势 $\Phi(\mathbf{x})$ 中自由下落的小房间，假设小房间下落的速度远小于光速，因此可以采用**非相对论的分析**。取房间质心的坐标为 $\mathbf{x}_0(t)$ ，由于在自由下落，所以它满足方程

$$\ddot{\mathbf{x}}_0 = -\nabla\Phi(\mathbf{x}_0). \quad (3.1)$$

下面考察房间中一个质量为 m 的小球，它相对于房间的运动速度也远小于光速，假定它相对于自由下落的房间质心的坐标为 \mathbf{y} ，而在地面上的人看来，小球的坐标为 $\mathbf{x}(t)$ ，因此

$$\mathbf{x}(t) = \mathbf{x}_0(t) + \mathbf{y}(t). \quad (3.2)$$

由于房间很小，所以小球在这个自由下落参考系中的坐标 $\mathbf{y}(t)$ 是一个小量！

我们可以在地面参考系中写出小球的作用量

$$S[\mathbf{x}(t)] = m \int dt \left[\frac{1}{2} \dot{\mathbf{x}}^2 - \Phi(\mathbf{x}) \right]. \quad (3.3)$$

代入 (3.2) 式, 并将 $\Phi(\mathbf{x}_0 + \mathbf{y})$ 按照小量 \mathbf{y} 展开, 即可得到

$$\begin{aligned} S[\mathbf{x}(t)] &= m \int dt \left[\frac{1}{2} (\dot{\mathbf{x}}_0 + \dot{\mathbf{y}})^2 - \Phi(\mathbf{x}) \right] \\ &= m \int dt \left[\frac{1}{2} \dot{\mathbf{x}}_0^2 + \frac{1}{2} \dot{\mathbf{y}}^2 + \dot{x}_0^i y^i - \Phi(\mathbf{x}_0) - \partial_i \Phi(\mathbf{x}_0) y^i \right. \\ &\quad \left. - \frac{1}{2} \partial_i \partial_j \Phi(\mathbf{x}_0) y^i y^j + \dots \right] \\ &= m \int dt \left[\frac{1}{2} \dot{\mathbf{x}}_0^2 - \Phi(\mathbf{x}_0) + \frac{1}{2} \dot{\mathbf{y}}^2 - \dot{x}_0^i y^i - \partial_i \Phi(\mathbf{x}_0) y^i \right. \\ &\quad \left. - \frac{1}{2} \partial_i \partial_j \Phi(\mathbf{x}_0) y^i y^j + \dots \right]. \end{aligned} \quad (3.4)$$

式中最后一个等号是对 $\dot{x}_0^i y^i$ 进行分部积分, 代入 $\dot{x}_0^i = -\partial_i \Phi(\mathbf{x}_0)$, 容易看到 y 的线性项正好消去了! 另外, \mathbf{x}_0 的运动是给定的, 所以与之相关的项可以丢掉, 进而即有小球在自由下落参考系中的作用量, 为

$$S[\mathbf{y}(t)] = m \int dt \left[\frac{1}{2} \dot{\mathbf{y}}^2 - \frac{1}{2} \partial_i \partial_j \Phi(\mathbf{x}_0) y^i y^j + \dots \right]. \quad (3.5)$$

式中 $\frac{1}{2} \partial_i \partial_j \Phi(\mathbf{x}_0) y^i y^j$ 代表的正是潮汐力的势能, 因为这一项正是没有完全抵消的引力的贡献。

以上结果告诉我们, 潮汐力的贡献是一个二阶小量。换言之, 对于足够小的自由下落参考系而言, 引力的效应直到一阶小量为止都完全抵消了, 只在二阶小量上对自由下落参考系里的小球作用量有修正!

不妨简单举例计算一下刻画潮汐力的张量 $\partial_i \partial_j \Phi(\mathbf{x}_0)$ 。比方说对于 $\Phi(\mathbf{x}) = -G \frac{M}{|\mathbf{x}|}$ (比如地球的引力势), 容易算得

$$\partial_i \partial_j \Phi(\mathbf{x}) = GM \frac{\delta_{ij} - 3e_i e_j}{|\mathbf{x}|^3}, \quad (3.6)$$

式中 $e_i = x^i / |\mathbf{x}|$ 。比如说在沿着 z 轴的 $\mathbf{x} = (0, 0, r)$ 点, 我们有

$$\partial_i \partial_j \Phi(\mathbf{x}) = \frac{GM}{r^3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \quad (3.7)$$

即是说, 在 $\mathbf{x} = (0, 0, r)$ 点附近, x, y 方向的潮汐力是指向这一点的, 而 z 方向的潮汐力是背离这一点的, 正如图 (3.1) 中 (b) 的小箭头所示。

3.1.3 局部惯性系与等效原理

上两小节中自由下落的足够小房间就是所谓的局部惯性系, 因为在这个参考系中, 狭义相对论的规律直到一阶小量为止都成立, 引力的效应只在二阶小量上体现为潮汐力的势

能。在上一小节中，为了分析的方便，我们假定了粒子运动的速度远小于光速，从而可以用非相对论近似。现在，我们要进行完整的相对论分析。

首先，假定在引力场中自由下落的粒子在地面坐标系（非局部惯性系）中的坐标为 $x^\mu(\sigma)$ 。假定跟着粒子一起下落的局部惯性系的坐标原点为 $x_0^\mu(\sigma)$ 。记粒子在此局部惯性系中的坐标为 $y_{x_0}^\mu(x(\sigma))$ ，由于 x_0 为坐标原点，所以当然有 $y_{x_0}^\mu(x_0) = 0$ 。根据上面的分析，只要局部惯性系足够小（从而可以假定 $y_{x_0}^\mu$ 很小），粒子的作用量就是标准的狭义相对论自由粒子作用量，即为

$$S[y_{x_0}(\sigma)] = -m \int d\tau = -m \int d\sigma \sqrt{-\eta_{\mu\nu} \frac{dy_{x_0}^\mu}{d\sigma} \Big|_{x_0} \frac{dy_{x_0}^\nu}{d\sigma} \Big|_{x_0}}. \quad (3.8)$$

引力的效应只在 $y_{x_0}^\mu$ 的二阶小量上对上述作用量有修正，因此在坐标原点 x_0 ，引力场的效应可以完全忽略。

下面，假设我们变换到地面上的非局部惯性系 $x^\mu(\sigma)$ ，则粒子的作用量将为

$$\begin{aligned} S[x_0(\sigma)] &= -m \int d\sigma \sqrt{-\eta_{\mu\nu} \frac{dy_{x_0}^\mu}{d\sigma} \Big|_{x_0} \frac{dy_{x_0}^\nu}{d\sigma} \Big|_{x_0}} \\ &= -m \int d\sigma \sqrt{-\eta_{\mu\nu} \frac{\partial y_{x_0}^\mu}{\partial x^\alpha} \Big|_{x_0} \frac{\partial y_{x_0}^\nu}{\partial x^\beta} \Big|_{x_0} \cdot \frac{dx_0^\alpha}{d\sigma} \frac{dx_0^\beta}{d\sigma}} \\ &= -m \int d\sigma \sqrt{-g_{\alpha\beta}(x_0) \frac{dx_0^\alpha}{d\sigma} \frac{dx_0^\beta}{d\sigma}}. \end{aligned} \quad (3.9)$$

式中 $g_{\alpha\beta}(x_0)$ 为（注意 $y_{x_0}^\mu$ 不只是 x 的函数，且它的定义本身依赖于 x_0 ）

$$g_{\alpha\beta}(x_0) = \eta_{\mu\nu} \frac{\partial y_{x_0}^\mu}{\partial x^\alpha} \Big|_{x_0} \frac{\partial y_{x_0}^\nu}{\partial x^\beta} \Big|_{x_0}. \quad (3.10)$$

但是我们说过，在非局部惯性系中，是有引力场的，这个引力场只可能由 $g_{\alpha\beta}(x_0)$ 来刻画！

假定我们让自由下落的 x_0 变动，即在任意一个 x 点都定义一个自由下落的局部惯性系，那就能得到一个一般地刻画引力场的 $g_{\alpha\beta}(x)$ 场。从而可以将引力场中自由粒子的作用量写为

$$S[x(\sigma)] = -m \int d\tau = -m \int d\sigma \sqrt{-g_{\alpha\beta}(x) \frac{dx^\alpha}{d\sigma} \frac{dx^\beta}{d\sigma}}. \quad (3.11)$$

(3.11) 式的写法中其实对于坐标系 x^μ 没有任何限制，我们当然也可以采用另外一个比方说 x'^μ 坐标系，这时候粒子的作用量有如下关系

$$\begin{aligned} S[x(\sigma)] &= -m \int d\sigma \sqrt{-g_{\mu\nu}(x) \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}} \\ &= -m \int d\sigma \sqrt{-g'_{\mu\nu}(x') \frac{dx'^\mu}{d\sigma} \frac{dx'^\nu}{d\sigma}}. \end{aligned} \quad (3.12)$$

很显然，刻画引力的 $g_{\mu\nu}(x)$ 场不同坐标下的变换关系为

$$g'_{\rho\sigma}(x') = g_{\mu\nu}(x) \frac{\partial x^\mu}{\partial x'^\rho} \frac{\partial x^\nu}{\partial x'^\sigma}. \quad (3.13)$$

这与前面第一章黎曼几何中度规场在不同坐标下的变换关系完全一样。

这强烈地暗示我们可以把刻画引力场的 $g_{\mu\nu}(x)$ 场看成是黎曼几何中的度规场，从而将时空看成是某种弯曲的黎曼几何，具有如下黎曼度规

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu. \quad (3.14)$$

它与在引力场中自由运动粒子固有时^①的关系是

$$-d\tau^2 = ds^2. \quad (3.15)$$

但引力场是黎曼度规这个看法要成立的话，还需要验证黎曼几何的局部平坦性质，即是说，需要验证总是存在局部坐标系（有时候称之为**黎曼正则坐标系**），使得其中度规场在给定点的泰勒展开直到一阶小量为止都是平坦的，只在二阶小量上受黎曼曲率张量的修正。

但根据第一章中的相关论证可知，局部平坦性本身正是坐标变换关系 (3.13) 的结果。实际上，这个体现局部平坦的黎曼正则坐标系正是等效原理中所说的局部惯性系，因为根据 (3.10) 式，变换到局部惯性系中的度规场在坐标原点 x_0 处的值正是平坦的闵可夫斯基度规张量 $\eta_{\mu\nu}$ ，而且我们也知道，在局部惯性系中，度规场刻画的引力的确只在二阶小量上（在 x_0 点的泰勒展开）有效应，也就是潮汐力。这意味着，潮汐力在数学上就是某种黎曼曲率，而引力场完全是一种黎曼几何，有引力场的时空一般来说是弯曲的。

所以所谓的爱因斯坦等效原理（有时候也称之为强等效原理），说的正是：**在引力场中，总是存在局部惯性系，而且局部惯性系正是黎曼几何中的黎曼正则坐标系**。由此可知，在局部惯性系的坐标原点，度规场对时空坐标的一阶偏导均等于零，而二阶偏导由于要反映黎曼曲率所以通常不为零。局部惯性系的存在反映了引力场可以由一种局部平坦的时空几何来描述，这样的几何当然就是黎曼几何。

另外，(3.12) 式告诉我们，对于引力场中的自由落体，其质量 m 果然是作为一个整体的常数因子而出现的，这正符合第一章所说的弱等效原理的要求。即是说，弱等效原理是爱因斯坦等效原理的一个推论，所有对弱等效原理的实验检验同时也是在检验爱因斯坦的等效原理。

由于黎曼度规 $g_{\mu\nu}(x)$ 描写了引力场，而引力场是一个动力学场，所以在广义相对论中，黎曼度规本身是动力学变量，有动力学自由度，需要寻找它们所满足的动力学微分方程，也就是后面的章节中将要讲述的爱因斯坦场方程。从动力学这个角度来说，广义相对论与黎曼几何可不是一回事，黎曼几何的黎曼度规是人为给定的，而广义相对论中的黎曼度规是通过动力学方程求解出来的，无需预先给定！

当然，广义相对论中的黎曼度规 $g_{\mu\nu}(x)$ 与标准黎曼几何中的黎曼度规还有一个区别，即：广义相对论中的黎曼度规作为一个 4×4 矩阵，它有一个负本征值和三个正本征值（这从定义式 (3.10) 可以看出），分别对于一个时间方向和三个空间方向。有时候称这种度规为闵氏符号度规。而标准黎曼几何中的黎曼度规通常是正定矩阵，也称之为欧氏符号度规。

3.2 引力场中的自由粒子

3.2.1 统一自由落体定律

在上一节中，我们已经看到，根据等效原理，引力场中自由落体的相对论作用量为

$$S[x(\sigma)] = -m \int d\tau = -m \int d\sigma \sqrt{-g_{\mu\nu}(x)} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} \quad (3.16)$$

但在前面第一章中，我们也给出了牛顿力学中自由落体的作用量，为

$$S = m \int dt \left[\frac{1}{2} \dot{\mathbf{x}}^2 - \Phi(\mathbf{x}) \right]. \quad (3.17)$$

如何统一这两个作用量呢？

实际上，牛顿力学中的自由落体是广义相对论自由落体的非相对论极限，为了看清楚这一点，我们只需取 $(i, j = 1, 2, 3)$ 为空间指标

$$g_{00}(x) = -(1 + 2\Phi(\mathbf{x})), \quad g_{0i}(x) = 0, \quad g_{ij}(x) = \delta_{ij}. \quad (3.18)$$

这基本上就是闵可夫斯基度规 $\eta_{\mu\nu}$ ，只不过用引力势 Φ 对 g_{00} 分量作了一个小修正，即假定¹

$$\Phi(x) \ll 1. \quad (3.19)$$

从而相对论作用量 (3.16) 可以重写成

$$\begin{aligned} S[x(\sigma)] &= -m \int d\tau = -m \int \sqrt{(1 + 2\Phi(\mathbf{x}))dt^2 - d\mathbf{x}^2} \\ &= -m \int dt \sqrt{1 + 2\Phi(\mathbf{x}) - \mathbf{v}^2} \\ &\approx m \int dt \left[\frac{1}{2} \mathbf{v}^2 - \Phi(\mathbf{x}) \right] \end{aligned} \quad (3.20)$$

上式最后一行我们利用了非相对论极限 $\mathbf{v} \ll 1, \Phi \ll 1$ 。可见，从广义相对论的自由落体作用量中，我们能够正确地得到非相对论极限下的自由落体作用量。这也是对作用量 (3.16) 正确性的一个检验。

3.2.2 测地线方程

根据作用量 (3.16) 的形式和最小作用量原理可知，引力场中的自由落体将走固有时极大的路径。数学上常常称这样的路径为起末两点间的测地线，或者更精确地称之为类时测地线。这是因为 $d\tau^2$ 和线元 ds^2 只差一个负号，因此人们完全可以将固有时看作是对粒子世界线上两点间曲线弧长的一个度量。固有时极大相当于说粒子走的是弧长距离最长的路径。

之所以粒子走的是最长路径而不是最短路径，是因为广义相对论中的度规是闵氏符号度规而不是欧氏符号度规。但这里的本质和欧氏符号度规下的最短路径其实是一样的。而

¹注意，我们采用了 $c=1$ 单位制，如果恢复 c ，那即是 $\Phi(x)/c^2 \ll 1$ 。

在黎曼几何中，两点间的最短路径就叫做测地线，它是平坦欧几里得空间直线概念的推广。与此类比，我们就称引力场中自由粒子走的是测地线。

而之所以强调是类时测地线，是因为有质量粒子走的一定是类时曲线，即只能从每一点的光锥内部穿过的曲线。或者说，曲线上每一点的四维速度 $u^\mu = \frac{dx^\mu}{d\tau}$ 都是类时向量，即满足

$$g_{\mu\nu}u^\mu u^\nu < 0. \quad (3.21)$$

实际上，由 $-d\tau^2 = g_{\mu\nu}dx^\mu dx^\nu$ ，很显然有 $g_{\mu\nu}u^\mu u^\nu = -1$ 。所有的概念都和上一章狭义相对论中的相应概念类似，只是将度规张量从闵可夫斯基度规 $\eta_{\mu\nu}$ 替换成了一般性的 $g_{\mu\nu}(x)$ 。

为了利用最小作用量原理导出类时测地线的方程，我们可以在 (3.16) 中记

$$L = \sqrt{-g_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}}, \quad (3.22)$$

因此当取固有时本身为世界线参数时，即 $\sigma = \tau$ 时，我们有 $L = 1$ ，即

$$-g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = 1. \quad (3.23)$$

现在，我们可以把相对论粒子的作用量 (3.16) 写成 $S[x(\sigma)] = -m \int d\sigma L$ ，为了导出粒子的运动方程，我们需要计算变分 $\delta S[x(\sigma)] = -m \int d\sigma \delta L$ 。利用

$$\begin{aligned} L\delta L &= \frac{1}{2}\delta(L^2) = -g_{\mu\nu} \frac{dx^\mu}{d\sigma} \delta\left(\frac{dx^\nu}{d\sigma}\right) - \frac{1}{2}\delta g_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} \\ &= -g_{\mu\rho} \frac{dx^\mu}{d\sigma} \frac{d(\delta x^\rho)}{d\sigma} - \frac{1}{2}(\partial_\rho g_{\mu\nu}) \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} \delta x^\rho. \end{aligned} \quad (3.24)$$

可以得到

$$\begin{aligned} \delta S[x(\sigma)] &= -m \int d\sigma \delta L = m \int d\sigma \frac{1}{L} \left[g_{\mu\rho} \frac{dx^\mu}{d\sigma} \frac{d(\delta x^\rho)}{d\sigma} + \frac{1}{2}(\partial_\rho g_{\mu\nu}) \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} \delta x^\rho \right] \\ &= m \int d\sigma \frac{d}{d\sigma} \left(\frac{1}{L} g_{\mu\rho} \frac{dx^\mu}{d\sigma} \delta x^\rho \right) - m \int d\sigma \left[\frac{d}{d\sigma} \left(\frac{1}{L} g_{\mu\rho} \frac{dx^\mu}{d\sigma} \right) - \frac{1}{2L}(\partial_\rho g_{\mu\nu}) \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} \right] \delta x^\rho \\ &= -m \int d\sigma \left[\frac{d}{d\sigma} \left(\frac{1}{L} g_{\mu\rho} \frac{dx^\mu}{d\sigma} \right) - \frac{1}{2L}(\partial_\rho g_{\mu\nu}) \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} \right] \delta x^\rho. \end{aligned} \quad (3.25)$$

式中最后一个等号是利用了路径两端是固定的（即在两端 $\delta x^\rho = 0$ ），从而全微分项的积分结果为零。进而根据最小作用量原理 $\delta S[x(\sigma)] = 0$ ，就可以得到运动微分方程

$$\frac{\delta S}{\delta x^\rho(\sigma)} = -m \left[\frac{d}{d\sigma} \left(\frac{1}{L} g_{\mu\rho} \frac{dx^\mu}{d\sigma} \right) - \frac{1}{2L}(\partial_\rho g_{\mu\nu}) \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} \right] = 0. \quad (3.26)$$

我们可以通过取 $\sigma = \tau$ 来简化这个运动方程，这时候 $L = 1$ ，从而即有

$$\begin{aligned} 0 &= \frac{d}{d\tau} \left(g_{\mu\rho} \frac{dx^\mu}{d\tau} \right) - \frac{1}{2}(\partial_\rho g_{\mu\nu}) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \\ &= g_{\mu\rho} \frac{d^2 x^\mu}{d\tau^2} + (\partial_\nu g_{\mu\rho}) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} - \frac{1}{2}(\partial_\rho g_{\mu\nu}) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \\ &= g_{\mu\rho} \frac{d^2 x^\mu}{d\tau^2} + \frac{1}{2}(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}. \end{aligned} \quad (3.27)$$

将最后一行乘上度规矩阵 $g_{\mu\rho}$ 的逆矩阵 $g^{\rho\sigma}$ (也是一个对称矩阵), 满足 $g_{\mu\rho}g^{\rho\sigma} = \delta_{\mu}^{\sigma}$, 即可以得到

$$\frac{d^2x^{\sigma}}{d\tau^2} + \Gamma_{\mu\nu}^{\sigma} \frac{dx^{\mu}}{d\tau} \frac{dx^{\nu}}{d\tau} = 0. \quad (3.28)$$

方程 (3.28) 就是所谓的测地线方程, 式中

$$\Gamma_{\mu\nu}^{\sigma} = \frac{1}{2}g^{\sigma\rho} (\partial_{\mu}g_{\nu\rho} + \partial_{\nu}g_{\mu\rho} - \partial_{\rho}g_{\mu\nu}) \quad (3.29)$$

称作克里斯托夫联络。很显然, $\Gamma_{\mu\nu}^{\sigma}$ 关于两个下指标对称, 即满足

$$\Gamma_{\mu\nu}^{\sigma} = \Gamma_{\nu\mu}^{\sigma}. \quad (3.30)$$

至于为什么叫联络, 后面的章节中会进一步解释。

测地线方程不依赖于粒子质量, 因此也可以推广到零质量粒子 (比如光子), 只不过这时候由于 $g_{\mu\nu}dx^{\mu}dx^{\nu} = 0$, 即固有时恒为零, 从而不能选固有时为世界线的参数。不过这时候依然可以选择某个所谓的仿射参数 σ , 使得下面方程成立

$$\begin{aligned} \frac{d^2x^{\rho}}{d\sigma^2} + \Gamma_{\mu\nu}^{\rho} \frac{dx^{\mu}}{d\sigma} \frac{dx^{\nu}}{d\sigma} &= 0 \\ g_{\mu\nu} \frac{dx^{\mu}}{d\sigma} \frac{dx^{\nu}}{d\sigma} &= 0. \end{aligned} \quad (3.31)$$

我们称这种测地线为零性测地线。

为了得到测地线方程的非相对论版本, 我们只需对非相对论落体作用量 (3.17) 变分, 即可得

$$\frac{d^2x^i}{dt^2} = -\partial_i\Phi. \quad (3.32)$$

另外, 根据 (3.20) 可知, 在 \mathbf{v} 和 Φ 的最低阶 (零阶) 近似上, 有

$$d\tau = dt. \quad (3.33)$$

所以这两个式子又可以写成 (在最低阶近似上),

$$\frac{d^2t}{d\tau^2} = 0, \quad \frac{d^2x^i}{d\tau^2} = -\partial_i\Phi \frac{dt}{d\tau} \frac{dt}{d\tau}. \quad (3.34)$$

与测地线方程 (3.28) 进行比较, 即可得到在非相对论极限下

$$\Gamma_{00}^i = \partial_i\Phi, \quad (3.35)$$

其余的联络分量均为零。 $\partial_i\Phi$ 正是引力场强度, 由此可知, 克里斯托夫联络正是对引力场强度的一个相对论刻画!

另外, 我们也知道, 潮汐力由 Φ 的二阶偏导 $\partial_i\partial_j\Phi$ 刻画, 也就是由 $\partial_j\Gamma_{00}^i$ 刻画。前面说过, 潮汐力的相对论性刻画就是黎曼曲率张量, 由此可知黎曼曲率张量必定和克里斯托夫联络对时空坐标的一阶偏导有关, 具体什么关系我们在后面的章节中再来研究。

3.2.3 克里斯托夫联络与等效原理

克里斯托夫联络的坐标变换

由于自由落体的作用量 (3.12) 不依赖于特定坐标系, 所以测地线方程 (3.28) 的形式当然也不依赖于特定坐标系。即是说, 从 x^μ 坐标变换到 x'^μ 坐标, $x^\mu(\tau) = x^\mu(x'(\tau))$, 依然有方程

$$\frac{d^2 x'^\rho}{d\tau^2} + \Gamma'_{\mu\nu} \frac{dx'^\mu}{d\tau} \frac{dx'^\nu}{d\tau} = 0. \quad (3.36)$$

另一方面

$$\begin{aligned} \frac{d^2 x^\sigma}{d\tau^2} &= \frac{d}{d\tau} \left(\frac{\partial x^\sigma}{\partial x'^\rho} \frac{dx'^\rho}{d\tau} \right) = \frac{\partial x^\sigma}{\partial x'^\rho} \frac{d^2 x'^\rho}{d\tau^2} + \frac{d}{d\tau} \left(\frac{\partial x^\sigma}{\partial x'^\mu} \right) \frac{dx'^\mu}{d\tau} \\ &= \frac{\partial x^\sigma}{\partial x'^\rho} \frac{d^2 x'^\rho}{d\tau^2} + \frac{\partial^2 x^\sigma}{\partial x'^\mu \partial x'^\nu} \frac{dx'^\mu}{d\tau} \frac{dx'^\nu}{d\tau} \\ &= \left(-\frac{\partial x^\sigma}{\partial x'^\rho} \Gamma'_{\mu\nu} + \frac{\partial^2 x^\sigma}{\partial x'^\mu \partial x'^\nu} \right) \frac{dx'^\mu}{d\tau} \frac{dx'^\nu}{d\tau} \end{aligned} \quad (3.37)$$

上式最后一行我们代入了 x' 坐标系中的测地线方程 (3.36)。

进而根据 x 坐标系中的测地线方程 (3.28), 我们有

$$\begin{aligned} 0 &= \frac{d^2 x^\sigma}{d\tau^2} + \Gamma_{\alpha\beta}^\sigma \frac{dx^\alpha}{d\tau} \frac{dx^\beta}{d\tau} \\ &= \frac{d^2 x^\sigma}{d\tau^2} + \Gamma_{\alpha\beta}^\sigma \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} \frac{dx'^\mu}{d\tau} \frac{dx'^\nu}{d\tau}, \end{aligned} \quad (3.38)$$

代入 (3.37) 式, 即可得

$$\begin{aligned} &\left(-\frac{\partial x^\sigma}{\partial x'^\rho} \Gamma'_{\mu\nu} + \frac{\partial^2 x^\sigma}{\partial x'^\mu \partial x'^\nu} + \Gamma_{\alpha\beta}^\sigma \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} \right) \frac{dx'^\mu}{d\tau} \frac{dx'^\nu}{d\tau} = 0 \\ \Rightarrow &-\frac{\partial x^\sigma}{\partial x'^\rho} \Gamma'_{\mu\nu} + \frac{\partial^2 x^\sigma}{\partial x'^\mu \partial x'^\nu} + \Gamma_{\alpha\beta}^\sigma \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} = 0. \end{aligned} \quad (3.39)$$

由此即可以得到克里斯托夫联络在不同坐标下的变换关系

$$\Gamma'_{\mu\nu}{}^\rho = \frac{\partial x'^\rho}{\partial x^\sigma} \left(\Gamma_{\alpha\beta}^\sigma \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} + \frac{\partial^2 x^\sigma}{\partial x'^\mu \partial x'^\nu} \right). \quad (3.40)$$

局部惯性系回顾

根据等效原理, 在任意时空点 P , 我们都可以选取一个局部惯性系 (也就是黎曼正则坐标系), 使得在此局部惯性系中

$$g_{\mu\nu}(P) = \eta_{\mu\nu}, \quad \partial_\rho g_{\mu\nu}|_P = 0. \quad (3.41)$$

当然, 通常来说, $\partial_\sigma \partial_\rho g_{\mu\nu}|_P \neq 0$, 因为黎曼曲率通常不为零。进而根据克里斯托夫联络的定义式 (3.29), 即可知, 在此局部惯性系中

$$\Gamma_{\mu\nu}^\rho(P) = 0. \quad (3.42)$$

而且通常来说 $\partial_\sigma \Gamma_{\mu\nu}^\rho|_P \neq 0$ 。即在局部惯性系的 P 点，不存在引力场（引力场完全被惯性力抵消了）！从而在局部惯性系中，自由落体的测地线方程 (3.28) 变为

$$\frac{d^2 x^\rho}{d\tau^2}|_P = 0. \quad (3.43)$$

即在局部惯性系的 P 点，自由落体作直线运动。从这个角度我们也可以理解为什么说测地线是直线在弯曲时空中的推广。

另一方面，根据克里斯托夫联络的定义 (3.29)，可以得到

$$g_{\rho\sigma} \Gamma_{\mu\nu}^\sigma = \frac{1}{2} (\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}). \quad (3.44)$$

将三个指标 ρ, μ, ν 进行轮换，即可得

$$g_{\mu\sigma} \Gamma_{\nu\rho}^\sigma = \frac{1}{2} (\partial_\nu g_{\rho\mu} + \partial_\rho g_{\nu\mu} - \partial_\mu g_{\nu\rho}), \quad (3.45)$$

$$g_{\nu\sigma} \Gamma_{\rho\mu}^\sigma = \frac{1}{2} (\partial_\rho g_{\mu\nu} + \partial_\mu g_{\rho\nu} - \partial_\nu g_{\rho\mu}). \quad (3.46)$$

将这两个式子相加，即可得

$$g_{\mu\sigma} \Gamma_{\nu\rho}^\sigma + g_{\nu\sigma} \Gamma_{\rho\mu}^\sigma = \partial_\rho g_{\mu\nu}. \quad (3.47)$$

利用 $\Gamma_{\mu\nu}^\rho$ 关于两个下指标对称，即可得

$$\partial_\rho g_{\mu\nu} - g_{\mu\sigma} \Gamma_{\rho\nu}^\sigma - g_{\sigma\nu} \Gamma_{\rho\mu}^\sigma = 0. \quad (3.48)$$

这个结果非常重要，在后面的章节中常用。

反过来，(3.48) 式以及 $\Gamma_{\mu\nu}^\rho$ 关于两个下指标对称的性质也唯一地决定了克里斯托夫联络。为了看清楚这一点，我们将 (3.48) 式的三个指标 ρ, μ, ν 进行轮换，从而得到

$$\partial_\mu g_{\nu\rho} - g_{\nu\sigma} \Gamma_{\mu\rho}^\sigma - g_{\sigma\rho} \Gamma_{\mu\nu}^\sigma = 0, \quad (3.49)$$

$$\partial_\nu g_{\rho\mu} - g_{\rho\sigma} \Gamma_{\nu\mu}^\sigma - g_{\sigma\mu} \Gamma_{\nu\rho}^\sigma = 0. \quad (3.50)$$

利用 $\Gamma_{\mu\nu}^\rho$ 关于两个下指标对称的性质计算 (3.49)+(3.50)-(3.48)，并将结果除以 2，即可得 (3.44) 式，从而即有克里斯托夫联络的表达式 (3.29)。

由 (3.48) 式可知，若在某个坐标系的 P 点有 $\Gamma_{\mu\nu}^\rho(P) = 0$ ，则也必定有 $\partial_\rho g_{\mu\nu}|_P = 0$ 。综上所述可知， $\partial_\rho g_{\mu\nu}|_P = 0$ 的要求和 $\Gamma_{\mu\nu}^\rho(P) = 0$ 的要求完全等价。从而即知局部惯性系的等价定义，就是使得下式成立的参考系

$$g_{\mu\nu}(P) = \eta_{\mu\nu}, \quad \Gamma_{\mu\nu}^\rho(P) = 0. \quad (3.51)$$

等效原理从物理上说，对于任意引力场，这样的坐标系总存在。而黎曼几何的局部平坦性质则是从数学角度证明，这样的坐标系总存在。实际上，正因为等效原理从物理上确保了这样的参考系总存在，所以广义相对论的时空几何才是黎曼几何！

3.3 等效原理的预言

3.3.1 引力红移

根据本章第一节的知识，引力场中的一架跟着物体一起运动的标准钟，它测到一个物理过程的固有时 $d\tau$ 为

$$d\tau = \sqrt{-g_{\mu\nu}dx^\mu dx^\nu}. \quad (3.52)$$

特别的，如果物体和钟都静止，则有

$$d\tau = \sqrt{-g_{00}(x)}dt \Rightarrow dt = d\tau(-g_{00})^{-1/2}, \quad (3.53)$$

式中 dt 就是这个物理过程的坐标时，这个式子告诉我们，引力场的存在影响时间的流逝！

但是，我们不能通过简单地测量某地物理过程的坐标时 dt ，并将它和固有时 $d\tau$ 进行比较来确定引力场如何影响时间。这是因为，引力场对一切事物的时间流逝施加了同样的影响。也即是说，如果标准钟指出某一物理过程在没有引力场并静止时进行了一秒钟，那在有引力场时它还是指出进行了一秒钟，因为标准钟和物理过程以同样的方式受引力场影响。换言之，同样的物理过程，无论它是否在引力场中，无论它在引力场的什么位置，它用当地的标准钟测出来的固有时总是一样的！因为引力场对时间的标准和对物理过程一样，施加了同样的影响。

简单来说，假设你正在黑洞的边上，那在空间无穷远处的观察者看来，很可能你的时间会变慢，你的一切物理过程都变得很缓慢，但是，对于当地的你来说，你发现不了任何异常，你也感觉不到任何变慢，因为引力场对你的物理过程和你的时间标准施加了同样的影响。

为了测出引力场如何影响时间，我们需要将引力场中两个不同地点的物理过程进行比较。例如，假设一个位于 \mathbf{x}_E 位置的观察者（不妨称之为发射者 (emitter)），正以 $d\tau_E$ 的固有时间间隔发射电磁波信号给位于 \mathbf{x}_R 位置的接收者 (receiver)。假设发射者和接收者均静止在引力场中，并且假定引力场是一个不随时间变化的静态引力场。那么信号从 \mathbf{x}_E 处传到 \mathbf{x}_R 处所需的坐标时 t 将是一个常数，因此相继到达接收者的两个信号的坐标时之差 dt_R 将等于它们离开发射者时的坐标时之差 dt_E ，即 $dt_R = dt_E$ 。由 (3.53) 式，

$$dt_R = dt_E = d\tau_E(-g_{00}(\mathbf{x}_E))^{-1/2}. \quad (3.54)$$

对于 \mathbf{x}_R 位置的接收者来说，他将测到信号的固有时间间隔 $d\tau_R$ 为

$$d\tau_R = (-g_{00}(\mathbf{x}_R))^{1/2}dt_R = (-g_{00}(\mathbf{x}_R))^{1/2}dt_E. \quad (3.55)$$

联立 (3.54) 式和 (3.55) 式，即有

$$\frac{d\tau_R}{d\tau_E} = \left(\frac{g_{00}(\mathbf{x}_R)}{g_{00}(\mathbf{x}_E)}\right)^{1/2}. \quad (3.56)$$

对于 $g_{00}(\mathbf{x}) = -(1 + 2\Phi(\mathbf{x}))$ 的牛顿引力场，我们有

$$\frac{d\tau_R}{d\tau_E} = \left(\frac{1 + 2\Phi(\mathbf{x}_R)}{1 + 2\Phi(\mathbf{x}_E)}\right)^{1/2}. \quad (3.57)$$

特别的, 假设 $\Phi(\mathbf{x}) = -G\frac{M}{|\mathbf{x}|}$, 假设发射者处于引力场中, 而接收者位于空间无穷远处, 从而 $\Phi(\mathbf{x}_R) \rightarrow 0$. 则 (3.57) 式告诉我们,

$$\frac{d\tau_R}{d\tau_E} = \left(1 - 2G\frac{M}{|\mathbf{x}_E|}\right)^{-1/2} > 1. \quad (3.58)$$

即是说, 在无穷远处的接收者看来, 信号的发射时间间隔变长了! 换言之, 在无穷远处的观察者看来, 引力场中的发射者的时间变慢了! 这就是所谓的引力场的时间膨胀效应, 也即是引力场使得时间变慢的效应。但要强调的是, 这个变慢是在无穷远处的观察者看来变慢了, 对于引力场中当地的观察者而言, 他不会感觉到任何异常, 也不会感觉到任何变慢。

由于电磁波的频率是时间周期的倒数, 因此, 假设记发射者发射电磁波的频率为 ν_E , 接收者接收到的电磁波频率为 ν_R , 则根据 (3.56) 式, 有

$$\frac{\nu_R}{\nu_E} = \left(\frac{g_{00}(\mathbf{x}_R)}{g_{00}(\mathbf{x}_E)}\right)^{-1/2}. \quad (3.59)$$

对于 $g_{00}(\mathbf{x}) = -(1 + 2\Phi(\mathbf{x}))$ 的牛顿引力场, 我们有

$$\frac{\nu_R}{\nu_E} = \left(\frac{1 + 2\Phi(\mathbf{x}_R)}{1 + 2\Phi(\mathbf{x}_E)}\right)^{-1/2}. \quad (3.60)$$

对于 $\Phi \ll 1$ 的弱引力场, 这个结果可以进一步近似成

$$\frac{\nu_R}{\nu_E} = 1 + \Phi(\mathbf{x}_E) - \Phi(\mathbf{x}_R). \quad (3.61)$$

或者引入 $\Delta\nu = \nu_R - \nu_E$, $\nu = \nu_E$, 进而将这个结果写成

$$\frac{\Delta\nu}{\nu} = \Phi(\mathbf{x}_E) - \Phi(\mathbf{x}_R). \quad (3.62)$$

特别的, 假设发射者处于 $\Phi(\mathbf{x}) = -G\frac{M}{|\mathbf{x}|}$ 的引力场中, 接收者位于空间无穷远处, 则上式就成为

$$\frac{\Delta\nu}{\nu} = \Phi(\mathbf{x}_E) = -G\frac{M}{|\mathbf{x}_E|} < 0. \quad (3.63)$$

即是说, 对于无穷远处的观察者而言, 引力场中的信号频率变小了! 或者说, 信号向红端移动了。这就是所谓的引力红移!

自从广义相对论预言了引力红移以后, 差不多过了近 50 年它才被实验证实! 以太阳为例, 记其半径为 R , 质量为 M , 则在太阳表面 (恢复了光速 c) $\Phi(R)/c^2 = GM/(Rc^2) \sim 10^{-6}$. 这就是太阳所产生的引力红移的数量级。但是在太阳表面发光的原子由于剧烈的热运动, 它还要受到一个标准的多普勒频移, 而且多普勒频移在量级上将完全盖过太阳的引力红移, 这就使得太阳的引力红移无法测量。对于地球而言, 其表面 $GM/(Rc^2) \sim 10^{-9}$, 所以地球的引力红移似乎更难测量。好在, 1958 年发现了穆斯堡尔效应。什么是穆斯堡尔效应呢? 正常来说, 晶体原子的发光谱线会被原子发光时的反冲加宽。但是穆斯堡尔发现, 在一定的条件下, 晶体中的原子将“锁链”在一起, 使得每个原子在发光时的反冲都转化为整个晶体的微不可见的反冲, 从而使得原子的发光谱线变得很尖细。1960 年, 庞德、雷布卡和斯奈德采用穆斯堡尔效应的实验方法, 测量由地面上高度相差 22.6 米的两点之间引力势的微小差别所造成的谱线频率的移动, 定量地验证了引力红移。结果表明实验值与理论值完全符合! 今天, 即使在我们日常使用的 GPS 系统中, 也已经考虑了引力红移所带来的修正了, 从这个意义上来说, 广义相对论早已走进了人们的日常生活。

3.3.2 光线偏折

虽然引力红移过了近 50 年才得到实验验证，但是等效原理的另一个预言在提出后却很快得到了验证，这就是光线偏折。

根据牛顿的引力理论，任何有质量的物体都要受到引力场的作用。但是，光是没有质量的，当然光有能量，那引力场会不会对光线产生作用呢？根据爱因斯坦的等效原理，答案是肯定的，引力场的存在会使得光线偏折。

为了回答引力场如何作用于光线，根据等效原理，我们只需分析光线在太空中一个加速参考系中如何运动。如图 (3.2) 所示，假设在太空中一个“向上”加速运动的小房间里，一个手持激光枪的人水平地开了一枪，在这个加速参考系中，激光枪打出的光线如何运动呢？很显然，由于小房间在向上加速，光线看起来就会往下偏折，如图 (3.2) 所示。

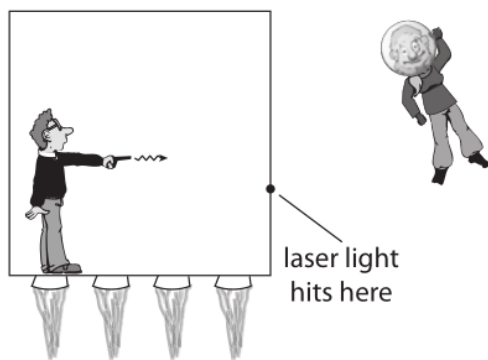


Figure 3.2: 加速参考系中，激光枪打出的光线会偏折。图片来自 A Zee, Einstein Gravity in a Nutshell

而根据等效原理，向上加速的参考系等效于一个向下的引力场（惯性力等效于引力），因此我们就可以得到结论，即在引力场中，光线会偏折！

实际上根据牛顿的理论，对于质量为 m 以光速 c 运动的粒子，引力场也会使得它偏折，有一个牛顿理论的偏折量。但是，这里我们没有具体计算广义相对论中光线的精确偏折量，在后面的章节中我们将完成这样的计算，我们将发现，广义相对论给出的偏折量是牛顿理论偏折量的两倍！对于光线刚好掠过太阳表面的情况，广义相对论给出的偏折量约为 1.75 弧秒。

要检验光线通过大质量物体附近发生偏折的程度，最好的机会莫过于在日全食时对太阳所在的天区进行照相。在日全食时拍摄若干照相底片，然后等若干时间之后，太阳远离了发生日食的天区，再次对该天区拍摄若干底片。通过将前后两组底片进行对比测算，就能确定星光被偏折的程度。

为了回应爱因斯坦关于光线偏折的预言，爱丁顿和戴森组织了两支考察队去观测 1919 年 5 月 29 日发生的日食，分别到非洲几内亚海湾的普林西比岛 (Principe) 和巴西北部的索布拉尔 (Sobral) 做了观测。两支考察队所得到的观测结果与爱因斯坦的预言符合得

很好。但是这一结果后来受到不少质疑，因为导致结果产生误差的因素很多，而爱丁顿的处理更远没有达到无懈可击。

光学观测的精度似乎到了极限，后来人们想到通过观测太阳对无线电波的偏折来检验广义相对论的预言。1974 年到 1975 年间，福马伦特 (A. B. Fomalont) 和什拉梅克 (R. A. Sramek) 利用甚长基线干涉技术，观测了太阳对三个射电源的偏折，最后 (1976 年) 得到太阳边缘处射电源的微波被偏折 1.761 ± 0.016 。终于，天文学家以误差小于百分之一的精度证实了广义相对论的预言。到 1991 年，更是利用多家天文台协同观测的技术，以万分之一的精度证实了广义相对论对光线弯曲的预言。只不过这时观测的不再是可见光而是无线电波。

4. 广义协变原理与微分同胚不变性

4.1 广义协变原理

从上一章的知识我们知道，可以用任意的坐标系来描写引力场中的自由粒子，从 x 坐标系变换到 x' 坐标系，度规场的变换关系为

$$g'_{\mu\nu}(x') = g_{\rho\sigma} \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu}. \quad (4.1)$$

根据本书第一章可知，引入任意的曲线坐标系，这正是高斯在研究内禀几何时引入的办法，这种办法后来被黎曼推广到了任意维的黎曼几何。等效原理既然告诉我们引力场存在时的时空几何就是黎曼几何，那在广义相对论中引入任意的坐标系就是一件必然的事情。

在物理上，之所以可以引入任意坐标系，原因在于等效原理允许我们引入相对于局部惯性系作任意加速运动的参考系，惯性力就是万有引力。因此在相对论性的引力理论中，一切参考系都是平权的，只不过引力场在不同参考系中的表现不同！在广义相对论的早期文献中，爱因斯坦也称这个结论为广义相对性原理，即将之看成是狭义相对性原理的进一步推广。狭义相对性原理说一切惯性系都平权，而广义相对性原理则进一步说一切参考系都平权！而不同参考系必然使用不同的时空坐标系，一切参考系平权就意味着，在广义相对论中，一切可能的坐标系都是平权的！

这就是所谓的广义协变原理，它告诉我们：广义相对论中的物理方程必须满足广义协变性，即在广义坐标变换 $x \rightarrow x'$ 下，方程的形式必须保持不变！

有些书上说，广义协变原理本身没有物理内容，因为任何物理定律都能写成广义协变的形式，即使牛顿定律也可以在任意的坐标系中表达出来。但是，诸如牛顿定律这样的理

论写成广义协变形式，其实都只是一种形式上的广义协变，因为在这样的广义协变写法中，度规场和克里斯托夫联络都是纯几何的系数，它们在不同坐标系中的变换关系是一种纯粹数学形式上的变换，的确没有任何物理内容。而广义相对论中的广义协变性却是有物理内容的，因为在广义相对论中，度规场和克里斯托夫联络不再是纯几何系数，它们描写了引力场，因此是有动力学的。要求这样的动力学场在广义坐标变换下协变，其实是对引力场的动力学规律施加了很强的约束！因此广义相对论中的广义协变原理是一条真正的物理原理。

根据高斯内禀几何学的研究，通常来说，任意一个坐标系 x 都只能覆盖时空的一个局部区域，而无法覆盖时空的整体，就好比球坐标系无法覆盖两维球面的南北极。因此通常又称任意的坐标系为局部坐标系。广义协变原理告诉我们，任何局部坐标系都是平权的。因此在广义相对论中有极大的选择坐标系的自由度，问题是，这些局部坐标是否是物理的呢？是否有物理含义呢？

对于这个问题，我们的回答是这样的：如果我们把局部坐标系看成是某些局域观察者所采用的坐标系，则只要我们赋予这些局域观察者物理意义，那么其相应的局部坐标系就有一定的物理意义，至少与局部坐标系对应的局部参考系有物理意义。

相反，由于可以进行任意的坐标变换，我们也可以作这样的处理：即不赋予局域观察者物理意义，进而认为任何局部坐标都只是一种数学描述，本身没有任何物理含义，而之所以能进行任意的坐标变换，也纯粹是因为坐标本身完全是对物理规律的描述上的冗余。在量子引力中，人们就是这样处理的！

其实，当我们将广义协变原理与最小作用量原理结合起来的时候，它才真正发挥出最大的威力。根据最小作用量原理，基本的物理方程应该通过要求某个作用量取极值来得出，又由于作用量是一个标量，从而为了使得物理方程广义协变，作用量就应该在广义坐标变换下保持不变。从而，结合最小作用量原理，我们就能把广义协变原理表述成：**在广义相对论中，物理系统的作用量必须在任意的 $x \rightarrow x'$ 的坐标变换下保持不变！尤其是，引力场本身的作用量必须在任意的 $x \rightarrow x'$ 变换下保持不变！**

4.2 张量场和微分同胚映射

为了构造满足广义协变原理的物理方程，我们需要让其中涉及的物理量在坐标变换下按照某种简单的规则变换。这样的变换规则就是 (4.1) 式变换规则的推广，这就需要引入弯曲时空中的张量场概念。

4.2.1 矢量场和张量场

简单来说，弯曲时空中的张量场就是将第二章引入的闵可夫斯基时空中的张量场概念推广到弯曲时空。还是让我们从所谓的 (1,0) 型张量场和 (0,1) 型张量场，也就是所谓的逆变矢量场和协变矢量场开始吧。

首先，在坐标变换 $x^\mu \rightarrow x'^\mu$ 之下，我们有如下变换

$$dx'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} dx^\nu, \quad \partial'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} \partial_\nu. \quad (4.2)$$

为书写简单起见，可以引入矩阵 $S^\mu{}_\nu$ ，

$$S^\mu{}_\nu(x) = \frac{\partial x'^\mu}{\partial x^\nu}, \quad (4.3)$$

特别的，对于闵可夫斯基时空中的洛伦兹变换，我们有 $S^\mu{}_\nu(x) = \Lambda^\mu{}_\nu$ ，为一个常数矩阵。但是，对于弯曲时空中的一般性坐标变换，矩阵 $S^\mu{}_\nu(x)$ 依赖于 x 。另外，根据

$$\frac{\partial x'^\mu}{\partial x^\nu} \frac{\partial x^\nu}{\partial x'^\rho} = \frac{\partial x'^\mu}{\partial x'^\rho} = \delta^\mu_\rho, \quad (4.4)$$

结果为单位矩阵，由此可知

$$(S^{-1})^\mu{}_\nu(x) = \frac{\partial x^\mu}{\partial x'^\nu}, \quad (4.5)$$

式中 S^{-1} 表示矩阵 S 的逆矩阵。利用 S 和 S^{-1} ，我们就能将变换关系 (4.2) 重写为

$$dx'^\mu = S^\mu{}_\nu(x) dx^\nu, \quad \partial'_\mu = (S^{-1})^\nu{}_\mu(x) \partial_\nu. \quad (4.6)$$

而度规场的变换关系 (4.1) 就能重写成

$$g'_{\mu\nu}(x') = g_{\rho\sigma}(x) (S^{-1})^\rho{}_\mu(x) (S^{-1})^\sigma{}_\nu(x). \quad (4.7)$$

将这个方程看作一个矩阵方程，并两边求逆矩阵，就可以得到

$$g'^{\mu\nu}(x') = S^\mu{}_\rho(x) S^\nu{}_\sigma(x) g^{\rho\sigma}(x), \quad (4.8)$$

式中 $g'^{\mu\nu}$ 为矩阵 $g_{\mu\nu}$ 的逆矩阵。

所谓的 (1,0) 型张量场 $A^\mu(x)$ ，或者说逆变矢量场 $A^\mu(x)$ ，就是指在坐标变换下和 dx^μ 的变换规则相同的一个场。换言之， $A^\mu(x)$ 在坐标变换下按如下规则变换

$$A'^\mu(x') = S^\mu{}_\nu(x) A^\nu(x). \quad (4.9)$$

而所谓的 (0,1) 型张量场 $B_\mu(x)$ ，或者说协变矢量场 $B_\mu(x)$ ，就是指在坐标变换下与 ∂_μ 的变换规则相同的一个场。换言之， $B_\mu(x)$ 在坐标变换下按如下规则变换

$$B'_\mu(x') = (S^{-1})^\nu{}_\mu(x) B_\nu(x). \quad (4.10)$$

而 $A^\mu(x)$ 与 $B_\mu(x)$ 进行指标缩并的结果则是一个标量场，即是说，它的变换规则为

$$A'^\mu(x') B'_\mu(x') = A^\mu(x) B_\mu(x). \quad (4.11)$$

更一般的，对于任何标量场 $\Phi(x)$ ，它在坐标变换之下的变换规则均为

$$\Phi'(x') = \Phi(x). \quad (4.12)$$

另外，根据度规场的变换关系 (4.7)、(4.8) 不难看出， $g_{\mu\nu}(x) A^\nu(x)$ 其实是一个协变矢量场，而 $g^{\mu\nu}(x) B_\nu(x)$ 其实是一个逆变矢量场。因此可以引入如下定义

$$A_\mu(x) \equiv g_{\mu\nu}(x) A^\nu(x), \quad B^\mu(x) \equiv g^{\mu\nu}(x) B_\nu(x). \quad (4.13)$$

即是说，我们可以利用度规场以及它的逆矩阵来将张量场的指标降下去或者升上来。

更一般地，我们可以引入有 k 个上指标 l 个下指标的 (k, l) 型张量场，记作 $T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l}(x)$ 。在坐标变换之下，它的每一个上指标都像 dx^μ 那样变，而每一个下指标都像 ∂_μ 那样变。特别的，度规场 $g_{\mu\nu}(x)$ 就是一个 $(0, 2)$ 型张量场，而其逆矩阵场 $g^{\mu\nu}(x)$ 就是一个 $(2, 0)$ 型张量场。总之，这里的一切都完全平行于第二章讲四维闵可夫斯基时空张量时的概念，只是现在用来升降指标的度规张量得是 $g_{\mu\nu}$ 而不是闵可夫斯基时空的 $\eta_{\mu\nu}$ ，另外，张量的每个指标得按照 $S^\mu{}_\nu$ 或者 $(S^{-1})^\mu{}_\nu$ 变，而不是按照洛伦兹变换的 $\Lambda^\mu{}_\nu$ 和 $(\Lambda^{-1})^\mu{}_\nu$ 变。

特别的，容易验证克龙内克记号 δ^μ_ν 是一个 $(1, 1)$ 型张量。因为

$$S^\rho{}_\mu (S^{-1})^\nu{}_\sigma \delta^\mu_\nu = S^\rho{}_\mu (S^{-1})^\mu{}_\sigma = \delta^\rho_\sigma. \quad (4.14)$$

很显然， δ^μ_ν 是一个在所有坐标系中分量都相同的张量。

下面举一个初看起来像是张量，但其实不是张量的例子，那就是克里斯托夫联络 $\Gamma^\rho_{\mu\nu}$ 。它初看起来像是一个 $(1, 2)$ 型张量，但是细看其在坐标变换下的变换规则

$$\begin{aligned} \Gamma'^\rho_{\mu\nu}(x') &= \frac{\partial x'^\rho}{\partial x^\sigma} \left(\Gamma^\sigma_{\alpha\beta} \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} + \frac{\partial^2 x^\sigma}{\partial x'^\mu \partial x'^\nu} \right) \\ &= S^\rho{}_\sigma \Gamma^\sigma_{\alpha\beta} (S^{-1})^\alpha{}_\mu (S^{-1})^\beta{}_\nu + S^\rho{}_\sigma \frac{\partial}{\partial x'^\mu} (S^{-1})^\sigma{}_\nu \\ &= S^\rho{}_\sigma \Gamma^\sigma_{\alpha\beta} (S^{-1})^\alpha{}_\mu (S^{-1})^\beta{}_\nu + S^\rho{}_\sigma (S^{-1})^\lambda{}_\mu \partial_\lambda (S^{-1})^\sigma{}_\nu, \end{aligned} \quad (4.15)$$

不难发现，它比张量的变换规则额外多出了一项，从而不是一个张量！

另外，对 $S^\rho{}_\sigma (S^{-1})^\sigma{}_\nu = \delta^\rho_\nu$ 式求导，可以得到

$$S^\rho{}_\sigma \partial_\lambda (S^{-1})^\sigma{}_\nu = -(S^{-1})^\sigma{}_\nu \partial_\lambda S^\rho{}_\sigma. \quad (4.16)$$

将这个结果代入 (4.15) 式，即可以得到

$$\Gamma'^\rho_{\mu\nu}(x') = S^\rho{}_\sigma \Gamma^\sigma_{\alpha\beta} (S^{-1})^\alpha{}_\mu (S^{-1})^\beta{}_\nu - (S^{-1})^\lambda{}_\mu (S^{-1})^\sigma{}_\nu \partial_\lambda S^\rho{}_\sigma. \quad (4.17)$$

有了张量的概念，我们就很容易构造一大类满足广义协变原理的方程：**任何方程，只要它是两个同样上下指标的张量的等式，则在广义坐标变换下，其形式都将保持不变。**例如，假设 $A^{\mu\nu}{}_\rho$ 和 $B^{\mu\nu}{}_\rho$ 是两个 $(2, 1)$ 型张量，又如果在 x^μ 坐标系里有 $A^{\mu\nu}{}_\rho = B^{\mu\nu}{}_\rho$ ，那么在 x'^μ 坐标系里，也必定有 $A'^{\mu\nu}{}_\rho = B'^{\mu\nu}{}_\rho$ 。特别的，如果在某个坐标系中某张量等于零，那么在所有的坐标系中，这个张量都必定等于零。

4.2.2 微分同胚映射和李导数

前面我们将 $x^\mu \rightarrow x'^\mu$ 的变换看作是同一个时空点在两组不同坐标下的坐标变换，通常人们称这种看法为**被动的观点**。与之相对应的，所谓的**主动观点**则是，将 x 和 x' 看成两个不同的时空点，将 $x \rightarrow x'$ 的变换看成是时空点 x 到时空点 x' 的映射，由于这种映射是一一对映，并且映射本身以及其逆映射均可微，所以也称作微分同胚映射。即是说，**被动观点中的坐标变换，在主动观点中，就是一个局部微分同胚映射。**

无穷小微分同胚映射

特别的, 我们可以考察如下无穷小微分同胚映射

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x). \quad (4.18)$$

其中 $\varepsilon^\mu(x)$ 是取值为无穷小量的函数。按照微分同胚映射的解释, 这个变换将 x 点映射到了无限靠近的 x' 点, 两点之间相差一个无穷小量 ε , 所以这是一个无穷小微分同胚映射。很显然, 在一阶近似上有

$$\frac{\partial x'^\mu}{\partial x^\nu} = \delta_\nu^\mu + \partial_\nu \varepsilon^\mu, \quad \frac{\partial x^\nu}{\partial x'^\mu} = \delta_\mu^\nu - \partial_\mu \varepsilon^\nu. \quad (4.19)$$

下面我们先来看一下度规张量在上述无穷小微分同胚变换下将如何变。很显然

$$\begin{aligned} g'_{\mu\nu}(x') &= g_{\rho\sigma}(x) \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} = g_{\rho\sigma}(x) (\delta_\mu^\rho - \partial_\mu \varepsilon^\rho) (\delta_\nu^\sigma - \partial_\nu \varepsilon^\sigma) \\ &= g_{\mu\nu}(x) - g_{\rho\nu} \partial_\mu \varepsilon^\rho - g_{\mu\sigma} \partial_\nu \varepsilon^\sigma = g_{\mu\nu}(x - \varepsilon) - g_{\rho\nu} \partial_\mu \varepsilon^\rho - g_{\mu\sigma} \partial_\nu \varepsilon^\sigma \\ &= g_{\mu\nu}(x') - \varepsilon^\rho \partial_\rho g_{\mu\nu} - g_{\rho\nu} \partial_\mu \varepsilon^\rho - g_{\mu\sigma} \partial_\nu \varepsilon^\sigma. \end{aligned} \quad (4.20)$$

将上式中的变量 x' 替换成变量 x (式中的 $x = x' - \varepsilon$ 则替换成 $x - \varepsilon$), 从而在一阶无穷小近似上, 我们可以将同在 x 点的两个张量 $g'_{\mu\nu}(x)$ 和 $g_{\mu\nu}(x)$ 作差, 即考察无穷小微分同胚变换前后 x 点度规场的改变量

$$\begin{aligned} \delta g_{\mu\nu}(x) &= g'_{\mu\nu}(x) - g_{\mu\nu}(x) \\ &= -[\varepsilon^\rho \partial_\rho g_{\mu\nu}(x) + g_{\rho\nu} \partial_\mu \varepsilon^\rho + g_{\mu\rho} \partial_\nu \varepsilon^\rho]. \end{aligned} \quad (4.21)$$

完全类似的, 也可以得到

$$\begin{aligned} \delta g^{\mu\nu}(x) &= g'^{\mu\nu}(x) - g^{\mu\nu}(x) \\ &= -[\varepsilon^\rho \partial_\rho g^{\mu\nu}(x) - g^{\rho\nu} \partial_\rho \varepsilon^\mu - g^{\mu\rho} \partial_\rho \varepsilon^\nu]. \end{aligned} \quad (4.22)$$

在广义相对论中, 考察度规场在微分同胚变换下的变化当然最为重要, 但我们也不能只关心度规场, 而是同时也要关心其它一些场 (比如标量场和矢量场) 在无穷小微分同胚 (4.18) 下如何变化。比如对于标量场 $\Phi(x)$, 由于有

$$\Phi'(x') = \Phi(x) = \Phi(x' - \varepsilon) = \Phi(x') - \varepsilon^\rho \partial_\rho \Phi, \quad (4.23)$$

所以在一阶无穷小的近似上, 很容易得到

$$\delta \Phi(x) = \Phi'(x) - \Phi(x) = -\varepsilon^\rho \partial_\rho \Phi. \quad (4.24)$$

再比如对于矢量场 $W^\mu(x)$, 在无穷小微分同胚 (4.18) 下, 有

$$\begin{aligned} W'^\mu(x') &= W^\rho(x) \frac{\partial x'^\mu}{\partial x^\rho} = W^\rho(x) (\delta_\rho^\mu + \partial_\rho \varepsilon^\mu) \\ &= W^\mu(x) + W^\rho \partial_\rho \varepsilon^\mu = W^\mu(x' - \varepsilon) + W^\rho \partial_\rho \varepsilon^\mu \\ &= W^\mu(x') - \varepsilon^\rho \partial_\rho W^\mu + W^\rho \partial_\rho \varepsilon^\mu. \end{aligned} \quad (4.25)$$

从而即有

$$\delta W^\mu(x) = W'^\mu(x) - W^\mu(x) = -[\varepsilon^\rho \partial_\rho W^\mu - W^\rho \partial_\rho \varepsilon^\mu]. \quad (4.26)$$

类似的，对于矢量场 $U_\mu(x)$ ，也可以得到

$$\delta U_\mu(x) = U'_\mu(x) - U_\mu(x) = -[\varepsilon^\rho \partial_\rho U_\mu + U_\rho \partial_\mu \varepsilon^\rho]. \quad (4.27)$$

完全类似的，对于任意二阶张量场 $W^{\mu\nu}(x)$ 和 $U_{\mu\nu}(x)$ ，我们可以得到

$$\begin{aligned} \delta W^{\mu\nu}(x) &= W'^{\mu\nu}(x) - W^{\mu\nu}(x) = -[\varepsilon^\rho \partial_\rho W^{\mu\nu} - W^{\rho\nu} \partial_\rho \varepsilon^\mu - W^{\mu\rho} \partial_\rho \varepsilon^\nu] \\ \delta U_{\mu\nu}(x) &= U'_{\mu\nu}(x) - U_{\mu\nu}(x) = -[\varepsilon^\rho \partial_\rho U_{\mu\nu}(x) + U_{\rho\nu} \partial_\mu \varepsilon^\rho + U_{\mu\rho} \partial_\nu \varepsilon^\rho]. \end{aligned} \quad (4.28)$$

很显然，前面推导的度规场在无穷小微分同胚下的变化规律只是这两个式子的一个应用。再比如，对于二阶张量场 $F^\mu{}_\nu(x)$ ，可以得到

$$\delta F^\mu{}_\nu(x) = F'^\mu{}_\nu(x) - F^\mu{}_\nu(x) = -[\varepsilon^\rho \partial_\rho F^\mu{}_\nu - F^\rho{}_\nu \partial_\rho \varepsilon^\mu + F^\mu{}_\rho \partial_\nu \varepsilon^\rho]. \quad (4.29)$$

好了，推导了这么多例子，我想读者应该已经找到规律了，那么当然就能写出下面这个一般性的式子

$$\begin{aligned} \delta T^{\mu_1 \mu_2 \dots \mu_k}{}_{\nu_1 \nu_2 \dots \nu_l}(x) &= T'^{\mu_1 \mu_2 \dots \mu_k}{}_{\nu_1 \nu_2 \dots \nu_l}(x) - T^{\mu_1 \mu_2 \dots \mu_k}{}_{\nu_1 \nu_2 \dots \nu_l}(x) \\ &= -[\varepsilon^\rho \partial_\rho T^{\mu_1 \mu_2 \dots \mu_k}{}_{\nu_1 \nu_2 \dots \nu_l} - T^{\rho \mu_2 \dots \mu_k}{}_{\nu_1 \nu_2 \dots \nu_l} \partial_\rho \varepsilon^{\mu_1} - T^{\mu_1 \rho \dots \mu_k}{}_{\nu_1 \nu_2 \dots \nu_l} \partial_\rho \varepsilon^{\mu_2} - \dots \\ &\quad + T^{\mu_1 \mu_2 \dots \mu_k}{}_{\rho \nu_2 \dots \nu_l} \partial_{\nu_1} \varepsilon^\rho + T^{\mu_1 \mu_2 \dots \mu_k}{}_{\nu_1 \rho \dots \nu_l} \partial_{\nu_2} \varepsilon^\rho + \dots]. \end{aligned} \quad (4.30)$$

由于按照定义 $\delta T^{\mu_1 \mu_2 \dots \mu_k}{}_{\nu_1 \nu_2 \dots \nu_l}(x)$ 为同在 x 点的两个张量之差，所以它也必定为一个张量，这就是**无穷小微分同胚变换的张量性质**。

下面，请读者验证无穷小微分同胚变换的 δ 运算满足莱布尼兹法则。比方说， $W^\mu U_\mu$ 缩并的结果是一个标量场，不难验证有

$$\delta(W^\mu U_\mu) = (\delta W^\mu) U_\mu + W^\mu (\delta U_\mu). \quad (4.31)$$

而如果不缩并，那 $W^\mu U_\nu$ 就是一个 (1,1) 型张量场，同样不难验证

$$\delta(W^\mu U_\nu) = (\delta W^\mu) U_\nu + W^\mu (\delta U_\nu). \quad (4.32)$$

类似的，还有比如

$$\delta(W^\mu U^\nu) = (\delta W^\mu) U^\nu + W^\mu (\delta U^\nu). \quad (4.33)$$

这些莱布尼兹法则对于任意阶张量之间的乘法（包括缩并）都成立！

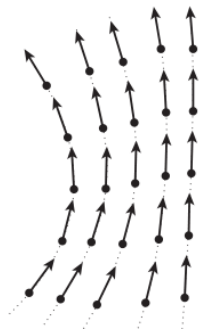


Figure 4.1: 矢量场形成的流线。

无穷小微分同胚映射与李导数

下面介绍数学家常用的理解无穷小微分同胚变换的方法。给定时空中的—个矢量场 $V^\mu(x)$ ，我们可以把它想象成时空点在“流动”形成的速度场，进而可以通过积分下面方程得到这种时空流体的流线

$$\frac{dx^\mu}{d\tau} = V^\mu(x(\tau)), \quad (4.34)$$

如图 (4.1) 所示。进一步，我们可以将这种时空流体的流动过程看成是一种随着参数 τ 连续进行的微分同胚映射。即是说，经过无穷小“时间” $\delta\tau$ 之后， x 点流到了邻近的 x' 点，

$$x'^\mu = x^\mu + \delta\tau V^\mu(x), \quad (4.35)$$

即有无穷小微分同胚

$$\varepsilon^\mu(x) = \delta\tau V^\mu(x). \quad (4.36)$$

因此时空流体的流动就把 x 点的场 $T^{\dots}(x)$ (\dots 表示任意的上指标和任意的下指标) 带到了无限邻近的 x' 点，变成 $T'^{\dots}(x')$ (称作将 x 点的场推前到 x' 点) 然后就可以和 x' 点原本的场 $T^{\dots}(x')$ 进行比较，也就是可以计算 $T^{\dots}(x') - T'^{\dots}(x') = -\delta T^{\dots}(x')$ 。这就是数学上所谓李导数的本质，沿着向量场 V^μ 的李导数记为 \mathcal{L}_V ，其具体定义为

$$\mathcal{L}_V T^{\dots}(x) = \frac{T^{\dots}(x') - T'^{\dots}(x')}{\delta\tau} = -\frac{\delta T^{\dots}}{\delta\tau}. \quad (4.37)$$

或者写作

$$\delta T^{\dots}(x) = -\delta\tau \mathcal{L}_V T^{\dots}(x). \quad (4.38)$$

注意到 $\varepsilon^\mu = \delta\tau V^\mu$ ，因此如果不强调沿着某个特定的矢量场 V^μ 作李导数的话，我们也可以将这个式子写成

$$\delta T^{\dots}(x) = -\mathcal{L}_\varepsilon T^{\dots}(x). \quad (4.39)$$

很显然，由于无穷小微分同胚变换的张量性质，所以任何张量场求李导数的结果必定依然是一个张量场，即 $\mathcal{L}_V T^{\dots}(x)$ 依然为张量场。

由李导数的这个定义，加上前面关于无穷小微分同胚变换的推导，很容易得到

$$\begin{aligned}\mathcal{L}_V W^\mu &= V^\rho \partial_\rho W^\mu - W^\rho \partial_\rho V^\mu, \\ \mathcal{L}_V U_\mu &= V^\rho \partial_\rho U_\mu + U_\rho \partial_\mu V^\rho.\end{aligned}\quad (4.40)$$

以及比方说对于二阶张量场，有

$$\begin{aligned}\mathcal{L}_V W^{\mu\nu}(x) &= V^\rho \partial_\rho W^{\mu\nu} - W^{\rho\nu} \partial_\rho V^\mu - W^{\mu\rho} \partial_\rho V^\nu \\ \mathcal{L}_V U_{\mu\nu}(x) &= V^\rho \partial_\rho U_{\mu\nu}(x) + U_{\rho\nu} \partial_\mu V^\rho + U_{\mu\rho} \partial_\nu V^\rho.\end{aligned}\quad (4.41)$$

诸如此类等等，总之，前面的所有结果都可以平行地移过来。特别的，李导数满足莱布尼兹法则！

特别的，我们有

$$\begin{aligned}\mathcal{L}_V T^{\mu_1 \mu_2 \dots \mu_k}_{\nu_1 \nu_2 \dots \nu_l}(x) &= V^\rho \partial_\rho T^{\mu_1 \mu_2 \dots \mu_k}_{\nu_1 \nu_2 \dots \nu_l} \\ &- T^{\rho \mu_2 \dots \mu_k}_{\nu_1 \nu_2 \dots \nu_l} \partial_\rho V^{\mu_1} - T^{\mu_1 \rho \dots \mu_k}_{\nu_1 \nu_2 \dots \nu_l} \partial_\rho V^{\mu_2} - \dots \\ &+ T^{\mu_1 \mu_2 \dots \mu_k}_{\rho \nu_2 \dots \nu_l} \partial_{\nu_1} V^\rho + T^{\mu_1 \mu_2 \dots \mu_k}_{\nu_1 \rho \dots \nu_l} \partial_{\nu_2} V^\rho + \dots\end{aligned}\quad (4.42)$$

由此容易验证

$$\mathcal{L}_{aV+bW} T^{\dots} = a \mathcal{L}_V T^{\dots} + b \mathcal{L}_W T^{\dots}, \quad (4.43)$$

式中 V, W 为两个矢量场， a, b 为两个任意的实常数。

4.3 协变微分

前面引入了张量场的概念，为了满足广义协变原理，我们希望物理方程中的各量都是张量。另一方面，一个物理方程中免不了出现物理量对时空坐标的导数，物理量是张量，然而正如我们将要看到的，一个张量场对时空坐标求导的结果就不再是一个张量场了，因为简单求导的结果是不协变的，因此这就破坏了广义协变原理。为了使广义协变原理始终成立，这一节我们将对时空导数的概念进行修改，引入称之为协变导数的概念。**协变导数依然可以看成是一种导数，它关于被求导量也是线性，并且同样满足求导的莱布尼兹法则。**引入协变导数的好处就在于，一个张量场的协变导数结果依然是一个张量场，即满足协变性。

为了看清楚张量场对时空坐标的简单导数不满足协变性，我们考察一个 $(1,0)$ 型张量场 $W^\mu(x)$ ，在坐标变换下，它的变换关系为

$$W'^\mu(x') = S^\mu_{\nu}(x) W^\nu(x). \quad (4.44)$$

在坐标变换下 $W^\mu(x)$ 的导数 $\partial_\lambda W^\mu(x)$ 应该变换为 $\partial'_\lambda W'^\mu(x')$ ，但是

$$\begin{aligned}\partial'_\lambda W'^\mu(x') &= \frac{\partial x^\rho}{\partial x'^\lambda} \frac{\partial}{\partial x^\rho} (S^\mu_{\nu}(x) W^\nu(x)) \\ &= (S^{-1})^\rho_{\lambda} S^\mu_{\nu} \partial_\rho W^\nu + ((S^{-1})^\rho_{\lambda} \partial_\rho S^\mu_{\nu}) W^\nu.\end{aligned}\quad (4.45)$$

显然, 这不满足张量在坐标变换下的协变规则, 它多出了第二行的第二项。

为了消去坐标变换下多出来的这一项, 我们在 $\partial_\lambda W^\mu(x)$ 上添加上一个和 W^ν 成线性的项 $\tilde{\Gamma}_{\lambda\nu}^\mu W^\nu$ ($\tilde{\Gamma}_{\lambda\nu}^\mu$ 为线性系数), 记结果为 $D_\lambda W^\mu$, 称为 W^μ 的协变导数,

$$D_\lambda W^\mu = \partial_\lambda W^\mu + \tilde{\Gamma}_{\lambda\nu}^\mu W^\nu. \quad (4.46)$$

我们要求 $D_\lambda W^\mu$ 满足张量的协变性, 即在坐标变换下按下式变换

$$\begin{aligned} D'_\lambda W'^\mu &= (S^{-1})^\rho{}_\lambda S^\mu{}_\nu D_\rho W^\nu \\ &= (S^{-1})^\rho{}_\lambda S^\mu{}_\nu \partial_\rho W^\nu + (S^{-1})^\rho{}_\lambda S^\mu{}_\nu \tilde{\Gamma}_{\rho\sigma}^\nu W^\sigma. \end{aligned} \quad (4.47)$$

由于

$$\begin{aligned} D'_\lambda W'^\mu &= \partial'_\lambda W'^\mu + \tilde{\Gamma}'^{\mu}_{\lambda\nu} W'^\nu \\ &= (S^{-1})^\rho{}_\lambda S^\mu{}_\nu \partial_\rho W^\nu + ((S^{-1})^\rho{}_\lambda \partial_\rho S^\mu{}_\nu) W^\nu + \tilde{\Gamma}'^{\mu}_{\lambda\nu} S^\nu{}_\sigma W^\sigma. \end{aligned} \quad (4.48)$$

因此, 为了满足 (4.47) 式, 必有

$$(S^{-1})^\rho{}_\lambda S^\mu{}_\nu \tilde{\Gamma}_{\rho\sigma}^\nu = (S^{-1})^\rho{}_\lambda \partial_\rho S^\mu{}_\sigma + \tilde{\Gamma}'^{\mu}_{\lambda\nu} S^\nu{}_\sigma \quad (4.49)$$

由此易得

$$\tilde{\Gamma}'^{\mu}_{\lambda\kappa} = (S^{-1})^\rho{}_\lambda (S^{-1})^\sigma{}_\kappa S^\mu{}_\nu \tilde{\Gamma}_{\rho\sigma}^\nu - (S^{-1})^\rho{}_\lambda (S^{-1})^\sigma{}_\kappa \partial_\rho S^\mu{}_\sigma. \quad (4.50)$$

人们称满足上式变换关系的线性系数 $\tilde{\Gamma}'^{\mu}_{\lambda\kappa}$ 为仿射联络。不难发现, 上式的变换关系与克里斯托夫联络的变换关系 (4.17) 完全相同。所以, 仿射联络 $\tilde{\Gamma}'^{\mu}_{\lambda\kappa}$ 是存在的, 因为它至少可以取作克里斯托夫联络 $\Gamma^\mu_{\lambda\kappa}$ 。实际上, 满足变换关系 (4.50) 的仿射联络不仅仅存在, 而且通常不唯一。仿射联络本身不是张量, 因为在坐标变换下它的变换关系中额外多出了一项。但是, 不难证明, **两个不同仿射联络的差却是一个 (1,2) 型张量**, 因为在坐标变换下两者变换关系中额外多出的项相互抵消了!

注意, 克里斯托夫联络关于两个下指标是对称的, 但是更一般的仿射联络可没有这个性质。

下面我们把协变导数推广到更一般的张量场。在这里关键是要利用协变导数满足莱布尼兹法则的性质。另外, 我们还要注意到, 对于一个标量场的协变导数就等于普通导数 (因为标量场的普通导数本身已经是协变的了), 即

$$D_\mu \Phi = \partial_\mu \Phi, \quad (4.51)$$

其中 Φ 为标量场。

比方说, 为了得到一个下指标矢量场 U_μ 的协变导数, 我们只要注意到 $U_\mu W^\mu$ 是一个标量场, 从而

$$D_\rho (U_\mu W^\mu) = \partial_\rho (U_\mu W^\mu). \quad (4.52)$$

另一方面，利用协变导数的莱布尼兹法则，有 $D_\rho(U_\mu W^\mu) = (D_\rho U_\mu)W^\mu + U_\mu(D_\rho W^\mu)$ 。由此不难验证，仅当 $D_\rho U_\mu$ 取如下形式时，(4.52) 式才能得以满足，

$$D_\rho U_\mu = \partial_\rho U_\mu - \tilde{\Gamma}_{\rho\mu}^\sigma U_\sigma. \quad (4.53)$$

注意，与 (4.46) 式不同，这里等式右边的第二项是一个减号。

为了得到更一般的张量，比方说 $T_\rho^{\mu\nu}$ 的协变导数，我们只需注意到 $T_\rho^{\mu\nu} W^\rho U_\mu V_\nu$ 是一个标量场，从而不难验证为了满足

$$D_\lambda(T_\rho^{\mu\nu} W^\rho U_\mu V_\nu) = \partial_\lambda(T_\rho^{\mu\nu} W^\rho U_\mu V_\nu), \quad (4.54)$$

而且同时满足协变导数的莱布尼兹法则，人们就必须取

$$D_\lambda T_\rho^{\mu\nu} = \partial_\lambda T_\rho^{\mu\nu} + \tilde{\Gamma}_{\lambda\sigma}^\mu T_\rho^{\sigma\nu} + \tilde{\Gamma}_{\lambda\sigma}^\nu T_\rho^{\mu\sigma} - \tilde{\Gamma}_{\lambda\rho}^\sigma T_\sigma^{\mu\nu}. \quad (4.55)$$

按照类似的办法，人们不难得到任意一个张量场的协变导数公式。

联络和协变导数也是规范场论中的基本概念，不过那里的联络和协变导数通常都是作用在取值于内部空间的向量场上，而广义相对论中的则是作用在取时空指标的张量场上。用数学家的话来说即是，规范场论的联络和协变导数通常定义在一般的向量丛上，而广义相对论中的联络和协变导数则是定义在切丛、余切丛以及它们的各种张量积上。

下面我们讨论沿着时空中一条曲线 $x^\mu(\tau)$ 的协变微分。假设时空中有矢量场 $A^\mu(x)$ ，我们将它限制在曲线 $x^\mu(\tau)$ 上，得到沿着曲线定义的矢量 $A^\mu(\tau)$ ，我们定义矢量 $A^\mu(\tau)$ 沿着曲线的协变微分 $\frac{DA^\mu}{D\tau}$ 为

$$\begin{aligned} \frac{DA^\mu}{D\tau} &= \frac{dx^\nu}{d\tau} D_\nu A^\mu = \frac{dx^\nu}{d\tau} (\partial_\nu A^\mu + \tilde{\Gamma}_{\nu\rho}^\mu A^\rho) \\ &= \frac{dA^\mu}{d\tau} + \tilde{\Gamma}_{\nu\rho}^\mu \frac{dx^\nu}{d\tau} A^\rho. \end{aligned} \quad (4.56)$$

显然，最后的结果并不需要 A^μ 在整个时空上有定义，只需要它沿着曲线 $x^\mu(\tau)$ 有定义就行。因此，对于任意沿着曲线 $x^\mu(\tau)$ 有定义的矢量 $A^\mu(\tau)$ ，我们可以直接定义它沿着曲线的协变微分为

$$\frac{DA^\mu}{D\tau} \equiv \frac{dA^\mu}{d\tau} + \tilde{\Gamma}_{\nu\rho}^\mu \frac{dx^\nu}{d\tau} A^\rho. \quad (4.57)$$

从引入这个定义的过程不难看出， $\frac{DA^\mu}{D\tau}$ 依然是一个矢量，也即是说在坐标变换下，它像一个矢量那样变换，

$$\frac{DA'^\mu}{D\tau} = \frac{\partial x'^\mu}{\partial x^\nu} \frac{DA^\nu}{D\tau}. \quad (4.58)$$

类似的，可以将协变矢量 $B_\mu(\tau)$ 沿着曲线 $x^\mu(\tau)$ 的协变微分定义成

$$\frac{DB_\mu}{D\tau} \equiv \frac{dB_\mu}{d\tau} - \tilde{\Gamma}_{\nu\mu}^\rho \frac{dx^\nu}{d\tau} B_\rho. \quad (4.59)$$

当然, $\frac{DB_\mu}{D\tau}$ 依然是一个协变矢量。同样的, 也可以定义任意张量 T 沿着曲线 $x^\mu(\tau)$ 的协变微分, 比如对于 (1,1) 张量 $T^\mu_\nu(\tau)$, 可以定义

$$\frac{DT^\mu_\nu}{D\tau} \equiv \frac{dT^\mu_\nu}{d\tau} + \tilde{\Gamma}^\mu_{\rho\lambda} \frac{dx^\rho}{d\tau} T^\lambda_\nu - \tilde{\Gamma}^\lambda_{\rho\nu} \frac{dx^\rho}{d\tau} T^\mu_\lambda. \quad (4.60)$$

当然, $\frac{DT^\mu_\nu}{D\tau}$ 依然是一个 (1,1) 张量。

特别的,

$$\frac{DA^\mu}{D\tau} = 0, \quad (4.61)$$

是一个协变方程, 因此它只要在一个坐标系中成立, 则在任意坐标系中均成立。通常称满足上述方程的 $A^\mu(\tau)$ 为矢量 A^μ 沿着曲线 $x^\mu(\tau)$ 的**平行移动**。上述方程也就称作矢量 A^μ 沿着曲线的平行移动方程, 很显然, 只要 $A^\mu(\tau)$ 在某个初始 $\tau=0$ 的值给定, 则通过求解平行移动方程, $A^\mu(\tau)$ 沿着整根曲线 $x^\mu(\tau)$ 的值就都给定了。完全类似的, 任何张量都可以用平行移动在一条曲线上定义, 只要这个张量沿着该曲线的协变微分等于零。

4.4 等效原理和微分同胚不变性

前面说过, 广义坐标变换在主动观点中可以看成是局部微分同胚映射。因此根据这个主动的观点, 我们也可以把**结合最小作用量原理的广义协变原理表述成: 在广义相对论中, 物理系统的作用量必须在任意 $x \rightarrow x'$ 的微分同胚映射下保持不变, 称之为微分同胚不变性。尤其是, 引力场本身的作用量必须具有微分同胚不变性。**

4.4.1 最小耦合原理

为了满足微分同胚不变性, 物理系统的作用量就只能由各种张量场以及它们的协变导数来构造。换言之, 物理系统的拉格朗日密度 \mathcal{L} 必然是由各种张量场以及它们的协变导数所构造出来的一个标量函数。这些张量场中既包括与引力耦合的物质场, 当然也包括刻画引力场的度规场 $g_{\mu\nu}(x)$ 。除了度规场以外, 协变导数本身还额外引入了一个场, 即仿射联络场 $\tilde{\Gamma}^\rho_{\mu\nu}$, 当然这个场本身并不是张量场。平直的闵可夫斯基时空当然不需要仿射联络场, 所以看起来似乎是, 仿射联络场 $\tilde{\Gamma}^\rho_{\mu\nu}$ 和度规场一样都是刻画引力场的。

但是, 如果在微分同胚不变性的基础上进一步应用等效原理, 就可以得出, **广义相对论中的仿射联络必须取克里斯托夫联络, 即必有 $\tilde{\Gamma}^\rho_{\mu\nu} = \Gamma^\rho_{\mu\nu}$** 。换言之, 广义相对论中的仿射联络完全由度规场决定, 从而刻画引力的唯一只有度规场。

具体的推理过程如下: 首先, 根据等效原理, 在局部惯性系中, 引力场被惯性力抵消了, 从而在局部惯性系中, 必有仿射联络等于零, 即 $\tilde{\Gamma}^\rho_{\mu\nu} = 0$ 。换言之, 在局部惯性系中, 协变导数 D_μ 就退化成通常的导数 ∂_μ 。另一方面, 在局部惯性系中, $g_{\mu\nu} = \eta_{\mu\nu}$, 根据克里斯托夫联络的表达式易知, 这时也必有克里斯托夫联络等于零, 即 $\Gamma^\rho_{\mu\nu} = 0$ 。从而可知, 在局部惯性系中 $T^\rho_{\mu\nu} = \tilde{\Gamma}^\rho_{\mu\nu} - \Gamma^\rho_{\mu\nu} = 0$ 。但是, 前面说过, 两个不同联络的差是一个张量, 从而 $T^\rho_{\mu\nu}$ 是一个张量。而一个张量在某个坐标系中等于零, 则在任何坐标系中都等于零, 从而可知 $T^\rho_{\mu\nu}$ 恒等于零。也即是说, 在任意坐标系中, 恒有 $\tilde{\Gamma}^\rho_{\mu\nu} = \Gamma^\rho_{\mu\nu}$ 。

因此，在广义相对论中的平行移动方程是

$$\frac{DA^\mu}{D\tau} = \frac{dA^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu \frac{dx^\nu}{d\tau} A^\rho = 0. \quad (4.62)$$

另一方面，当把参数 τ 取作固有时我们就可以定义四维速度 u^μ

$$u^\mu = \frac{dx^\mu}{d\tau}, \quad (4.63)$$

根据固有时间的方程 $-d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu$ ，我们有

$$g_{\mu\nu} u^\mu u^\nu = -1 \Rightarrow u^\mu u_\mu = -1. \quad (4.64)$$

进而当曲线参数 τ 是固有时时，我们就可以把平行移动方程重写为

$$\frac{DA^\mu}{D\tau} = \frac{dA^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu u^\nu A^\rho = 0. \quad (4.65)$$

特别的，四维速度本身的平行移动方程是，

$$\begin{aligned} \frac{Du^\mu}{D\tau} &= \frac{du^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu u^\nu u^\rho = 0 \\ \Leftrightarrow \frac{Du^\mu}{D\tau} &= \frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\nu\rho}^\mu \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0. \end{aligned} \quad (4.66)$$

结果正是引力场中自由下落粒子的测地线方程。所以，测地线方程其实就是粒子四维速度的平行移动方程。

从本书第三章的相关结论我们知道，克里斯托夫联络满足如下等式

$$\partial_\rho g_{\mu\nu} - g_{\mu\sigma} \Gamma_{\rho\nu}^\sigma - g_{\sigma\nu} \Gamma_{\rho\mu}^\sigma = 0. \quad (4.67)$$

在广义相对论中，结合协变导数的定义可知，上式可以重写成

$$D_\rho g_{\mu\nu} = 0. \quad (4.68)$$

通常称这个结果为协变导数与度规场的相容性条件。不是任何协变导数都满足这个相容性条件，但是广义相对论中的协变导数自动满足这个相容性条件。实际上，根据第三章的相关讨论也不难知道，相容性条件再加上联络是对称联络的要求，反过来也唯一决定了广义相对论中的联络为克里斯托夫联络。

另一种得到协变导数与度规场相容性条件的方法是，先取局部惯性系，注意在局部惯性系的原点，联络为零，从而即有 $D_\rho g_{\mu\nu} = \partial_\rho \eta_{\mu\nu} = 0$ 。但是 $D_\rho g_{\mu\nu}$ 是协变的，它在局部惯性系中为零，则在任何坐标系中都恒为零，从而即有相容性条件。完全类似于这个推理，我们也不难得到

$$D_\rho g^{\mu\nu} = 0, \quad D_\rho \delta_\nu^\mu = 0. \quad (4.69)$$

等效原理加上广义协变原理就基本上决定了物质场（即除引力场之外其他一切产生引力的物质的场）与引力场的相互作用形式，或者说决定了物质场与引力场耦合的形式。为

了看清楚这一点, 我们首先可以选取一个局部惯性系, 由于在局部惯性系的原点引力场完全被抵消了, 因此可知, 物质场的拉格朗日密度在这一点上必定与平直的闵可夫斯基时空中的形式完全相同, 特别的, 这时候拉格朗日密度中物质场对时空的导数必定就是普通的偏导 ∂_μ 。然后, 我们从局部惯性系坐标变换到任意坐标系, 根据广义协变性, 这时候物质场作用量中的时空度规 $\eta_{\mu\nu}$ 就应该替换成更一般的 $g_{\mu\nu}$, 同样, 为了满足广义协变性, 物质场作用量中的偏导 ∂_μ 就应该替换成协变导数 D_μ 。另一方面, 上文我们已经看到, 广义相对论中的协变导数取决于克里斯托夫联络, 而它完全由度规场 $g_{\mu\nu}$ 决定, 因此这种替换手续就决定了物质场与描写引力的度规场 $g_{\mu\nu}$ 的耦合形式。人们通常称这种由平直的闵可夫斯基时空过渡到弯曲时空的替换手续为物质场与引力场的**最小耦合原理**。很显然, 最小耦合原理基本上可以看作是等效原理和广义协变原理的一个推论。

4.4.2 微分同胚不变性与能量动量张量

无穷小微分同胚再研究

前面已经看到, 在无穷小局域微分同胚 $x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x)$ 之下, 张量场 $T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l}(x)$ 的改变量为

$$\begin{aligned} & \delta T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l}(x) \\ &= - \left[\varepsilon^\rho \partial_\rho T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l} - T^{\rho\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l} \partial_\rho \varepsilon^{\mu_1} - T^{\mu_1\rho\dots\mu_k}_{\nu_1\nu_2\dots\nu_l} \partial_\rho \varepsilon^{\mu_2} - \dots \right. \\ & \quad \left. + T^{\mu_1\mu_2\dots\mu_k}_{\rho\nu_2\dots\nu_l} \partial_{\nu_1} \varepsilon^\rho + T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\rho\dots\nu_l} \partial_{\nu_2} \varepsilon^\rho + \dots \right]. \end{aligned} \quad (4.70)$$

但是由于广义相对论中的联络为克里斯托夫联络, 而克里斯托夫联络是一个对称联络, 满足 $\Gamma_{\mu\nu}^\rho = \Gamma_{\nu\mu}^\rho$, 由此不难直接证明, 可以将上面 $\delta T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l}(x)$ 中的普通导数通通改成协变导数, 即有

$$\begin{aligned} & \delta T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l}(x) \\ &= - \left[\varepsilon^\rho D_\rho T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l} - T^{\rho\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l} D_\rho \varepsilon^{\mu_1} - T^{\mu_1\rho\dots\mu_k}_{\nu_1\nu_2\dots\nu_l} D_\rho \varepsilon^{\mu_2} - \dots \right. \\ & \quad \left. + T^{\mu_1\mu_2\dots\mu_k}_{\rho\nu_2\dots\nu_l} D_{\nu_1} \varepsilon^\rho + T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\rho\dots\nu_l} D_{\nu_2} \varepsilon^\rho + \dots \right]. \end{aligned} \quad (4.71)$$

同样的, 对于李导数, 我们也有

$$\begin{aligned} & \mathcal{L}_V T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l}(x) = V^\rho D_\rho T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l} \\ & - T^{\rho\mu_2\dots\mu_k}_{\nu_1\nu_2\dots\nu_l} D_\rho V^{\mu_1} - T^{\mu_1\rho\dots\mu_k}_{\nu_1\nu_2\dots\nu_l} D_\rho V^{\mu_2} - \dots \\ & + T^{\mu_1\mu_2\dots\mu_k}_{\rho\nu_2\dots\nu_l} D_{\nu_1} V^\rho + T^{\mu_1\mu_2\dots\mu_k}_{\nu_1\rho\dots\nu_l} D_{\nu_2} V^\rho + \dots \end{aligned} \quad (4.72)$$

特别的, 我们可以将度规场在无穷小微分同胚下的改变量 $\delta g_{\mu\nu}$ 重写成

$$\delta g_{\mu\nu} = -[\varepsilon^\rho D_\rho g_{\mu\nu} + g_{\rho\nu} D_\mu \varepsilon^\rho + g_{\mu\rho} D_\nu \varepsilon^\rho]. \quad (4.73)$$

进一步利用度规场与协变导数的相容性条件, 即有

$$\delta g_{\mu\nu} = -[D_\mu \varepsilon_\nu + D_\nu \varepsilon_\mu]. \quad (4.74)$$

完全类似的，也有

$$\delta g^{\mu\nu} = D^\mu \varepsilon^\nu + D^\nu \varepsilon^\mu. \quad (4.75)$$

同样，我们也可以将度规场的李导数重写成

$$\mathcal{L}_\varepsilon g_{\mu\nu} = D_\mu \varepsilon_\nu + D_\nu \varepsilon_\mu. \quad (4.76)$$

能量动量张量

下面我们来考察最小耦合原理和微分同胚不变性的一个重要推论。为此，我们先回顾一下本书第二章讲述诺特定理和能量-动量张量时的一个结论。在第二章中，我们说过，在平直的闵可夫斯基时空中，在如下无穷小局域时空变换下，

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x), \quad (4.77)$$

任何物质系统作用量 S_m 的改变量必定可以写成

$$\delta S_m = \int dv T^{\mu\nu}(\eta) \partial_\mu \varepsilon_\nu(x) = \frac{1}{2} \int dv T^{\mu\nu}(\eta) (\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu). \quad (4.78)$$

其中 $T^{\mu\nu}(\eta)$ 是一个对称张量，称之为能量-动量张量， η 表示闵可夫斯基时空的度规张量 $\eta_{\mu\nu}$ ， $T^{\mu\nu}(\eta)$ 的完整含义是定义在闵可夫斯基时空的能量动量张量，这是我们在第二章中引入的。式中 dv 为四维时空体积元，对于闵可夫斯基时空 $dv = d^4x$ 。因此，根据最小耦合原理的替换规则，可以得到：在任意弯曲时空中，在无穷小局域微分同胚 $x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x)$ 之下，如果不考虑度规场 $g_{\mu\nu}$ 在微分同胚下的改变，而仅仅只考虑物质场的微分同胚变换，则任何物质系统作用量 S_m 的改变量必定可以写成

$$\delta S_m = \frac{1}{2} \int dv T^{\mu\nu}(g) (D_\mu \varepsilon_\nu + D_\nu \varepsilon_\mu), \quad (4.79)$$

其中 $T^{\mu\nu}(g)$ 是 $T^{\mu\nu}(\eta)$ 经过最小耦合原理的替换规则后得到的弯曲时空能量动量张量。式中 dv 表示弯曲时空的四维体积元，它和平直时空的体积元 d^4x 有所差别， dv 的具体形式我们稍后给出。当然，平直时空中的能动量守恒方程，现在就应该替换成

$$\partial_\mu T^{\mu\nu} = 0 \Rightarrow D_\mu T^{\mu\nu} = 0. \quad (4.80)$$

另一方面，由于和引力场耦合，所以弯曲时空中的物质场作用量 S_m 当然与度规场 $g_{\mu\nu}$ 有关，不妨将物质场作用量对 $g_{\mu\nu}$ 的变分记为如下形式

$$\delta S_m = \frac{1}{2} \int dv \Theta^{\mu\nu} \delta g_{\mu\nu}, \quad (4.81)$$

式中 $\Theta^{\mu\nu}$ 为某个对称的二阶张量。下面我们将会证明， $\Theta^{\mu\nu}$ 其实就等于物质场的能量动量张量 $T^{\mu\nu}$ ，即

$$\Theta^{\mu\nu} = T^{\mu\nu}. \quad (4.82)$$

为了证明这一点，我们需要用到物质场作用量的微分同胚不变性。为此，考察无穷小微分同胚 $x^\mu \rightarrow x^\mu + \varepsilon^\mu(x)$ ，由于在此微分同胚下**物质场和度规场都要发生改变**，从而物质场作用量的改变量必为

$$\begin{aligned}\delta S_m &= \frac{1}{2} \int dv \Theta^{\mu\nu} \delta g_{\mu\nu} + \frac{1}{2} \int dv T^{\mu\nu} (D_\mu \varepsilon_\nu + D_\nu \varepsilon_\mu) \\ &= \frac{1}{2} \int dv [(-\Theta^{\mu\nu} + T^{\mu\nu})(D_\mu \varepsilon_\nu + D_\nu \varepsilon_\mu)],\end{aligned}\quad (4.83)$$

式中我们已经代入了 (4.74) 式的结果。进一步根据物质场作用量的微分同胚不变性，必有 $\delta S_m = 0$ ，从而必有

$$\delta S_m = 0 \Rightarrow \Theta^{\mu\nu} = T^{\mu\nu}.\quad (4.84)$$

综上，我们即得到物质场能量动量张量的一个等价定义，即：物质场能动量张量等于弯曲时空中物质场作用量对度规张量 $g_{\mu\nu}$ 的泛函导数，即

$$\delta S_m = \frac{1}{2} \int dv T^{\mu\nu} \delta g_{\mu\nu}.\quad (4.85)$$

值得强调的是，物质场作用量的微分同胚不变性并不要求场位形满足场的运动微分方程（场方程），但是，能动量张量的守恒方程 $D_\mu T^{\mu\nu} = 0$ 要求其中的物质场满足场方程。当一个场位形满足场方程时，人们也称之为在壳的 (On Shell)，而对于不满足场方程的场位形，人们就称之为离壳的 (Off Shell)。微分同胚不变性并不要求物质场在壳，但是能动量守恒要求物质场在壳。

物质场的微分同胚不变性即使场离壳也成立！但是，**当物质场在壳时，微分同胚不变性本身就能推论出能动量张量的守恒方程**。为了看清楚这一点，我们不妨将物质场抽象地记作 Φ ，将它在无穷小微分同胚下的改变量记作 $\delta\Phi$ 。在无穷小微分同胚之下，由于物质场和度规场都要变换，所以物质场作用量的改变量为

$$\begin{aligned}\delta S_m &= \frac{1}{2} \int dv T^{\mu\nu} \delta g_{\mu\nu} + \int d^4x \frac{\delta S_m}{\delta \Phi} \delta \Phi(x) \\ &= -\frac{1}{2} \int dv T^{\mu\nu} (D_\mu \varepsilon_\nu + D_\nu \varepsilon_\mu) + \int d^4x \frac{\delta S_m}{\delta \Phi} \delta \Phi(x).\end{aligned}\quad (4.86)$$

下面假设物质场在壳，从而根据最小作用量原理，必有场方程

$$\frac{\delta S_m}{\delta \Phi} = 0.\quad (4.87)$$

因此，这时候物质场作用量在无穷小微分同胚之下的改变量就等于

$$\delta S_m = -\frac{1}{2} \int dv T^{\mu\nu} (D_\mu \varepsilon_\nu + D_\nu \varepsilon_\mu) = - \int dv T^{\mu\nu} D_\mu \varepsilon_\nu.\quad (4.88)$$

利用 $D_\mu (T^{\mu\nu} \varepsilon_\nu) = (D_\mu T^{\mu\nu}) \varepsilon_\nu + T^{\mu\nu} D_\mu \varepsilon_\nu$ 进行分部积分，再利用高斯定理 $\int dv D_\mu A^\mu = \int_\infty dS_\mu A^\mu$ ，并假设在无穷远边界上 $\varepsilon_\nu = 0$ ，从而即有

$$\delta S_m = \int dv (D_\mu T^{\mu\nu}) \varepsilon_\nu.\quad (4.89)$$

当然，物质场在壳时，作用量同样会有微分同胚不变性，也即 $\delta S_m = 0$ ，由此就可以得出

$$D_\mu T^{\mu\nu} = 0.\quad (4.90)$$

正是能动量张量守恒方程。

四维时空体积元

下面我们给出四维时空体积元 dv 的具体形式。首先，我们注意到四维体积元一定是不依赖于特定坐标系的，即是说 dv 的表达式得在广义坐标变换之下保持不变。

其次，根据多元微积分知识可以知道，在广义坐标变换 $x^\mu \rightarrow x'^\mu$ 之下，

$$d^4x' = \left| \frac{\partial x'}{\partial x} \right| d^4x, \quad (4.91)$$

式中 $\left| \frac{\partial x'}{\partial x} \right|$ 为坐标变换的雅可比行列式，即 $\left| \frac{\partial x'}{\partial x} \right|$ 为 4×4 矩阵 $S^\mu{}_\nu = \frac{\partial x'^\mu}{\partial x^\nu}$ 的行列式。即是说， d^4x 在广义坐标变换之下并不能保持不变，从而它并不能作为四维弯曲时空的体积元。当然，如果我们考察的是平直的闵可夫斯基时空，那由于洛伦兹变换的行列式等于 1，所以这时候 d^4x 就是洛伦兹不变的，从而就可以作为平直时空的体积元。

为了找到一个广义坐标变换不变的四维体积元，我们注意到度规场在广义坐标变换之下按照下式变换，

$$g'_{\mu\nu} = \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}. \quad (4.92)$$

从而不难看到， 4×4 度规矩阵的行列式 $g = \det(g_{\mu\nu})$ 在坐标变换之下将按如下规则变换

$$g'(x') = \left| \frac{\partial x'}{\partial x} \right|^{-2} g(x). \quad (4.93)$$

再注意到度规矩阵 $g_{\mu\nu}$ 为闵氏符号的，即它有三个正本征值，一个负本征值，从而不难知道行列式 g 必为负数。进而根据 (4.91) 式和 (4.93) 式不难看出， $\sqrt{-g}d^4x$ 是广义坐标变换不变的，即有

$$\sqrt{-g'}d^4x' = \sqrt{-g}d^4x. \quad (4.94)$$

从而四维弯曲时空的体积元就可以定义为 $\sqrt{-g}d^4x$ ，即

$$dv = \sqrt{-g}d^4x. \quad (4.95)$$

特别的，对于闵可夫斯基时空，则有 $g = -1$ ，从而即有 $dv = d^4x$ ，正好和通常的体积元定义一致。

能动量张量举例

下面列举几个物质系统的能动量张量。

例 1，引力场中的多粒子体系

第一个例子是引力场中的多粒子体系。这个物质系统的作用量为

$$S_m = - \sum_n m_n \int d\tau_n = - \sum_n m_n \int d\tau_n \sqrt{-g_{\mu\nu}(x_n)} \frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n}. \quad (4.96)$$

式中 x_n^μ 为粒子 n 的时空坐标， τ_n 为它的固有时。

不难计算 S_m 对度规张量的变分, 为

$$\begin{aligned}\delta S_m &= -\sum_n m_n \int d\tau_n \delta \sqrt{-g_{\mu\nu}(x_n) \frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n}} \\ &= \sum_n m_n \int d\tau_n \frac{1}{2} \frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n} \delta g_{\mu\nu}(x_n) \\ &= \frac{1}{2} \int \sqrt{-g} d^4x \left[\frac{1}{\sqrt{-g}} \sum_n m_n \int d\tau_n \frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n} \delta^4(x - x_n(\tau_n)) \right] \delta g_{\mu\nu}(x).\end{aligned}$$

式中第二行我们利用了 $\sqrt{-g_{\mu\nu}(x_n) \frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n}} = 1$ 。将上面的结果与能动量张量的定义式 (4.85) 进行比较, 不难得到

$$T^{\mu\nu}(x) = \frac{1}{\sqrt{-g}} \sum_n m_n \int d\tau_n \frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n} \delta^4(x - x_n(\tau_n)). \quad (4.97)$$

与第二章中的相关结果比较不难发现, 这正是弯曲时空中多粒子体系的能动量张量。

例 2, 弯曲时空中的标量场

我们要考察的第二个例子是弯曲时空中的标量场。作为场论系统, 这个物质系统的作用量必定可以写成如下形式

$$S_m = \int dv \mathcal{L} = \int \sqrt{-g} d^4x \mathcal{L}, \quad (4.98)$$

式中 \mathcal{L} 表示拉格朗日密度。对于弯曲时空中的标量场 ϕ , 根据最小耦合原理, 我们可以将相应的拉格朗日密度取成

$$\mathcal{L} = -\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - \mathcal{U}(\phi). \quad (4.99)$$

为了计算物质场作用量对度规张量的变分, 我们需要用到如下引理。**引理:** 对于任何矩阵 M , 总有

$$\text{Tr}(M^{-1} \delta M) = \delta \ln \det M. \quad (4.100)$$

式中 $\det M$ 表示矩阵 M 的行列式, Tr 代表矩阵求迹, 即对角元之和。

上面引理的证明非常直接了当,

$$\begin{aligned}\delta \ln \det M &= \ln \det(M + \delta M) - \ln \det M \\ &= \ln \frac{\det(M + \delta M)}{\det M} = \ln \det(M^{-1}(M + \delta M)) \\ &= \ln \det(1 + M^{-1} \delta M) \\ &= \ln(1 + \text{Tr}(M^{-1} \delta M)) = \text{Tr}(M^{-1} \delta M).\end{aligned} \quad (4.101)$$

后面两行的等号都是表示在忽略高阶无穷小的意义上相等。

取引理中的矩阵 M 为度规矩阵 $g_{\mu\nu}$, 则根据引理不难得到

$$\delta g = g g^{\mu\nu} \delta g_{\mu\nu}. \quad (4.102)$$

另外，将矩阵方程 $g^{\mu\nu}g_{\nu\rho} = \delta_{\rho}^{\mu}$ 两边微分，不难得到 $(\delta g^{\mu\nu})g_{\nu\rho} + g^{\mu\nu}\delta g_{\nu\rho} = 0$ ，由此即得

$$\delta g^{\mu\nu} = -g^{\mu\rho}\delta g_{\rho\sigma}g^{\sigma\nu}. \quad (4.103)$$

当然，这实际上就是如下矩阵方程的一个简单例子

$$\delta M^{-1} = -M^{-1}(\delta M)M^{-1}, \quad (4.104)$$

式中 M 表示任何可逆矩阵。

利用上一段的两个结果，不难计算物质场作用量 S_m 对度规场的变分

$$\begin{aligned} \delta S_m &= \int d^4x(\delta\sqrt{-g})\mathcal{L} + \int \sqrt{-g}d^4x\delta\mathcal{L} \\ &= \int \sqrt{-g}d^4x\left(\frac{1}{2}g^{\mu\nu}\mathcal{L}\delta g_{\mu\nu} + \delta\mathcal{L}\right). \end{aligned} \quad (4.105)$$

对于前面给出的标量场拉格朗日密度，不难算得它对度规场的变分为

$$\delta\mathcal{L} = \frac{1}{2}\partial^{\mu}\phi\partial^{\nu}\phi\delta g_{\mu\nu}. \quad (4.106)$$

综上，即得

$$\delta S_m = \frac{1}{2}\int d^4x(g^{\mu\nu}\mathcal{L} + \partial^{\mu}\phi\partial^{\nu}\phi)\delta g_{\mu\nu}. \quad (4.107)$$

从而可知标量场的能量动量张量为

$$\begin{aligned} T^{\mu\nu} &= \partial^{\mu}\phi\partial^{\nu}\phi + g^{\mu\nu}\mathcal{L} \\ &= \partial^{\mu}\phi\partial^{\nu}\phi - \frac{1}{2}g^{\mu\nu}\partial_{\rho}\phi\partial^{\rho}\phi - g^{\mu\nu}\mathcal{U}(\phi). \end{aligned} \quad (4.108)$$

与第二章中的相关结果比较，不难发现，这正是弯曲时空中标量场的能量动量张量。

注记：三种能量动量张量

实际上，在理论物理学中，有三种定义不相同的能量动量张量。首先，时空平移对称性所对应的守恒流称之为正则能量动量张量，正则能量动量张量不一定是对称张量。不过，对于每一个正则能量动量张量，人们都可以构造一个相应的对称守恒张量，称之为 Belinfante-Rosenfeld 张量。Belinfante-Rosenfeld 张量当然也是能量动量张量，所以这就有了两种能量动量张量。

第三种能量动量张量是所谓的希尔伯特能量动量张量，它也就是前面我们引入的对称张量 $\Theta^{\mu\nu}$ ，它的定义就是物质场作用量对度规场的泛函导数。

在本书第二章以及本章(第四章)的相关讨论中，我们实际上是澄清了这三种能量动量张量之间的关系。首先，在第二章中，通过所谓的诺特技巧(Noether's trick)，我们可以利用无穷小局域洛伦兹变换引入对称的 Belinfante-Rosenfeld 张量，这样就以诺特定理的形式统一了正则能量动量张量和 Belinfante-Rosenfeld 张量。

而在本章中，通过利用微分同胚不变性，我们又进一步证明了，只要物质场与引力场之间的耦合是最小耦合，那么弯曲时空中的 Belinfante-Rosenfeld 张量就严格等于希尔伯特能量动量张量。这样我们就彻底澄清了三种能量动量张量之间的关系。

4.5 弯曲时空中的高斯定理

前面的相关讨论中我们用到了弯曲时空中的高斯定理，而且后面的章节中也会用到这个定理。本节我们对此进行一个详细的讨论和证明。

引理：首先我们讨论一个引理，即克里斯托夫联络的指标收缩能够写成如下形式

$$\Gamma_{\mu\nu}^{\mu} = \frac{1}{\sqrt{|g|}} \partial_{\nu} \sqrt{|g|}, \quad (4.109)$$

式中 $g = \det g_{\mu\nu}$ 。

证明：首先，根据克里斯托夫联络的定义式，我们有

$$\Gamma_{\mu\nu}^{\mu} = \frac{1}{2} g^{\mu\rho} \partial_{\nu} g_{\mu\rho} = \frac{1}{2} \text{Tr}(g^{-1} \partial_{\nu} g). \quad (4.110)$$

式中 g 代表矩阵 $(g_{\mu\nu})$ 。利用 $\text{Tr}(M^{-1} \delta M) = \delta \ln \det M$ ，可以进一步得到

$$\begin{aligned} \Gamma_{\mu\nu}^{\mu} &= \frac{1}{2} \text{Tr}(g^{-1} \partial_{\nu} g) = \frac{1}{2} \partial_{\nu} \ln |\det g| \\ &= \frac{1}{2} \partial_{\nu} \ln |g| = \partial_{\nu} \ln(\sqrt{|g|}) = \frac{1}{\sqrt{|g|}} \partial_{\nu} \sqrt{|g|}. \end{aligned} \quad (4.111)$$

式中第二行和第一行有所不同，其中的 $g = \det(g_{\mu\nu})$ 。

弯曲时空的高斯定理：假设 M 是一个 $D = n$ 维的弯曲时空，假设它有一个边界 ∂M 。将 ∂M 的指向时空外侧的法向量记作 n^{ν} ， n^{ν} 为一个单位向量，满足 $g_{\mu\nu} n^{\mu} n^{\nu} = 1$ 。则对于 M 上的任何向量场 A^{μ} ，我们有

$$\int_M d^n x \sqrt{|g|} D_{\mu} A^{\mu} = \int_{\partial M} d^{n-1} x \sqrt{|h|} n_{\mu} A^{\mu} = \int_{\partial M} dS_{\mu} A^{\mu}. \quad (4.112)$$

式中 h_{ij} 是边界 ∂M 上的诱导度规， $h = \det(h_{ij})$ ， $dS_{\mu} \equiv d^{n-1} x \sqrt{|h|} \cdot n_{\mu}$ 表示边界上的面积元矢量。

证明：引用上面的引理，即有

$$\begin{aligned} \sqrt{|g|} D_{\mu} A^{\mu} &= \sqrt{|g|} (\partial_{\mu} A^{\mu} + \Gamma_{\mu\nu}^{\mu} A^{\nu}) = \sqrt{|g|} (\partial_{\mu} A^{\mu} + X^{\nu} \frac{1}{\sqrt{|g|}} \partial_{\nu} \sqrt{|g|}) \\ &= \partial_{\mu} (\sqrt{|g|} A^{\mu}). \end{aligned} \quad (4.113)$$

也即是说，一个向量场的协变散度可以重新表达成普通散度的形式！因此

$$\int_M d^n x \sqrt{|g|} D_{\mu} A^{\mu} = \int_M d^n x \partial_{\mu} (\sqrt{|g|} A^{\mu}). \quad (4.114)$$

然后我们当然就可以用普通的高斯定理。

为了看得更清楚，我们不妨假设 ∂M 可以局部地由 $x^n = \text{常数}$ 的超曲面刻画，进而可以将度规写成如下形式

$$g_{\mu\nu} = \begin{pmatrix} h_{ij} & 0 \\ 0 & N^2 \end{pmatrix}, \quad (4.115)$$

式中 h_{ij} 即为边界 ∂M 上的诱导度规, N 为某函数。从而局部地有法向量 $n^\mu = (0, 0, \dots, 1/N)$, 满足 $g_{\mu\nu}n^\mu n^\nu = 1$. 当然也有 $n_\mu = g_{\mu\nu}n^\nu = (0, 0, \dots, N)$.

利用上一段的结果即有

$$\int_M d^n x \partial_\mu (\sqrt{|g|} A^\mu) = \int_{\partial M} d^{n-1} x (\sqrt{|h|} N^2 A^n) = \int_{\partial M} d^{n-1} x \sqrt{|h|} n_\mu A^\mu.$$

式中第一个等号就是在 ∂M 由 $x^n = \text{常数}$ 刻画的局部坐标中应用了普通的高斯定理。结合上面这个结果和 (4.114) 式, 即有

$$\int_M d^n x \sqrt{|g|} D_\mu A^\mu = \int_{\partial M} d^{n-1} x \sqrt{|h|} n_\mu A^\mu. \quad (4.116)$$

定理得证。

拉普拉斯算符: 上述证明过程有一个副产品, 即对于一个标量场 Φ , 拉普拉斯算符对其作用为

$$D_\mu D^\mu \Phi = D_\mu (\partial^\mu \Phi) = \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} \partial^\mu \Phi) = \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} g^{\mu\nu} \partial_\nu \Phi). \quad (4.117)$$

举个例子, 比如平坦的三维欧几里德空间, 这时候在三维直角坐标中, 协变导数当然就是普通的导数, 所以上面协变形式的拉普拉斯算符实际上就是普通的三维空间的拉普拉斯算符! 但是, 协变的形式不止适用于直角坐标, 而是可以适用于任何坐标。下面, 转入到球坐标, 这时候度规为

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2. \quad (4.118)$$

度规张量的行列式为 $|g| = r^4 \sin^2 \theta$, 度规张量的逆也是显然的

$$g^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^{-2} & 0 \\ 0 & 0 & r^{-2} \sin^{-2} \theta \end{pmatrix}. \quad (4.119)$$

从而根据上一段的公式, 容易求得

$$\begin{aligned} & \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} g^{\mu\nu} \partial_\nu \Phi) \\ &= \left[\frac{1}{r^2} \partial_r (r^2 \partial_r) + \frac{1}{r^2 \sin \theta} \partial_\theta (\sin \theta \partial_\theta) + \frac{1}{r^2 \sin^2 \theta} \partial_\phi^2 \right] \Phi. \end{aligned} \quad (4.120)$$

这正是球坐标中拉普拉斯算符的表达式。

5. 黎曼曲率张量和爱因斯坦场方程

我们已经看到如何通过最小耦合原理得到弯曲时空中的物质场作用量，下一步当然是要寻找引力场本身的作用量。从第二章关于经典场论的讨论可以知道，这样一个引力场的作用量只能包括两个度规场导数（高次导数不能出现，否则就会有 Ostrogradsky 不稳定性），要么以一阶导数的“平方”的形式，要么以度规场的二阶导数项的形式，在作用量原理中这两种形式是等价的，因为一阶导数的“平方”分部积分以后就会得到一个二阶导数项。与此同时，微分同胚不变性也要求引力场作用量在广义坐标变换之下保持不变。

但是，要同时满足以上两个要求却并不容易。首先，如果度规场的导数是普通导数的话，那就不容易满足广义协变性的要求。但是，如果将度规场的导数取作协变导数，那由于 $D_\rho g_{\mu\nu} = 0$ ，因此又不容易得到一个非零的作用量。所以寻找引力场的作用量是一个高度非平凡的问题。结果表明，解决这个高度非平凡的问题需要用到黎曼的伟大贡献，那就是黎曼曲率张量。

5.1 黎曼曲率张量

我们已经看到，引力是时空的弯曲，但是，如何刻画时空的弯曲程度呢？换言之，如何刻画时空的曲率呢？为了讨论这个问题我们不妨一般性地假定时空是 D 维的。

在第一章的最后我们已经看到，为了刻画 D 维弯曲时空，我们需要寻找这样一个量，它满足：第一，它得包括度规场的二阶导数项，因为在黎曼正则坐标系（局部惯性系）的原点，度规场的一阶导数是零，要刻画这一点的曲率就需要引入度规场的二阶导数。第二，这个刻画要在所有的坐标系中均成立，也就是说，这个量得是协变的，它得是一个张量。

第三，在平直的闵可夫斯基时空中，这个量要恒等于零，因为它是刻画时空弯曲程度的量。第四，这个量得有 $\frac{1}{12}D^2(D^2 - 1)$ 个独立分量。正是黎曼找到了这个量的具体表达式，这就是所谓的黎曼曲率张量。

5.1.1 寻找黎曼曲率张量

首先让我们来回顾一下平直空间的无穷小平移如何影响一个函数。假设在平直时空从 x 点移动到邻近的 $x + \delta x$ 点，则某个函数 $f(x)$ 的值将变为 $f(x + \delta x)$ 。很显然，在忽略高阶无穷小的意义上，我们有

$$f(x + \delta x) = (1 + \delta x^\mu \partial_\mu) f(x), \quad (5.1)$$

即是说， $(1 + \delta x^\mu \partial_\mu)$ 是无穷小平移算符，当将它作用在函数 $f(x)$ 上，就得到 $f(x + \delta x)$ 。

现在，我们要考察的是弯曲时空的移动，这时候当然就要将平移算符中的偏导运算替换成协变导数，所以弯曲时空的无穷小移动算符就应该是

$$(1 + \delta x^\mu D_\mu). \quad (5.2)$$

很显然的是，平直时空中的平移与具体的移动路径无关，从同一起点经不同路径到达同一终点，结果总是相同的。但是，正如我们将要看到的，在弯曲时空中，移动的结果依赖于具体的移动路径。这种依赖正好反应了时空弯曲的曲率。

为了看清楚这一点，让我们考察在弯曲时空中沿着 $\delta_1 x$ 和 $\delta_2 x$ 两个不同方向将一个矢量 S_ρ 从两个对角中的左下角移动到右上角，如图 (5.1) 所示。现在我们有两条不同的移动路径，先沿着 $\delta_1 x$ 移动，再沿着 $\delta_2 x$ 移动，得到的结果是 $(1 + (\delta_2 x)^\mu D_\mu)(1 + (\delta_1 x)^\nu D_\nu) S_\rho$ ，相反，先沿着 $\delta_2 x$ 移动，再沿着 $\delta_1 x$ 移动，得到的结果将是 $(1 + (\delta_1 x)^\nu D_\nu)(1 + (\delta_2 x)^\mu D_\mu) S_\rho$ ，两条不同移动路径的差值是

$$\begin{aligned} & (1 + (\delta_2 x)^\mu D_\mu)(1 + (\delta_1 x)^\nu D_\nu) S_\rho - (1 + (\delta_1 x)^\nu D_\nu)(1 + (\delta_2 x)^\mu D_\mu) S_\rho \\ &= (\delta_2 x)^\mu (\delta_1 x)^\nu [D_\mu, D_\nu] S_\rho, \end{aligned} \quad (5.3)$$

式中 $[D_\mu, D_\nu] = D_\mu D_\nu - D_\nu D_\mu$ 称作协变导数算符的对易子。我们认为两条不同移动路径所得到的结果之差 $(\delta_2 x)^\mu (\delta_1 x)^\nu [D_\mu, D_\nu] S_\rho$ 正好可以用来刻画时空曲率。

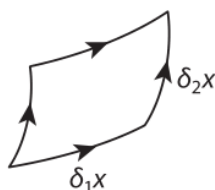


Figure 5.1: 在弯曲时空中沿着 $\delta_1 x$ 和 $\delta_2 x$ 两个不同方向将一个矢量从两个对角中的左下角移动到右上角。

由于 $(\delta_2 x)^\mu (\delta_1 x)^\nu$ 正比于图 (5.1) 中无穷小四边形的面积，它是我们选取移动路径的结果，所以和时空曲率无关，另外，时空曲率当然也应该和我们具体移动的矢量 S_ρ 无关。所以正确的说应该是，我们认为 $[D_\mu, D_\nu]S_\rho$ 模去矢量 S_ρ 的信息就可以用来刻画时空曲率。 $[D_\mu, D_\nu]S_\rho$ 是一个张量，而且显然与 S_ρ 成线性关系，所以我们可以假设

$$[D_\mu, D_\nu]S_\rho = -R^\sigma_{\rho\mu\nu}S_\sigma, \quad (5.4)$$

式中的负号只是为了使得 $R^\sigma_{\rho\mu\nu}$ 与通常的约定相符。根据上面的讨论可以知道，上式中的张量 $R^\sigma_{\rho\mu\nu}$ 正好可以用来刻画时空曲率，实际上，它就是所谓的黎曼曲率张量。

为了得到 $R^\sigma_{\rho\mu\nu}$ 的表达式，我们进行如下计算：首先，

$$D_\mu D_\nu S_\rho = \partial_\mu (D_\nu S_\rho) - \Gamma^\sigma_{\mu\nu} D_\sigma S_\rho - \Gamma^\sigma_{\mu\rho} D_\nu S_\sigma. \quad (5.5)$$

而我们真正要计算的是 $[D_\mu, D_\nu]S_\rho$ ，为此我们需要将 (5.5) 式的结果减去将它的 μ 、 ν 指标互换之后的结果。注意到 (5.5) 式右边的第二项关于 μ 、 ν 指标是对称的，因此直接被减掉了。而对于 (5.5) 式右边第一项和第三项中的 $D_\nu S_\rho$ ，我们可以直接代入协变导数的定义，由此不难得到

$$\begin{aligned} [D_\mu, D_\nu]S_\rho &= \partial_\mu (\partial_\nu S_\rho - \Gamma^\sigma_{\nu\rho} S_\sigma) - \Gamma^\sigma_{\mu\rho} (\partial_\nu S_\sigma - \Gamma^\lambda_{\nu\sigma} S_\lambda) - (\mu \leftrightarrow \nu) \\ &= -(\partial_\mu \Gamma^\sigma_{\nu\rho}) S_\sigma - \Gamma^\sigma_{\nu\rho} \partial_\mu S_\sigma - \Gamma^\sigma_{\mu\rho} \partial_\nu S_\sigma + \Gamma^\sigma_{\mu\rho} \Gamma^\lambda_{\nu\sigma} S_\lambda - (\mu \leftrightarrow \nu) \\ &= -(\partial_\mu \Gamma^\sigma_{\nu\rho} - \partial_\nu \Gamma^\sigma_{\mu\rho} + \Gamma^\sigma_{\mu\lambda} \Gamma^\lambda_{\nu\rho} - \Gamma^\sigma_{\nu\lambda} \Gamma^\lambda_{\mu\rho}) S_\sigma. \end{aligned} \quad (5.6)$$

与 $R^\sigma_{\rho\mu\nu}$ 的定义式 (5.4) 进行比较，即有

$$R^\sigma_{\rho\mu\nu} = \partial_\mu \Gamma^\sigma_{\nu\rho} - \partial_\nu \Gamma^\sigma_{\mu\rho} + \Gamma^\sigma_{\mu\lambda} \Gamma^\lambda_{\nu\rho} - \Gamma^\sigma_{\nu\lambda} \Gamma^\lambda_{\mu\rho}. \quad (5.7)$$

这就是黎曼曲率张量的表达式。

下面我们来看一下，(5.7) 式给出的黎曼曲率张量满不满足前面我们对刻画时空曲率的量的四条要求。先看它满不满足前三条要求，回答是满足的！首先，根据定义 $R^\sigma_{\rho\mu\nu}$ 是一个张量，因此是协变的。其次，从 (5.7) 式可以清楚地看到， $R^\sigma_{\rho\mu\nu}$ 中包含克里斯托夫联络的一阶偏导，而克里斯托夫联络本身又是度规场的一阶偏导，所以 $R^\sigma_{\rho\mu\nu}$ 中的确包含度规场的二阶偏导，这个要求是满足的。第三，对于平直的闵可夫斯基时空，我们可以在全局上选取惯性参考系，在这样的参考系中，恒有 $\Gamma^\rho_{\mu\nu} \equiv 0$ ，代入 (5.7) 式即有 $R^\sigma_{\rho\mu\nu} = 0$ ，所以这第三条要求也是满足的。相反，在弯曲的时空中，一般来说 $R^\sigma_{\rho\mu\nu}$ 将不等于零。因为虽然这时候可以取局部惯性系（也即是局部平坦坐标系，或者说黎曼正则坐标系），在局部惯性系的原点（即 P 点）有 $\Gamma^\rho_{\mu\nu}(P) = 0$ ，但是 $\partial_\lambda \Gamma^\rho_{\mu\nu}|_P \neq 0$ ，所以 $R^\sigma_{\rho\mu\nu}(P) \neq 0$ ，而是等于

$$R^\sigma_{\rho\mu\nu}(P) = \partial_\mu \Gamma^\sigma_{\nu\rho}|_P - \partial_\nu \Gamma^\sigma_{\mu\rho}|_P. \quad (5.8)$$

对第四条要求的验证我们稍后进行。实际上，仅仅只需要包含度规的二阶偏导以及是一个张量这两个要求就可以唯一确定黎曼曲率张量了，因为黎曼曲率张量其实是同时满足这两个要求的唯一量，关于这个结论的证明，我们放在本章的附录中。正如后文将要看到的，黎曼曲率张量这种唯一性强烈地限制了引力场的作用量！

5.1.2 黎曼曲率张量的代数性质

下面我们来研究黎曼曲率张量的一些代数性质，同时验证它满足对时空曲率的第四条要求，即它有 $\frac{1}{2}D^2(D^2 - 1)$ 个独立分量。

为了把这些代数性质说清楚，我们引入如下定义：对于任意高阶张量 $T_{\mu_1\mu_2\dots\mu_n\dots}$ ，我们定义

$$T_{(\mu_1\mu_2\dots\mu_n)\dots} = \frac{1}{n!} (T_{\mu_1\mu_2\dots\mu_n\dots} + \text{对所有的关于 } \mu_1, \mu_2, \dots, \mu_n \text{ 的置换求和}).$$

我们称 $T_{(\mu_1\mu_2\dots\mu_n)\dots}$ 为张量 $T_{\mu_1\mu_2\dots\mu_n\dots}$ 关于前 n 个指标的全对称部分。类似的，我们定义

$$T_{[\mu_1\mu_2\dots\mu_n]\dots} = \frac{1}{n!} (T_{\mu_1\mu_2\dots\mu_n\dots} + \text{对所有的关于 } \mu_1, \mu_2, \dots, \mu_n \text{ 的置换交错求和}).$$

所谓交错求和，指的是，对于偶置换就直接取加号，而对于奇置换则在表达式前面添上一个负号再求和。我们称 $T_{[\mu_1\mu_2\dots\mu_n]\dots}$ 为张量 $T_{\mu_1\mu_2\dots\mu_n\dots}$ 关于前 n 个指标的全反对称部分。比如对于二阶张量 $A_{\mu\nu}$ ，我们有

$$A_{(\mu\nu)} = \frac{1}{2}(A_{\mu\nu} + A_{\nu\mu}), \quad A_{[\mu\nu]} = \frac{1}{2}(A_{\mu\nu} - A_{\nu\mu}). \quad (5.9)$$

由于广义坐标变换不会改变一个张量关于其指标间的各种对称性，所以为了看清楚黎曼曲率张量的代数性质，我们不妨选取一个特殊的广义坐标系，即选取局部惯性系。并且，相比于原来的 $R^\sigma_{\rho\mu\nu}$ ，我们不妨考察与之密切相关的 $R_{\sigma\rho\mu\nu} = g_{\sigma\lambda}R^\lambda_{\rho\mu\nu}$ 。根据 (5.7) 式，同时再利用在局部惯性系的原点，有 $\Gamma_{\mu\nu}^\rho(P) = 0$, $\partial_\rho g_{\mu\nu}|_P = 0$, $\partial_\rho g^{\mu\nu}|_P = 0$ ，从而不难得到

$$R_{\sigma\rho\mu\nu}(P) = \frac{1}{2}(\partial_\mu\partial_\rho g_{\sigma\nu} - \partial_\nu\partial_\rho g_{\sigma\mu} - \partial_\mu\partial_\sigma g_{\rho\nu} + \partial_\nu\partial_\sigma g_{\rho\mu})|_P. \quad (5.10)$$

根据上面的 (5.10) 式，很容易看出， $R_{\sigma\rho\mu\nu}$ 满足如下代数性质：

(A). 反对称性，即

$$R_{\sigma\rho\mu\nu} = -R_{\rho\sigma\mu\nu}, \quad R_{\sigma\rho\mu\nu} = -R_{\sigma\rho\nu\mu}, \quad (5.11)$$

即 $R_{\sigma\rho\mu\nu}$ 关于前两个指标反对称，也关于后两个指标反对称，或者说

$$R_{\rho\sigma\mu\nu} = R_{[\rho\sigma][\mu\nu]}. \quad (5.12)$$

(B). 对称性，即

$$R_{\sigma\rho\mu\nu} = R_{\mu\nu\sigma\rho}. \quad (5.13)$$

(C). 循环性，即

$$R_{\sigma\rho\mu\nu} + R_{\sigma\mu\nu\rho} + R_{\sigma\nu\rho\mu} = 0. \quad (5.14)$$

结合反对称性，我们也可以将这条性质写成 $R_{\sigma[\rho\mu\nu]} = 0$ 。再结合对称性，又可以进一步写成

$$R_{[\sigma\rho\mu\nu]} = 0. \quad (5.15)$$

下面我们来数一下黎曼曲率张量 $R_{\sigma\rho\mu\nu}$ 一共有多少个独立分量。首先, 根据反对称性和对称性, $R_{\rho\sigma\mu\nu} = R_{[\rho\sigma][\mu\nu]}$, 可以看成是以 $[\rho\sigma]$ 和 $[\mu\nu]$ 为行列“指标”的一个对称矩阵, 而由于反对称性, 每一个“指标”有 $D(D-1)/2$ 种独立可能性。根据 $N \times N$ 对称矩阵有 $\frac{1}{2}N(N+1)$ 个独立矩阵元, 进而可知, $R_{\rho\sigma\mu\nu} = R_{[\rho\sigma][\mu\nu]}$ 有

$$\frac{1}{2} \left[\frac{1}{2} D(D-1) \right] \left[\frac{1}{2} D(D-1) + 1 \right] = \frac{1}{8} D(D-1)(D^2 - D + 2) \quad (5.16)$$

个独立分量。最后, 根据循环性条件, $R_{[\sigma\rho\mu\nu]} = 0$, 这是一个全反对称条件, 它相当于加上了 $D(D-1)(D-2)(D-3)/4!$ 个限制, 从而剩下的 $R_{\rho\sigma\mu\nu}$ 的独立分量个数等于

$$C_D = \frac{1}{8} D(D-1)(D^2 - D + 2) - \frac{1}{24} D(D-1)(D-2)(D-3) \quad (5.17)$$

并项后得到

$$C_D = \frac{1}{12} D^2 (D^2 - 1), \quad (5.18)$$

正好符合关于曲率张量的第四条要求。特别的, 在 $D=4$ 维时空中, 黎曼曲率张量有 20 个独立分量。而在 $D=2$ 维时空中, 它只有唯一一个独立分量, 也就是两维曲面的高斯曲率。

5.2 爱因斯坦-希尔伯特作用量和爱因斯坦场方程

5.2.1 爱因斯坦-希尔伯特作用量

下面我们可以来探讨引力场本身的作用量了。对这个作用量的最重要要求就是微分同胚不变性, 其次还要求作用量中只含两个度规场导数, 下面我们会看到这两个要求几乎唯一确定了引力场的作用量。

首先, 我们注意到四维体积元 $dv = d^4x \sqrt{-g}$ 是坐标变换不变的, 因为为了满足微分同胚不变性, 我们可以把引力场作用量写成 $S_g = \int dv A(x)$, 其中 $A(x)$ 表示一个标量。其次, 标量 $A(x)$ 中只含两个度规场导数, 而我们又知道两个度规场导数构造出来的几乎唯一的张量就是黎曼曲率张量, 所以, 标量 $A(x)$ 应该由黎曼曲率张量 $R_{\rho\sigma\mu\nu}$ 通过指标缩并而来! 注意, 只是将一个 $R_{\rho\sigma\mu\nu}$ 进行指标缩并, 不是将两个或者更多个 $R_{\rho\sigma\mu\nu}$ 缩并在一起, 后者将包含四个以上度规场偏导, 从而不符合要求。

根据黎曼曲率张量的代数性质 (反对称性和对称性), 不难看出, 由 $R_{\rho\sigma\mu\nu}$ 缩并得来的唯一独立二阶张量为

$$R_{\sigma\nu} = g^{\rho\mu} R_{\rho\sigma\mu\nu}, \quad (5.19)$$

其它缩并方式要么得到零, 要么得到 $-R_{\sigma\nu}$ 。 $R_{\sigma\nu}$ 称作里奇张量, 由黎曼曲率张量的对称性, 不难看出, **里奇张量为一个对称张量**, 而且刚刚已经说了, 它具有**唯一性**。

由里奇张量进一步缩并, 得到的唯一标量为

$$R = g^{\sigma\nu} R_{\sigma\nu}, \quad (5.20)$$

称作曲率标量。很明显，曲率标量本质上是由黎曼曲率张量缩并得来的唯一标量，具有唯一性。

结合黎曼曲率张量的唯一性以及缩并的唯一性，可以知道，包含两个度规偏导的唯一标量，就是曲率标量 R 。从而可以知道，引力场的作用量必定可以写成如下形式

$$S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x \sqrt{-g} R, \quad (5.21)$$

式中 κ 为某个待定常数。 S_{EH} 就是所谓的爱因斯坦-希尔伯特作用量。

5.2.2 爱因斯坦场方程

因此，考虑到引力场，整个物理系统的作用量就应该是

$$S = S_{\text{EH}} + S_m, \quad (5.22)$$

S_m 就是除引力场之外的物质的作用量。有了作用量就可以根据最小作用量原理得到引力场的场方程。为此需要将作用量对度规场 $g_{\mu\nu}$ 进行变分，记这个变分为 $\delta S = \delta S_{\text{EH}} + \delta S_m$ ，则场方程由 $\delta S = 0$ 给出。根据上一章的知识，物质作用量对度规场的变分就是

$$\delta S_m = \frac{1}{2} \int d^4x \sqrt{-g} T^{\mu\nu} \delta g_{\mu\nu}, \quad (5.23)$$

式中 $T^{\mu\nu}$ 就是物质的能量动量张量。即是说，为了得到引力场的场方程，关键是要计算爱因斯坦-希尔伯特作用量对度规场的变分。

下面让我们来计算这个变分，注意到 $S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x \sqrt{-g} R_{\mu\nu} g^{\mu\nu}$ ，从而即有

$$\delta S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x [\sqrt{-g} R_{\mu\nu} \delta g^{\mu\nu} + \sqrt{-g} (\delta R_{\mu\nu}) g^{\mu\nu} + \delta(\sqrt{-g}) R]. \quad (5.24)$$

利用 $\delta(\sqrt{-g}) = \frac{1}{2} \sqrt{-g} g^{\mu\nu} \delta g_{\mu\nu}$ 以及 $\delta g^{\mu\nu} = -g^{\mu\rho} \delta g_{\rho\sigma} g^{\sigma\nu}$ ，即可以得到

$$\delta S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x \sqrt{-g} [- (R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R) \delta g_{\mu\nu} + (\delta R_{\mu\nu}) g^{\mu\nu}]. \quad (5.25)$$

所以最后要算的就是 $\delta R_{\mu\nu}$ 。

$R_{\mu\nu}$ 是由黎曼曲率张量 $R^\rho_{\mu\sigma\nu}$ 将指标 ρ, σ 缩并而来，所以要算 $\delta R_{\mu\nu}$ ，我们可以先算 $\delta R^\rho_{\mu\sigma\nu}$ 。为了计算 $\delta R^\rho_{\mu\sigma\nu}$ ，我们可以选择 $\Gamma^\rho_{\mu\nu}|_P = 0$ 的局部惯性系，利用 (5.8) 式，从而即有

$$\delta R^\rho_{\mu\sigma\nu}|_P = \partial_\sigma (\delta \Gamma^\rho_{\nu\mu})|_P - \partial_\nu (\delta \Gamma^\rho_{\sigma\mu})|_P. \quad (5.26)$$

注意，克里斯托夫联络的变分 $\delta \Gamma^\rho_{\mu\nu}$ 实际上就是改变之后的联络减去变分之前的联络，所以 $\delta \Gamma^\rho_{\mu\nu}$ 的实质是两个联络之差。前面我们讲过，联络本身不是张量，但是**联络之差是张量**，所以 $\delta \Gamma^\rho_{\mu\nu}$ 其实是一个张量。又由于在局部惯性系中，有 $\Gamma^\rho_{\mu\nu}|_P = 0$ ，所以协变导数就等于普通的导数，所以我们可以将 $\delta R^\rho_{\mu\sigma\nu}|_P$ 的表达式重写为

$$\delta R^\rho_{\mu\sigma\nu}|_P = D_\sigma (\delta \Gamma^\rho_{\nu\mu})|_P - D_\nu (\delta \Gamma^\rho_{\sigma\mu})|_P. \quad (5.27)$$

但这种写法显然是协变的，因此它并不依赖于原来的局部惯性系，而是在任何坐标系中都成立，因此在任何坐标系中均有

$$\delta R^{\rho}_{\mu\sigma\nu} = D_{\sigma}(\delta\Gamma^{\rho}_{\nu\mu}) - D_{\nu}(\delta\Gamma^{\rho}_{\sigma\mu}). \quad (5.28)$$

这个式子有时候也称作 Palatini 恒等式。利用这个恒等式，即有

$$\delta R_{\mu\nu} = D_{\rho}(\delta\Gamma^{\rho}_{\nu\mu}) - D_{\nu}(\delta\Gamma^{\rho}_{\rho\mu}). \quad (5.29)$$

利用以上结果，容易看出 (5.25) 式中的 $\delta R_{\mu\nu}$ 项可以化成

$$\begin{aligned} \frac{1}{2\kappa} \int d^4x \sqrt{-g} (\delta R_{\mu\nu}) g^{\mu\nu} &= \frac{1}{2\kappa} \int dv [D_{\rho}(\delta\Gamma^{\rho}_{\nu\mu}) - D_{\nu}(\delta\Gamma^{\rho}_{\rho\mu})] g^{\mu\nu} \\ &= \frac{1}{2\kappa} \int dv [D_{\rho}(g^{\mu\nu} \delta\Gamma^{\rho}_{\nu\mu}) - D_{\nu}(g^{\mu\nu} \delta\Gamma^{\rho}_{\rho\mu})]. \end{aligned} \quad (5.30)$$

式中已经利用了 $D_{\rho}g^{\mu\nu} = 0$ 。再利用高斯定理 $\int dv D_{\mu}A^{\mu} = \int_{\infty} dS_{\mu}A^{\mu}$ ，很容易看到 (5.30) 式其实是一个时空无穷远处的表面积分项，由于在利用最小作用量原理时我们总是假定场地形在时空无穷远处的变分等于零，从而可知**这个无穷远处的表面项实际上等于零**。

表面项贡献等于零的这个结论我们这里就这么说了，但实际上这里有一个非常微妙的细节问题，这个微妙的细节是 Gibbons 和 Hawking 等人解决的，但我们现在还不具备讨论这个细节的数学基础，后面章节适当的时候我们会再次回到这里。

根据上面这个结果，我们就可以将 (5.25) 式简化成

$$\delta S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x \sqrt{-g} \left[-(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R) \delta g_{\mu\nu} \right]. \quad (5.31)$$

结合能量动量张量的定义式 (5.23) 式，即有

$$\delta S = \frac{1}{2\kappa} \int d^4x \sqrt{-g} \left[-(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R - \kappa T^{\mu\nu}) \delta g_{\mu\nu} \right]. \quad (5.32)$$

根据最小作用量原理，即有引力场的场方程

$$R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = \kappa T^{\mu\nu}. \quad (5.33)$$

这个方程即是著名的**爱因斯坦场方程**。我们看到，广义相对论中只有一个常数 κ ，而牛顿引力中也有一个常数，即牛顿万有引力常数 G ，既然两个理论都是描述万有引力，那 κ 必定与 G 有关。为了定出 κ 与 G 的关系，我们需要取广义相对论在引力场不太强时的牛顿引力近似，并将近似的结果和牛顿引力理论进行比较，进而确定 κ 与 G 的定量关系，不过，我们将把这个步骤放在后面的章节中进行。这里不妨先把结果写出来，结果就是 $\kappa = 8\pi G$ 。

$\kappa = 8\pi G$ 这个结果是合理的，这可以从量纲分析看出来，也即是说 κ 其实和 G 有相同的量纲。为了看清楚这一点，我们注意到黎曼度规是 $ds^2 = g_{\mu\nu}dx^{\mu}dx^{\nu}$ ，等式左边的 ds^2 具有长度平方的量纲，记作 $[L]^2$ ($[]$ 表示取量纲)，等式右边 $dx^{\mu}dx^{\nu}$ 也具有长度平方量纲 $[L]^2$ ，因此 $g_{\mu\nu}$ 是无量纲的。标量曲率 R 由度规张量构造出来，但其中包含了二阶导数，因此是 $[L]^{-2}$ 量纲。又注意到 d^4x 是 $[L]^4$ 量纲，因此为了使得爱因斯坦-希尔伯特作用量

$S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x \sqrt{-g} R$ 具有作用量的正确量纲，即能量乘以长度的量纲 $[E][L]$ ，常数 κ 必须具有如下量纲

$$[\kappa][E][L] = [L]^4[L]^{-2} \Leftrightarrow [\kappa] = [L]/[E]. \quad (5.34)$$

另一方面，根据牛顿万有引力势能 GMm/r 具有能量量纲，而 Mm 是能量平方量纲，所以 G 的量纲也是 $[G] = [L]/[E]$ 。所以 κ 与 G 量纲相同，从而必定与 G 成比例关系，所以 $\kappa = 8\pi G$ 是合理的。

5.2.3 微分同胚不变性与比安基恒等式

我们知道，爱因斯坦-希尔伯特作用量具有微分同胚不变性。下面来看看这会告诉我们什么进一步的知识。为此，考虑如下无穷小微分同胚

$$\delta g_{\mu\nu} = -(D_\mu \varepsilon_\nu + D_\nu \varepsilon_\mu), \quad (5.35)$$

并假定在时空无穷远处 ε_μ 趋于零，从而在时空无穷远处 $\delta g_{\mu\nu} = 0$ 。在进行具体考察之前，让我们先引入一个概念，叫**爱因斯坦张量**，记作 $G_{\mu\nu}$ ，它的定义是

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R. \quad (5.36)$$

这是一个对称张量。

根据 (5.31) 式，不难得到 S_{EH} 在上述无穷小微分同胚下的改变量为，

$$\begin{aligned} \delta S_{\text{EH}} &= \frac{1}{2\kappa} \int d^4x \sqrt{-g} \left[-(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R) \delta g_{\mu\nu} \right] \\ &= \frac{1}{2\kappa} \int d^4x \sqrt{-g} \left[G^{\mu\nu} (D_\mu \varepsilon_\nu + D_\nu \varepsilon_\mu) \right] \\ &= \frac{1}{\kappa} \int d^4x \sqrt{-g} G^{\mu\nu} D_\mu \varepsilon_\nu \\ &= -\frac{1}{\kappa} \int d^4x \sqrt{-g} (D_\mu G^{\mu\nu}) \varepsilon_\nu. \end{aligned} \quad (5.37)$$

式中最后一行我们进行了分部积分，并注意到了无穷远处的边界项为零。但是， S_{EH} 是微分同胚不变的，因此在无穷小微分同胚下，其改变量必定等于零，即 $\delta S_{\text{EH}} = 0$ ，因此根据上式，这就意味着爱因斯坦张量必然满足如下恒等式

$$D_\mu G^{\mu\nu} = 0. \quad (5.38)$$

上述恒等式的成立确保了爱因斯坦场方程是自洽的。为了看清楚这一点，我们利用爱因斯坦张量将场方程重写为

$$G_{\mu\nu} = \kappa T_{\mu\nu}. \quad (5.39)$$

由于能动量守恒，爱因斯坦方程的右边必然满足 $D_\mu T^{\mu\nu} = 0$ ，因此方程的左边也必然要满足 $D_\mu G^{\mu\nu} = 0$ ，这正是上面根据微分同胚不变性导出来的恒等式。实际上，根据上一章的

相关知识和刚才的推导可知，能动量守恒和上述恒等式 (5.38)，它们都是微分同胚不变性的必然结果。

(5.38) 式还有另一种不同的证明方式。为此我们取局部惯性系，则根据前面的 (5.10) 式，我们有

$$R_{\sigma\rho\mu\nu}|_P = \frac{1}{2}(\partial_\mu\partial_\rho g_{\sigma\nu} - \partial_\nu\partial_\rho g_{\sigma\mu} - \partial_\mu\partial_\sigma g_{\rho\nu} + \partial_\nu\partial_\sigma g_{\rho\mu})|_P. \quad (5.40)$$

从而

$$D_\lambda R_{\sigma\rho\mu\nu}|_P = \partial_\lambda R_{\sigma\rho\mu\nu}|_P = \frac{1}{2}\partial_\lambda(\partial_\mu\partial_\rho g_{\sigma\nu} - \partial_\nu\partial_\rho g_{\sigma\mu} - \partial_\mu\partial_\sigma g_{\rho\nu} + \partial_\nu\partial_\sigma g_{\rho\mu})|_P.$$

对前三个指标轮换后求和，有

$$\begin{aligned} & D_\lambda R_{\sigma\rho\mu\nu}|_P + D_\sigma R_{\rho\lambda\mu\nu}|_P + D_\rho R_{\lambda\sigma\mu\nu}|_P \\ &= \frac{1}{2}(\partial_\lambda\partial_\mu\partial_\rho g_{\sigma\nu} - \partial_\lambda\partial_\nu\partial_\rho g_{\sigma\mu} - \partial_\lambda\partial_\mu\partial_\sigma g_{\rho\nu} + \partial_\lambda\partial_\nu\partial_\sigma g_{\rho\mu} \\ &\quad + \partial_\sigma\partial_\mu\partial_\lambda g_{\rho\nu} - \partial_\sigma\partial_\nu\partial_\lambda g_{\rho\mu} - \partial_\sigma\partial_\mu\partial_\rho g_{\lambda\nu} + \partial_\sigma\partial_\nu\partial_\rho g_{\lambda\mu} \\ &\quad + \partial_\rho\partial_\mu\partial_\sigma g_{\lambda\nu} - \partial_\rho\partial_\nu\partial_\sigma g_{\lambda\mu} - \partial_\rho\partial_\mu\partial_\lambda g_{\sigma\nu} + \partial_\rho\partial_\nu\partial_\lambda g_{\sigma\mu})|_P \\ &= 0. \end{aligned} \quad (5.41)$$

由于 P 点是时空中的任意给定点，所以将上面结果写得更清楚一点，即是

$$D_\lambda R_{\sigma\rho\mu\nu} + D_\sigma R_{\rho\lambda\mu\nu} + D_\rho R_{\lambda\sigma\mu\nu} = 0. \quad (5.42)$$

结合黎曼曲率张量的代数性质，我们也可以把这个结果写成

$$D_{[\lambda} R_{\sigma\rho]\mu\nu} = 0, \quad (5.43)$$

称之为**比安基恒等式**。

比安基恒等式其实是下面算符恒等式的一个自然结果

$$[D_\lambda, [D_\sigma, D_\rho]] + [D_\sigma, [D_\rho, D_\lambda]] + [D_\rho, [D_\lambda, D_\sigma]] = 0. \quad (5.44)$$

这个算符恒等式直接展开就能证明。把这个算符恒等式作用在任意的矢量场 S_ρ 上，再利用黎曼曲率张量的定义，就能导出比安基恒等式，细致的推导有些繁琐，而且由于我们已经给出比安基恒等式的一个证明了，所以这里不再推了。但是，道理其实很简单，首先，从前面黎曼曲率张量的引入过程可以看到，黎曼曲率张量起源于协变导数的不可对易性，所以 $[D_\sigma, D_\rho] \sim R_{\sigma\rho\mu\nu}$ ，而 $[D_\lambda, [D_\sigma, D_\rho]]$ 即是对黎曼曲率张量再进行一次协变导数，即 $[D_\lambda, [D_\sigma, D_\rho]] \sim D_\lambda R_{\sigma\rho\mu\nu}$ ，从而就不难看出上面的算符恒等式等价于比安基恒等式。

利用比安基恒等式，我们就能给出 (5.38) 的又一个证明。证明如下

$$\begin{aligned} 0 &= g^{\mu\lambda} g^{\nu\sigma} (D_\lambda R_{\rho\sigma\mu\nu} + D_\rho R_{\sigma\lambda\mu\nu} + D_\sigma R_{\lambda\rho\mu\nu}) \\ &= D^\mu R_{\rho\mu} - D_\rho R + D^\nu R_{\rho\nu} \\ &= 2D^\mu (R_{\mu\rho} - \frac{1}{2}g_{\mu\rho}R) = 2D^\mu G_{\mu\rho}. \end{aligned} \quad (5.45)$$

正好是 (5.38) 式。

5.2.4 广义相对论作为一个有效理论

宇宙学常数项

前面我们通过微分同胚不变性等要求确定了爱因斯坦-希尔伯特作用量。但前面的推理其实还有一个小漏洞,那就是其实微分同胚不变的项不只有标量曲率给出的爱因斯坦-希尔伯特作用量,简单的体积项 $\int d^4x\sqrt{-g}$ 也是微分同胚不变的,当然,这一项不含有度规场的导数,因此单独这一项当然不能构成引力场的作用量,但是,在已经有了一个含度规偏导的标曲率项的前提下, $\int d^4x\sqrt{-g}$ 这一微分同胚不变的项就没有理论不加进来了。基于这个考虑,我们其实应该把更一般的引力场作用量 S_g 写成如下形式

$$S_g = \frac{1}{2\kappa} \int d^4x\sqrt{-g}(R - 2\Lambda), \quad (5.46)$$

式中 Λ 为某个常数,称作宇宙学常数。不难得出按照新的作用量 S_g 变分而来的引力场方程,为

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu}. \quad (5.47)$$

在作用量 S_g 中,宇宙学常数 Λ 是和标曲率 R 加在一起的,因此 Λ 的量纲必定与 R 相同,从而 $[\Lambda] = [R] = [L]^{-2}$ 。另外,天文观测表明,人类生存的这个宇宙的宇宙学常数大于零,不妨令

$$\Lambda = \frac{1}{L^2}, \quad (5.48)$$

式中 L 为某个标准长度,不妨称之为宇宙学长度,可以理解为宇宙的半径。假设我们真正关心的是某一片时空区域的引力场,而不是关心整个大的宇宙,假设这个时空区域的曲率半径量级为 l ,根据曲率的定义,这也就是说标曲率 R 在 $1/l^2$ 的量级,即

$$R \sim \frac{1}{l^2}. \quad (5.49)$$

很显然,只要 $l \ll L$,那标曲率这一项对作用量的贡献将远远超过宇宙学常数项,那这时候就可以忽略宇宙学常数项了。这就是为什么我们常常只考虑爱因斯坦-希尔伯特作用量的原因。

处理宇宙学常数的另一种方式是,将之从爱因斯坦场方程 (5.47) 的左边移项到右边,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa(T_{\mu\nu} - \frac{\Lambda}{\kappa}g_{\mu\nu}). \quad (5.50)$$

即将宇宙学常数看成是一种物质,很显然这种物质的能动张量 $T_{\mu\nu}^\Lambda$ 为

$$T_{\mu\nu}^\Lambda = -\frac{\Lambda}{\kappa}g_{\mu\nu} = -\frac{\Lambda}{8\pi G}g_{\mu\nu}. \quad (5.51)$$

换言之,这种物质的能量密度 $\rho_\Lambda = -T_{\Lambda 0}^0$ 为

$$\rho_\Lambda = \frac{\Lambda}{8\pi G}, \quad (5.52)$$

也就是说具有常数能量密度。

什么物质的能量密度为常数呢？这样的东西我们目前只知道一种，那就是物质场作为量子场的真空零点能。所以，物质场的真空零点能必然对宇宙学常数有贡献，甚至就是宇宙学常数。但问题是，天文观测告诉我们，如果采取合适的单位制（即取普朗克常数 $\hbar = 1$ 的单位制），那 $\rho_\Lambda \sim (10^{-3}eV)^4$ ，但是物质场真空零点能密度即使以粒子物理标准模型的能量标度来计算也大约是 $(200GeV)^4$ ，两者完全不在一个量级上，不仅不在一个量级上，甚至是天差地远。这是一个很严重的问题，通常的观点是认为，量子场论的计算在宇宙学的尺度上失效了，至于为什么失效，正确的量子理论计算是什么，这些都依然是理论物理最为重大的未解难题。这也就是著名的宇宙学常数问题。

广义相对论作为一个有效理论

前面对爱因斯坦-希尔伯特作用量的一个要求是，作用量只包含两个度规场导数，不包括高阶导数。这样的要求对于一个基本的量子场论是合理的，因为否则的话系统的能量将没有下界，也就是会出现 Ostrogradsky 不稳定性。但是，广义相对论作为一个基本理论是不可重整的（学过量子场论的读者不妨用量纲的观点思考一下这是为什么），因此现代的观点是将它看作是一个有效理论，相应的作用量也是有效作用量。有效作用量是要包含量子修正的，而量子修正就会带来超过二阶的高阶导数修正项。

那么，这些量子效应应该怎么考虑呢？根据场论的路径积分量子化，量子效应可以通过将相因子 $\exp(iS/\hbar)$ 对所有的场位形进行积分（称作泛函积分或者路径积分）来考察，其中 S 就是场论系统的作用量，对于广义相对论来说 S 可以取作 S_{EH} 。因此，考虑量子引力的一种粗糙方法是将如下相因子进行泛函积分

$$\exp\left(i\frac{1}{16\pi G\hbar}\int d^4x\sqrt{-g}R\right). \quad (5.53)$$

可见，引力的量子效应依赖于常数 $G\hbar$ ， G 的量纲是 $[G] = [L]/[E]$ ，而 \hbar 的量纲和作用量量纲相同，也就是 $[\hbar] = [E][T] = [E][L]$ （式中 T 表示时间），所以 $[G\hbar] = [L]^2$ 。不妨令

$$G\hbar = l_p^2, \quad (5.54)$$

这里 l_p 是一个刻画量子引力的基本长度，称作普朗克长度。当然，以上量纲分析是在光速 $c = 1$ 的单位制中进行的，如果恢复光速单位，那么就有

$$G\hbar/c^3 = l_p^2. \quad (5.55)$$

代入这几个基本常数的测量值，就能得到 l_p 大约是 $10^{-35}m$ 。

考虑到量子引力引起的修正，引力场的作用量就不能只包括爱因斯坦-希尔伯特作用量了，而是可以加进更多的高阶导数项，比如我们可以取引力作用量 S 为

$$S/\hbar = \frac{1}{16\pi l_p^2}\int d^4x\sqrt{-g}R + \lambda\int d^4x\sqrt{-g}R_{\mu\nu}R^{\mu\nu} + \chi\int d^4x\sqrt{-g}R^3\dots \quad (5.56)$$

其中等式右边的第一项就是爱因斯坦-希尔伯特作用量，等式右边的第二项和第三项代表量子引力带来的高阶导数修正项，它们分别依赖于参数 λ 和 χ 。类似的修正项当然很多，

我们只是写出其中两项作为代表。量纲分析一下（注意 S/\hbar 无量纲），不难发现 λ 的量纲为 $[\lambda] = [L]^0$ ， χ 的量纲为 $[\chi] = [L]^2$ 。由于量子引力只有一个有量纲的常数 l_p ，所以可以令

$$\lambda \sim 1, \quad \chi \sim l_p^2. \quad (5.57)$$

假设我们考察的时空尺度为 l ，也即是说时空的曲率半径在 l 的量级，从而

$$R \sim R_{\mu\nu} \sim \frac{1}{l^2}. \quad (5.58)$$

因此上面考虑到量子引力修正的引力场作用量 (5.56) 的典型大小为

$$S/\hbar \sim \int dv \frac{1}{l_p^2 l^2} + \int dv \frac{1}{l^4} + \int dv \frac{l_p^2}{l^6} + \dots \quad (5.59)$$

等式右边的第一项代表爱因斯坦-希尔伯特作用量的贡献，第二项和第三项代表量子引力带来的高阶导数修正。很显然，只要满足如下条件，那量子引力修正项相比于爱因斯坦-希尔伯特作用量的贡献来说就可以忽略。这个条件即是

$$l \gg l_p, \quad (5.60)$$

即是说，在时空尺度远远大于普朗克尺度时，我们只需考虑爱因斯坦-希尔伯特作用量，而量子引力带来的高阶导数修正项则完全都可以忽略！

量子引力同样会对物质场与引力场的耦合产生修正，使之不再是最小耦合。以标量场为例，不考虑量子引力的修正，它与引力场之间的最小耦合作用量 S_{mm} 为

$$S_{mm} = \int \sqrt{-g} d^4x \left[-\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - \mathcal{U}(\phi) \right]. \quad (5.61)$$

考虑到量子引力修正，就需要加上各种修正项，比如

$$S_m = \int \sqrt{-g} d^4x \left[-\frac{1}{2} (1 + \alpha l_p^2 R) g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - \mathcal{U}(\phi) + \dots \right], \quad (5.62)$$

根据量纲分析，式中的参数 α 是无量纲的。很明显，和上一段的分析一样，当 $l \gg l_p$ 时，这些修正项相对都可以忽略，从而使得物质场与引力场的耦合回到最小耦合。

以上分析也说明，爱因斯坦广义相对论的成立是有条件的，条件就是时空尺度远远大于普朗克尺度，如果时空尺度接近普朗克尺度，那这时候无穷无尽的高阶导数修正项都将变得比爱因斯坦-希尔伯特作用量以及物质场的最小耦合作用量更为重要。而我们当然无法考虑无穷无尽的修正项，因此这实际上意味着，当时空尺度接近普朗克尺度时，时空作为黎曼几何的这幅几何图像和分析方法是完全失效的。

总结一下：宇宙学常数问题告诉我们，通常量子场论的框架在宇宙大小的尺度上是失效的，而刚才的分析又告诉我们，广义相对论的框架在普朗克尺度上是失效的。所以今天的物理学理论有两个失效的尺度，一个是大尺度（称之为红外尺度）的宇宙学尺度，一个是小尺度（称之为紫外尺度）的普朗克尺度。至于在这两个失效的地方，正确的理论是什么，我们今天依然不知道。不过人们相信最大的时空尺度和最小的时空尺度很可能是有联系的，称之为紫外-红外对应，通常人们相信正确的量子引力理论将能够告诉我们紫外-红外具体如何对应，并同时解决当前理论的两个失效尺度的问题。

5.3 附录: 曲率张量的唯一性

下面我们来证明黎曼曲率张量 $R^\sigma_{\rho\mu\nu}$ 是唯一能从度规张量以及它的一阶和二阶偏导中构造出来, 且对二阶偏导是线性的张量。

为了证明这一点, 首先让我们回想一下第四章中给出的克里斯托夫联络在不同坐标下的变换关系

$$\Gamma'^\rho_{\mu\nu} = S^\rho_\sigma \Gamma^\sigma_{\alpha\beta} (S^{-1})^\alpha_\mu (S^{-1})^\beta_\nu - (S^{-1})^\alpha_\mu (S^{-1})^\beta_\nu \partial_\alpha S^\rho_\beta. \quad (5.63)$$

写得更清楚一点即是

$$\Gamma'^\rho_{\mu\nu} = \frac{\partial x'^\rho}{\partial x^\sigma} \Gamma^\sigma_{\alpha\beta} \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} - \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} \frac{\partial^2 x'^\rho}{\partial x^\alpha \partial x^\beta}. \quad (5.64)$$

其次, 我们在某个任意的 P 点取局部惯性系, 使得 $\Gamma^p_{\mu\nu}|_P = 0$ 。此外, 我们仅限于考察 P 点的不同局部惯性系之间的坐标变换, 也即保持克里斯托夫联络在 P 点为零的那一类坐标变换。根据 (5.64) 式, 这也就是满足下式的一类坐标变换

$$\frac{\partial^2 x'^\rho}{\partial x^\alpha \partial x^\beta} \Big|_P = 0, \quad (5.65)$$

这里 $x \rightarrow x'$ 表示坐标变换。很显然, 一个一般坐标变换下的张量在这一类特殊的坐标变换下也必然要像张量那样变。

由于克里斯托夫联络在 P 点为零, 所以根据第三章的知识, 度规张量的任何一阶偏导在 P 点都必定为零。所以根据命题的要求, 我们要构造的张量一定只是度规张量二阶偏导的线性组合, 或者等价地, 必定只是克里斯托夫联络一阶偏导的线性组合。

下面我们来考察克里斯托夫联络在 P 点的一阶偏导, 由于在所考察的坐标变换下, $\Gamma'^\rho_{\mu\nu}|_P = \Gamma^\rho_{\mu\nu}|_P = 0$, 所以根据 (5.64) 式, 有

$$\frac{\partial \Gamma'^\rho_{\mu\nu}}{\partial x'^\eta} \Big|_P = \frac{\partial x'^\rho}{\partial x^\sigma} \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} \frac{\partial x^\kappa}{\partial x'^\eta} \frac{\partial \Gamma^\sigma_{\alpha\beta}}{\partial x^\kappa} \Big|_P - \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} \frac{\partial x^\kappa}{\partial x'^\eta} \frac{\partial^3 x'^\rho}{\partial x^\kappa \partial x^\alpha \partial x^\beta} \Big|_P. \quad (5.66)$$

这显然不是一个张量变换关系, 那么, 我们能取什么样的 $\partial \Gamma / \partial x$ 的线性组合, 使得结果在上述坐标变换下像一个张量那样变呢? 很明显, 我们必须能够消除这个变换中多出来的非齐次项 $\frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} \frac{\partial x^\kappa}{\partial x'^\eta} \frac{\partial^3 x'^\rho}{\partial x^\kappa \partial x^\alpha \partial x^\beta} \Big|_P$ 。

然而, 这个非齐次项关于指标 μ, ν, η 是完全任意的, 唯一的限制就是, 它关于这三个指标对称。因此, 要取 $\partial \Gamma / \partial x$ 的线性组合, 并使得它在一切满足 (5.65) 式的坐标变换下像一个张量一样变, 唯一的方法就是对 κ 和 α 指标 (或者等价地对 κ 和 β) 进行反对称化, 也就是考虑

$$T^\sigma_{\beta\kappa\alpha} = \frac{\partial \Gamma^\sigma_{\alpha\beta}}{\partial x^\kappa} - \frac{\partial \Gamma^\sigma_{\kappa\beta}}{\partial x^\alpha}. \quad (5.67)$$

从而 (5.66) 式就变成

$$T^\rho_{\nu\eta\mu} \Big|_P = \frac{\partial x'^\rho}{\partial x^\sigma} \frac{\partial x^\alpha}{\partial x'^\mu} \frac{\partial x^\beta}{\partial x'^\nu} \frac{\partial x^\kappa}{\partial x'^\eta} T^\sigma_{\beta\kappa\alpha} \Big|_P. \quad (5.68)$$

这就说明，在克里斯托夫联络 Γ 为零的点，我们所要构造的张量必定为上面的 $T_{\rho\mu\nu}^{\sigma}$ 。但在另外一方面，从 (5.67) 的表达式可以清楚地看到，在局部惯性系的 P 点， $T_{\rho\mu\nu}^{\sigma}$ 其实就是黎曼曲率张量 $R_{\rho\mu\nu}^{\sigma}$ ，即

$$T_{\rho\mu\nu}^{\sigma}|_P = R_{\rho\mu\nu}^{\sigma}|_P. \quad (5.69)$$

而作为张量等式，这就说明，我们所要构造的张量在任何坐标系中都必定为黎曼曲率张量 $R_{\rho\mu\nu}^{\sigma}$ 。这就证明了黎曼曲率张量的唯一性。

6. 坐标无关的数学语言初步

本章要初步解决一个数学问题。那就是，前面的章节中，我们都是通过引入局部坐标来进行数学分析的，然后再通过协变性和张量的概念来将这种局部的分析拼合成一个整体，这当然是一种极端重要的数学思想。但是，局部坐标本身却是人为的和任意的，真正的几何和物理应该不依赖于局部坐标，那么问题就是，如何才能用一种坐标无关的方式来讨论黎曼几何？

实际上，现代数学中有精确的概念来刻画广义相对论的时空模型，那就是所谓微分流形 (manifold) 的概念，不过本书并不需要微分流形的精确数学定义，感兴趣的读者可以参阅 Loring W. Tu 那本大受好评的 *An Introduction to Manifolds*，这本书的第二章就给出了流形的精确数学定义，所需的点集拓扑知识这本书也用附录的形式包含了。

简言之，流形就是一个可以任意选取局部坐标，并能用局部坐标来进行微积分计算，但真正的几何与拓扑却又不依赖于局部坐标具体选择的拓扑空间。在广义相对论中，时空就是一个流形，是一个带有赝黎曼度规张量结构的赝黎曼流形。所谓赝黎曼度规指的是广义相对论中的度规张量是闵可夫斯基符号度规，而不是标准黎曼几何的欧几里德符号度规。

本章假设讨论一般的 $D = n$ 维时空流形，记为 M 。

6.1 切向量场，余切向量场，张量场

6.1.1 切向量场

让我们先来讨论如何以坐标无关的方式处理时空中的逆变向量场。不妨把时空流形 M 上所有可能逆变向量场的集合记为 $\mathcal{X}(M)$ 。 $\mathcal{X}(M)$ 是一个无穷维向量空间，这是因为向

量场本身是时空点 x 的函数，这样不同向量场进行线性组合的时候组合系数也可以是点 x 的函数，具体来说，假设有两个任意的逆变向量场 $X^\mu(x)$ 和 $Y^\mu(x)$ ，则下面的线性组合也是一个向量场

$$a(x)X^\mu(x) + b(x)Y^\mu(x), \quad (6.1)$$

式中 $a(x)$ 和 $b(x)$ 是时空中的两个任意标量函数。由于组合系数可以依赖于 x ，所以这就带来无穷的组合可能性，使得 $\mathcal{X}(M)$ 成为一个无穷维向量空间。

为了使得向量空间变成有限维，我们可以把向量场限制在某个固定时空点 p 上，这时候向量场当然就成为 p 点的一个普通向量，我们把 p 点所有向量所构成的向量空间记为 V_p 。很明显，由于时空为 n 维，所以向量空间 V_p 也为 n 维向量空间。

但是上述对向量场以及 V_p 的讨论需要先取一个局部坐标 x ， p 点的向量 $X^\mu(x)|_p$ 依赖于局部坐标 x ，常常记为 $X_p^\mu(x)$ 。我们知道， $X_p^\mu(x)$ 在局部坐标变换 $x^\mu \rightarrow x'^\mu$ 之下，按如下方式变换

$$X_p'^\mu(x') = S^\mu{}_\nu(x)X_p^\nu(x), \quad (6.2)$$

式中变换矩阵 $S^\mu{}_\nu$ 为，

$$S^\mu{}_\nu(x) = \frac{\partial x'^\mu}{\partial x^\nu}. \quad (6.3)$$

这当然是以一种局部坐标依赖的方式讨论向量场以及 p 点的 V_p 。注意，这里 x 和 x' 描述的是同一时空点 p 。

而正如我们熟知的，在线性代数中，向量是坐标无关的，具体来说，在线性代数中，我们将一个向量 \mathbf{A} 写成

$$\mathbf{A} = A^\mu \mathbf{e}_\mu, \quad (6.4)$$

\mathbf{e}_μ 称作第 μ 个基矢， A^μ 就是向量 \mathbf{A} 在这一组基矢下的分量形式。而在线性代数中，所谓坐标变换就是换一组基矢，这时候向量的分量形式也会跟着变，但是向量 \mathbf{A} 本身是坐标无关的。

可见，要用坐标无关的方式来讨论向量场 X^μ ，我们就需要在向量空间 V_p 中引入基矢，并让坐标变换下基矢的改变正好补偿向量场分量形式的改变，以使得最终向量场本身与具体坐标无关。当然，现在相比线性代数复杂的是，变换矩阵 $S^\mu{}_\nu(x)$ 本身不是一个常数，而是依赖于时空坐标 x ，这就使得基矢 \mathbf{e}_μ 无法随意引入。但好在，我们注意到

$$\partial'_\mu|_p = \frac{\partial x^\nu}{\partial x'^\mu} \partial_\nu|_p \Rightarrow \partial'_\mu|_p = (S^{-1})^\nu{}_\mu(x) \partial_\nu|_p. \quad (6.5)$$

不难发现这个变换正好能补偿分量形式的变换 (6.2)。因此我们想到取

$$\mathbf{e}_\mu = \partial_\mu|_p, \quad (6.6)$$

反正根据线性代数的知识, 基矢 \mathbf{e}_μ 本身是一个抽象的客体, 我们对它并没有太多的限制, 因此当我们将 p 点的偏导算符 $\partial_\mu|_p$ 也看作一个抽象客体的时候, 当然就可以取 $\mathbf{e}_\mu = \partial_\mu|_p$ 。最后我们发现

$$X_p'^\mu(x')\partial'_\mu|_p = X_p^\mu(x)\partial_\mu|_p \quad (6.7)$$

是坐标无关的。由于最终的表达式是坐标无关的, 因此我们可以把 p 点处的这个向量记作 X_p ,

$$X_p = X_p^\mu(x)\partial_\mu|_p. \quad (6.8)$$

最后, 我们也可以将限制于 p 点这一条件去掉, 得到向量场 (而不是 p 点的向量) 的坐标无关表达式 X ,

$$X = X^\mu(x)\partial_\mu. \quad (6.9)$$

这么处理的一个额外好处是, 它使得向量场 X 具有了双重身份, 当将 $\partial_\mu|_p$ 看作抽象的基矢量时, X_p 是线性空间 V_p 中的一个向量。而当将 ∂_μ 看作偏导运算时, X 就成了一个偏导算符。因此, 给定时空中的标量函数 $f(x)$, 我们可以把偏导算符 X 作用上去

$$X(f) = X^\mu\partial_\mu f(x), \quad (6.10)$$

很显然, 这表示的是函数 $f(x)$ 沿着 X^μ 方向的方向导数。作为偏导算符, X 当然满足求导的一系列法则, 比方说满足莱布尼兹法则, 即, 对于时空中两个任意函数 $f(x), g(x)$, 我们有

$$X(fg) = X(f)g + fX(g). \quad (6.11)$$

给定向量场 $X = X^\mu\partial_\mu$ 、 $Y = Y^\mu\partial_\mu$, 我们可以计算这两个偏导算符的对易子, $[X, Y] = XY - YX$, 称作这两个向量场的李括号, 并且显然有

$$[X, Y] = -[Y, X]. \quad (6.12)$$

为了计算出 $[X, Y]$, 我们可以将它作用在标量函数 $f(x)$ 上,

$$[X, Y](f) = X(Y(f)) - Y(X(f)) = [X^\nu\partial_\nu(Y^\mu) - Y^\nu\partial_\nu(X^\mu)]\partial_\mu f. \quad (6.13)$$

由此不难看出, $[X, Y]$ 依然是一个向量场, 这个向量场的 μ 分量 (记作 $[X, Y]^\mu$) 为

$$[X, Y]^\mu = X^\nu\partial_\nu Y^\mu - Y^\nu\partial_\nu X^\mu. \quad (6.14)$$

与第四章引入的李导数进行比较, 不难发现

$$[X, Y]^\mu = \mathcal{L}_X Y^\mu. \quad (6.15)$$

人们也常常以一种坐标无关的方式来讨论李导数，即定义

$$\mathcal{L}_X Y \equiv (\mathcal{L}_X Y^\mu) \partial_\mu. \quad (6.16)$$

如此一来，即有

$$\mathcal{L}_X Y = [X, Y]. \quad (6.17)$$

两个向量场的李括号，结果依然是一个向量场。而且很容易证明，向量场的李括号满足所谓的雅可比恒等式，即任给三个向量场 X, Y, Z ，我们有

$$[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0. \quad (6.18)$$

定义了李括号运算并且满足雅可比恒等式的向量空间称作李代数。所以，时空中所有向量场的集合 $\mathcal{X}(M)$ 在李括号下构成一个无穷维李代数。

时空流形上的向量场直观上看起来像什么样子呢？为了弄清楚这个问题，不妨举一个例子，假设我们的时空流形是一个两维球面，即 $M = S^2$ ，而我们可以超脱出这个时空之外来观察它，具体来说，就是设想把这个两维球面放在三维空间中来观察它。这时候时空中的向量场直观上应该怎么样呢？很显然它们应该躺在这个两维球面上，换言之，在球面的每一点，向量场都应该和球面相切，看起来大体如图 (6.1) 所示。这个例子告诉我们，

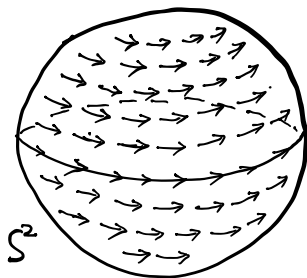


Figure 6.1: 球面上的切向量场。

时空中诸如 $X = X^\mu \partial_\mu$ 这样的向量场，从时空之外的更高维度来看，它们在每一点都与时空流形相切，因此数学上也把这样的向量场称作切向量场。给定一个时空点 p ，该点处所有可能切向量构成一个向量空间，称作 p 点的切空间，当然它就是前面的向量空间 V_p 。

对于时空流形中一条以 λ 为参数的曲线 $x^\mu(\lambda)$ ，我们也可以定义其切矢量 T ，用分量形式来说即是 $T^\mu = \frac{dx^\mu}{d\lambda}$ ，因此用坐标无关的写法即是

$$T = T^\mu \partial_\mu = \frac{dx^\mu}{d\lambda} \partial_\mu = \frac{\partial}{\partial \lambda} = \partial_\lambda. \quad (6.19)$$

当然这个矢量只沿着曲线才有定义，因此不是一个向量场。特别的，如果这条曲线是粒子运动的世界线，并且将粒子的固有时 τ 取作世界线参数，那粒子的四维速度矢量 u 即是

$$u = \frac{dx^\mu}{d\tau} \partial_\mu = \frac{\partial}{\partial \tau} = \partial_\tau. \quad (6.20)$$

以上，我们只讨论了如何用坐标无关的方式来处理上指标的逆变矢量场。结论就是，这样的逆变矢量场其实应该理解为时空流形上的切向量场。那诸如 $\omega_\mu(x)$ 这样的下指标的协变矢量场用坐标无关的方式又该如何处理呢？

6.1.2 余切向量场

我们可以用完全类似的方式处理时空中的协变矢量场。为此只需要注意到 $\omega_\mu(x)$ 在坐标变换下按如下方式变换

$$\omega'_\mu(x') = (S^{-1})^\nu{}_\mu(x)\omega_\nu(x), \quad (6.21)$$

而坐标微分 dx^μ 按如下方式变换

$$dx'^\mu = S^\mu{}_\nu(x)dx^\nu, \quad (6.22)$$

从而下式是一个坐标无关的向量场

$$\omega = \omega_\mu(x)dx^\mu, \quad (6.23)$$

称作时空中的余切向量场。

ω 限制在时空点 p 得到的向量叫做余切向量, 记作 ω_p , p 点所有余切向量构成的空间称作 p 点的余切空间, 记为 V_p^* 。很显然

$$\omega_p = \omega_\mu(x)|_p dx^\mu|_p, \quad (6.24)$$

也即是说, $dx^\mu|_p$ 作为一个抽象的代数客体, 可以看成是余切空间 V_p^* 的一组基矢量。

利用向量场的缩并, 我们可以把 p 点的余切向量 ω_p 和切向量 X_p 缩并在一起, 形成一个坐标无关的量, 也即

$$(\omega_\mu X^\mu)|_p, \quad (6.25)$$

很显然, 给定 ω_p , 上面这个表达式对于切向量 X_p 是线性依赖的, 可以看作是 X_p 的线性函数。换言之, 我们可以将余切向量 ω_p 看作是切向量的线性函数, 即定义

$$\omega_p(X_p) = (\omega_\mu X^\mu)|_p, \quad (6.26)$$

很明显, 这满足线性函数的性质, 即对于任意 X_p, Y_p , 以及组合系数 a_p, b_p , 我们有

$$\omega_p(a_p X_p + b_p Y_p) = a_p \omega_p(X_p) + b_p \omega_p(Y_p). \quad (6.27)$$

但是, 表达式 (6.25) 关于 ω_p 和 X_p 完全对称, 因此给定 X_p , 这个表达式关于 ω_p 也是线性依赖的。所以我们可以把切向量 X_p 看成是余切向量的线性函数, 即定义

$$X_p(\omega_p) = (\omega_\mu X^\mu)|_p = \omega_p(X_p). \quad (6.28)$$

它满足线性函数的性质, 即对于任意余切向量 ω_p , 和 $\phi_p = (\phi_\mu dx^\mu)|_p$, 有

$$X_p(a_p \omega_p + b_p \phi_p) = a_p X_p(\omega_p) + b_p X_p(\phi_p). \quad (6.29)$$

以上分析告诉我们, 余切空间 V_p^* 可以看成是切空间 V_p 上的所有线性函数所构成的向量空间, 反过来也一样, 切空间 V_p 也可以看作是余切空间 V_p^* 上所有线性函数所构成的向量空间, 因此, 数学上常常称这两个向量空间互为对偶空间, 记作

$$V_p \leftrightarrow V_p^*. \quad (6.30)$$

给定余切向量 $\omega_p = \omega_\mu|_p dx^\mu|_p$ 和切向量 $X_p = X_p^\mu \partial_\mu|_p$, 按照线性函数的理解, 我们有

$$\begin{aligned}\omega_p(X_p) &= \omega_p(X_p^\nu \partial_\nu|_p) = X_p^\nu \omega_p(\partial_\nu|_p) \\ &= X_p^\nu \omega_\mu|_p dx^\mu|_p(\partial_\nu|_p),\end{aligned}\quad (6.31)$$

与 $\omega_p(X_p)$ 的定义式 (6.26) 相比较, 即有

$$dx^\mu|_p(\partial_\nu|_p) = \delta_\nu^\mu. \quad (6.32)$$

当然, 同样也有

$$\partial_\nu|_p(dx^\mu|_p) = dx^\mu|_p(\partial_\nu|_p) = \delta_\nu^\mu. \quad (6.33)$$

因此通常称 $dx^\mu|_p$ 和 $\partial_\mu|_p$ 互为对偶基矢量。

注意到 (6.32) 式和 (6.33) 等号右边与 p 无关, 因此我们可以去除限制于 p 点这一要求, 得到

$$dx^\mu(\partial_\nu) = \delta_\nu^\mu. \quad (6.34)$$

进而我们也可以在线性函数的定义式 (6.26) 和 (6.28) 中去除限制于 p 点这一要求, 进而定义时空中的标量 $\omega(X) = X(\omega)$ 如下

$$\omega(X) = \omega_\mu(x)X^\mu(x) = X(\omega). \quad (6.35)$$

也即是说, 我们可以把余切向量场看作是切向量场的线性函数, 满足如下线性关系, 即对于任意标量函数 $f(x)$ 有

$$\omega(fX) = f\omega(X). \quad (6.36)$$

同样, 切向量场也是余切向量场的线性函数。

6.1.3 张量场

下面我们来讨论如何处理时空流形上的一般张量场。设要处理的张量场为 (k, l) 阶张量场

$$T^{\mu_1 \dots \mu_k}{}_{\nu_1 \dots \nu_l}(x). \quad (6.37)$$

与切向量场一样, 所有 (k, l) 阶张量场的集合也构成一个无穷维线性空间。

与对切向量场的讨论一样, 我们将 $T^{\mu_1 \dots \mu_k}{}_{\nu_1 \dots \nu_l}(x)$ 限制在时空点 p , 即考察 p 点的 (k, l) 张量 (不是张量场) $T^{\mu_1 \dots \mu_k}{}_{\nu_1 \dots \nu_l}|_p$, 所有这样的张量当然构成一个有限维线性空间。我们把这个线性空间的基矢量记为

$$(\partial_{\mu_1} \otimes \dots \otimes \partial_{\mu_k} \otimes dx^{\nu_1} \otimes \dots \otimes dx^{\nu_l})|_p. \quad (6.38)$$

式中 \otimes 只是一个抽象的记号, 用来把多重切空间基矢和多重余切空间基矢并置在一起, 通常称这一记号为张量积。显然, 这样的独立基矢一共有 n^{k+l} 个, 所以 p 点的 (k, l) 张量所构成的线性空间为 n^{k+l} 维。记 $T^{\mu_1 \dots \mu_k}{}_{\nu_1 \dots \nu_l}|_p$ 的坐标无关表达为 T_p , 很显然

$$T_p = T^{\mu_1 \dots \mu_k}{}_{\nu_1 \dots \nu_l}|_p (\partial_{\mu_1} \otimes \dots \otimes \partial_{\mu_k} \otimes dx^{\nu_1} \otimes \dots \otimes dx^{\nu_l})|_p. \quad (6.39)$$

通常把所有可能 T_p 的线性空间称作 k 个 V_p 和 l 个 V_p^* 的张量积空间, 记为

$$V_p \otimes \dots \otimes V_p \otimes V_p^* \otimes \dots \otimes V_p^*. \quad (6.40)$$

去除掉限制于 p 点这一条件, 即可以得到 (k, l) 张量场的坐标无关表达, 为

$$T = T^{\mu_1 \dots \mu_k}{}_{\nu_1 \dots \nu_l}(x) (\partial_{\mu_1} \otimes \dots \otimes \partial_{\mu_k} \otimes dx^{\nu_1} \otimes \dots \otimes dx^{\nu_l}). \quad (6.41)$$

当然, 和前面切向量场以及余切向量场的情况类似, 我们也可以把张量场理解成某种线性函数。举例来说, 比如对于 $(0, l)$ 阶张量场 $\psi_{\nu_1 \dots \nu_l}$, 我们可以把它理解为切向量场的 l 重线性函数, 也即是, 对于 l 个切向量场 X_1, X_2, \dots, X_l , 我们定义

$$\psi(X_1, X_2, \dots, X_l) = \psi_{\nu_1 \nu_2 \dots \nu_l} X_1^{\nu_1} X_2^{\nu_2} \dots X_l^{\nu_l}. \quad (6.42)$$

特别的, 度规场是一个 $(0, 2)$ 型张量场, 记作 $g = g_{\mu\nu} dx^\mu \otimes dx^\nu$, 它可以理解成切向量场的二重线性函数, 即

$$g(X, Y) = g_{\mu\nu} X^\mu Y^\nu. \quad (6.43)$$

由于度规场的指标对称性质, 显然有

$$g(X, Y) = g(Y, X). \quad (6.44)$$

利用度规场 g , 我们可以把任意切向量场映射为一个余切向量场, 具体来说, 对于任意切向量场 X , 我们可以定义余切向量场

$$\begin{aligned} g(X, \cdot) &= g_{\mu\nu} dx^\mu \otimes dx^\nu (X, \cdot) = g_{\mu\nu} dx^\mu (X) dx^\nu (\cdot) \\ &= g_{\mu\nu} X^\mu dx^\nu (\cdot) = X_\nu dx^\nu (\cdot), \end{aligned} \quad (6.45)$$

式中 \cdot 表示一个可以插入任何切向量场的空槽。很显然, 这就是我们早已经熟悉的降指标操作。同样, 利用 $g^{\mu\nu} \partial_\mu \otimes \partial_\nu$, 我们也可以将任意余切向量场映射为切向量场, 也就是我们熟悉的升指标操作。

另一种处理度规场的方式是引入向量场内积的概念。对于任意两个向量场 X 和 Y , 我们定义其内积 $\langle X, Y \rangle$ 如下,

$$\langle X, Y \rangle \equiv g(X, Y) = g_{\mu\nu} X^\mu Y^\nu. \quad (6.46)$$

不难看出这样定义的内积的确满足对向量内积的通常要求, 具体来说即是满足: 1. **对称性**, 即有

$$\langle X, Y \rangle = \langle Y, X \rangle. \quad (6.47)$$

2. **双线性**，即对于任意三个向量场 X, Y, Z ，以及任意两个标量函数 $a(x)$ 和 $b(x)$ ，均有

$$\langle X, aY + bZ \rangle = a\langle X, Y \rangle + b\langle X, Z \rangle. \quad (6.48)$$

注意，这里我们并不要求正定性，即不要求 $\langle X, X \rangle \geq 0$ ，这是因为广义相对论中的度规是闵氏符号度规，而不是正定的欧氏符号度规。

根据内积的双线性，我们又有

$$\langle X, Y \rangle = \langle X^\mu \partial_\mu, Y^\nu \partial_\nu \rangle = X^\mu Y^\nu \langle \partial_\mu, \partial_\nu \rangle, \quad (6.49)$$

与上面内积的定义式 (6.46) 比较，即有

$$g_{\mu\nu} = \langle \partial_\mu, \partial_\nu \rangle, \quad (6.50)$$

即是说，度规张量的各分量其实是切向量场相应基矢之间的内积。

6.2 李导数与微分同胚群

这一节我们来讨论时空流形 M 的微分同胚变换以及李导数概念。假设 Φ 为时空流形 M 上的一个微分同胚映射，它把任意的 x 点映射到某个 x' 点，即 $\Phi: x \rightarrow x'$ 。由于这个映射是可微的一一对映，所以可以把映射的箭头倒过来，得到一个逆映射 Φ^{-1} ， $\Phi^{-1}: x' \rightarrow x$ 。很明显 Φ 和 Φ^{-1} 的复合映射 (记作 $\Phi^{-1} \circ \Phi$) 是一个恒等映射 (记为 id)，所以

$$\Phi^{-1} \circ \Phi = \Phi \circ \Phi^{-1} = id. \quad (6.51)$$

记流形 M 上所有可能微分同胚映射的全体为 $\text{Diff}(M)$ 。上面那一段实际上引入了 $\text{Diff}(M)$ 不同元素间 (即不同微分同胚映射间) 的一种“乘法”，即映射的复合，具体来说，假设有两个微分同胚映射 Φ_1 和 Φ_2 ，我们称这两个映射的复合 (先进行 Φ_1 再进行 Φ_2)，即 $\Phi_2 \circ \Phi_1$ ，为 Φ_1 和 Φ_2 “相乘”。很显然 $\Phi_2 \circ \Phi_1$ 依然是一个微分同胚映射，所以 $\text{Diff}(M)$ 在映射复合的“乘法”下是封闭的。而且上一段也说明了， $\text{Diff}(M)$ 的每一个元素在映射复合的“乘法”下都存在一个逆元素。数学上称具备这些性质的集合为一个群。所以，流形 M 上所有可能微分同胚映射的集合构成一个群，称作 M 的微分同胚群，也即是 $\text{Diff}(M)$ 。

回想一下第四章中关于无穷小微分同胚映射与李导数的讨论可以知道，李导数与连续的微分同胚映射密切相关。具体来说，假设有某个切向量场 X ，那我们就可以在流形 M 上画出一些参数为 τ 的流线，并让 X 是这些流线的切向量，即满足 $X = \partial_\tau$ 。如此一来，时空点沿着这些流线“流动”的过程就生成了一簇依赖于连续参数 τ 的微分同胚映射 Φ_τ ，它当然是由切向量场 X 生成的。由于每个切向量场都能生成一簇单参微分同胚映射，所以也称切向量场为微分同胚群的生成元。

现在，设 $\tau = 0$ 时的 $\Phi_0 = id$ ，则一个张量场沿着向量场 X 的李导数 \mathcal{L}_X ，就是 τ 取无穷小量 ε 时，张量场在 Φ_τ 作用前后的无穷小改变量与 ε 的比值。有时也称李导数 \mathcal{L}_X 为微分同胚群在张量场上作用的生成元。

下面我们先回想一下李导数的几个性质：首先，任给向量场 X 和 Y ，以及常数 c_1, c_2 ，有

$$\mathcal{L}_{c_1X+c_2Y} = c_1\mathcal{L}_X + c_2\mathcal{L}_Y. \quad (6.52)$$

注意，这个性质只对常数 c_1, c_2 成立， c_1, c_2 为标量函数的话结果并不成立，这和上一节讨论的张量场作为线性函数的那种线性关系是不同的。其次，李导数满足莱布尼兹法则，包括对两个张量场的缩并运算也同样满足莱布尼兹法则。第三，对于任意标量函数 $f(x)$ 的李导数为

$$\mathcal{L}_X f = X^\rho \partial_\rho f = X(f). \quad (6.53)$$

第四，对于切向量场 Y 的李导数为

$$\mathcal{L}_X Y = [X, Y]. \quad (6.54)$$

前三条性质来自于第四章的相关讨论，第四条性质是本章前面讲过的。实际上，反过来，这些性质也完全确定了任意张量场的李导数。

下面我们证明李导数的一个重要代数关系，即

$$[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_{[X, Y]}. \quad (6.55)$$

要证明这个等式，只需证明它在任意张量场上的作用都是成立的。

具体证明如下：第一步，我们证明上式作用在标量函数 $f(x)$ 上是成立的，因为等式左边的作用为

$$[\mathcal{L}_X, \mathcal{L}_Y]f = (\mathcal{L}_X \mathcal{L}_Y - \mathcal{L}_Y \mathcal{L}_X)f = X(Y(f)) - Y(X(f)). \quad (6.56)$$

而等式右边的作用也为

$$\mathcal{L}_{[X, Y]}f = [X, Y](f) = X(Y(f)) - Y(X(f)). \quad (6.57)$$

所以左右两边相等。

第二步，我们证明 (6.55) 在任意切向量场 Z 上的作用是成立的，因为等式左边的作用为

$$[\mathcal{L}_X, \mathcal{L}_Y]Z = (\mathcal{L}_X \mathcal{L}_Y - \mathcal{L}_Y \mathcal{L}_X)Z = [X, [Y, Z]] - [Y, [X, Z]]. \quad (6.58)$$

而等式右边的作用为

$$\mathcal{L}_{[X, Y]}Z = [[X, Y], Z] = -[Z, [X, Y]]. \quad (6.59)$$

又由切向量场的雅可比恒等式，我们有

$$\begin{aligned} [X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] &= 0 \\ \Rightarrow [X, [Y, Z]] - [Y, [X, Z]] &= -[Z, [X, Y]], \end{aligned} \quad (6.60)$$

所以 (6.55) 式在 Z 上的作用成立。

第三步, 我们再利用李导数的莱布尼兹法则, 将 (6.55) 式推广到对任意余切向量场 $\omega = \omega_\mu dx^\mu$ 的作用情况。为此只需注意到, 根据两个缩并张量的莱布尼兹法则

$$\mathcal{L}_X(\omega(Z)) = (\mathcal{L}_X\omega)(Z) + \omega(\mathcal{L}_XZ), \quad (6.61)$$

从而可以得到

$$[\mathcal{L}_X, \mathcal{L}_Y](\omega(Z)) = ([\mathcal{L}_X, \mathcal{L}_Y]\omega)(Z) + \omega([\mathcal{L}_X, \mathcal{L}_Y]Z). \quad (6.62)$$

注意, $[\mathcal{L}_X, \mathcal{L}_Y]$ 关于 X, Y 反对称, 所以在推导过程中, 诸如 $(\mathcal{L}_Y\omega)(\mathcal{L}_XZ) + (\mathcal{L}_X\omega)(\mathcal{L}_YZ)$ 这样的交叉项都会自动消掉。注意到 $\omega(Z)$ 是一个标量, 而 Z 是一个切向量场, 所以根据证明的前两步, 可以把 (6.62) 式重写为

$$\mathcal{L}_{[X,Y]}(\omega(Z)) = ([\mathcal{L}_X, \mathcal{L}_Y]\omega)(Z) + \omega(\mathcal{L}_{[X,Y]}Z). \quad (6.63)$$

另一方面, 根据莱布尼兹法则, 我们又有

$$\mathcal{L}_{[X,Y]}(\omega(Z)) = (\mathcal{L}_{[X,Y]}\omega)(Z) + \omega(\mathcal{L}_{[X,Y]}Z). \quad (6.64)$$

将上面两个式子进行比较, 可得, 对于任意 Z , 有

$$([\mathcal{L}_X, \mathcal{L}_Y]\omega)(Z) = (\mathcal{L}_{[X,Y]}\omega)(Z). \quad (6.65)$$

由于 Z 任意, 所以这就说明

$$[\mathcal{L}_X, \mathcal{L}_Y]\omega = \mathcal{L}_{[X,Y]}\omega. \quad (6.66)$$

最后一步, 完全类似于第三步, 可以证明 (6.55) 式在任意张量场上的作用都成立。举例来说, 比方要证明 (6.55) 式在 $(0,2)$ 型张量场 $\theta = \theta_{\mu\nu} dx^\mu \otimes dx^\nu$ 上的作用成立, 则我们只需考察标量 $\theta(Z, W) = \theta_Z(W)$, 其中 $\theta_Z = (\theta_{\mu\nu} Z^\mu) dx^\nu = \theta(Z, \cdot)$ 为一个余切向量场, 然后对 $\theta_Z(W)$ 引用第三步的证明, 得到 (6.55) 式在 $\theta_Z = \theta(Z, \cdot)$ 上的作用成立, 最后再对 $\theta(Z, \cdot)$ 再次引用第三步的证明过程, 得到 (6.55) 式在 θ 上的作用成立, 这样就完成了所需的证明。对于更一般的任意张量场情形, 可以考虑将类似的推理过程写成数学归纳法的形式, 并对张量场的阶数进行归纳, 从而完成证明。

由 (6.55) 不难验证, 李导数也满足雅可比恒等式, 证明如下

$$\begin{aligned} & [[\mathcal{L}_X, \mathcal{L}_Y], \mathcal{L}_Z] + [[\mathcal{L}_Y, \mathcal{L}_Z], \mathcal{L}_X] + [[\mathcal{L}_Z, \mathcal{L}_X], \mathcal{L}_Y] \\ &= [\mathcal{L}_{[X,Y]}, \mathcal{L}_Z] + [\mathcal{L}_{[Y,Z]}, \mathcal{L}_X] + [\mathcal{L}_{[Z,X]}, \mathcal{L}_Y] \\ &= \mathcal{L}_{[[X,Y],Z]} + \mathcal{L}_{[[Y,Z],X]} + \mathcal{L}_{[[Z,X],Y]} \\ &= \mathcal{L}_{[[X,Y],Z] + [[Y,Z],X] + [[Z,X],Y]} = \mathcal{L}_0 = 0. \end{aligned} \quad (6.67)$$

上式的最后一行利用了切向量场的雅可比恒等式。这个结果说明, 微分同胚群在张量场上作用的所有可能生成元 (即李导数) 构成一个李代数。事实上, 上面的 (6.55) 式告诉我们,

李导数所构成的这个李代数与切向量场所构成的李代数是同态的, 所以当然也是一个无穷维李代数。

切向量场是微分同胚群 $\text{Diff}(M)$ 的生成元, 而所有连续微分同胚映射构成的群是一个连续群, 或者说李群, 一个数学上熟知的事实是, 李群的生成元必然构成李代数, 微分同胚群所对应的李代数就是切向量场在李括号运算下所构成的那个无穷维李代数。反过来, 李代数对应的群必定是李群, 无穷维李代数对应的李群就是无穷维李群。因此, 流形 M 的微分同胚群 $\text{Diff}(M)$ 必定为无穷维李群。这个无穷维李群在张量场上的作用称之为微分同胚群的表示, 李导数就是这表示的生成元。因此李导数当然也会构成一个无穷维李代数, 并且这个李代数必然与微分同胚群本身的李代数同态。这就是对上述关于李导数的两个结果 (6.55) 和 (6.67) 本质性的理解。

6.3 协变导数, 黎曼曲率张量

为了用一种坐标无关的方式处理协变导数, 我们可以将某切向量场 X^μ 与协变导数 D_μ 进行缩并, 定义沿着 X 方向的协变导数 D_X

$$D_X \equiv X^\mu D_\mu. \quad (6.68)$$

很显然, D_X 作用在标量函数 $f(x)$ 上有

$$D_X f = X^\mu D_\mu f = X^\mu \partial_\mu f = X(f). \quad (6.69)$$

为了将协变导数作用在某个切向量场 Y 上, 我们定义

$$D_X Y \equiv (D_X Y^\mu) \partial_\mu, \quad (6.70)$$

因此作用的结果依然是一个切向量场。

很显然, D_X 对于向量场 X 是线性依赖的, 即任给两个切向量场 X 和 Y , 以及任给两个标量函数 $a(x), b(x)$, 均有

$$D_{aX+bY} = aD_X + bD_Y. \quad (6.71)$$

特别的, 假设沿着某粒子的世界线 $x^\mu(\tau)$ (粒子的四速度为 $u = \partial_\tau$) 平行移动某个向量场 X , 则根据第四章中引入的向量场平行移动方程, 我们有

$$\frac{DX}{D\tau} = u^\mu D_\mu X = D_u X = 0. \quad (6.72)$$

这就是坐标无关写法下的平行移动方程。特别的, 如果我们平行移动的对象是四速度本身, 那就得到测地线方程

$$D_u u = 0. \quad (6.73)$$

这就是坐标无关语言中的测地线方程。

任给两个切向量场 X 和 Y , 我们可以进行如下推导

$$\begin{aligned}
 D_X Y - D_Y X &= [X^\nu (D_\nu Y^\mu) - Y^\nu (D_\nu X^\mu)] \partial_\mu \\
 &= [X^\nu (\partial_\nu Y^\mu) + X^\nu \Gamma_{\nu\rho}^\mu Y^\rho - Y^\nu (\partial_\nu X^\mu) - Y^\nu \Gamma_{\nu\rho}^\mu X^\rho] \partial_\mu \\
 &= [X^\nu (\partial_\nu Y^\mu) - Y^\nu (\partial_\nu X^\mu) + X^\rho \Gamma_{\rho\nu}^\mu Y^\nu - Y^\nu \Gamma_{\nu\rho}^\mu X^\rho] \partial_\mu \\
 &= [X^\nu (\partial_\nu Y^\mu) - Y^\nu (\partial_\nu X^\mu)] \partial_\mu = [X, Y].
 \end{aligned} \tag{6.74}$$

上式倒数第二行我们利用了克里斯托夫联络是一个对称联络, 满足 $\Gamma_{\rho\nu}^\mu = \Gamma_{\nu\rho}^\mu$ 。也即是说, 我们有

$$D_X Y - D_Y X = [X, Y]. \tag{6.75}$$

当然, 正如上面推导过程所显示的, 以上结果完全因为广义相对论中的克里斯托夫联络是一个对称联络。如果我们是用一个不对称的仿射联络 $\tilde{\Gamma}_{\mu\nu}^\rho$ 来定义协变导数, 那就没有上面的结果了, 这时候我们可以考察

$$T(X, Y) \equiv D_X Y - D_Y X - [X, Y], \tag{6.76}$$

完全类似于前面的推导, 不难得到

$$T(X, Y) = X^\mu Y^\nu [\tilde{\Gamma}_{\mu\nu}^\rho - \tilde{\Gamma}_{\nu\mu}^\rho] \partial_\rho. \tag{6.77}$$

根据第四章中给出的仿射联络在坐标变换之下的变换关系不难证明, 虽然 $\tilde{\Gamma}_{\mu\nu}^\rho$ 不是一个张量, 但是 $[\tilde{\Gamma}_{\mu\nu}^\rho - \tilde{\Gamma}_{\nu\mu}^\rho]$ 却是一个张量, 通常称作挠率张量, 上面给出的 $T(X, Y)$ 实际上就是挠率张量的坐标无关化处理。当然, 在广义相对论中, 由于克里斯托夫联络是对称的, 因此**挠率张量恒等于零**, 也就是恒有 (6.75) 式成立。

由于度规张量的协变导数为零, 即 $D_X g_{\mu\nu} = 0$ (即是所谓的联络与度规张量的相容性条件), 从而

$$\begin{aligned}
 X(\langle Y, Z \rangle) &= X(g_{\mu\nu} Y^\mu Z^\nu) = D_X (g_{\mu\nu} Y^\mu Z^\nu) \\
 &= (D_X g_{\mu\nu}) Y^\mu Z^\nu + g_{\mu\nu} (D_X Y^\mu) Z^\nu + g_{\mu\nu} Y^\mu (D_X Z^\nu) \\
 &= g_{\mu\nu} (D_X Y^\mu) Z^\nu + g_{\mu\nu} Y^\mu (D_X Z^\nu) \\
 &= \langle D_X Y, Z \rangle + \langle Y, D_X Z \rangle.
 \end{aligned} \tag{6.78}$$

也即

$$X(\langle Y, Z \rangle) = \langle D_X Y, Z \rangle + \langle Y, D_X Z \rangle. \tag{6.79}$$

这就是联络与度规张量相容性条件的坐标无关化处理, 注意, 这个结果只对克里斯托夫联络成立。

为了用坐标无关的方式处理黎曼曲率张量, 完全类似于上一章引入黎曼曲率张量时的推导, 我们注意到

$$[D_\mu, D_\nu] Y^\rho = R^\rho{}_{\sigma\mu\nu} Y^\sigma. \tag{6.80}$$

即是说, 黎曼曲率张量完全取决于协变导数的算符对易子 $[D_\mu, D_\nu]$ 。为了用坐标无关的方式处理这个算符对易子, 我们将它和两个切向量场 Z, W 进行缩并, 即定义如下曲率算子

$$R(Z, W) \equiv Z^\mu W^\nu [D_\mu, D_\nu]. \quad (6.81)$$

显然这样定义的曲率算子对于 Z, W 都是线性依赖的, 因此是切向量场的二重线性算符函数。并且很显然有

$$R(Z, W) = -R(W, Z). \quad (6.82)$$

下面我们来证明 $R(Z, W)$ 可以重写为如下形式

$$R(Z, W) = D_Z D_W - D_W D_Z - D_{[Z, W]}. \quad (6.83)$$

为了证明这一点, 显然只需要证明如下等式成立

$$D_Z D_W - D_W D_Z = Z^\mu W^\nu [D_\mu, D_\nu] + D_{[Z, W]}. \quad (6.84)$$

根据算符的作用规则, 可以证明如下

$$\begin{aligned} D_Z D_W - D_W D_Z &= D_Z W^\nu D_\nu - D_W Z^\mu D_\mu \\ &= (D_Z W^\nu) D_\nu - (D_W Z^\mu) D_\mu + W^\nu D_Z D_\nu - Z^\mu D_W D_\mu \\ &= D_{D_Z W} - D_{D_W Z} + Z^\mu W^\nu [D_\mu, D_\nu] \\ &= D_{D_Z W - D_W Z} + Z^\mu W^\nu [D_\mu, D_\nu] \\ &= D_{[Z, W]} + Z^\mu W^\nu [D_\mu, D_\nu]. \end{aligned} \quad (6.85)$$

上式的最后一个等号利用了无挠条件 (6.75)。值得说明的是, 为了得到 (6.83) 式我们虽然利用了无挠条件, 但这完全是因为我们对协变导数的定义是通过分量形式来引入的, 这和数学书中的处理有些微妙的区别, 结果就是, 如果完全按照数学家的定义方式, (6.83) 式其实并不需要无挠条件成立, 因此可以推广到克里斯托夫联络之外的任何仿射联络。

注意, 我们不能把曲率算子定义成 $D_Z D_W - D_W D_Z$, 因为任给一个标量函数 $f(x)$,

$$D_{fZ} D_W - D_W D_{fZ} \neq f(x) (D_Z D_W - D_W D_Z), \quad (6.86)$$

从而如果这样定义的话, 那曲率算子就不是切向量场的二重线性算符函数, 这是我们所不希望的。

为了将曲率算子和分量形式的黎曼曲率张量具体联系起来, 我们将它作用在切向量场 Y 上, 即定义

$$R(Z, W)Y \equiv (R(Z, W)Y^\rho) \partial_\rho. \quad (6.87)$$

从而有

$$\begin{aligned} R(Z, W)Y &= (R(Z, W)Y^\rho) \partial_\rho = (Z^\mu W^\nu [D_\mu, D_\nu] Y^\rho) \partial_\rho \\ &= (Y^\sigma Z^\mu W^\nu R^\rho_{\sigma\mu\nu}) \partial_\rho. \end{aligned} \quad (6.88)$$

进而有

$$\begin{aligned}\langle X, R(Z, W)Y \rangle &= \langle X^\rho \partial_\rho, (Y^\sigma Z^\mu W^\nu R^\lambda{}_{\sigma\mu\nu}) \partial_\lambda \rangle \\ &= X^\rho Y^\sigma Z^\mu W^\nu R^\lambda{}_{\sigma\mu\nu} \langle \partial_\rho, \partial_\lambda \rangle \\ &= X^\rho Y^\sigma Z^\mu W^\nu R_{\rho\sigma\mu\nu}.\end{aligned}\quad (6.89)$$

显然，这是切向量场的一个四重线性函数，通常记为 $R(X, Y, Z, W)$ ，即

$$R(X, Y, Z, W) \equiv X^\rho Y^\sigma Z^\mu W^\nu R_{\rho\sigma\mu\nu} = \langle X, R(Z, W)Y \rangle. \quad (6.90)$$

至此我们就完成了黎曼曲率张量的坐标无关化处理。根据黎曼曲率张量的代数性质，我们有

$$\begin{aligned}R(X, Y, Z, W) &= -R(Y, X, Z, W) \\ R(X, Y, Z, W) &= -R(X, Y, W, Z) \\ R(X, Y, Z, W) &= R(Z, W, X, Y) \\ R(X, Y, Z, W) + R(X, Z, W, Y) + R(X, W, Y, Z) &= 0.\end{aligned}\quad (6.91)$$

根据上面性质，我们可以定义一个切向量场上的对称型 $Q(X, Y)$

$$Q(X, Y) = R(X, Y, X, Y), \quad (6.92)$$

称作黎曼曲率张量的相配对称型。并且，由于曲率张量的代数性质，数学上可以证明（见附录）， $Q(X, Y)$ 完全确定了黎曼曲率张量。

如果 Π 是 p 点切空间 V_p 的一个二维子空间，设 $\{v_1, v_2\}$ 是 Π 的任意一组基矢量，我们定义 Π 的截面曲率为

$$K(\Pi) = \frac{R(v_1, v_2, v_1, v_2)}{|v_1 \wedge v_2|^2}, \quad (6.93)$$

式中 $|v_1 \wedge v_2|^2 = \langle v_1, v_1 \rangle \langle v_2, v_2 \rangle - \langle v_1, v_2 \rangle^2$ ，可以理解为切向量 v_1, v_2 所构成的平行四边形面积平方。利用黎曼曲率张量的代数性质 (6.91) 可以直接证明：截面曲率 $K(\Pi)$ 只依赖于两维切平面 Π ，与 Π 上基矢量 $\{v_1, v_2\}$ 的具体选取无关。

为了证明 $K(\Pi)$ 与基矢的选取无关，我们在 Π 平面上重新选择一组基矢 $\{w_1, w_2\}$ ，则存在实数 a, b, c, d 使得

$$w_1 = av_1 + bv_2, \quad w_2 = cv_1 + dv_2, \quad ad - bc \neq 0. \quad (6.94)$$

此时，直接的计算表明

$$|w_1 \wedge w_2|^2 = (ad - bc)^2 |v_1 \wedge v_2|^2. \quad (6.95)$$

另一方面，根据 $R(X, Y, Z, W)$ 关于 X, Y, Z, W 的四重线性性，以及代数性质 $R(Y, X, Z, W) = -R(X, Y, Z, W)$ 和 $R(X, Y, W, Z) = -R(X, Y, Z, W)$ ，可得

$$\begin{aligned}R(w_1, w_2, w_1, w_2) &= R(av_1 + bv_2, cv_1 + dv_2, av_1 + bv_2, cv_1 + dv_2) \\ &= (ad - bc)R(v_1, v_2, av_1 + bv_2, cv_1 + dv_2) \\ &= (ad - bc)^2 R(v_1, v_2, v_1, v_2).\end{aligned}\quad (6.96)$$

从而

$$\frac{R(w_1, w_2, w_1, w_2)}{|w_1 \wedge w_2|^2} = \frac{R(v_1, v_2, v_1, v_2)}{|v_1 \wedge v_2|^2}, \quad (6.97)$$

这就证明了截面曲率与 Π 平面上基矢的具体选取无关。

由于对称型 $Q(v_1, v_2)$ 完全确定了黎曼曲率张量，从而可知：知道了 V_p 所有 2- 平面 Π 的截面曲率 $K(\Pi)$ ，就等价于知道了 p 点的黎曼曲率张量。另外，如果时空流形 M 是一张两维曲面，那截面曲率其实就是这张两维曲面的高斯曲率。

除了黎曼曲率张量以外，上一章我们还定义了里奇张量，它的坐标无关化处理是作为切向量场上的二重线性函数 $\text{Ric}(X, Y)$ ，定义如下

$$\text{Ric}(X, Y) = R_{\mu\nu} X^\mu Y^\nu. \quad (6.98)$$

由于里奇张量是一个对称张量，所以显然有

$$\text{Ric}(X, Y) = \text{Ric}(Y, X). \quad (6.99)$$

6.4 * 附录：一个并非重点的证明

这里我们给出相配对称型 $Q(X, Y) = R(X, Y, X, Y)$ 完全决定黎曼曲率张量 $R(X, Y, Z, W)$ 的证明。

首先，利用 $R(X, Y, Z, W)$ 关于 X, Y, Z, W 的线性性以及 $R(X, Y, Z, W) = R(Z, W, X, Y)$ 的性质，不难得到

$$R(X + Z, Y, X + Z, Y) = R(X, Y, X, Y) + R(Z, Y, Z, Y) + 2R(X, Y, Z, Y),$$

即是说，

$$R(X, Y, Z, Y) = \frac{1}{2} [Q(X + Z, Y) - Q(X, Y) - Q(Z, Y)], \quad (6.100)$$

记 $F(X, Y, Z) = R(X, Y, Z, Y)$ ，称之为相配三变量函数，因此上式说明**相配三变量函数完全由相配对称型决定**。

再次利用 $R(X, Y, Z, W)$ 关于 X, Y, Z, W 的线性性，不难得到

$$\begin{aligned} R(X, Y + W, Z, Y + W) &= R(X, Y, Z, Y) + R(X, W, Z, W) \\ &\quad + [R(X, Y, Z, W) + R(X, W, Z, Y)]. \end{aligned} \quad (6.101)$$

记相配四变量函数 $G(X, Y, Z, W)$ 为

$$G(X, Y, Z, W) \equiv R(X, Y, Z, W) + R(X, W, Z, Y), \quad (6.102)$$

则 (6.101) 式告诉我们

$$G(X, Y, Z, W) = F(X, Y + W, Z) - F(X, Y, Z) - F(X, W, Z), \quad (6.103)$$

因此，相配四变量函数完全由相配三变量函数决定，进而也就完全由相配对称型决定。

在 (6.102) 式中，分别将变量 Y, Z, W 替换成 Z, W, Y ，即得

$$G(X, Z, W, Y) = R(X, Z, W, Y) + R(X, Y, W, Z). \quad (6.104)$$

用 (6.102) 式减去 (6.104) 式，并注意到 $R(X, Y, W, Z) = -R(X, Y, Z, W)$ 以及 $R(X, Z, W, Y) = -R(X, Z, Y, W)$ ，即得

$$\begin{aligned} G(X, Y, Z, W) - G(X, Z, W, Y) = \\ 2R(X, Y, Z, W) + R(X, W, Z, Y) + R(X, Z, Y, W). \end{aligned} \quad (6.105)$$

根据

$$\begin{aligned} R(X, W, Z, Y) + R(X, Z, Y, W) + R(X, Y, W, Z) = 0 \\ \Rightarrow R(X, W, Z, Y) + R(X, Z, Y, W) = -R(X, Y, W, Z) = R(X, Y, Z, W). \end{aligned}$$

代入 (6.105) 式，即得

$$G(X, Y, Z, W) - G(X, Z, W, Y) = 3R(X, Y, Z, W), \quad (6.106)$$

这个式子说明，黎曼曲率张量 $R(X, Y, Z, W)$ 完全由相配四变量函数决定，进而也就完全由相配对称型决定。证明完成。

7. 曲率的意义

前面我们引入了黎曼曲率张量，它是两维曲面高斯曲率的高维推广，是一个内禀量，相应的黎曼几何也是一种内禀几何。所谓内禀几何，意思就是黎曼几何所描述的高维时空流形不需要嵌入到一个更高维的欧氏空间中，而是仅仅只需要一个在时空本身定义的度规张量就可以描述。但这种内禀的描述就使得时空弯曲变成了一个很抽象很不好理解的东西，进而使得黎曼曲率张量作为一个描述时空曲率的量显得非常抽象。从内禀几何的角度，曲率的几何含义是什么呢？在第一章中，我们给出过一个回答，在黎曼正则坐标系中，曲率张量就是对局部平坦的二阶偏离。本章我们将进一步用内禀的方式讨论黎曼曲率张量的几何含义。

另一方面，第三章中提到过，对潮汐力的相对论性刻画就是黎曼曲率张量。本章我们将进一步搞清楚这个潮汐力与黎曼曲率张量之间的关系，并进而弄清楚黎曼曲率张量在广义相对论中的物理含义。特别的，我们将清楚地看到爱因斯坦场方程如何在牛顿近似中回到牛顿引力理论。

而且，上一章我们引入了坐标无关的数学语言。本章我们将看到，有些问题的处理、有些公式的推导，用这种坐标无关的数学语言来进行将会简洁得多。

7.1 测地线偏离方程与牛顿近似

7.1.1 测地线偏离方程

为了看清楚黎曼曲率张量如何刻画时空弯曲，我们首先注意到，弯曲时空中的测地线就是平直时空中直线概念的推广，原因有二：第一，在平直的闵可夫斯基时空，自由粒子

走的是直线，而在弯曲时空中，自由粒子走的是测地线；第二，在第二章关于光锥与世界线的讨论中，我们看到，闵可夫斯基时空中走直线的粒子，具有最大的固有时，也即是说，直线对应固有时的极值，而根据第三章中对类时测地线的定义，弯曲时空中的类时测地线正是对应固有时的极大值。因此，测地线就是弯曲时空中的“直线”。

这就告诉我们一种刻画时空弯曲的方法：在时空中任意取初始时相互平行的两根类时测地线，如果它们始终保持平行，那就说明时空是平直的。相反，如果随着粒子往前飞行，这两根测地线之间的相对距离在变化（比方说两者离得越来越近），那就说明时空本身是弯曲的。

上述想法可以转化成精确的数学公式。为此我们考虑依赖于参数 σ 的一簇测地线 $x^\mu(\tau, \sigma)$ (不同的 σ 对应不同的测地线)， τ 为这些测地线的固有时参数（或者对于类光的零性测地线簇， τ 就是仿射参数）。记 $u^\mu(\tau, \sigma)$ 为这些测地线的切向量，即

$$u^\mu(\tau, \sigma) = \frac{\partial}{\partial \tau} x^\mu(\tau, \sigma), \quad (7.1)$$

用坐标无关的写法则是

$$u(\tau, \sigma) = u^\mu(\tau, \sigma) \partial_\mu = \partial_\tau. \quad (7.2)$$

满足测地线方程

$$D_u u = 0. \quad (7.3)$$

如图 (7.1) 所示，记两根无限靠近的测地线之间的相对偏离为 $\delta\sigma \xi^\mu(\tau, \sigma)$ ，式中 $\delta\sigma$ 为

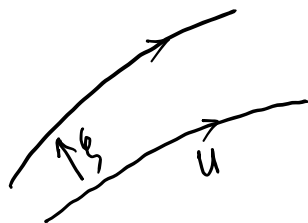


Figure 7.1: 测地线单参簇与偏离矢量 ξ .

无穷小量，向量 $\xi^\mu(\tau, \sigma)$ 称作偏离向量，它由下式给出

$$\xi^\mu(\tau, \sigma) = \frac{\partial}{\partial \sigma} x^\mu(\tau, \sigma), \quad (7.4)$$

写成坐标无关的形式即是

$$\xi(\tau, \sigma) = \xi^\mu(\tau, \sigma) \partial_\mu = \partial_\sigma, \quad (7.5)$$

式中 $\partial_\sigma = \frac{\partial}{\partial \sigma}$ 。由于 τ 和 σ 为两个相互独立的参数，因此 $\partial_\tau \partial_\sigma = \partial_\sigma \partial_\tau$ ，从而

$$[u, \xi] = 0. \quad (7.6)$$

利用克里斯托夫联络的无挠条件 $D_u \xi - D_\xi u = [u, \xi]$, 即有

$$D_u \xi = D_\xi u. \quad (7.7)$$

利用上面的 (7.3) 式和 (7.7) 式, 再利用 $R(u, \xi) = D_u D_\xi - D_\xi D_u - D_{[u, \xi]} = D_u D_\xi - D_\xi D_u$, 即可以进行如下推导

$$\begin{aligned} \frac{D^2 \xi}{D\tau^2} &\equiv D_u D_u \xi = D_u D_\xi u \\ &= D_\xi D_u u + R(u, \xi)u = R(u, \xi)u. \end{aligned} \quad (7.8)$$

即有测地线偏离方程

$$\frac{D^2 \xi}{D\tau^2} = R(u, \xi)u. \quad (7.9)$$

或者也可以根据上一章的知识, 将之写成局部坐标中的形式,

$$\frac{D^2 \xi^\mu}{D\tau^2} = R^\mu{}_{\alpha\beta\nu} u^\alpha u^\beta \xi^\nu. \quad (7.10)$$

实际上, 对于类时测地线簇, 我们可以进一步要求偏离矢量与测地线的切向量 u 正交。为了看清楚这一点, 我们注意到 τ 为测地线的固有时, 因此切向量 u 其实就是四速度, 从而满足

$$g_{\mu\nu} u^\mu u^\nu = -1 \Rightarrow \langle u, u \rangle = -1. \quad (7.11)$$

从而我们可以重新定义一个偏离矢量 η

$$\eta = \xi + \langle \xi, u \rangle u. \quad (7.12)$$

很明显 η 与 u 正交, 即满足

$$\langle \eta, u \rangle = 0. \quad (7.13)$$

另外, 利用 (7.6) 式, 也有

$$\begin{aligned} [u, \eta] &= [u, \xi + \langle \xi, u \rangle u] \\ &= [u, \langle \xi, u \rangle u] = u(\langle \xi, u \rangle)u, \end{aligned} \quad (7.14)$$

再利用 $u(\langle \xi, u \rangle) = D_u(\langle \xi, u \rangle) = \langle D_u \xi, u \rangle = \langle D_\xi u, u \rangle = \frac{1}{2} \xi(\langle u, u \rangle) = \frac{1}{2} \xi(-1) = 0$ (这里我们已经用到了 (7.3) 式和 (7.7) 式), 即有

$$[u, \eta] = 0. \quad (7.15)$$

问题是, 这样重新定义的 η 有权作为偏离矢量吗? 即 η 满足测地线偏离方程吗? 回答是肯定的, 因为测地线偏离方程的成立只需要 $D_u u = 0$ 和 $[u, \eta] = 0$ 成立。当然我们也

可以直接验证 η 满足测地线偏离方程, 为此只需注意到

$$\begin{aligned}
 D_u D_u \eta &= D_u D_u \xi + D_u D_u (\langle \xi, u \rangle u) \\
 &= D_u D_u \xi + D_u (\langle D_u \xi, u \rangle u) \\
 &= D_u D_u \xi + \langle D_u D_u \xi, u \rangle u \\
 &= R(u, \xi) u + \langle R(u, \xi) u, u \rangle u \\
 &= R(u, \xi) u + R(u, u, u, \xi) u.
 \end{aligned} \tag{7.16}$$

式中我们已经代入了 ξ 所满足的测地线偏离方程 (7.9)。又注意到曲率算子 $R(X, Y)$ 关于 X, Y 为线性依赖, 而且满足 $R(X, Y) = -R(Y, X)$, 从而 $R(u, u) = 0$, 从而

$$R(u, \eta) = R(u, \xi). \tag{7.17}$$

另外, 根据 $R(X, Y, Z, W)$ 关于 X, Y 反对称的性质, 可知 $R(u, u, u, \xi) = 0$. 将这些代入 (7.16) 式, 即有

$$\frac{D^2 \eta}{D\tau^2} \equiv D_u D_u \eta = R(u, \eta) u. \tag{7.18}$$

从而, 重新定义的和 u 正交的偏离矢量 η 同样满足测地线偏离方程。

注意到测地线偏离方程是关于偏离矢量的一个线性微分方程, 所以不妨引入一个线性微分算子 \hat{J}_u , 其在偏离矢量 η 上的作用如下

$$\hat{J}_u \eta \equiv -\frac{D^2 \eta}{D\tau^2} + R(u, \eta) u. \tag{7.19}$$

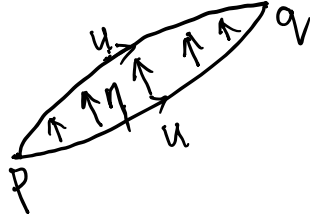
如此一来测地线偏离方程就可以简写为

$$\hat{J}_u \eta = 0. \tag{7.20}$$

测地线偏离方程清楚地告诉我们, 黎曼曲率张量决定了两根邻近测地线之间相对距离的变化情况, 因此黎曼曲率张量是对时空弯曲情况的一种刻画。

将测地线偏离方程限制在一条基准测地线 γ 上所得的结果也称作**雅可比方程**, 相应的偏离矢量 ξ (或者 η) 也称作沿 γ 定义的雅可比场。如果基准测地线 γ 是一簇临近测地线中的一条, 那我们当然有测地线偏离方程。反过来的情况却不一定成立, 即是说, 沿着 γ 有一个雅可比场并不意味着 γ 经过变分 $\delta x^\mu = \delta \sigma \eta^\mu$ 以后得到的临近世界线一定是测地线, 因为雅可比方程只是关于偏离矢量 η 的线性方程, 因此, **它只意味着 γ 变分以后的临近曲线在一阶近似上满足测地线方程。**

如果基准测地线 γ 上存在两点 p, q , 使得雅可比场在这两点的值为零 (但在这两点之间不为零), 即满足 $\eta(p) = \eta(q) = 0$, 则称这两点为 γ 的**一对共轭点**, 如图 (7.2) 所示。根据雅可比场的定义可以知道, 如果测地线 γ 上存在一对共轭点 p, q , 则有时候 (但不是必然) 意味着存在一簇从共同的 p 点发出, 又汇聚于 q 点的测地线单参簇, 使得 γ 是这一簇测地线中的一条。比方说, 两维球面上的测地线就是大圆, 因此地球仪上所有经线圈都是测地线, 它们构成了测地线簇, 而南极和北极就是这测地线上的一对共轭点。共轭点 p, q

Figure 7.2: p, q 为一对共轭点.

的存在必然意味着的是，原测地线 γ 可以沿着雅可比场 η 作一个小的变动 $\delta x^\mu = \delta \sigma \eta^\mu$ ，变动以后的曲线依然连接 p, q 两点，并且在一阶近似上依然满足测地线方程。

前面我们说过，类时测地线对应固有时的极大值，其实这个说法只在固有时较小时才成立，如果两点的时空距离足够远，从而使得固有时足够大，那这个说法一般来说是值得商榷的。后面我们将会看到，共轭点的存在对于决定类时测地线 γ 是否为两点之间固有时取极大值的曲线有重要意义。

为了进一步考察沿着测地线 γ 的雅可比方程，下面取定一个沿着 γ 自由下落的局部惯性系，则根据局部惯性系的定义，沿着 γ 有

$$g_{\mu\nu}|_\gamma = \eta_{\mu\nu}, \quad \Gamma_{\mu\nu}^\rho|_\gamma = 0. \quad (7.21)$$

由于沿着整个 γ 均有 $\Gamma_{\mu\nu}^\rho|_\gamma = 0$ ，所以当然也有 $\partial_\tau \Gamma_{\mu\nu}^\rho|_\gamma = 0$ ，从而即有

$$\frac{D^2 \eta^\mu}{D\tau^2} = \frac{d^2 \eta^\mu}{d\tau^2}. \quad (7.22)$$

因此在这样的坐标系中，雅可比方程可以重写为

$$\hat{J}_u \eta^\mu = -\frac{d^2 \eta^\mu}{d\tau^2} + R^\mu{}_{\alpha\beta\nu} u^\alpha u^\beta \eta^\nu = 0. \quad (7.23)$$

在这个沿着 γ 自由下落的局部惯性系中，粒子的四维速度 u^μ 满足

$$u^\mu = (1, 0, 0, 0), \quad (7.24)$$

即 $u^0 = 1, u^i = 0, i = 1, 2, 3$ ，因此与之正交的偏离矢量 η^μ 必定满足

$$\eta^0 = 0, \quad (7.25)$$

从而雅可比方程的 0 分量变成

$$\begin{aligned} \hat{J}_u \eta^0 &= -\frac{d^2 \eta^0}{d\tau^2} + R^0{}_{\alpha\beta\nu} u^\alpha u^\beta \eta^\nu = 0 \\ \Rightarrow \hat{J}_u \eta^0 &= 0 + R^0{}_{00\nu} u^0 u^0 \eta^\nu = 0 \\ \Rightarrow \hat{J}_u \eta^0 &= -R_{000\nu} \eta^\nu = 0, \end{aligned} \quad (7.26)$$

由于 $R_{000\nu} = 0$, 所以最后是一个自动成立的平凡结果。类似的, 可以得到雅可比方程的非平凡空间分量 ($i, j = 1, 2, 3$), 有

$$\begin{aligned}\hat{J}_u \eta^i &= -\frac{d^2 \eta^i}{d\tau^2} + R^i{}_{\alpha\beta\nu} u^\alpha u^\beta \eta^\nu = 0 \\ \Rightarrow \hat{J}_u \eta^i &= -\frac{d^2 \eta^i}{d\tau^2} + R_{i00j} \eta^j = 0.\end{aligned}\quad (7.27)$$

即是说, 在沿着测地线 γ 自由下落的局部惯性系中, 决定雅可比方程的线性微分算子 \hat{J}_u 可以简单理解为如下 3×3 矩阵值的微分算子

$$\hat{J}_u = -\delta_{ij} \frac{d^2}{d\tau^2} + V_{ij}, \quad \text{式中 } V_{ij} = R_{i00j} = -R_{0i0j}.\quad (7.28)$$

由于 $R_{0i0j} = R_{0j0i}$, 所以 $V_{ij} = V_{ji}$, 从而 \hat{J}_u 作为 3×3 矩阵是一个实对称矩阵。进而类比于量子力学里的哈密顿算符可以知道, 作为作用在“波函数” $\eta^j(\tau)$ 上的算符, 算符 \hat{J}_u 是一个厄密算符!

7.1.2 测地线偏离方程与潮汐力

引力场中自由下落的粒子所走的就是类时测地线, 因此偏离矢量 η 就可以理解为两个邻近的自由下落粒子之间的相对位置矢量, 从而 $D^2 \eta / D\tau^2$ 就是这两个粒子的四维相对加速度, 不妨记为 a^μ , 从而根据测地线偏离方程

$$a^\mu = \frac{D^2 \eta^\mu}{D\tau^2} = R^\mu{}_{\alpha\beta\nu} u^\alpha u^\beta \eta^\nu.\quad (7.29)$$

另一方面, 假设以两个粒子中的一个为坐标原点, 取定一个沿着粒子所走的测地线自由下落的局部惯性系, 那么根据第三章的知识, 另一个邻近粒子在这个局部惯性系中感受到的力就是潮汐力, 从而相对加速度 a^μ 其实就是协变的潮汐加速度。根据上面式子可知, 黎曼曲率张量完全决定了这个协变的潮汐加速度, 换言之, 潮汐加速度正是黎曼曲率张量的物理反映。

下面, 取这无限邻近的两个粒子之一所走的测地线为 γ , 并取定一个沿着 γ 自由下落的局部惯性系, 则根据上一小节的分析可知, 协变潮汐加速度可以写成

$$a^\mu = \frac{D^2 \eta^\mu}{D\tau^2} = \frac{d^2 \eta^\mu}{d\tau^2}.\quad (7.30)$$

又由于在此坐标系中 $\eta^0 = 0$, 从而潮汐加速度也满足 $a^0 = 0$ 。进一步, 对于潮汐加速度的空间分量 $i, j = 1, 2, 3$, 有

$$a^i = R^i{}_{00j} \eta^j.\quad (7.31)$$

下面, 假设引力场不强, 并且粒子的运动速度远小于光速, 那这时候第三章中对潮汐力的牛顿力学分析就近似成立了。在牛顿近似中, 粒子固有时约等于坐标时, 即 $\tau \approx x^0 = t$, 从而上面的潮汐加速度 a^i 就是普通牛顿力学中的潮汐加速度 (相对加速度)。

另一方面, 根据牛顿力学, 自由落体的加速度为

$$\frac{d^2 x^i}{dt^2} = -\partial_i \Phi(\mathbf{x}).\quad (7.32)$$

式中 $\Phi(\mathbf{x})$ 为牛顿万有引力势。将上式对 x^i 变分, 并取 $\delta x^i = \eta^i$ 为偏离矢量, 即可以得到两个邻近的自由下落粒子之间的相对加速度 (即潮汐加速度)

$$\begin{aligned} a^i &= \frac{d^2 \eta^i}{dt^2} = \frac{d^2 \delta x^i}{dt^2} = -(\partial_i \partial_j \Phi) \delta x^j \\ &= -(\partial_i \partial_j \Phi) \eta^j. \end{aligned} \quad (7.33)$$

将这个结果与 (7.31) 式进行比较, 即有, 在牛顿近似下

$$R^i{}_{00j} \approx -(\partial_i \partial_j \Phi). \quad (7.34)$$

或者也可以写成

$$R^i{}_{0j0} \approx (\partial_i \partial_j \Phi). \quad (7.35)$$

在第三章中我们讲过 $\partial_i \partial_j \Phi$ 完全刻画了牛顿力学中的潮汐力, (7.35) 式告诉我们, 黎曼曲率张量正是对这一刻画的相对论协变推广。

(7.35) 式还能告诉我们一个重要结论, 即里奇张量的 00 分量在牛顿近似下为

$$R_{00} \approx \nabla^2 \Phi. \quad (7.36)$$

为了看清楚这一点, 我们只需注意到, 根据黎曼曲率张量指标的对称性, 有 $R^\mu{}_{000} = 0$, 从而

$$R_{00} = R^\mu{}_{0\mu 0} = R^0{}_{000} + R^i{}_{0i0} = R^i{}_{0i0} \approx \nabla^2 \Phi, \quad (7.37)$$

上式的最后一个等于号我们代入了 (7.35) 式。

7.1.3 爱因斯坦场方程的牛顿近似

有了上一小节的准备, 我们就可以来讨论爱因斯坦场方程的牛顿近似了, 我们将清楚地看到爱因斯坦场方程如何在牛顿近似中回到牛顿引力理论。

为了对爱因斯坦场方程进行牛顿近似, 我们首先需要对方程右边的物质能动量张量进行非相对论近似。为此, 我们先回想一下第四章中导出的多质点体系的能动量张量, 为

$$\begin{aligned} T^{\mu\nu}(x) &= \frac{1}{\sqrt{-g}} \sum_n m_n \int d\tau_n \frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n} \delta^4(x - x_n(\tau_n)) \\ &= \frac{1}{\sqrt{-g}} \sum_n m_n \int d\sigma_n \frac{dx_n^\mu}{d\sigma_n} \frac{dx_n^\nu}{d\sigma_n} \delta^4(x - x_n(\sigma_n)). \end{aligned} \quad (7.38)$$

式中我们已经利用变量代换将粒子世界线参数换成了任意的 σ_n 。仿照第二章中的相关做法, 根据 σ_n 的任意性, 取 $\sigma_n = x_n^0$, 并做完对 σ_n 的积分以后, 就可以得到

$$T^{\mu\nu} = \frac{1}{\sqrt{-g}} \left[\sum_n \frac{p_n^\mu p_n^\nu}{p_n^0} \delta^3(\mathbf{x} - \mathbf{x}_n(t)) \right]. \quad (7.39)$$

式中 $p_n^\mu = m_n \frac{dx_n^\mu}{d\tau_n}$ 为粒子的四维动量, 并且根据四维速度满足的 $g_{\mu\nu} \frac{dx_n^\mu}{d\tau_n} \frac{dx_n^\nu}{d\tau_n} = -1$, 有

$$-g_{\mu\nu} p_n^\mu p_n^\nu = m_n^2. \quad (7.40)$$

以单个粒子为例, 回想一下第三章中曾简略地提及如何对广义相对论进行牛顿近似, 为此只需假定粒子的速度为一阶小量, 满足

$$\mathbf{v} \ll 1, \quad (7.41)$$

并取度规张量为 ($i, j = 1, 2, 3$ 为空间指标)

$$g_{00}(x) = -(1 + 2\Phi(\mathbf{x})), \quad g_{0i}(x) = 0, \quad g_{ij}(x) = \delta_{ij}. \quad (7.42)$$

其中小量

$$\Phi(x) \ll 1. \quad (7.43)$$

在零阶近似上, $g_{\mu\nu}$ 就是闵可夫斯基度规 $\eta_{\mu\nu}$, 从而 $g = \det(g_{\mu\nu})$ 的零阶近似就是 -1 . 粒子的作用量可以近似为

$$\begin{aligned} S &= -m \int d\tau = -m \int \sqrt{(1 + 2\Phi(\mathbf{x}))dt^2 - d\mathbf{x}^2} \\ &= -m \int dt \sqrt{1 + 2\Phi(\mathbf{x}) - \mathbf{v}^2} \\ &\approx m \int dt \left[\frac{1}{2} \mathbf{v}^2 - \Phi(\mathbf{x}) \right]. \end{aligned} \quad (7.44)$$

由此可见 \mathbf{v}^2 与牛顿引力势 Φ 是同阶小量, \mathbf{v} 为一阶小量, 因此 Φ 就是二阶小量。

但是对于一个质量为 M 的质点, 其产生的牛顿引力势为 $\Phi = G \frac{M}{r}$, 我们将 M 和 r 理解为有限大小的零阶量, 因此 Φ 为二阶小量就意味着牛顿引力常数 G 应该理解为二阶小量。

对于多质点系统, 由以上假设, 即有

$$p_n^i = p_n^0 v_n^i \ll p_n^0. \quad (7.45)$$

或者说, 相对于零阶的 p_n^0 , 三维动量 p_n^i 是一阶小量。从而容易看出 (7.40) 式的零阶近似为

$$(p_n^0)^2 \approx m_n^2 \Leftrightarrow p_n^0 \approx m_n. \quad (7.46)$$

进而根据 (7.39) 式不难看出: T^{0i} 为一阶小量, T^{ij} 为二阶小量, 以及在零阶近似上, 有

$$T_{00} \approx T^{00} \approx \sum_n m_n \delta^3(\mathbf{x} - \mathbf{x}_n(t)) \equiv \rho, \quad (7.47)$$

即是说, 能量密度就约等于质量密度 ρ , 这实际上就是爱因斯坦的质能关系。同样在零阶近似上, 我们也有

$$T^{0i} \approx T^{ij} \approx 0, \quad (7.48)$$

因为它们都是小量。进一步 (在零阶近似上), 也有

$$T \equiv g_{\mu\nu} T^{\mu\nu} \approx -T^{00} \approx -\rho. \quad (7.49)$$

为了得到爱因斯坦场方程的牛顿近似，我们将场方程 $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa T_{\mu\nu}$ 进行指标缩并，进而得到

$$R - \frac{1}{2}4R = \kappa T \Leftrightarrow R = -\kappa T \approx \kappa\rho. \quad (7.50)$$

从而可以将爱因斯坦场方程近似为

$$R_{\mu\nu} \approx \kappa\left(T_{\mu\nu} + \frac{1}{2}g_{\mu\nu}\rho\right). \quad (7.51)$$

取 00 分量，并利用上一节得到的 $R_{00} \approx \nabla^2\Phi$ 以及 $T_{00} \approx \rho$, $g_{00} \approx -1$, 即有

$$\nabla^2\Phi \approx \frac{1}{2}\kappa\rho. \quad (7.52)$$

这就是场方程的牛顿近似。将之与下面的牛顿引力场方程进行比较

$$\nabla^2\Phi = 4\pi G\rho, \quad (7.53)$$

即可以定出常数 κ 为

$$\kappa = 8\pi G. \quad (7.54)$$

最终的这个结果我们在引入场方程的那一章中就已经知道了，这里只是给出了这个结果的具体推导过程。这个推导本身也说明了，爱因斯坦场方程在牛顿近似上的确能回到牛顿的引力理论，这本身也是对广义相对论的一个最基本的理论上的检验。

如果将上述近似过程用于带宇宙学常数的爱因斯坦场方程

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}, \quad (7.55)$$

则在最低阶近似上，我们将得到

$$\nabla^2\Phi = 4\pi G\rho - \Lambda. \quad (7.56)$$

这是带宇宙学常数的牛顿引力场方程。如果考察的是单个质量为 M 的质点所产生的引力场，则可以取 $\rho = M\delta^3(\mathbf{x})$, 进而由上述方程的解可以得到引力加速度 \mathbf{g} , 为

$$\mathbf{g} = -\nabla\Phi = \left(-\frac{GM}{r^2} + \frac{\Lambda r}{3}\right)\mathbf{e}_r, \quad (7.57)$$

式中 \mathbf{e}_r 为径向方向的单位矢量。从这个结果可以看到，如果宇宙学常数为负，则它将诱导出一个随距离线性增加的吸引力，而如果宇宙学常数为正，则它所诱导出来的就是一个随距离线性增加的**排斥力**。天文观测表明，我们的宇宙在加速膨胀，这通常被认为是由于我们的宇宙具有一个正的宇宙学常数，它所产生的排斥力使得宇宙加速膨胀。

7.2 * 粒子固有时的一阶和二阶变分公式

第三章中说过，有质量粒子所走的测地线是使得固有时取极大值的路径，但这种说法其实是有漏洞的，这一节我们将具体分析漏洞出在哪里，以及这漏洞和时空曲率有什么关系。具体来说，假设记时空流形为 M ，则本节要考察的问题是：给定时空中的一段世界线 $\gamma: [a, b] \rightarrow M$ ，如何确定在连接 $\gamma(a) = p$ 到 $\gamma(b) = q$ 的所有可能世界线中，原来的 γ 是否具有极大的固有时？

为此我们需要考察 γ 的变分，也就是考察与 γ 邻近的那些可能世界线。具体来说，我们可以把 γ 嵌入到一个单参世界线簇 $\{\gamma_\sigma\}_{0 \leq \sigma \leq \varepsilon}$ 中去，使得 $\gamma = \gamma_0$ (称作基准世界线)，而且这个单参簇的两个端点 p, q 固定，即满足

$$\gamma_\sigma(a) = \gamma(a) = p, \quad \gamma_\sigma(b) = \gamma(b) = q, \quad \forall \sigma \in [0, \varepsilon], \quad (7.58)$$

如图 (7.3) 所示。我们要验证的是，如下定义的函数 $L(\sigma)$ 在 $\sigma = 0$ 处是否具有局部极大

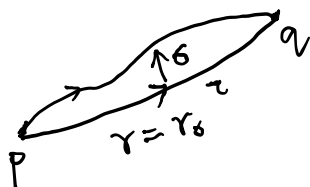


Figure 7.3: 世界线 γ 的单参变分.

值¹,

$$L(\sigma) = \text{世界线段 } \gamma_\sigma \text{ 的固有时长}. \quad (7.60)$$

如果 $L'(0) = 0$, $L''(0) < 0$ (撇号表示对 σ 求导), 则函数 $L(\sigma)$ 在 $\sigma = 0$ 处有局部极大值。如果对所有以 γ 为基准世界线的单参簇 $\{\gamma_\sigma\}$, 均有 $L'(0) = 0$, $L''(0) < 0$, 那就表示 γ 这条世界线与邻近于 γ 的所有其它世界线相比, γ 的固有时都是极大的。所以, 分析的关键是要找到 $L'(0)$, 和 $L''(0)$ 的公式, 也就是所谓的粒子固有时的一阶和二阶变分公式。

记 τ 为世界线本身的参数, 则单参世界线簇就是 $\gamma_\sigma(\tau)$, 它们共同的起点是 $\tau = a$, 共同的终点是 $\tau = b$ 。这个单参世界线簇在局部坐标中的表达是 $x^\mu(\tau, \sigma)$ 。与上一节考察测地线偏离方程不同的是, 现在要求这些世界线在相同的 τ “时间” 之内, 从共同的起点 $\gamma(a) = p$ 到达共同的终点 $\gamma(b) = q$, 因此我们无法要求 τ 为单参簇中每一条世界线的固有时, 不过, 我们可以取 τ 为基准世界线 $\gamma = \gamma_0$ 的固有时。

定义世界线 $\gamma_\sigma(\tau)$ 的切矢量为 $u = \partial_\tau$,

$$u = \partial_\tau = \frac{\partial x^\mu}{\partial \tau} \partial_\mu, \quad (7.61)$$

¹显然, $L(\sigma)$ 和粒子作用量 S 的关系为

$$S = -mL. \quad (7.59)$$

因此 $u^\mu = \frac{\partial x^\mu}{\partial \tau}$ 。由于粒子世界线必定从光锥内部穿过，从而必定为类时曲线，即必定满足

$$\langle u, u \rangle < 0. \quad (7.62)$$

又由于 τ 为 $\sigma = 0$ 这条世界线的固有时，所以 $u|_{\sigma=0}$ 必定为这条基准世界线的四维速度，从而有

$$\langle u, u \rangle|_{\sigma=0} = -1. \quad (7.63)$$

注意， $\sigma \neq 0$ 时可没有这个结果。由此我们可以写出函数 $L(\sigma)$ ，为

$$\begin{aligned} L(\sigma) &= \int_a^b d\tau \sqrt{-g_{\mu\nu} \frac{\partial x^\mu}{\partial \tau} \frac{\partial x^\nu}{\partial \tau}} \\ &= \int_a^b d\tau \sqrt{-\langle u, u \rangle} \equiv \int_a^b d\tau |u|. \end{aligned} \quad (7.64)$$

式中 $|u| \equiv \sqrt{-\langle u, u \rangle}$ ，特别的 $|u|_{\sigma=0} = 1$ 。

为了考察 $x^\mu(\tau, \sigma)$ 随着参数 σ 的变分，我们定义 $\delta x^\mu \equiv \delta \sigma \xi^\mu$ ，式中

$$\xi^\mu = \frac{\partial x^\mu}{\partial \sigma}. \quad (7.65)$$

当然， ξ^μ 的坐标无关化表达就是 $\xi = \partial_\sigma$ 。由于单参簇有共同的固定端点，所以显然有 $\xi^\mu(a) = \xi^\mu(b) = 0$ ，即

$$\xi(a) = \xi(b) = 0. \quad (7.66)$$

另外，由于 σ 和 τ 相互独立，所以显然有

$$[\xi, u] = 0. \quad (7.67)$$

进而利用无挠条件，即有 $D_\xi u = D_u \xi$ 。

利用 $|u|^2 = -\langle u, u \rangle$ 可得 $2|u|\xi(|u|) = \xi(|u|^2) = -\xi(\langle u, u \rangle) = -2\langle D_\xi u, u \rangle$ ，即有

$$\xi(|u|) = -\frac{1}{|u|} \langle D_\xi u, u \rangle = -\frac{1}{|u|} \langle D_u \xi, u \rangle. \quad (7.68)$$

进而可得

$$\begin{aligned} L'(\sigma) &= \int_a^b d\tau \frac{\partial}{\partial \sigma} |u| = \int_a^b d\tau \xi(|u|) \\ &= -\int_a^b d\tau \frac{1}{|u|} \langle D_u \xi, u \rangle. \end{aligned} \quad (7.69)$$

特别的，在 $\sigma = 0$ 处，

$$\left(\frac{1}{|u|} \langle D_u \xi, u \rangle \right)|_{\sigma=0} = \left(\langle D_u \xi, u \rangle \right)|_{\sigma=0} = u(\langle \xi, u \rangle)|_{\sigma=0} - \langle \xi, D_u u \rangle|_{\sigma=0}. \quad (7.70)$$

所以

$$\begin{aligned} L'(0) &= - \int_a^b d\tau \left[\partial_\tau (\langle \xi, u \rangle) - \langle \xi, D_u u \rangle \right]_{\sigma=0} \\ &= - (\langle \xi, u \rangle)|_a^b + \int_a^b d\tau \langle \xi, D_u u \rangle|_{\sigma=0}. \end{aligned} \quad (7.71)$$

利用路径两端的 $\xi = 0$, 即得

$$L'(0) = \int_a^b d\tau \langle \xi, D_u u \rangle|_{\sigma=0}. \quad (7.72)$$

如果对于任意单参簇 (从而任意的 ξ), 均有 $L'(0) = 0$, 则 $\sigma = 0$ 的基准世界线必定满足

$$D_u u = 0, \quad (7.73)$$

这正是测地线方程! 所以上面推导只不过是使用一种坐标无关的方式从固有时一阶变分为零导出了测地线方程, 当然, 这也就是重新处理了一遍最小作用量原理。

二阶变分

下面来计算 $L''(0)$, 由本节开头的叙述可以知道, 仅当 $L'(0) = 0$ 时, 计算 $L''(0)$ 才有意义。于是下面我们可以假设 $\sigma = 0$ 的基准世界线 $\gamma = \gamma_0$ 为一条类时测地线。

由 (7.69) 式, 可得

$$L''(\sigma) = - \int_a^b d\tau \frac{\partial}{\partial \sigma} \left(\frac{1}{|u|} \langle D_u \xi, u \rangle \right). \quad (7.74)$$

利用 (7.68) 式, 可以把式中的被积项表达为

$$\begin{aligned} & \xi \left(\frac{1}{|u|} \langle D_u \xi, u \rangle \right) \\ &= \frac{1}{|u|^3} \langle D_u \xi, u \rangle^2 + \frac{1}{|u|} \langle D_u \xi, D_\xi u \rangle + \frac{1}{|u|} \langle D_\xi D_u \xi, u \rangle \\ &= \frac{1}{|u|^3} [u(\langle \xi, u \rangle) - \langle \xi, D_u u \rangle]^2 + \frac{1}{|u|} \langle D_u \xi, D_u \xi \rangle \\ & \quad + \frac{1}{|u|} [\langle R(\xi, u) \xi, u \rangle + \langle D_u D_\xi \xi, u \rangle], \end{aligned} \quad (7.75)$$

最后一行我们利用了 $[\xi, u] = 0$, 从而 $R(\xi, u) = D_\xi D_u - D_u D_\xi$ 。

然后, 在上面结果中取 $\sigma = 0$, 并注意到这时候有测地线方程以及 $|u| = 1$, 即

$$D_u u = 0, \quad |u| = 1. \quad (7.76)$$

也即有

$$\langle D_u D_\xi \xi, u \rangle = u(\langle D_\xi \xi, u \rangle) = \frac{\partial}{\partial \tau} (\langle D_\xi \xi, u \rangle). \quad (7.77)$$

最后即得到

$$\begin{aligned} L''(0) &= - \int_a^b d\tau \left[(u \langle \xi, u \rangle)^2 + \langle D_u \xi, D_u \xi \rangle + \langle R(\xi, u) \xi, u \rangle + \frac{\partial}{\partial \tau} \langle D_\xi \xi, u \rangle \right] \\ &= - \int_a^b d\tau \left[(u \langle \xi, u \rangle)^2 + \langle D_u \xi, D_u \xi \rangle + \langle R(\xi, u) \xi, u \rangle \right] \\ & \quad - \langle D_\xi \xi, u \rangle|_a^b, \end{aligned} \quad (7.78)$$

式中 $\langle D_\xi \xi, u \rangle|_a^b$ 为端点项, 但是由于世界线单参簇的端点固定 (即在端点处恒有 $\xi = 0$), 所以这个端点项其实恒等于零。再利用 $\langle R(\xi, u)\xi, u \rangle = \langle u, R(\xi, u)\xi \rangle = R(u, \xi, \xi, u) = R(\xi, u, u, \xi) = \langle \xi, R(u, \xi)u \rangle$, 即将上面结果重写成

$$L''(0) = - \int_a^b d\tau \left[(u\langle \xi, u \rangle)^2 + \langle D_u \xi, D_u \xi \rangle + \langle \xi, R(u, \xi)u \rangle \right]. \quad (7.79)$$

重新定义一个变分矢量 η

$$\eta = \xi + \langle \xi, u \rangle u, \quad (7.80)$$

显然, η 与 u 正交, 即满足 $\langle \eta, u \rangle = 0$, 同时 η 在世界线的两个端点处也为零。注意 $\sigma = 0$ 时有测地线方程 $D_u u = 0$, 从而易得 $D_u \eta = D_u \xi + \langle D_u \xi, u \rangle u$, 又注意到 $\langle u, u \rangle = -1$, 进而不难得到

$$\langle D_u \eta, D_u \eta \rangle = \langle D_u \xi, D_u \xi \rangle + \langle D_u \xi, u \rangle^2, \quad (7.81)$$

又由于 $\langle D_u \xi, u \rangle = u\langle \xi, u \rangle - \langle \xi, D_u u \rangle = u\langle \xi, u \rangle$, 从而

$$\langle D_u \eta, D_u \eta \rangle = \langle D_u \xi, D_u \xi \rangle + (u\langle \xi, u \rangle)^2. \quad (7.82)$$

另外, 利用黎曼曲率张量的线性性和反对称性, 可得

$$\langle \xi, R(u, \xi)u \rangle = R(\xi, u, u, \xi) = R(\eta, u, u, \eta) = \langle \eta, R(u, \eta)u \rangle. \quad (7.83)$$

将 (7.82) 式和 (7.83) 式代入 $L''(0)$ 的表达式 (7.79), 即得

$$L''(0) = - \int_a^b d\tau \left[\langle D_u \eta, D_u \eta \rangle + \langle \eta, R(u, \eta)u \rangle \right]. \quad (7.84)$$

这就是最终的沿着类时测地线的二阶变分公式。

仔细看一看二阶变分的这个公式, 我们发现重点是其中再次出现了起决定性作用的时空曲率 $R(u, \eta, u, \eta)$, 这清楚地说明曲率是在二阶扰动 (二阶变分) 的意义上影响自由下落粒子作用量的。

为了更进一步看清二阶变分公式的物理含义, 我们可以完全类似于前面关于测地线偏离方程的情形, 在沿着测地线 γ 自由下落的局部惯性系中讨论这个二阶变分公式。比方说, 这时候我们也有 $\eta^0 = 0$, $D_u \eta^i = \frac{d\eta^i}{d\tau}$ 。则最终可以将 (7.84) 式在此局部惯性系中重写成

$$L''(0) = - \int d\tau \left[\left(\frac{d\vec{\eta}}{d\tau} \right)^2 + R_{i00j} \eta^i \eta^j \right]. \quad (7.85)$$

若进一步取上面这个式子的牛顿近似, 并记 $\eta^i \equiv y^i$, 则有

$$\begin{aligned} \frac{1}{2} L''(0) &\approx - \int dt \left[\frac{1}{2} \left(\frac{d\mathbf{y}}{dt} \right)^2 + \frac{1}{2} R_{i00j} y^i y^j \right] \\ &\approx - \int dt \left[\frac{1}{2} \dot{\mathbf{y}}^2 - \frac{1}{2} (\partial_i \partial_j \Phi) y^i y^j \right]. \end{aligned} \quad (7.86)$$

上式第二行代入了 (7.34) 式。很显然, 这正是第三章引入潮汐力概念时根据牛顿自由落体作用量所得到的结果。

我们可以根据上述结果 (7.85) 进一步讨论在什么情况下 $L''(0) < 0$ 。首先, 假设基准测地线 $\gamma = \gamma_0$ 足够短, 即粒子从 p 端点运动到 q 端点的固有时足够短, 不妨记这固有时为 τ , 则很显然 $\frac{d\vec{\eta}}{d\tau}$ 在 $\Delta\vec{\eta}/\tau$ 的量级, 这里 $\Delta\vec{\eta}$ 为 p, q 之间的相对空间位置矢量。进而可以作如下估算:

$$\int d\tau \left(\frac{d\vec{\eta}}{d\tau}\right)^2 \sim \tau \left(\frac{\Delta\vec{\eta}}{\tau}\right)^2 = \frac{(\Delta\vec{\eta})^2}{\tau}. \quad (7.87)$$

类似的,

$$\int d\tau [R_{i00j} \eta^i \eta^j] \sim C(\eta, \eta) \tau, \quad (7.88)$$

式中 $C(\eta, \eta) \sim R_{i00j} \eta^i \eta^j$ 。即是说, 我们有

$$L''(0) \sim -\left[\frac{(\Delta\vec{\eta})^2}{\tau} + C(\eta, \eta)\tau\right]. \quad (7.89)$$

很显然, 当 τ 足够小时, 括号中的第一项占主导地位, 第二项相比可以忽略。即当 τ 足够小时, 我们有

$$L''(0) \sim -\frac{(\Delta\vec{\eta})^2}{\tau} < 0. \quad (7.90)$$

结论就是: 当端点 p 和端点 q 离得足够近, 从而 p 到 q 的固有时足够短时, 必定有

$$L''(0) < 0, \quad (7.91)$$

这时候从 p 到 q 的类时测地线有极大的固有时!

但是, 如果从 p 到 q 的固有时很长, 那上述结论就不一定成立了, 因为这时候二阶变分公式 (7.84) 中曲率项的贡献就变得很重要, 从而使得 $L''(0) < 0$ 不一定成立了。临界的情况当然就是, 可能存在某些变分矢量 η , 使得 $L''(0) = 0$ 。

为了更好地分析什么情况下会出现 $L''(0) = 0$, 我们不妨改写一下二阶变分公式 (7.84)。为此我们可以利用 $\langle D_u \eta, D_u \eta \rangle = u \langle \eta, D_u \eta \rangle - \langle \eta, D_u D_u \eta \rangle = \partial_\tau \langle \eta, D_u \eta \rangle - \langle \eta, \frac{D^2 \eta}{D\tau^2} \cdot \rangle$ 进而对 (7.84) 式右边的第一项进行分部积分, 得到

$$\begin{aligned} L''(0) &= -\int_a^b d\tau \left[-\langle \eta, \frac{D^2 \eta}{D\tau^2} \rangle + \langle \eta, R(u, \eta)u \rangle \right] - \langle \eta, D_u \eta \rangle \Big|_a^b \\ &= -\int_a^b d\tau \langle \eta, -\frac{D^2 \eta}{D\tau^2} + R(u, \eta)u \rangle. \end{aligned} \quad (7.92)$$

式中第二行我们利用了路径的两个端点处 $\eta = 0$ 。进一步, 利用微分算子 \hat{J}_u 的定义 (7.19), 就可以把这个结果简写为

$$L''(0) = -\int_a^b d\tau \langle \eta, \hat{J}_u \eta \rangle. \quad (7.93)$$

很显然, 当变分矢量 η 同时是沿着基准测地线定义的雅可比场时 (即满足测地线偏离方程 $\hat{J}_u \eta = 0$ 时), 则有 $L''(0) = 0$ 。也即是说, 只要 p, q 之间存在非零的雅可比场, 那就能使得 $L''(0) = 0$ 的临界情况成立。进而我们可以得到如下重要定理。

为了进一步简化问题，我们也可以在沿着 γ 自由下落的局部惯性系中进行分析，这时候上一段的结果就变成，

$$\begin{aligned} L''(0) &= - \int_a^b d\tau \left[\left(\frac{d\vec{\eta}}{d\tau} \right)^2 + R_{i00j} \eta^i \eta^j \right] \\ &= - \int_a^b d\tau \eta^i \left(-\delta_{ij} \frac{d^2}{d\tau^2} + R_{i00j} \right) \eta^j. \end{aligned} \quad (7.94)$$

式中已经利用了 $\eta^i(a) = \eta^i(b) = 0$ 。根据前面讨论雅可比方程时定义的 (7.28)，这个结果显然与 (7.93) 式一致。

(7.94) 式的形式激励我们将它类比于一个量子力学能量泛函的变分问题。为此我们考虑定义在区间 $[a, b]$ 上的一维量子力学问题，取 τ 为一维变量，取量子力学波函数 $\psi(\tau) = (\eta^1(\tau), \eta^2(\tau), \eta^3(\tau))^T$ (T 表示矩阵转置)，并且波函数在区间两端为零 $\psi(a) = \psi(b) = 0$ 。同时，我们可以将变分矢量 $\eta^i(\tau)$ 进行适当的归一化，进而给出波函数归一化条件

$$1 = \langle \psi | \psi \rangle = \int_a^b d\tau \psi^T(\tau) \psi(\tau). \quad (7.95)$$

进一步，取系统的哈密顿算符 H 为

$$H = -\delta_{ij} \frac{d^2}{d\tau^2} + V_{ij} = \hat{J}_u, \quad \text{式中 } V_{ij} = R_{i00j} = -R_{0i0j}. \quad (7.96)$$

则 (7.94) 式可以写成

$$L''(0) = -\langle \psi | H | \psi \rangle = -E(\psi), \quad (7.97)$$

式中能量泛函 $E(\psi)$ 为

$$E(\psi) = \langle \psi | H | \psi \rangle = \int_a^b d\tau \psi^T(\tau) H \psi(\tau). \quad (7.98)$$

显然，要分析 $L''(0)$ 是否恒小于零，关键是要分析能量泛函 $E(\psi)$ 的最小值，也就是上述量子力学系统基态的能量，只要这个基态能量大于零，那即有 $L''(0) < 0$ ，反过来也一样。不妨固定区间 $[a, b]$ 的 a 端，将基态能量看成 b 的函数，记作 E_b ，即

$$E_b = \min_{\psi} \{E(\psi)\}. \quad (7.99)$$

由前面对固有时足够小时 (即 $b-a$ 足够小) 的分析可知，当 $b-a$ 足够小时，能量泛函 $E(\psi)$ 是正定的，从而必有 $E_b > 0$ 。

进一步，可以论证基态能量 E_b 关于 b 是单调递减的，原因在于，一个更小区间的基态波函数可以自然地作为一个更大区间的能量泛函变分的试探波函数。具体来说，假设 $a < b_1 < b$ ，假设 $\psi_0(\tau)$ 为 $[a, b_1]$ 区间上的基态波函数，则我们可以自然地定义 $[a, b]$ 区间上的试探波函数 $\psi(\tau)$ 为

$$\psi(\tau) = \begin{cases} \psi_0(\tau), & \tau \in [a, b_1] \\ 0, & \tau \in [b_1, b] \end{cases}. \quad (7.100)$$

真正基态的能量当然比试探波函数对应的能量要低，从而必有 $E_b \leq E_{b_1}$ 。

进而结合上面两段的结果可知，要让某个大的区间 $[a, b]$ 上的 $E_b \leq 0$ ，**充要条件**是，存在某个小一些的临界 c ， $a < c < b$ ，使得 $E_c = 0$ ，即是说 $H\psi(\tau) = 0$ 在 $[a, c]$ 区间上有非零解。由于 $H = \hat{J}_u$ ，因此这也意味着雅可比方程 $\hat{J}_u \eta^i = 0$ 在 $[a, c]$ 区间上有满足 $\eta(a) = \eta(c) = 0$ 的非零解。这也就说明， $\gamma(c)$ 是 $\gamma(a)$ 的共轭点。结论就是，在某个大区间 $[a, b]$ 上， $E_b \leq 0$ (进而相应基态使得 $L''(0) \geq 0$) 的**充要条件**是，在 $\gamma(b)$ 和 $\gamma(a)$ 之间存在 $\gamma(a)$ 的一个共轭点 $\gamma(c)$ 。进而即有如下重要定理。

重要定理： 设 γ 是一条连接时空中 p, q 两点的类时曲线，对应固有时参数区间 $[a, b]$ ， $\gamma(a) = p$ 、 $\gamma(b) = q$ 。则 γ 的固有时取其任意单参变分 (固定 p, q 点) 极大值的充要条件是， γ 是一条类时测地线，并且在 p, q 两点之间不存在任何点 p 的共轭点。

证明概要： 首先，假设 γ 不是一条测地线，则根据一阶变分公式 (7.72)，我们总可以选取一个合适的变分矢量 ξ ，使得 $L'(0) > 0$ ，从而 γ 的固有时就不是极大值。

其次，假定 γ 是一条类时测地线。则根据上面的讨论，存在变分矢量 η 使得 $L''(0) \geq 0$ 的充要条件是在 p, q 两点之间存在点 p 的共轭点 r 。不仅如此，还可以进一步证明，假设存在共轭点 r ，则必定存在变分矢量 η 使得 $L''(0) > 0$ (去掉了等于号)。因此这种情况下， γ 的固有时也不是取极大值。

证明如下，假设在 p, q 两点之间存在点 p 的共轭点 r 。则根据共轭点的定义，在 p, r 之间存在一个非零的雅可比场 η_0 ，它满足雅可比方程 $\hat{J}_u \eta_0 = 0$ ，同时满足 $\eta_0(p) = \eta_0(r) = 0$ 。因此这时候我们可以定义如下非零变分矢量 η

$$\eta = \begin{cases} \eta_0, & \text{在 } p, r \text{ 之间} \\ 0, & \text{在 } r, q \text{ 之间} \end{cases} \quad (7.101)$$

显然，对于这个变分矢量有 $L''(0) = 0$ 。进而我们只要将这个变分矢量在拐角 r 处进行适当的圆滑化 (如图 (7.4) 所示)，以降低 (7.94) 式中 $(\frac{d\eta}{d\tau})^2$ 项在拐角处的贡献，进而就可以得到一个新的使得 $L''(0) > 0$ 的单参变分。



Figure 7.4: 连接 p, q 的类时测地线 γ 上有点 p 的共轭点 r ，可以将变分矢量在拐角 r 处适当圆滑化，得到一条新的类时曲线 γ' ， γ' 的固有时长于 γ 。

反过来, 假设 γ 是一条类时测地线, 与此同时在 p, q 之间不存在任何点 p 的共轭点。则这时候根据上述讨论的充分必要性可知, 必有 $E_b > 0$, 进而对于任意单参变分, 均有 $L''(0) < 0$, 从而这时候 γ 的固有时就的确取极大值。**证明完成。**

上述定理对于彭罗斯和霍金的奇性定理的证明来说至关重要, 不过, 就本章来说, 这个定理的重要性在于, 它是关于固有时二阶变分公式, 进而关于时空曲率意义的一个重要说明。

8. 施瓦西解以及广义相对论经典检验

爱因斯坦场方程是一个关于度规场 $g_{\mu\nu}$ 的高度非线性的方程，非常难以精确求解。然而，1915 年爱因斯坦发表他的场方程以后，身处一战战场的施瓦西很快就得到了第一个不平凡精确解，并在 1915 年 12 月 22 日将结果寄给了爱因斯坦，这就是本章将要讲述的施瓦西解。

施瓦西得到精确解的关键点在于充分利用对称性来约束 $g_{\mu\nu}$ 的独立分量个数。具体来说，施瓦西假设考察的是一个空间部分具有球对称性的时空。其次，施瓦西假设考察的是没有物质分布的真空部分的解，因此施瓦西解可以描述一个质点或者一个球对称星体外部的引力场。

本章将要讲述的第二部分内容有关广义相对论的一些经典检验，主要是具体推导引力场中的光线偏折以及推导广义相对论效应下的行星近日点进动，以和天文观测进行比较。

8.1 球对称性与度规

时空对称性的系统性考量需要用到基林矢量场的概念，这个我们将在后面的章节中介绍。本章我们仅仅是初识对称性分析的威力，对于本章的目的而言，广义坐标协变性结合球对称性的直观考量就已经足够了。

如果没有对称性的约束，那么度规场 $g_{\mu\nu}(x)$ 作为一个二阶对称张量，它将有 $4 \cdot (4 + 1) / 2 = 10$ 个独立分量，考虑到广义坐标的任意性，其中有 4 个分量是冗余的，剩下 6 个独立分量是真正需要通过爱因斯坦场方程来确定的，作为广义坐标的函数，每个独立分量

又是一个含 4 个自变量的函数。然而，正如我们将要看到的，在球对称性的约束下，真正的独立分量将只有两个，而且每个独立分量仅仅是含两个自变量的函数。

考察球对称性的方便坐标当然是球坐标，在球坐标下，具有球对称性（即在绕球心的旋转下保持不变）的函数只能是时间 t 和径向坐标 r 这两个自变量的函数，而具有球对称性的二次微元就只有 dt^2 、 $dt dr$ 、 dr^2 以及 $d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$ ，其中 $d\Omega^2$ 就是二维单位球面 S^2 上的标准度规。所以具有球对称性的最一般度规将具有如下形式，

$$ds^2 = -U(t,r)dt^2 - 2V(t,r)dt dr + W(t,r)dr^2 + (X(t,r))^2 d\Omega^2. \quad (8.1)$$

但是，以上并没有考虑广义坐标协变性，考虑到这个，则我们可以重新定义径向坐标 $r' = X(t,r)$ ，进而 $dr' = \partial_t X(t,r)dt + \partial_r X(t,r)dr$ 。由此可以消去 (8.1) 中的 r 和 dr ，进而将度规写成 $ds^2 = -U'(t,r')dt^2 - 2V'(t,r')dt dr' + W'(t,r')dr'^2 + r'^2 d\Omega^2$ 的形式。为了方便起见，不妨去掉所有的 ' 号，重命名各个变量，进而即知具有球对称性的一般度规必定可以写成

$$ds^2 = -U(t,r)dt^2 - 2V(t,r)dt dr + W(t,r)dr^2 + r^2 d\Omega^2. \quad (8.2)$$

这里的 U, V, W 当然和 (8.1) 中的 U, V, W 不是一回事，我们只是用了同样的记号而已。

下面我们将表面，广义协变性能够进一步去除 (8.2) 式中的 $-2V(t,r)dt dr$ 项。为此我们只需引入某个合适的函数 $\Phi(t,r)$ ，并定义新的时间变量 $t' = \Phi(t,r)$ ，从而 $dt' = \partial_t \Phi(t,r)dt + \partial_r \Phi(t,r)dr$ 。其中函数 $\Phi(t,r)$ 是这么决定的，我们要求 $dt' = \partial_t \Phi(t,r)dt + \partial_r \Phi(t,r)dr = \lambda(t,r)(Udt + Vdr)$ ，式中 $\lambda(t,r)$ 为某待定函数。很显然，只要能确定 $\lambda(t,r)$ ，那 $\Phi(t,r)$ 就跟着确定了。为了找到 $\lambda(t,r)$ ，我们注意到 $\lambda U = \partial_t \Phi(t,r)$ ， $\lambda V = \partial_r \Phi(t,r)$ ，从而 $\partial_r(\lambda U) = \partial_t(\lambda V)$ ，根据这个式子可知，给定 $U(t,r)$ 和 $V(t,r)$ ，以及给定 $\lambda(t,r)$ 在 0 时刻的初始值 $\lambda(0,r)$ ，那我们就可以通过将 $\partial_t(\lambda V) = \partial_r(\lambda U)$ 对 t 积分进而确定任意时刻 t 的 $\lambda(t,r)$ ，总之，我们总能找到合适的函数 $\lambda(t,r)$ ，使得 $dt' = \lambda(t,r)(Udt + Vdr)$ 。

进而即有 $dt'^2 = \lambda^2(U^2 dt^2 + 2UV dt dr + V^2 dr^2) = \lambda^2[U(Udt^2 + 2Vdt dr) + V^2 dr^2]$ ，即 $-Udt^2 - 2Vdt dr = -\frac{1}{\lambda^2 U} dt'^2 + \frac{V^2}{U} dr^2$ 。代入 (8.2) 式并用新变量 t' 来消去 t ，即知度规一定能够写成 $ds^2 = -\frac{1}{\lambda^2 U} dt'^2 + (\frac{V^2}{U} + W) dr^2 + r^2 d\Omega^2$ 。去掉 t' 上的 ' 号，并重命名各个变量，即知球对称时空的度规一定能够写成如下形式

$$ds^2 = -A(t,r)dt^2 + B(t,r)dr^2 + r^2 d\Omega^2. \quad (8.3)$$

这就是我们最终要找的结果，显然它只有两个独立分量，分别由函数 $A(t,r)$ 和函数 $B(t,r)$ 所刻画。

上式有非常清楚的几何含义，它告诉我们：给定一个时刻 t ，空间几何叶层化为一些同心球面，这些球面的面积为 $4\pi r^2$ ，而两个相邻球面之间的空间距离为 $\sqrt{B} dr$ 。而对于一个位于给定 r 处的观察者来说，在无穷小时间 dt 之内，他所经历的固有时流逝为 $\sqrt{A} dt$ 。

8.2 联络与曲率的计算

8.2.1 如何计算克里斯托夫联络

给出一个待定的度规场 $g_{\mu\nu}(x)$ ，为了代爱因斯坦场方程，就需要计算里奇曲率张量和曲率标量，为此我们又需要首先计算出克里斯托夫联络。当然可以直接代克里斯托夫联络

的定义式计算，但这很容易出错漏，下面我们介绍一种计算克里斯托夫联络的有用技巧。

考虑时空中的一条世界线 $x^\mu(\lambda)$ ， λ 为世界线的参数，考虑如下作用量

$$S = \int d\lambda \frac{1}{2} g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}. \quad (8.4)$$

当然，这个作用量并不是一个真实粒子的作用量，它只是我们用来辅助计算克里斯托夫联络的东西，之所以用这个辅助作用量而不用真实粒子的作用量，是因为这个作用量的变分比粒子作用量的变分要好算很多。

对上面这个作用量变分，不难得到

$$\begin{aligned} \delta S &= - \int d\lambda \left[g_{\sigma\rho} \frac{d^2 x^\sigma}{d\lambda^2} + \frac{1}{2} (\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}) \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} \right] \delta x^\rho \\ &= - \int d\lambda g_{\sigma\rho} \left[\frac{d^2 x^\sigma}{d\lambda^2} + \Gamma_{\mu\nu}^\sigma \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} \right] \delta x^\rho. \end{aligned} \quad (8.5)$$

利用最小作用量原理 $\delta S[x(\lambda)] = 0$ ，即可得到运动方程

$$\frac{d^2 x^\sigma}{d\lambda^2} + \Gamma_{\mu\nu}^\sigma \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} = 0. \quad (8.6)$$

注意，如果 λ 是固有时参数，那这个运动方程就正好是测地线方程。

当然，对于我们的目的来说，并不一定需要将 λ 取作固有时（当然最终取成固有时也可以）。我们的关键技巧在于，给出一个 $g_{\mu\nu}$ ，我们就把它代入到辅助作用量 (8.4) 中，然后对这个辅助作用量变分得出运动方程，最后将所得的方程与标准形式 (8.6) 进行比较，从这个比较中就能读出所有非零的联络系数。

相比于直接代联络的定义式进行计算，这个方法至少有两个好处：一是在实际应用中常常会有很多联络系数分量取零，但是不算一遍人们又无从知道它们是零，而在上述方法中，取零的那些联络系数对计算过程自动就没有贡献；第二，联络系数的独立分量数目太多，一个分量一个分量地代定义式计算就会太繁琐（指手算，而不是用数学软件来推导），而且很容易遗漏，但是上述方法是同时算出所有的非零联络系数，繁琐程度就大大降低了。

下面举例说明上述方法的运用。比如考虑二维球面，其度规为

$$ds^2 = d\theta^2 + \sin^2 \theta d\phi^2. \quad (8.7)$$

由此即有辅助作用量

$$S = \int d\lambda \frac{1}{2} \left[\left(\frac{d\theta}{d\lambda} \right)^2 + \sin^2 \theta \left(\frac{d\phi}{d\lambda} \right)^2 \right]. \quad (8.8)$$

对这个作用量变分，就能得到如下两个运动微分方程

$$\begin{aligned} \frac{d^2 \theta}{d\lambda^2} - \sin \theta \cos \theta \left(\frac{d\phi}{d\lambda} \right)^2 &= 0 \\ \frac{d}{d\lambda} \left(\sin^2 \theta \frac{d\phi}{d\lambda} \right) &= 0. \end{aligned} \quad (8.9)$$

其中第一个方程已经是 (8.6) 式那样的标准形式，将第 2 个方程也化成标准形式，得

$$\frac{d^2 \phi}{d\lambda^2} + 2 \cot \theta \frac{d\theta}{d\lambda} \frac{d\phi}{d\lambda} = 0. \quad (8.10)$$

将这些标准形式的方程与定义式 (8.6) 进行比较, 就能读出非零的联络系数, 为

$$\Gamma_{\phi\phi}^{\theta} = -\sin\theta \cos\theta, \quad \Gamma_{\theta\phi}^{\phi} = \Gamma_{\phi\theta}^{\phi} = \cot\theta. \quad (8.11)$$

其余联络系数都为零。

为了检验这些联络系数是否算对, 我们不妨进一步计算一下黎曼曲率张量。根据黎曼曲率张量的代数性质, 在两维中它只有一个独立的非零分量, 为

$$R_{\phi\theta\theta\phi}^{\theta} = \partial_{\theta}\Gamma_{\phi\phi}^{\theta} - \partial_{\phi}\Gamma_{\theta\phi}^{\theta} + \Gamma_{\theta i}^{\theta}\Gamma_{\phi\phi}^i - \Gamma_{\phi i}^{\theta}\Gamma_{\theta\phi}^i. \quad (8.12)$$

式中 $i = \theta, \phi$ 。注意到有不少联络系数为零, 然后再代入上面的那些非零联络系数, 即可得

$$\begin{aligned} R_{\phi\theta\theta\phi}^{\theta} &= \partial_{\theta}\Gamma_{\phi\phi}^{\theta} - \Gamma_{\phi\phi}^{\theta}\Gamma_{\theta\phi}^{\phi} \\ &= (\sin^2\theta - \cos^2\theta) - (-\sin\theta \cos\theta)(\cot\theta) = \sin^2\theta. \end{aligned} \quad (8.13)$$

进而即有

$$R_{\theta\phi\theta\phi} = g_{\theta\theta}R_{\phi\theta\theta\phi}^{\theta} = \sin^2\theta. \quad (8.14)$$

进而可以算得截面曲率, 为

$$K = \frac{R_{\theta\phi\theta\phi}}{g_{\theta\theta}g_{\phi\phi}} = \frac{\sin^2\theta}{\sin^2\theta} = 1. \quad (8.15)$$

二维曲面的截面曲率也叫高斯曲率, 单位球面的高斯曲率当然为 1, 与这里的计算完全吻合。

还可以算得里奇张量的各分量为

$$\begin{aligned} R_{\phi\phi} &= R_{\phi\theta\theta\phi}^{\theta} = \sin^2\theta, \quad R_{\theta\theta} = R_{\theta\phi\phi\theta}^{\phi} = g^{\phi\phi}R_{\phi\theta\theta\phi}^{\theta} = 1. \\ R_{\theta\phi} &= R_{\phi\theta} = 0. \end{aligned} \quad (8.16)$$

从而可以进一步算得里奇标量, 为

$$R = g^{\theta\theta}R_{\theta\theta} + g^{\phi\phi}R_{\phi\phi} = 2. \quad (8.17)$$

的确, 二维曲面的里奇标量就应该是高斯曲率的两倍, 这是由它们各自的定义决定的。

8.2.2 计算球对称时空的联络和曲率

前面说过, 球对称时空的度规必定可以写成如下形式

$$ds^2 = -A(t, r)dt^2 + B(t, r)dr^2 + r^2d\Omega^2. \quad (8.18)$$

然后我们又介绍了计算克里斯托夫联络的一种技巧, 下面不妨用这种技巧计算一下与上述度规 (8.18) 相应的克里斯托夫联络。

首先, 辅助计算的作用量为

$$S = \frac{1}{2} \int d\lambda \left[-A \left(\frac{dt}{d\lambda} \right)^2 + B \left(\frac{dr}{d\lambda} \right)^2 + r^2 \left(\frac{d\theta}{d\lambda} \right)^2 + r^2 \sin^2\theta \left(\frac{d\phi}{d\lambda} \right)^2 \right]. \quad (8.19)$$

对这个作用量进行变分, 可得如下运动方程 (式中 $\dot{A} \equiv \frac{\partial A}{\partial t}$, $A' \equiv \frac{\partial A}{\partial r}$, B 与此类似)

$$\begin{aligned} \frac{d}{d\lambda} \left(A \frac{dt}{d\lambda} \right) - \frac{1}{2} \dot{A} \left(\frac{dt}{d\lambda} \right)^2 + \frac{1}{2} \dot{B} \left(\frac{dr}{d\lambda} \right)^2 &= 0 \\ -\frac{1}{2} A' \left(\frac{dt}{d\lambda} \right)^2 - \frac{d}{d\lambda} \left(B \frac{dr}{d\lambda} \right) + \frac{1}{2} B' \left(\frac{dr}{d\lambda} \right)^2 + r \left(\frac{d\theta}{d\lambda} \right)^2 + r \sin^2 \theta \left(\frac{d\phi}{d\lambda} \right)^2 &= 0 \\ -\frac{d}{d\lambda} \left(r^2 \frac{d\theta}{d\lambda} \right) + r^2 \sin \theta \cos \theta \left(\frac{d\phi}{d\lambda} \right)^2 &= 0 \\ -\frac{d}{d\lambda} \left(r^2 \sin^2 \theta \frac{d\phi}{d\lambda} \right) &= 0. \end{aligned} \quad (8.20)$$

注意到 t, r, θ, ϕ 均是世界线参数 λ 的函数, 由此即可将这四个方程化成如下标准形式

$$\begin{aligned} \frac{d^2 t}{d\lambda^2} + \frac{1}{2A} \dot{A} \left(\frac{dt}{d\lambda} \right)^2 + \frac{A'}{A} \frac{dt}{d\lambda} \frac{dr}{d\lambda} + \frac{1}{2A} \dot{B} \left(\frac{dr}{d\lambda} \right)^2 &= 0 \\ \frac{d^2 r}{d\lambda^2} + \frac{1}{2B} B' \left(\frac{dr}{d\lambda} \right)^2 + \frac{\dot{B}}{B} \frac{dt}{d\lambda} \frac{dr}{d\lambda} + \frac{1}{2B} A' \left(\frac{dt}{d\lambda} \right)^2 - \frac{r}{B} \left(\frac{d\theta}{d\lambda} \right)^2 - \frac{r \sin^2 \theta}{B} \left(\frac{d\phi}{d\lambda} \right)^2 &= 0 \\ \frac{d^2 \theta}{d\lambda^2} + \frac{2}{r} \frac{dr}{d\lambda} \frac{d\theta}{d\lambda} - \sin \theta \cos \theta \left(\frac{d\phi}{d\lambda} \right)^2 &= 0 \\ \frac{d^2 \phi}{d\lambda^2} + \frac{2}{r} \frac{dr}{d\lambda} \frac{d\phi}{d\lambda} + 2 \cot \theta \frac{d\theta}{d\lambda} \frac{d\phi}{d\lambda} &= 0. \end{aligned}$$

从中可以读出所有非零的联络系数, 为

$$\begin{aligned} \Gamma_{tt}^t &= \frac{1}{2A} \dot{A}, & \Gamma_{tr}^t &= \Gamma_{rt}^t = \frac{1}{2A} A', & \Gamma_{rr}^t &= \frac{1}{2A} \dot{B} \\ \Gamma_{rr}^r &= \frac{1}{2B} B', & \Gamma_{tr}^r &= \Gamma_{rt}^r = \frac{1}{2B} \dot{B}, & \Gamma_{tt}^r &= \frac{1}{2B} A' \\ \Gamma_{\theta\theta}^r &= -\frac{r}{B}, & \Gamma_{\phi\phi}^r &= -\frac{r \sin^2 \theta}{B} \\ \Gamma_{r\theta}^\theta &= \Gamma_{\theta r}^\theta = \frac{1}{r}, & \Gamma_{\phi\phi}^\theta &= -\sin \theta \cos \theta \\ \Gamma_{r\phi}^\phi &= \Gamma_{\phi r}^\phi = \frac{1}{r}, & \Gamma_{\theta\phi}^\phi &= \Gamma_{\phi\theta}^\phi = \cot \theta. \end{aligned} \quad (8.21)$$

下面计算里奇曲率张量, 利用里奇张量的定义式

$$R_{\mu\nu} = R^\rho_{\mu\rho\nu} = \partial_\rho \Gamma_{\nu\mu}^\rho - \partial_\nu \Gamma_{\rho\mu}^\rho + \Gamma_{\rho\sigma}^\rho \Gamma_{\nu\mu}^\sigma - \Gamma_{\nu\sigma}^\rho \Gamma_{\rho\mu}^\sigma. \quad (8.22)$$

利用 (8.21) 式给出的联络系数, 经过一个略微繁琐但却直接的计算, 可以得到

$$\begin{aligned} R_{tt} &= \frac{A''}{2B} + \frac{A'}{rB} - \frac{A'}{4B} \left(\frac{A'}{A} + \frac{B'}{B} \right) - \frac{\ddot{B}}{2B} + \frac{\dot{B}}{4B} \left(\frac{\dot{A}}{A} + \frac{\dot{B}}{B} \right) \\ R_{rr} &= -\frac{A''}{2A} + \frac{B'}{rB} + \frac{A'}{4A} \left(\frac{A'}{A} + \frac{B'}{B} \right) + \frac{\ddot{B}}{2A} - \frac{\dot{B}}{4A} \left(\frac{\dot{A}}{A} + \frac{\dot{B}}{B} \right) \\ R_{\theta\theta} &= 1 - \left(\frac{r}{B} \right)' - \frac{r}{2B} \left(\frac{A'}{A} + \frac{B'}{B} \right) \\ R_{\phi\phi} &= \sin^2 \theta R_{\theta\theta}. \end{aligned} \quad (8.23)$$

以及

$$R_{tr} = R_{rt} = \frac{\dot{B}}{rB}. \quad (8.24)$$

里奇张量的其它分量都是零。另外，由 (8.23) 式容易得到

$$R_{tt}/A + R_{rr}/B = \frac{1}{rB} \left(\frac{A'}{A} + \frac{B'}{B} \right). \quad (8.25)$$

8.3 施瓦西解

8.3.1 真空爱因斯坦方程

前面提到过，施瓦西解是描述一个球对称星体的外部时空，从而它描述的是没有物质分布的真空部分，这时候相应的爱因斯坦场方程可以简化。为了看清楚这种简化，我们首先把标准形式的爱因斯坦场方程 $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi GT_{\mu\nu}$ 进行指标缩并，进而得到（注意是四维时空） $-R = 8\pi GT$ ，其中 $T = T^\mu_\mu$ ，由此就可以把四维时空的爱因斯坦方程改写成如下形式

$$R_{\mu\nu} = 8\pi G \left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T \right). \quad (8.26)$$

对于没有物质分布的真空部分，有 $T_{\mu\nu} = T = 0$ ，从而即有如下真空的爱因斯坦方程

$$R_{\mu\nu} = 0. \quad (8.27)$$

这绝对不意味着，在真空中时空一定是平坦的，时空如果要平坦，则必须满足更严格的条件 $R_{\mu\nu\rho\sigma} = 0$ ，而不只是里奇张量等于零。

8.3.2 施瓦西解

下面我们求解球对称时空的真空爱因斯坦方程。首先，根据 (8.24) 式和真空方程的 $R_{rr} = 0$ ，我们有

$$\dot{B} = 0. \quad (8.28)$$

也即是说，球对称时空度规 (8.18) 里的函数 B 其实并不依赖于时间 t ，而仅仅是 r 的函数，可以写作 $B(r)$ 。

其次，根据真空方程的 $R_{tt} = R_{rr} = 0$ ，我们有 $R_{tt}/A + R_{rr}/B = 0$ ，进而由 (8.25) 式可以得到

$$\frac{A'}{A} + \frac{B'}{B} = 0, \quad (8.29)$$

由此可知 $(\ln(AB))' = 0$ ，积分即得

$$AB = f(t), \quad (8.30)$$

式中 $f(t)$ 为 t 的某个待定函数。进一步，将 (8.29) 式代入 (8.23) 式的 $R_{\theta\theta}$ 分量，并利用真空方程的 $R_{\theta\theta} = 0$ ，即可以得到

$$R_{\theta\theta} = 1 - \left(\frac{r}{B} \right)' = 0. \quad (8.31)$$

积分, 即可得到

$$B(r) = \left(1 + \frac{c}{r}\right)^{-1}, \quad (8.32)$$

式中 c 为积分常数。结合 (8.32) 式和 (8.30) 式, 即有

$$A(t, r) = f(t) \left(1 + \frac{c}{r}\right). \quad (8.33)$$

上面的求解过程没有用到 $R_{tt} = R_{rr} = 0$, 而只用到了 $R_{tt}/A + R_{rr}/B = 0$, 所以我们实际上还要检验比方说 $R_{rr} = 0$ 是否满足? 代入 (8.32) 式和 (8.33) 式不难发现, 的确满足! 看起来这是一个巧合, 因为方程的数目似乎比变量的数目要多。但这其实不是巧合, 而是微分同胚不变性 (或者说广义协变性) 的必然结果。由于微分同胚不变性, 我们必然有恒等式 $D^\mu G_{\mu\nu} = 0$, 这个恒等式确保了在前面的求解已经完成必然自动地有 $R_{rr} = 0$ 。

所以球对称的真空解必定具有如下度规

$$ds^2 = -f(t) \left(1 + \frac{c}{r}\right) dt^2 + \left(1 + \frac{c}{r}\right)^{-1} dr^2 + r^2 d\Omega^2.$$

将时间变量重新定义为 $\sqrt{f(t)} dt \rightarrow dt$ 即可以吸收掉函数 $f(t)$, 进而将球对称真空解写成

$$ds^2 = -\left(1 + \frac{c}{r}\right) dt^2 + \left(1 + \frac{c}{r}\right)^{-1} dr^2 + r^2 d\Omega^2. \quad (8.34)$$

我们看到, 球对称的真空解必定是一个度规不依赖于时间的静态时空, 更准确地说是球对称时空的度规必定由上式给出, 这个结论也称作**伯克霍夫定理**。注意上述这个解在 $r = 0$ 处是奇异的, 好在我们本来关心的也是球对称星体的外面, 从而不会取 $r = 0$ 。

伯克霍夫定理的一个推论是, 对于一个质量均匀分布的空心球壳, 其空腔内部的时空必定是平坦时空。这是因为, 首先空腔内部满足真空爱因斯坦方程, 其次, 系统有球对称性。所以空腔内部的时空度规必定由上面的 (8.34) 式描述。最后, 由于 $r = 0$ 的球心也在空腔内部, 所以这时候 (8.34) 式中的待定常数 c 必定等于零, 否则 $r = 0$ 处的时空就奇异了! 如此一来, 空腔内部的时空度规就必定是平坦度规。或者说, 质量均匀分布的球壳空腔内部必定没有引力场! 这个结论在牛顿引力理论中当然是人们熟知的。

为了看清楚 (8.34) 式中常数 c 的物理含义, 我们注意到在牛顿近似中 $g_{00} = -(1 + 2\Phi)$, 而对于一个球对称的星体, 牛顿万有引力势 $\Phi = -\frac{GM}{r}$, 其中 M 为星体的总质量。将牛顿近似的这个结论与 (8.34) 式中的 $g_{00} = -\left(1 + \frac{c}{r}\right)$ 进行比较, 即可知 $c = -2GM$ 。从而解 (8.34) 可以重写成

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega^2. \quad (8.35)$$

这就是著名的**施瓦西解**, 它描述一个质量为 M 的球对称星体的外部时空。

虽然我们是一次性同时讲述施瓦西解和伯克霍夫定理, 但在历史上伯克霍夫定理当然出现在施瓦西解之后, 施瓦西最初解方程的时候额外假设了 A, B 不依赖于时间 t , 是伯克霍夫解除了这一条假设。

伯克霍夫定理告诉我们, 即使引力源在收缩、膨胀、甚至沿着径向在振荡, 只要它保持球对称性, 那它外部的时空度规就必定是不依赖于时间的 (8.34) 式。简言之, 并没有球

对称的引力波 (通常所说的球面波并不一定是球对称的, 球对称引力波还要求波的极化方向也保持球对称)! 在电磁学中, 我们也有一个类似的结论, 即真空中保持球对称性的电场必定是库伦场, 没有球对称的电磁波。这些结论和下面的事实密切相关, 即在多极矩展开中, 电磁辐射的领头阶是偶极辐射, 没有零极辐射, 类似的, 引力辐射的领头阶是四极辐射, 没有零极辐射, 甚至也没有偶极辐射。能量动量的零极矩就是质量, 偶极矩就是角动量, 因此引力波没有零极辐射和伯克霍夫定理密切相关, 而没有偶极辐射则和爱因斯坦方程存在具有非零角动量的旋转解 (即所谓克尔解) 密切相关。并且这些事实也和所谓的黑洞无毛定理 (也叫黑洞三毛) 相关, 所谓的黑洞无毛即是说四维时空中的无论什么黑洞都可以由三个参数 (即“三根毛”) 唯一确定, 这三个参数即是质量、电荷、以及角动量。其它毛正好对应引力辐射的四极或者四极以上的高极矩 (或者电磁辐射的偶极以及偶极以上的高极矩), 而这些高极矩是存在引力辐射的, 因此在黑洞形成过程中, 这些高极的毛都辐射掉了, 最后只剩下不会辐射的三根毛。

从上述解中可以看到, $2GM$ 具有长度量纲, 可以理解为某种引力半径, 称作施瓦西半径, 记为 r_g

$$r_g \equiv 2GM. \quad (8.36)$$

利用 r_g , 施瓦西解就可以写成

$$ds^2 = -\left(1 - \frac{r_g}{r}\right) dt^2 + \left(1 - \frac{r_g}{r}\right)^{-1} dr^2 + r^2 d\Omega^2. \quad (8.37)$$

对于太阳, 其施瓦西半径 $r_g \sim 3km$, 这与太阳半径相比是一个非常微小的量。实际上, 对于各种典型的天文学客体, 其施瓦西半径 r_g 与物质分布半径 (简称半径) R 的比值 r_g/R 大约为, 10^{-9} (地球)、 10^{-6} (太阳)、 10^{-4} (白矮星), 以及 10^{-1} (中子星), 总的来说都是一个小量, 这就是后文的一些微扰计算中常常把含 r_g 的项看作微扰小量的原因。

通过重新定义径向变量 $r \rightarrow \rho$,

$$r = \rho \left(1 + \frac{r_g}{4\rho}\right)^2. \quad (8.38)$$

我们也可以把施瓦西解 (8.35) 写成如下各向同性的形式

$$ds^2 = -\frac{\left(1 - \frac{r_g}{4\rho}\right)^2}{\left(1 + \frac{r_g}{4\rho}\right)^2} dt^2 + \left(1 + \frac{r_g}{4\rho}\right)^4 (d\rho^2 + \rho^2 d\theta^2 + \rho^2 \sin^2 \theta d\phi^2). \quad (8.39)$$

8.4 粒子在施瓦西时空中的运动

本节我们主要研究粒子 (包括有质量的粒子和无质量的光子) 在施瓦西时空中的运动。通过这个研究, 我们将推导出光线在球对称引力场中的偏折公式, 至于它和天文观测的比较则是我们在第三章讲述等效原理的时候就已经讲过了的。我们还将推导出行星绕太阳运动时由于广义相对论效应所带来的近日点进动, 尤其是, 我们将计算出水星的近日点进动并和天文观测进行比较, 在历史上, 这双方的吻合也是对广义相对论理论的经典检验之一。

8.4.1 粒子在球对称时空中的运动

有质量粒子

(8.20) 式实际上已经得到了粒子在球对称时空中的测地线方程 (只需取参数 λ 为有质量粒子的固有时参数或者无质量粒子的仿射参数)。比方对于有质量粒子, 这时候 $\lambda = \tau$ 。而对于施瓦西解, 注意到最终时空是静态的, 即 $\dot{A} = \dot{B} = 0$, 所以这组方程简化为

$$\frac{d}{d\tau} \left(A \frac{dt}{d\tau} \right) = 0 \quad (8.40)$$

$$-\frac{1}{2} A' \left(\frac{dt}{d\tau} \right)^2 - \frac{d}{d\tau} \left(B \frac{dr}{d\tau} \right) + \frac{1}{2} B' \left(\frac{dr}{d\tau} \right)^2 + r \left(\frac{d\theta}{d\tau} \right)^2 + r \sin^2 \theta \left(\frac{d\phi}{d\tau} \right)^2 = 0 \quad (8.41)$$

$$-\frac{d}{d\tau} \left(r^2 \frac{d\theta}{d\tau} \right) + r^2 \sin \theta \cos \theta \left(\frac{d\phi}{d\tau} \right)^2 = 0 \quad (8.42)$$

$$-\frac{d}{d\tau} \left(r^2 \sin^2 \theta \frac{d\phi}{d\tau} \right) = 0. \quad (8.43)$$

另外, 对于有质量粒子, 其四维速度与自身的内积等于 -1 , 从而必有

$$A(r) \left(\frac{dt}{d\tau} \right)^2 - B(r) \left(\frac{dr}{d\tau} \right)^2 - r^2 \left(\frac{d\theta}{d\tau} \right)^2 - r^2 \sin^2 \theta \left(\frac{d\phi}{d\tau} \right)^2 = 1. \quad (8.44)$$

方程 (8.40) 和方程 (8.43) 的解为

$$A \frac{dt}{d\tau} = \varepsilon, \quad r^2 \sin^2 \theta \frac{d\phi}{d\tau} = l. \quad (8.45)$$

式中 ε 和 l 均为常数。记粒子的四维速度矢量为 $u^\mu = \frac{dx^\mu}{d\tau}$, 由于 $A \frac{dt}{d\tau} = -g_{00} u^0 = -u_0$, 所以 $-u_0 = \varepsilon$ 为守恒量所反映的, 正是粒子的能量守恒, 具体来说守恒量 ε 是每单位质量的能量。而 $r^2 \sin^2 \theta \frac{d\phi}{d\tau} = g_{\phi\phi} u^\phi = u_\phi$, 因此 $u_\phi = l$ 所反映的, 正是粒子环绕竖直轴的角动量守恒, l 就是每单位质量的角动量, 之所以有角动量守恒, 当然是因为系统有球对称性。

另外, 方程 (8.42) 的解可以取为 $\theta(\tau) = \frac{\pi}{2}$, 即粒子实际上是在黄道面上运动。将这些结果代入方程 (8.44), 可以得到

$$\frac{\varepsilon^2}{A} - B(r) \left(\frac{dr}{d\tau} \right)^2 - \frac{l^2}{r^2} = 1. \quad (8.46)$$

取 $\theta(\tau) = \frac{\pi}{2}$, 并将这个方程对径向坐标 r 求导, 即可以自动满足方程 (8.41)。因此, (8.41) 式无需再独立求解, 而我们最终需要进一步处理的方程仅仅只有方程 (8.46)。不难发现这个方程可以重写成如下形式

$$\frac{1}{2} \left(\frac{dr}{d\tau} \right)^2 + V_{\text{eff}}(r) = 0, \quad (8.47)$$

式中

$$V_{\text{eff}}(r) = \frac{1}{2B} \left(1 + \frac{l^2}{r^2} \right) - \frac{\varepsilon^2}{2AB}. \quad (8.48)$$

在数学形式上, 我们可以将 (8.47) 理解为, 一个总能量为零的非相对论性的牛顿质点在一个 $V_{\text{eff}}(r)$ 的一维势场中运动, 其中 τ 可以理解为时间。

很多时候, 我们仅仅关心粒子在黄道平面上的运动轨道, 这时候我们可以利用 $\frac{d\phi}{d\tau} = \frac{l}{r^2}$, 以及利用 $\frac{dr}{d\tau}/\frac{d\phi}{d\tau} = \frac{dr}{d\phi}$ 消去固有时参数 τ , 进而将 (8.47) 式比上 $(\frac{d\phi}{d\tau})^2$, 即可以得到粒子的轨道方程

$$\left(\frac{dr}{d\phi}\right)^2 + \frac{1}{B}\left(\frac{r^4}{l^2} + r^2\right) - \frac{\varepsilon^2}{l^2} \frac{r^4}{AB} = 0. \quad (8.49)$$

给定 $A(r)$ 和 $B(r)$, 原则上我们只要将这个式子积分, 就能求出轨道方程 $r(\phi)$ 。

实际上, 人们常常应用如下技巧, 即引入变量代换 $u = 1/r$, 从而 $\frac{dr}{d\phi} = -\frac{1}{u^2} \frac{du}{d\phi}$, 代入上面的轨道方程, 即有

$$\left(\frac{du}{d\phi}\right)^2 + \frac{1}{B}\left(\frac{1}{l^2} + u^2\right) - \frac{\varepsilon^2}{l^2} \frac{1}{AB} = 0. \quad (8.50)$$

当然, 有时候人们关心粒子运动的坐标时 t 与 r 之间的关系 $t(r)$, 那这时候就可以利用 $A\frac{dt}{d\tau} = \varepsilon$, 以及 $\frac{dr}{d\tau}/\frac{dt}{d\tau} = \frac{dr}{dt}$ 来消去固有时参数 τ , 并得到 $t(r)$ 所满足的微分方程。

光子

如果我们考察的不是一个有质量粒子, 而是一个无质量的光子。这时候测地线方程当然是一样的, 只需将固有时参数 τ 替换为仿射参数 λ 。唯一不同的是, 方程 (8.44) 应该替换成如下类光四矢量条件

$$A(r)\left(\frac{dt}{d\lambda}\right)^2 - B(r)\left(\frac{dr}{d\lambda}\right)^2 - r^2\left(\frac{d\theta}{d\lambda}\right)^2 - r^2 \sin^2 \theta \left(\frac{d\phi}{d\lambda}\right)^2 = 0. \quad (8.51)$$

这时候当然也有能量守恒和角动量守恒, 取光子在黄道平面 $\theta = \pi/2$ 内运动, 相应的能量守恒和角动量守恒即为 $A\frac{dt}{d\lambda} = \varepsilon$, $r^2\frac{d\phi}{d\lambda} = l$ 。相应的, 方程 (8.46) 就应该替换成

$$\frac{\varepsilon^2}{A} - B(r)\left(\frac{dr}{d\lambda}\right)^2 - \frac{l^2}{r^2} = 0. \quad (8.52)$$

这个方程可以重写为

$$\left(\frac{dr}{d\lambda}\right)^2 + \frac{1}{B}\left(\frac{l^2}{r^2}\right) - \frac{\varepsilon^2}{AB} = 0. \quad (8.53)$$

它可以类似地看作为一个牛顿粒子在一个一维势场中的运动。

当然, 更多时候, 我们更关心的, 其实是光子在黄道平面上的轨道方程, 为此我们只需利用 $\frac{dr}{d\lambda}/\frac{d\phi}{d\lambda} = \frac{dr}{d\phi}$ 消去仿射参数 λ , 即可以得到

$$\left(\frac{dr}{d\phi}\right)^2 + \frac{r^2}{B} - \left(\frac{\varepsilon}{l}\right)^2 \frac{r^4}{AB} = 0. \quad (8.54)$$

很显然, 光子的轨道仅仅依赖于两个守恒量的比值 ε/l 。为了看清楚这个比值的几何含义, 我们取 $r \rightarrow \infty$, 这时候时空几何将趋于平坦的闵可夫斯基几何, 即 $A \rightarrow 1$, $B \rightarrow 1$, 从而上述轨道方程趋于

$$\left(\frac{dr}{d\phi}\right)^2 \simeq \left(\frac{\varepsilon}{l}\right)^2 r^4, \quad (8.55)$$

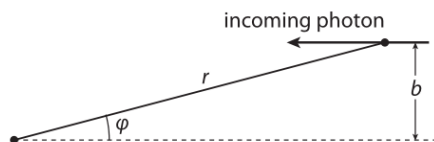


Figure 8.1: 从无穷远处入射的光子具有瞄准参数 $r\phi \simeq b = l/\varepsilon$.

积分, 即可得 $1 \simeq \frac{\varepsilon}{l} r\phi \Leftrightarrow r\phi \simeq l/\varepsilon$ 。如图 (8.1) 所示, 这意味着 $b = l/\varepsilon$ 为无穷远处入射的光子被引力场散射时的瞄准参数。

引入新变量 $u = 1/r$, 从而 $\frac{dr}{d\phi} = -\frac{1}{u^2} \frac{du}{d\phi}$, 代入上面的轨道方程 (8.54), 即有

$$\left(\frac{du}{d\phi}\right)^2 + \frac{1}{B}u^2 - \frac{1}{b^2} \frac{1}{AB} = 0. \quad (8.56)$$

与有质量粒子的轨道方程 (8.50) 相比, 光子的轨道方程仅仅依赖于瞄准参数 b , 也就是仅仅依赖于 l 与 ε 的比值, 而不依赖于 l 或 ε 本身。这是因为, 对于光而言, 其仿射参数 λ 的选取并不唯一, 而是可以相差一个重标度, 在仿射参数的重标度之下 $A \frac{dt}{d\lambda} = \varepsilon$ 当然也要重标度, 类似的 l 也要重标度, 因此对于光, 其 ε 和 l 本身都不是物理的, 但是, 这两者的比值是物理的, 不依赖于仿射参数的选取。

8.4.2 光线偏折再回顾

前面在讨论等效原理的相关章节中, 我们已经说过, 等效原理的一个预言是, 光线在引力场中会被散射, 从而发生偏折。我们也说过, 如果把光看成一个有质量的小颗粒, 那么即使根据牛顿的引力理论, 也能计算出光线在引力场中的一个偏折量。问题是, 广义相对论计算出来的光线偏折量是牛顿理论的 2 倍。下面我们就来推导, 这个 2 倍到底是怎么来的。

为此, 我们在轨道方程 (8.56) 中代入具体的施瓦西解, 即 $AB = 1$ 以及 $B^{-1} = 1 - \frac{2GM}{r} = 1 - r_g u$, 从而即有光线在施瓦西解中的轨道方程

$$\left(\frac{du}{d\phi}\right)^2 + u^2 - r_g u^3 = \frac{1}{b^2}. \quad (8.57)$$

它在 $r \rightarrow \infty$, 从而 $u \rightarrow 0$ 处的渐近解就是我们已经知道的 $u \simeq \frac{1}{b}\phi$ 。

下面我们来近似求解轨道方程 (8.57), 为此, 我们将这个方程再求导一次, 即有

$$\frac{d^2u}{d\phi^2} + u = \frac{3}{2}r_g u^2. \quad (8.58)$$

将球形星体的存在看作一个微扰, 即将右边的 $\frac{3}{2}r_g u^2$ 看作微扰项。首先, 我们忽略这一项, 从而相应的方程即是标准的谐振子方程 $\frac{d^2u}{d\phi^2} + u = 0$, 这个方程的标准解即是 $u = \frac{1}{b} \sin \phi$ (系数由 $u \rightarrow 0$ 的渐近解确定), 或者说 $b = r \sin \phi$, 很显然这描述的是直线前进的光线。为了反映微扰项的影响, 我们设完整的解为 $u = \frac{1}{b} \sin \phi + u_1$, 其中 u_1 为微扰带来的修正。代入方程 (8.58) 并保留一阶微扰近似, 即有

$$\frac{d^2u_1}{d\phi^2} + u_1 \simeq \frac{3}{2}(r_g/b^2) \sin^2 \phi. \quad (8.59)$$

这个方程的解为

$$u_1 \simeq \frac{r_g}{b^2} \left(1 - \frac{1}{2} \sin^2 \phi\right) \Leftrightarrow u \simeq \frac{1}{b} \sin \phi + \frac{r_g}{b^2} \left(1 - \frac{1}{2} \sin^2 \phi\right). \quad (8.60)$$

如图 (8.2) 所示, 为了求出光线在引力场中的偏折量 $\Delta\phi$, 我们取 $r \rightarrow \infty$, 即 $u \rightarrow 0$, 从而有

$$\frac{1}{b} \sin \phi(r = \infty) + \frac{r_g}{b^2} \left(1 - \frac{1}{2} \sin^2 \phi(r = \infty)\right) \simeq 0. \quad (8.61)$$

这个方程有两个根, 其中一个是非物理的, 另一个近似是 $\sin \phi(r = \infty) \simeq -r_g/b$, 注意这是

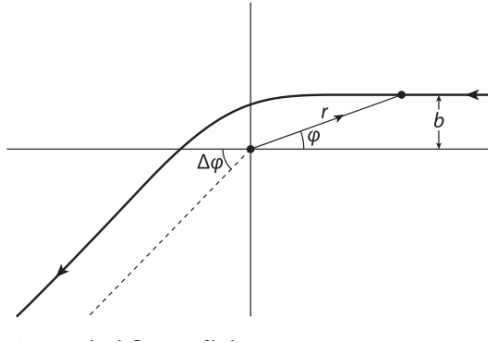


Figure 8.2: 光线偏折.

一个小量, 所以这个解告诉我们: 当光线从无穷远处入射进来的时候, $\phi(r = \infty) \simeq -r_g/b$, 而当光线出射到无穷远处去的时候 $\phi(r = \infty) \simeq \pi + r_g/b$, 所以, 光线的偏折角 $\Delta\phi$ 为

$$\Delta\phi \simeq 2r_g/b = 4GM/b. \quad (8.62)$$

这就是广义相对论给出的光线偏折公式, 这个结果恰恰是牛顿结果的 2 倍。对于太阳, 当我们代入太阳质量, 并取 b 为太阳半径, 即可以得到星光从太阳边缘掠过时的偏折量, 大约为 $1.75''$ 。

8.4.3 水星近日点进动

类似的分析也可以应用于有质量粒子, 当然, 一个重要的区别是, 光线并没有束缚轨道, 因此只需分析散射问题, 但是, 有质量粒子是有束缚轨道的, 就好比行星绕太阳的运动, 实际上我们通常对这些束缚轨道更感兴趣。

为了分析有质量粒子围绕球对称星体的轨道运动, 我们在轨道方程 (8.50) 中代入具体的施瓦西解, 即 $AB = 1$ 以及 $B^{-1} = 1 - \frac{2GM}{r} = 1 - r_g u$, 从而即可得到

$$\left(\frac{du}{d\phi}\right)^2 + \left(\frac{1}{l^2} + u^2\right)(1 - r_g u) = \frac{\epsilon^2}{l^2}. \quad (8.63)$$

重新整理一下, 或者也可以写成

$$\left(\frac{du}{d\phi}\right)^2 + u^2 - \frac{r_g}{l^2} u - r_g u^3 = \frac{\epsilon^2 - 1}{l^2}. \quad (8.64)$$

为了简单起见, 引入记号 $r_g/l^2 \equiv 2\sigma$ (即 $GM/l^2 \equiv \sigma$), $\varepsilon^2 - 1 \equiv 2E$, 则上式即是

$$\left(\frac{du}{d\phi}\right)^2 + u^2 - 2\sigma u - r_g u^3 = 2E/l^2. \quad (8.65)$$

现在, 让我们回顾一下标准的牛顿力学中的粒子在万有引力场中的运动, 这时候, 我们有能量守恒方程

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 + \frac{l^2}{2r^2} - \frac{GM}{r} = E. \quad (8.66)$$

式中 E 为单位质量的非相对论性粒子的总能量, l 和上面一样是每单位质量的角动量。当然, 我们还有角动量守恒方程 $\frac{d\phi}{dt} = \frac{l}{r^2}$ 。完全类似于前面的操作可以消去时间变量 t , 并进行变量代换 $u = 1/r$, 最后我们可以得到

$$\left(\frac{du}{d\phi}\right)^2 + u^2 - 2\sigma u = 2E/l^2. \quad (8.67)$$

这就是广义相对论的 (8.65) 式在牛顿力学中的对应物。从这个对应可以清楚地看到, 广义相对论中定义的 $E = (\varepsilon^2 - 1)/2$ 就是单位质量粒子非相对论性总能量在相对论中的推广, 特别的, 这个能量是扣除了自由粒子静止能量的。

为了求解方程 (8.67), 我们将它再次对 ϕ 求导, 即可得

$$\frac{d^2u}{d\phi^2} + u = \sigma. \quad (8.68)$$

这是一个谐振子方程, 其通解可以写成

$$u_0 = \sigma(1 + e \cos(\phi - \phi_0)), \quad (8.69)$$

式中 e 和 ϕ_0 为两个积分常数。为方便起见, 人们常常选 $\phi_0 = 0$ 。所以, (8.67) 式的解必定具有如下形式

$$u_0 = \sigma(1 + e \cos \phi), \quad (8.70)$$

反代入 (8.67) 式, 可得 e 满足

$$\sigma^2(e^2 - 1) = 2E/l^2. \quad (8.71)$$

(8.70) 式所描写的就是极坐标下的一个标准的二次曲线, e 就是它的离心率, 对于椭圆轨道 (束缚轨道), $e < 1$, 从 (8.71) 式可以看出, 这相当于要求 $E < 0$ 。所以在牛顿力学中, $E < 0$ 对应束缚轨道, $E \geq 0$ 对应的即是散射轨道。

给定单位质量粒子的角动量 l 和能量 E , 我们就能确定轨道参数 σ 和 e 。但是, 在天文学上, 椭圆轨道的 e 不难观测, 反而是 E 和 l 难以直接确定, 为了用天文学上更容易观测的量来表达椭圆方程中的参数 σ , 我们注意到, 在近日点, r 最小, u_0 最大, 从而对应 $\phi = 0$ 。对于束缚的椭圆轨道, 假设半长轴为 a (这是天文学上容易观测的量), 则近日点的距离为 $a(1 - e)$, 从而由 (8.70) 式, 可得

$$\frac{1}{a(1 - e)} = \sigma(1 + e) \Rightarrow \sigma = \frac{1}{a(1 - e^2)}. \quad (8.72)$$

下面,回到完整的广义相对论轨道方程 (8.65),与牛顿力学的方程 (8.67) 相比,我们发现左边多了一个 $-r_g u^3$ 项,完全类似于光线偏折情形的讨论,我们把这一项看作对牛顿引力的一个微扰,从而即可以用微扰论的方法来求解 (8.65)。具体来说,我们设 $u = u_0 + u_1$, 其中 $u_0 = \sigma(1 + e \cos \phi)$ 即为上面的牛顿解,而 u_1 是微扰修正,将这个假设代入轨道方程 (8.65),展开到微扰的一阶项,即可得

$$-\sin(\phi) \frac{du_1}{d\phi} + \cos(\phi) u_1 = \frac{r_g \sigma^2}{2e} (1 + e \cos \phi)^3. \quad (8.73)$$

上述方程左边关于 u_1 是线性的,而右边的驱动项则是 $1, \cos \phi, \cos^2 \phi, \cos^3 \phi$ 这四者的一个线性组合,注意没有 $\sin \phi$ 项。为了比较左右两边,不失一般性地,我们可以设

$$u_1 = \alpha + \beta \cdot \cos \phi + \gamma \cdot \cos^2 \phi + \delta \cdot \phi \sin \phi, \quad (8.74)$$

特别的, u_1 中不可能包含 $\phi \cos \phi$ 这样的项,因为这样的项对方程 (8.73) 左边的贡献和右边的驱动项根本无法匹配起来。将上面的假设代入方程 (8.73) 并比较左右两边 $1, \cos \phi, \cos^2 \phi, \cos^3 \phi$ 这四项的组合系数,就可以定出 $\alpha, \beta, \gamma, \delta$, 它们当然都是 r_g 的一阶小量,特别的, $\delta = r_g \frac{3}{2} e \sigma^2$ 。从而,在各组合项领头阶近似上,我们可以把完整的解 $u = u_0 + u_1$ 近似成

$$\begin{aligned} u &\simeq \sigma(1 + e \cos \phi + r_g \frac{3}{2} e \sigma \phi \sin \phi) \\ &\simeq \sigma \left\{ 1 + e \cos \left[\left(1 - \frac{3}{2} r_g \sigma \right) \phi \right] \right\} \end{aligned} \quad (8.75)$$

从最后的结论我们可以看到,每次粒子的径向坐标(比方说近日点坐标)达到其在 $\phi = 0$ 时的同样值,它的角度 ϕ 都得转过 $2\pi / (1 - \frac{3}{2} r_g \sigma) \simeq 2\pi + 3\pi \sigma r_g$ (注意 r_g 是小量),也即是说,这个角度相对于 2π 角往前进了

$$\Delta\phi \simeq 3\pi \sigma r_g = 6\pi \left(\frac{GM}{l} \right)^2. \quad (8.76)$$

这就是束缚粒子轨道的近日点在引力场中每个周期进动的角度,如图 (8.3) 所示。

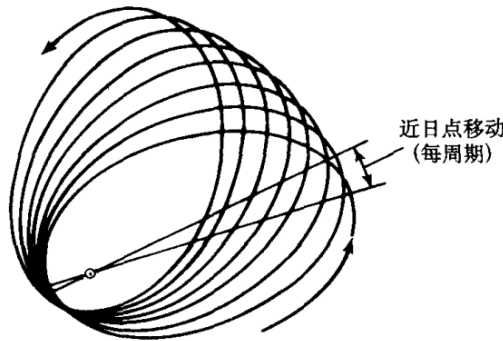


Figure 8.3: 近日点进动.

由于是束缚轨道,利用 (8.72) 式,我们也可以将上述结果表示成

$$\Delta\phi \simeq 3\pi \sigma r_g = 6\pi \frac{GM}{a(1-e^2)} = 6\pi \frac{GM}{a(1-e^2)c^2}. \quad (8.77)$$

在上式最后一个等号中，我们通过量纲分析恢复了光速 c 。对于水星绕太阳运动，这个大小大约是每周期 $0.1035''$ ，或者每世纪 $42.98''$ (42.98 弧秒)。这个值远比实际上观测到的水星近日点的总近动值 (大约是每世纪 $570''$) 要小。这是因为，对水星近日点进动作用最大的不是广义相对论效应，而是其它行星对水星的影响，特别是金星、木星和地球的影响，它们三个加起来的贡献约占总观测量的百分之九十一。但是，在历史上，人们发现水星近日点进动的观测量在扣除了其它行星的影响之后，依然有大约每世纪 $43''$ 的值无法用其它已知理论解释，而这个值正好可以成功地用广义相对论效应来解释。正因为如此，我们才常说，水星近日点进动是广义相对论的一个经典检验。

9. 黑洞基础

米歇尔和拉普拉斯曾经论证，天体的引力可以大到连光都无法逃逸，因此这样的天体可能是看不见的。这就是人类最早对黑洞的想象，当然他们的论证依赖于牛顿力学，尤其是它假设光是一种有质量的粒子，因此这个论证在细节上当然是不对的。正确的论证需要用到广义相对论，尤其是从施瓦西解中，我们可以论证当星体的半径小于其施瓦西半径 $r_g = 2GM$ 时，它就会成为一个黑洞。所以，广义相对论允许黑洞的存在，然而理论允许并不等于黑洞真的会存在，因为如何论证星体的半径真的可以小于其施瓦西半径是一个问题，从而论证黑洞真的会存在是一个远为复杂的问题，对此，我们推迟到后面的章节再进行一些讨论。本节先讨论施瓦西解如何描述一个被允许存在的黑洞，称作施瓦西黑洞。

关于黑洞，最重要的一件事情也许是存在一个事件视界，本章将以施瓦西黑洞为例，详细讨论黑洞视界的方方面面。

本章要讨论的第二个问题是一组性质非常好的坐标系，称作克鲁斯卡坐标。用克鲁斯卡坐标来描述施瓦西解就称作施瓦西解的克鲁斯卡延拓。除了施瓦西解的延拓之外，我们还会使用统一的办法讨论 Reissner-Nordstrom 黑洞 (RN 黑洞) 的克鲁斯卡延拓。

本节要讨论的第三个问题有关于时空的整体结构，尤其是因果结构。我们将讨论，如何在不破坏因果结构的情况下，像画地图一样，将整个时空画在一张图上。这样的时空地图即是由彭罗斯和卡特引入的彭罗斯图。本节将详细讨论平坦的闵可夫斯基时空的彭罗斯图，以及施瓦西黑洞和 RN 黑洞的彭罗斯图。

9.1 黑洞视界

9.1.1 视界的定义

事件视界

首先写出施瓦西解,

$$ds^2 = -\left(1 - \frac{r_g}{r}\right) dt^2 + \left(1 - \frac{r_g}{r}\right)^{-1} dr^2 + r^2 d\Omega^2. \quad (9.1)$$

用上一章的记号即是 $A(r) = \left(1 - \frac{r_g}{r}\right)$, $B(r) = \left(1 - \frac{r_g}{r}\right)^{-1}$, 从而 $AB = 1$ 。显然, 函数 $A(r)$ 在 $r = r_g$ 处是一个一阶零点, 即在 r_g 附近, 有

$$A(r) \sim (2\kappa) \cdot (r - r_g) + \dots \quad (9.2)$$

显然, 在这里 $\kappa = 1/(2r_g) = 1/(4GM)$, 注意它和前面章节中出现的符号 κ 不是同一个东西。

在上一章中, 我们都是假设星体的半径大于甚至远大于其施瓦西半径 $r_g = 2GM$, 然后施瓦西解描述的是星体的外部时空, 所以在上一章中, 恒有 $r > r_g$ 。现在, 我们假设星体的半径小于 r_g , 甚至极端一点, 我们假设星体是 $r = 0$ 处的一个质点, 然后问, 现在施瓦西解描述的时空有何奇异之处?

最显眼的一个奇异之处, 从 (9.1) 式可以看到: 当 $r > r_g$ 时, $g_{tt} = -\left(1 - \frac{r_g}{r}\right) < 0$, 而 $g_{rr} = \left(1 - \frac{r_g}{r}\right)^{-1} > 0$, 所以, t 是时间坐标, 而 r 是空间坐标, 这和我们的预期一致。但是, 当 $r < r_g$ 时却有, $g_{tt} > 0$, 而 $g_{rr} < 0$, 所以, 这时候实际上 r 变成了时间坐标, 而 t 变成了空间坐标。即是说, 在 $r = r_g$ 处, 时间和径向空间互换了!

值得注意的是, 这种时空互换并不意味着观察者在穿过 $r = r_g$ 处的球面往中心掉落时其时钟会变成米尺而米尺会变成时钟。不会有这种戏剧性的事件发生, 之所以我们有这种错觉, 是因为我们选择了施瓦西 (t, r) 坐标的原因, 这个坐标在 $r = r_g$ 处是奇异的, 因为 $g_{tt}(r_g) = 0, g_{rr}(r_g) = \infty$, 所以这个坐标可以覆盖 $r < r_g$ 的区域, 也可以覆盖 $r > r_g$ 的区域, 但却不能同时覆盖两者, 如果你让它同时覆盖这两者, 那就会产生上面的错觉。后文我们将会讨论能够同时覆盖两个区域的时空坐标。

但是, 在经典物理中, 时间和空间性质有根本性的不同, 空间是可以前进也可以后退的, 但是时间却是单向的! 这就说明, 对于施瓦西解描述的时空而言, 一旦过了 $r = r_g$ 处的球面, 身处 $r < r_g$ 的区域, 则由于 r 方向变成了时间方向, 因此沿着 r 方向走只可能有两种选择 (这两种选择互为对方的时间反演): 要么 r 减少的方向为时间前进方向, 这时所有粒子都单向地朝着 r 减少的方向前进, 直至 $r = 0$ 的时间终点, 这就是黑洞, 这时 $r < r_g$ 的区域就叫黑洞区域; 要么 r 增加的方向为时间前进方向, 这时候 $r = 0$ 就变成了时间起点, 所有粒子都从这里出发单向地朝 $r = r_g$ 的球面前进, 然而穿过这个球面跑到外面的区域, 这就是所谓的白洞! 当然, 白洞通常被认为不是物理的, 因为黑洞可以由恒星发生引力坍缩而形成, 但却并没有形成白洞的物理机制。

为了将施瓦西解和黑洞的定义 (光无法逃逸) 联系起来, 我们考察光在施瓦西时空中

的运动, 由上一章的式子 (并代入 $AB = 1$),

$$\begin{aligned} \left(\frac{dr}{d\lambda}\right)^2 + \frac{1}{B}\left(\frac{l^2}{r^2}\right) - \frac{\varepsilon^2}{AB} &= 0 \\ \Rightarrow \left(\frac{dr}{d\lambda}\right)^2 + A(r)\left(\frac{l^2}{r^2}\right) - \varepsilon^2 &= 0. \end{aligned} \quad (9.3)$$

在 $r = r_g$ 处, 由于有 $A(r_g) = 0$, 所以显然 $\frac{dr}{d\lambda}|_{r_g} = \pm\varepsilon$, 这说明在跟着光子一起运动的参考系看来, $r = r_g$ 处并没有什么特殊的地方, 因为光子在此处的径向速度很正常。所以, 从跟着光子一起运动的参考系来看, 光子 (其它物质粒子也一样) 可以正常地穿过 $r = r_g$ 的球面。

下面, 我们从空间无穷远处的静止观察者的角度来看 $r = r_g$ 处。从 (9.1) 式可以看到, 对于无穷远处的静止观察者而言, 时间 t 其实就是其经历的固有时。结合上一章的式子 $\frac{dt}{d\lambda} = \frac{\varepsilon}{A}$ 和上面的 (9.3) 式, 可以得到

$$\left(\frac{dr}{dt}\right)^2 + A^2\left[A\frac{1}{r^2}\frac{l^2}{\varepsilon^2} - 1\right] = 0.$$

所以, 在 $r = r_g$ 附近, 有 $\left(\frac{dr}{dt}\right)^2 \sim A^2 \sim (2\kappa)^2(r - r_g)^2$ 。也即是说, 在无穷远处的静止观察者看来, 光子在 r_g 附近会运动得越来越慢, 甚至在 $r = r_g$ 处, 光速会减小到零, 所以, 无穷远处的静止观察者无法看到光穿越 $r = r_g$ 这个球面。不仅无法看到光穿越它, 类似的分析同样告诉我们, 无穷远处的静止观察者也无法看到任何物质粒子穿越 $r = r_g$ 这个球面。由于看不到光穿越 $r = r_g$, 所以 $r < r_g$ 的区域对于无穷远处的静止观察者而言当然是黑暗的, 光无法从这个区域逃逸到无穷远处去, 这就是黑洞! $r = r_g$ 处的球面随着时间演化在时空中扫过的柱形世界面就称作**黑洞事件视界!**它就是**无穷远处的静止观察者能够看到的时空区域的边界, 越过边界就是黑洞区域。**

反过来也一样, 无穷远处的观察者并不能看到外面的粒子真正掉进黑洞区域, 在无穷远处的观察者看来, 当粒子不断接近视界面时, 其速度会越来越缓慢, 直至无限趋近于零, 因此他并不能看到粒子真正掉进去, 只能看到粒子越来越慢越来越接近 $r = r_g$ 的视界面, 在无穷远处的观察者看来, 粒子需要无穷长的时间才能真正到达视界面。

视界面还有另外一种定义, 为了说清楚这种定义 (对于施瓦西解), 不妨考察一簇由方程 $f(r) = c$ 所描述的超曲面 (c 为这一簇超曲面的参数), 其中 $f(r)$ 是径向坐标的某个函数, 比方说如果 $f(r) = r - r_g$, 那么视界面就对应 $c = 0$ 。这样的超曲面簇自然有一个余法向量 n_μ ,

$$n_\mu = \partial_\mu f(r). \quad (9.4)$$

比方说对于 $f(r) = r - r_g$, 在 (t, r, θ, ϕ) 的坐标系中

$$n_\mu = (0, 1, 0, 0), \quad (9.5)$$

即是说, 它只有径向分量 $n_r = 1$, 其余分量都是零。这个余法向量在视界面上是一个类光向量, 因为,

$$g^{\mu\nu}n_\mu n_\nu = g^{rr}(\partial_r f(r))^2 = \left(1 - \frac{r_g}{r}\right)(\partial_r f(r))^2, \quad (9.6)$$

这个表达式限制在 $r = r_g$ 的视界面上显然为零, 满足 $(g^{\mu\nu}n_\mu n_\nu)|_{r_g} = 0$ 。

从上面的计算可以看出, 在视界面的外面, 即 $r > r_g$ 时, 余法向量类空, 即满足 $g^{\mu\nu}n_\mu n_\nu > 0$, 这表示沿着 $f(r) = c$ 超曲面的法向是一个空间方向, 沿着这个方向可以前进也可以后退。但是, 在视界面以内, 即 $r < r_g$ 的黑洞区域, 余法向量类时, 即满足 $g^{\mu\nu}n_\mu n_\nu < 0$, 这表示这时候 $f(r) = c$ 超曲面的法向是一个时间方向, 因此有单向性, 对于黑洞, 只能沿着法向朝 r 减小的方向前进。因此, 人们也称黑洞区域的 $f(r) = c$ 超曲面为**单向膜**, 而**黑洞事件视界, 就是单向膜区开始的那个超曲面**。

所以, 黑洞存在一个事件视界面, 它其实有非常特殊的性质。但是 (9.1) 式的 (t, r, θ, ϕ) 坐标系无法仔细地考察这些性质, 因为这个坐标系在 $r = r_g$ 处是奇异的, 它的 $g_{tt}|_{r_g} = 0$ 而 $g_{rr}|_{r_g} = \infty$ 。这些奇异性导致, 比方说, 时空上的余切向量 $n_\mu = (0, 1, 0, 0)$ 在视界面上的对偶向量为零, 即 $n^\mu|_{r_g} = 0$ (因为 $g^{rr}|_{r_g} = 0$), 类似的, 切向量 $\xi = \partial_t$, 即 $\xi^\mu = (1, 0, 0, 0)$ 在视界面上的对偶向量也为零 (因为 $g_{tt}|_{r_g} = 0$)。一个非零向量的对偶向量却是零, 这当然是病态的, 这种病态源自于 (t, r) 坐标在 $r = r_g$ 处的奇异性。

后文将会讨论一些在视界面上也定义良好的其它坐标系, 比如说后文将要研究的近视几何中的 (V, U) 光锥坐标系, 以及后面将要详细讨论的克鲁斯卡坐标系。

无限红移面

施瓦西黑洞的事件视界还有一层含义, 和引力红移有关。为了看清楚这一层含义, 让我们回想一下第三章中对引力红移的讨论。根据第三章中的相关讨论可以知道, 在一个静态时空中, 位于 \mathbf{x}_E 位置的信号发射者发射电磁波的频率 ν_E , 和位于 \mathbf{x}_R 位置的信号接收者接收到的电磁波频率 ν_R 之间满足如下关系

$$\frac{\nu_R}{\nu_E} = \left(\frac{g_{00}(\mathbf{x}_R)}{g_{00}(\mathbf{x}_E)} \right)^{-1/2}.$$

对于施瓦西时空, 即是

$$\frac{\nu_R}{\nu_E} = \left(\frac{1 - \frac{r_g}{r_R}}{1 - \frac{r_g}{r_E}} \right)^{-1/2}.$$

特别的, 如果信号接收者位于空间无穷远处, 则有

$$\nu_R = \nu_E \sqrt{1 - \frac{r_g}{r_E}}. \quad (9.7)$$

很显然, 无穷远接收者接收到的信号频率 ν_R 要比发射频率 ν_E 低, 在从发射地传播到接收地的过程中, 信号频率向更低频移动了, 这就是所谓的引力红移, 因为低频信号就是“更红”的信号。

从 (9.7) 式容易看出, 如果信号发射者越靠近黑洞视界, 则对于无穷远处的接收者而言, 它发出的信号就红移得越厉害。换句话说, 在无穷远观察者看来, 一事物越接近视界面, 则它的频率就红移得越低, 它的时间过程看起来就越慢, 最后, 当事物无限接近视界面时, 在无穷远观察者看来, 它的频率就无限接近于零了, 称之为发生了无限红移, 当然, 与此同时, 它的时间过程看起来也无限变慢, 正由于这样的无限变慢, 因此在无穷远观察者看来, 一个粒子要想真正达到视界面, 需要无穷长的时间!

无穷远观察者永远也不会看到一个粒子真正掉进黑洞之中！他只会看到它越来越红越来越慢，越来越接近视界面，因此施瓦西黑洞的事件视界同时又是所谓的无限红移面。

当然，对于更一般的黑洞，比如后面的章节中要讲述的克尔黑洞，无限红移面与事件视界就不是完全重合的一回事了。

9.1.2 近视界几何

为了搞清楚黑洞视界的性质，不妨将注意力集中在 $r = r_g$ 的视界附近，根据前面的分析，在视界附近 $A(r) \sim (2\kappa)(r - r_g)$ ，因此施瓦西解可以近似为（称作近视界几何）

$$ds^2 \sim -(2\kappa)(r - r_g)dt^2 + (2\kappa)^{-1}(r - r_g)^{-1}dr^2 + r_g^2 d\Omega^2. \quad (9.8)$$

重新定义径向坐标 ρ ，使得 $d\rho = (2\kappa)^{-1/2}(r - r_g)^{-1/2}dr$ ，积分即

$$\rho = (2\kappa)^{-1/2}2(r - r_g)^{1/2} \Leftrightarrow \rho^2 = 4r_g(r - r_g). \quad (9.9)$$

黑洞视界 $r = r_g$ 对应于 $\rho = 0$ 。利用这个新的径向坐标 ρ ，就可以将近视界几何重写成

$$ds^2 \sim -\kappa^2 \rho^2 dt^2 + d\rho^2 + r_g^2 d\Omega^2. \quad (9.10)$$

不妨忽略角度部分，仅仅考察径向和时间向这个 1+1 维的时空，从而近视界几何为

$$ds^2 = -\kappa^2 \rho^2 dt^2 + d\rho^2. \quad (9.11)$$

这样一个 1+1 维时空又称作 Rindler 时空。所以，施瓦西黑洞的近视界几何是一个 Rindler 时空。

为了搞清楚 Rindler 时空的本质，引入新的坐标 V, U ，

$$V = \rho e^{\kappa t}, \quad U = -\rho e^{-\kappa t}, \quad (9.12)$$

不难验证，在 (V, U) 坐标之下，Rindler 时空度规正好是

$$ds^2 = -dVdU. \quad (9.13)$$

但这正是平坦的 1+1 维闵可夫斯基时空，为了看清这一点可以令 $V = T + X$ ， $U = T - X$ ，从而使得度规变成标准的 $ds^2 = -dT^2 + dX^2$ ，换言之， V, U 也是平坦的 1+1 维闵可夫斯基时空的光锥坐标。由于我们是研究近视界几何得到上面的式子，所以它清楚地说明了黑洞视界面附近的时空并没有奇异性，它就是很正常的时空，所有的奇异性都来自于原来的 (t, r) 坐标。

当然，从 (9.12) 式可以看出，严格来说，Rindler 时空只是闵可夫斯基时空的四分之一，也就是 $V > 0, U < 0$ 的那四分之一，如图 (9.1) 所示（右侧满足 $X > |T|$ 的那四分之一）。

为了看清楚原来位于 $\rho = 0$ 处的黑洞视界在新的 (V, U) 坐标下如何描述，我们根据 (9.12) 式，得

$$VU = -\rho^2, \quad U/V = -e^{-2\kappa t}. \quad (9.14)$$

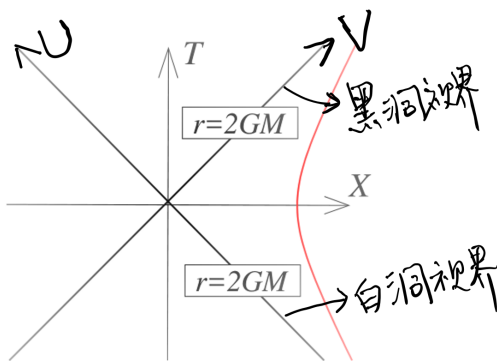


Figure 9.1: 施瓦西黑洞的近视界几何，红线表示位于黑洞外面 ($r > r_g$) 一个恒定 r 处的一个观察者.

由 $VU = -\rho^2$ 可见， $\rho = 0$ 现在对应 $V = 0$ 或 $U = 0$ 这两个坐标轴，在图 (9.1) 中也就是与垂直轴成 45 度角的那两条斜线，这就是对黑洞视界的新的描述，注意，现在的 (V, U) 坐标即使在视界面上也同样定义良好。

当然严格来说， $U = 0, V > 0$ 半轴对应的才是黑洞视界，而 $V = 0, U < 0$ 半轴对应的实际上是我们延拓出来的白洞的视界，后文会进一步进行相关讨论。

从 $U/V = -e^{-2\kappa t}$ 可见， t 等于常数的曲面现在对应 (V, U) 坐标系中过原点的一条斜率为常数的直线。特别的， $V = 0$ 对应 $t = -\infty$ ，而 $U = 0$ 对应 $t = \infty$ 。也就是说，在这种描述下，视界与 $t = \pm\infty$ 的超曲面重合，对此的物理解释是，从无穷远处的静止观察者看来（还记得吗？ t 是这种观察者的固有时），粒子要花无限长的时间才能到达视界。

将角度部分也包括进来，那么近视界几何即是

$$ds^2 = -dVdU + r_g^2 d\Omega^2. \quad (9.15)$$

假设把这个度规限制在 $U = 0, V > 0$ 的黑洞视界面上这个三维超曲面上，我们得到的度规即是

$$ds^2|_{U=0} = r_g^2 d\Omega^2 = r_g^2 (d\theta^2 + \sin^2 \theta d\phi^2), \quad (9.16)$$

很显然，这是一个退化的度规，它没有 dV^2 项，也没有 $dVd\theta$ 以及 $dVd\phi$ 这样的交叉项。换言之，视界面上的度规是退化的，其号差可以示意性地记为 $0++$ ，其中 $++$ 对应 $d\theta^2$ 和 $d\phi^2$ 之前的正定度规， 0 表示 dV 相关项的退化。一个具有 $0++$ 号差的超曲面就称之为一个**类光超曲面**，类光主要指沿着这个超曲面的度规退化方向走 $ds^2 = 0$ ，这当然也就是类光。所以，**黑洞的事件视界是一个类光超曲面！**

完全类似的讨论可以知道，白洞视界面也是一个类光超曲面，总之， $V = 0$ 和 $U = 0$ 的两个超曲面都是类光超曲面。

以 $U = 0$ 的类光超曲面为例，给定其上一点，其切空间的三个独立切向量为 $\partial_V, \partial_\theta, \partial_\phi$ ，根据与坐标无关的数学语言， $\langle \partial_\mu, \partial_\nu \rangle = g_{\mu\nu}$ ，从而不难得出

$$\langle \partial_V, \partial_V \rangle|_{U=0} = 0, \quad \langle \partial_V, \partial_\theta \rangle = \langle \partial_U, \partial_\phi \rangle = 0. \quad (9.17)$$

也即是说, 切向量 ∂_V 是类光向量! 并且 ∂_V 这个切向量与整个切空间都正交, 所以它其实也是 $U=0$ 的类光超曲面的法向量! 类似的, ∂_U 是 $V=0$ 的类光超曲面的类光切向量, 当然也是其法向量。这正是类光超曲面的特殊之处, 它的类光切向量也正是其法向量。

有一个定理告诉我们, 所有的类光超曲面都可以看成是由类光测地线铺成的, 证明可以参见 E. Witten 的经典综述文章 Light Rays, Singularities, and All That, 附录 E。这篇文章有中文译本, 叫《光线、奇点, 以及其它》, 是由我们一些人翻译的, 网上能搜到。从上面对于黑洞视界的这个描述也能清楚地看到, $U=0, V>0$ 的黑洞视界是由给定 θ, ϕ 值的类光测地线生成的, V 就是这些类光测地线的仿射参数! $V=0$ 的视界面也有类似结论。

9.1.3 基林 (Killing) 视界

从施瓦西解的表达式 (9.1) 容易看出, 这个解的度规表达式在如下 t 平移的单参微分同胚变换下保持不变

$$t \rightarrow t + c, \quad (9.18)$$

其中平移量 c 为任意常数。注意, 广义相对论虽然允许任意的微分同胚变换, 但是, 一般来说, 度规表达式在微分同胚变换下是要变的。因此, t 平移对于施瓦西解来说是特殊的, 它是施瓦西解的一个对称性, 而不仅仅是像其它微分同胚变换一样只是广义相对论理论 (也就是场方程) 的对称性, 它的对称从方程延伸到了解。

很明显, t 平移可以用如下切向量场 ξ 生成,

$$\xi = \partial_t. \quad (9.19)$$

这样的对应于度规对称性的切向量场, 就称作基林 (Killing) 向量场, 后面的章节中会进一步深入讨论基林 (Killing) 向量场以及它们和时空对称性的关系。总之, 施瓦西解有一个基林向量场 $\xi = \partial_t$ 。

很显然, 这个基林向量场满足

$$\langle \xi, \xi \rangle = \langle \partial_t, \partial_t \rangle = g_{tt} = -(1 - \frac{r_g}{r}). \quad (9.20)$$

显然, ξ 在 $r > r_g$ 处是类时的, 特别的, ξ 在空间无穷远处是类时的, 是无穷远处的一个类时基林向量场, 在无穷远处 ξ 归一化为 $\langle \xi, \xi \rangle|_{\infty} = -1$ 。但是, 在 $r < r_g$ 处, ξ 类空! 特别的, 在 $r = r_g$ 的视界面上, ξ 类光, 满足 $\langle \xi, \xi \rangle|_{r_g} = 0$ 。为了考察基林向量 ξ 在视界面上的行为, 下面我们来考察它在近视界几何光锥坐标 (V, U) 中的表达式。

假设在近视界几何中考察一个位于给定 ρ 位置的粒子随时间的演化, 不妨设时间平移了一个 c 吧, 即 $t \rightarrow t + c$ 。则根据 (9.12) 式, 有

$$V \rightarrow Ve^{kc}, \quad U \rightarrow Ue^{-kc}. \quad (9.21)$$

由于 V, U 也是平坦时空的光锥坐标, 所以这个式子的含义其实就是一个洛伦兹推动 (boost), 也即是说, 在 (V, U) 坐标中看来, 这个粒子的时间演化其实是一个洛伦兹推动 (boost)。

取 $c = \delta t$ 为无穷小时间平移, 则从 (9.21) 式容易看出,

$$\delta V = \kappa V \delta t, \quad \delta U = -\kappa U \delta t. \quad (9.22)$$

从而即可以得到基林向量场 ξ 在 (V, U) 坐标中的表达式, 为

$$\xi = \partial_t = \frac{\partial V}{\partial t} \frac{\partial}{\partial V} + \frac{\partial U}{\partial t} \frac{\partial}{\partial U} = \kappa \left(V \frac{\partial}{\partial V} - U \frac{\partial}{\partial U} \right). \quad (9.23)$$

用 (V, U) 坐标的分量形式来表达即是

$$\xi^\mu = \kappa(V, -U). \quad (9.24)$$

特别的, 对于 (V, U) 坐标的原点, $V = U = 0$, 从而 $\xi = 0$, 这样的点称为**基林分岔点** (bifurcation point), 当然, 如果加上角度部分的坐标, 那这样的点实际上就是一个两维球面, 称作**分岔球**。很显然, 这里整个 $V = 0$ 或 $U = 0$ 的视界面都是由分岔球上所发出的类光测地线扫出的。

在 $U = 0$ 的视界面上, 基林向量 $\xi^\mu = \kappa(V, 0)$, 由于近视界面度规为 $-dVdU$, 所以这显然是 $U = 0$ 视界面的一个类光切向量, 同时, 因为 $U = 0$ 的视界面为类光超曲面, 所以这时 ξ^μ 也为这个超曲面的类光法向量。类似的, 在 $V = 0$ 的视界面上 (此时 $\xi^\mu = \kappa(0, -U)$), 基林向量 ξ 也是一个类光切向量, 当然也是类光法向量。

如果一个类光超曲面以一个基林向量为类光法向量, 则我们就称之为一个**基林视界**, 所以, 施瓦西黑洞的事件视界面同时也是基林视界。

再强调一下, 所谓的**基林视界**, 其一般定义是, 首先它是一个类光超曲面, 其次, 其类光法向量 (也是一个切向量) 刚好是一个基林向量 (不妨一般性地记为 K^μ), 当一个超曲面同时满足这两点时, 就称之为一个基林视界。所以, 从定义来看, 基林视界和事件视界的定义是相互独立的, 两者并不一定重合。不过, 对于施瓦西黑洞而言, 其事件视界正好也是其基林视界! 而且相应基林向量为在无穷远处类时的 ξ , 即 $K = \xi$ 。

实际上, 霍金和卡特曾经证明过一个定理: **在稳态时空中, 事件视界就是类时基林视界**。当然, 这里的类时指的是这个基林向量场在空间无穷远处类时! 施瓦西时空就是这个定理的一个典型例子。

根据前面提及的 E.Witten 文章里的定理, 作为一个类光超曲面, 基林视界当然也是由类光测地线铺成的, 不防记这些类光测地线的切向量场为 l^μ , 它满足如下测地线方程

$$l^\mu D_\mu l^\nu = 0. \quad (9.25)$$

l^μ 当然是一个类光切向量, 因此它也是基林视界的法向量场, 因此在基林视界上, 它必定和同样作为法向量场的基林向量场 K^μ 成比例, 即存在函数 $\chi(x)$, 使得

$$K^\mu = \chi(x) l^\mu. \quad (9.26)$$

注意 l^μ 只在基林视界上有定义, 从而这个式子也只在基林视界上成立。

从而根据 (9.25) 式, 不难得到, 在基林视界上, K^μ 满足方程 $K^\mu D_\mu K^\nu = \chi(x) l^\mu D_\mu (\chi(x) l^\nu) = \chi(x) (l^\mu \partial_\mu \chi(x)) l^\nu = (l^\mu \partial_\mu \chi(x)) K^\nu$, 记式中的 $l^\mu \partial_\mu \chi(x) = w(x)$, 即有方程

$$K^\mu D_\mu K^\nu = w(x) K^\nu. \quad (9.27)$$

这个方程的本质其实依然是测地线方程, 不过测地线的参数不是仿射参数而已, 这个式子当然也只在基林视界上成立。

稍后我们将证明, 对于施瓦西黑洞的视界面 (以及类似的一类球对称黑洞), 如果取基林向量 $K = \xi = \partial_t$, 则必定有 $w(x) = \pm \kappa$, 这里的 κ 也就是前面引入的 $1/(2r_g) = 1/(4GM)$ 。换言之, 对于施瓦西黑洞的视界面 (以及类似的一类球对称黑洞), 我们有

$$\xi^\mu D_\mu \xi^\nu = \pm \kappa \xi^\nu. \quad (9.28)$$

这个结果依赖于 ξ 的归一化, 因为如果将 ξ 替换成 $c\xi$, 那上式中的 κ 就应该替换成 $c\kappa$ 。这里是取 $\xi = \partial_t$, 因此是将 ξ 的模长平方在空间无穷远处归一为负一, 即 $\langle \xi, \xi \rangle|_\infty = -1$ 。而之所以有正负号的不确定性, 是因为如果将 ξ 替换成时间反演版的 $-\xi$, 那 $w(x)$ 就将变成 $-w(x)$, 而对于施瓦西黑洞而言, $\xi = \partial_t$ 和它的时间反演是地位平等的, 都可以作为类时基林向量场。值得说明的是, 通常人们称方程 (9.28) 中的 κ 为表面引力, 因此本章前面引入的 $\kappa = 1/(2r_g)$ 当然就是施瓦西黑洞视界上的表面引力。

下面, 我们来证明定义了表面引力的 (9.28) 式。由于这个式子只定义在基林视界上, 所以要证明它显然只需要考察基林视界附近的情况, 从而只需研究近视界几何就足够了! 为此, 首先注意到, 在 (V, U) 坐标中, 近视界几何的度规是平凡的 $-dVdU$, 从而可知在这个坐标中, 联络系数都等于零, 从而在 (V, U) 坐标中协变导数就等于普通的导数。所以在基林视界附近我们有

$$\begin{aligned} \xi^\mu D_\mu \xi^\nu &= \xi^\mu \partial_\mu \xi^\nu \\ &= \kappa \left(V \frac{\partial}{\partial V} - U \frac{\partial}{\partial U} \right) \kappa(V, -U) = \kappa^2(V, U). \end{aligned} \quad (9.29)$$

所以, 在 $U = 0$ 的基林视界上 (此时 $\xi^\nu = \kappa(V, 0)$), 我们有

$$\xi^\mu D_\mu \xi^\nu = \kappa^2(V, 0) = \kappa \xi^\nu. \quad (9.30)$$

而在 $V = 0$ 的基林视界上 (此时 $\xi^\nu = \kappa(0, -U)$), 我们有

$$\xi^\mu D_\mu \xi^\nu = \kappa^2(0, U) = -\kappa \xi^\nu. \quad (9.31)$$

这就证明了 (9.28) 式。

不仅如此, 在后面的章节中我们还将证明, 对于任意时空的任意基林视界, 只要它可以由其分岔球上的类光测地线扫出, 则 (9.27) 式中的 $w(x)$ 的平方 (w^2) 沿着整个基林视界都必定为常数! 从而在基林视界上 $w(x)$ 只能和一个常数 (就记作 κ) 相差正负号, 所以可以一般地记 $w(x) = \pm \kappa$, 进而一般性地将方程 (9.27) 重写为

$$K^\mu D_\mu K^\nu = \pm \kappa K^\nu. \quad (9.32)$$

对于类时基林视界，通常还要求 K^μ 满足归一化条件 $\langle K, K \rangle|_\infty = -1$ 。注意，这个方程和上一段证明的结论的区别在于，这个方程的成立条件要广泛得多，而上一段只是它在施瓦西视界情形的特例。

9.1.4 表面引力与霍金温度

黑洞视界是一个物质有去无回的单向膜区开始的地方。然而，1974年，霍金发现，在考虑了量子效应以后，黑洞也能往外辐射，它有一个非零的温度，这就是霍金温度。

为了推导出霍金温度的表达式，设想在空间无穷远处，将一个量子场 $\phi(x)$ 与黑洞耦合，设想量子场与黑洞达成了热平衡，温度为 T_{BH} ，温度的负一次方记作 $\beta = 1/T_{BH}$ （假设取玻尔兹曼常数 $k_B = 1$ 的单位制）。记量子场的哈密顿量为 H ，则量子场的配分函数为

$$\mathcal{Z} = \text{Tr}(e^{-\beta H}). \quad (9.33)$$

人们可以用路径积分的方法计算这个配分函数，为此，我们需要将时间延拓为欧氏时间 t_E

$$t \rightarrow -it_E, \quad (9.34)$$

并将这个欧氏时间紧致化为一个周长为 $\hbar\beta$ 的圆周，也就是要进行如下的周期等同

$$t_E \sim t_E + \hbar\beta. \quad (9.35)$$

然后在这个时间紧致化了的欧氏时空中对量子场进行欧氏路径积分，结果就是上面的配分函数。

很显然，要推导出霍金温度的公式，就是要确定上面周期等同的 β 具体为多少？为此，让我们来考察上述欧氏化对近视界几何的影响。

延拓到欧氏时间 t_E 以后，近视界几何的 Rindler 度规就变成了

$$ds^2 = \kappa^2 \rho^2 dt_E^2 + d\rho^2. \quad (9.36)$$

现在，我们定义变量 $\theta_E = \kappa t_E$ ，并将之理解成一个角度变量，那么上面的度规就成为

$$ds^2 = \rho^2 d\theta_E^2 + d\rho^2. \quad (9.37)$$

很显然，如果 θ_E 以 2π 为周期，那这个度规就是两维欧氏平面的极坐标度规，特别的，径向坐标 $\rho = 0$ 的原点没有任何奇异性。而如果 θ_E 不以 2π 为周期，而是其它的某个周期，那上面这个度规描述的就是一个圆锥面， $\rho = 0$ 就是这个圆锥的顶点，它当然是奇异的，这种奇异性也称作锥形奇异性。但是， $\rho = 0$ 对应的是黑洞视界，它并不是一个时空奇异的地方。因此，为了消除锥形奇异性，我们只能取 θ_E 的周期为 2π ，也即是

$$\kappa t_E \sim \kappa t_E + 2\pi \Leftrightarrow t_E \sim t_E + 2\pi/\kappa. \quad (9.38)$$

所以， t_E 的紧化周期 $\hbar\beta$ 必须为 $2\pi/\kappa$ ，也即是

$$\beta = \frac{2\pi}{\hbar\kappa} \Leftrightarrow T_{BH} = \frac{\hbar\kappa}{2\pi}. \quad (9.39)$$

如果恢复玻尔兹曼常数，即是

$$T_{BH} = \frac{\kappa \hbar}{2\pi k_B}. \quad (9.40)$$

以上就是著名的霍金温度，它由表面引力决定，表达式中 \hbar 的存在意味着，这个温度完全是一个量子效应，在经典物理的层次上，一个黑洞是没有什么温度概念的。另外，值得说明的是，霍金温度之下施瓦西黑洞与量子场的热平衡其实是不稳定的，因为施瓦西黑洞会吸收物质，从而使得质量增加（即能量增加），但是温度却降低（注意 $\kappa \propto 1/M$ ），进而使得能量会进一步向施瓦西黑洞传输，也就是说，施瓦西黑洞具有负比热。

9.2 克鲁斯卡延拓

9.2.1 克鲁斯卡延拓的一般原理

假设我们考察如下球对称时空

$$ds^2 = -f(r)dt^2 + \frac{1}{f(r)}dr^2 + r^2d\Omega^2, \quad (9.41)$$

如果用之前的记号即是考察 $A(r) = f(r), B(r) = 1/f(r)$ 的球对称时空，施瓦西时空是其特例。按照前面的分析，黑洞视界必定对应函数 $f(r)$ 的零点（如果 $f(r)$ 没有零点，那相应的时空就不存在黑洞）。假设 $f(r)$ 在 $r = r_h$ 处有一个零点，即在 r_h 附近，有

$$f(r) \sim (2\kappa)(r - r_h) + \dots \quad (9.42)$$

式中 2κ 为泰勒展开的系数。很显然，在这个零点位置， (t, r) 坐标是奇异的，因此它不能同时覆盖 $r < r_h$ 的区域和 $r > r_h$ 的区域。下面我们来寻找能够同时覆盖这两个区域的一组坐标系。

为此，我们将 (9.41) 式的度规改写成如下形式

$$ds^2 = -f(r)\left(dt + \frac{dr}{f(r)}\right)\left(dt - \frac{dr}{f(r)}\right) + r^2d\Omega^2. \quad (9.43)$$

定义新的径向坐标 r_* ，使得 $dr_* = \frac{dr}{f(r)}$ ，也即

$$r_* = \int \frac{dr}{f(r)}, \quad (9.44)$$

显然，函数 $f(r)$ 的零点就对应这个积分表达式的极点。下面，定义新的坐标 v, u

$$v = t + r_*, \quad u = t - r_*, \quad (9.45)$$

从而可以将度规 (9.43) 重写成

$$ds^2 = -f(r)dvdu + r^2d\Omega^2. \quad (9.46)$$

很明显， v, u 为两个类光坐标。值得注意的是， (v, u) 坐标在 $f(r)$ 的零点 $r = r_h$ 处依然是奇异的，因为 (v, u) 坐标下的度规 (9.46) 在这个零点处是病态的（有 $g_{vu}(r_h) = 0$ ）。

当然, (9.45) 式也可以重写成

$$v + u = 2t, \quad v - u = 2r_*. \quad (9.47)$$

先将注意力集中在 $v - u = 2r_*$ 上, 它也即是

$$v - u = 2 \int \frac{dr}{f(r)}. \quad (9.48)$$

现在考察这个积分表达式的极点 (也就是 $f(r)$ 的零点) 所控制的一个邻域, 并在这个邻域上将极点的积分分离出来, 即写成

$$\begin{aligned} v - u &= \int \left[\frac{1}{\kappa(r - r_h)} + g(r) \right] dr \\ &= \frac{1}{\kappa} \ln \left| \frac{r - r_h}{r_h} \right| + G(r), \end{aligned} \quad (9.49)$$

式中 $g(r)$ 表示分离出极点以后的一个不含极点的被积项, $G(r)$ 是它的积分, 当然由于被积函数没有极点, 所以 $G(r)$ 一定是一个在这个邻域上光滑的函数。由 (9.49) 式, 即有

$$e^{\kappa(v-u)} = \left| \left(\frac{r - r_h}{r_h} \right) \right| e^{\kappa G(r)}. \quad (9.50)$$

下面来改进我们的类光坐标, 将之变成类似于近视界几何中的那种光锥坐标 (V, U) , 使之在 $r = r_h$ 处不再奇异。首先, 根据上面的 (9.50) 式, 我们有, 当 $r < r_h$ 时 (黑洞或者白洞区域)

$$e^{\kappa(v-u)} = - \left(\frac{r - r_h}{r_h} \right) e^{\kappa G(r)}. \quad (9.51)$$

对于黑洞区域, 由于粒子总是朝着视界里面一去不复返的 (即走向未来), 所以它应该处于这种 (V, U) 坐标的未来部分, 反之白洞区域则应该处于过去部分, 比方说, 黑洞区域和白洞区域可以分别处于光锥坐标 (V, U) 的第一象限和第三象限, 即上部 (未来) 和下部 (过去)。为此我们可以进行如下操作: 我们可以在 $r < r_h$ 的邻域上定义新的坐标 V, U 如下

$$V = \pm e^{\kappa v}, \quad U = \pm e^{-\kappa u}, \quad (9.52)$$

从而使得 (利用了 $v + u = 2t$)

$$VU = - \left(\frac{r - r_h}{r_h} \right) e^{\kappa G(r)}, \quad U/V = e^{-\kappa(v+u)} = e^{-2\kappa t}. \quad (9.53)$$

很显然, (9.52) 式定义的两可能性的确分别覆盖了 (V, U) 坐标系的一、三象限。

同样根据上面的 (9.50) 式, 我们有, 当 $r > r_h$ 时 (外部区域)

$$e^{\kappa(v-u)} = \left(\frac{r - r_h}{r_h} \right) e^{\kappa G(r)}. \quad (9.54)$$

当然, 这个黑洞和白洞外部的区域只可能占据剩下的二、四象限了 (也即左右两侧)。为此我们需要进行如下操作: 我们可以在 $r > r_h$ 的邻域上定义新的坐标 V, U 如下

$$V = \pm e^{\kappa v}, \quad U = \mp e^{-\kappa u}, \quad (9.55)$$

以使得

$$VU = -\left(\frac{r-r_h}{r_h}\right)e^{\kappa G(r)}, \quad U/V = -e^{-\kappa(v+u)} = -e^{-2\kappa t}. \quad (9.56)$$

很显然, (9.55) 式定义的可能性的确分别覆盖了 (V, U) 坐标系的二、四象限。

从 (9.53) 式和 (9.56) 式可以看到, $r = r_h$ 的视界面正好对应 $V = 0$ 或 $U = 0$ 的两个坐标轴, 并且它正好分别和 $t = \pm\infty$ 的超曲面重合, 总之, 这些都类似于前面在研究近视界几何时所得到的结论。

不难验证, 无论 $r > r_h$ 还是 $r < r_h$, 均有

$$dVdU = -\kappa^2 VU dvdu = \kappa^2 \left(\frac{r-r_h}{r_h}\right)e^{\kappa G(r)} dvdu. \quad (9.57)$$

从而能够将时空度规 (9.46) 重写为

$$ds^2 = -\frac{r_h}{\kappa^2} \frac{f(r)}{r-r_h} e^{-\kappa G(r)} dVdU + r^2 d\Omega^2. \quad (9.58)$$

式中的变量 r 应该理解为通过反解 $VU = -\left(\frac{r-r_h}{r_h}\right)e^{\kappa G(r)}$ 式所得到的关于变量 VU 的表达式。很明显, 在这个式子中, $f(r)$ 的零点被分母上的 $r-r_h$ 消去了, 从而度规在 $r = r_h$ 处不再病态, 因此 (V, U) 坐标系在 $r = r_h$ 处不奇异。

总之, (9.52) 式和 (9.55) 式给出 V, U 的四种可能定义, 它们分别覆盖 (V, U) 坐标系的四个象限, 但是在每一个象限里面, 与 V, U 对应的 v, u 的取值范围都是 $(-\infty, +\infty)$, 所以用 v, u 坐标来看的话, 每一个象限都是一个完整的时空, 四片完整的时空拼合在一起就构成整个大的由 (V, U) 坐标描述的时空。而从 (v, u) 所描写的一个象限延拓到完整的 (V, U) 所描写的四个象限的过程, 就叫做**克鲁斯卡延拓**, (V, U) 坐标有时也称作**克鲁斯卡坐标**, 如图 (9.2) 所示。值得强调的是, 克鲁斯卡坐标在 $r = r_h$ 的视界面上是定义良好的, 没有奇

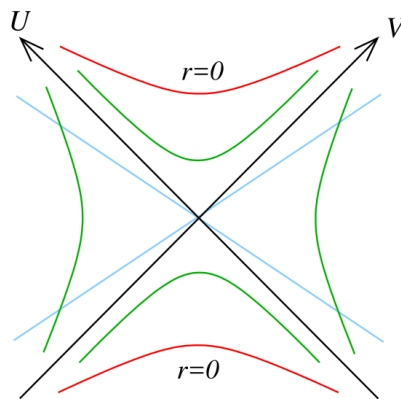


Figure 9.2: 施瓦西黑洞的克鲁斯卡延拓。图中 U, V 轴对应 $r = r_h$ 的黑洞视界, 红线是 $r = 0$ 处的时空奇性, 绿线表示 r 为常数的超曲面, 而蓝线表示 t 为常数的超曲面。

异性。

举例来说, 比如对于 $V = +e^{\kappa v}$, $U = -e^{-\kappa u}$ 所描述的第四象限 (即右侧的四分之一), 它对应的当然是 $r > r_h$ 的黑洞外部时空, $V = 0$ 或 $U = 0$ 的事件视界将它和黑洞以及白洞内部分隔开来, 当然, 严格来说 $U = 0, V > 0$ 半轴对应的才是黑洞视界, 因为它处于 (V, U) 坐标系的未来部分, 是所谓的未来事件视界。而 $V = 0, U < 0$ 半轴对应的视界是白洞视界, 因为它处于 (V, U) 坐标系的过去部分, 是过去事件视界。

根据上面这个分析, 我们就容易知道, 黑洞视界所分隔的一、四象限中, 第四象限对应黑洞外部, 第一象限所描述的则是黑洞内部 (即上部的四分之一)。类似的, 白洞视界所分隔的三、四象限中, 第四象限是白洞外部, 而第三象限是白洞内部 (即底部的四分之一)。所以, 正如我们所期望的, 在 (V, U) 坐标中, 处于未来的第一象限描述的是黑洞内部, 而处于过去的第三象限描述的则是白洞内部, 而第四象限和第二象限所描述的则是黑洞和白洞的共同外部区域。至于第二象限和第四象限这两个外部区域之间是什么关系, 我们稍后再作分析。

上述时空延拓过程的奇妙之处在于, 你可以从爱因斯坦方程在某个时空区域的一个解开始, 通过变换坐标进行延拓, 得到一个更大的时空区域上的解。注意, 对于新坐标所描述的更大时空区域, 你无需再次求解爱因斯坦场方程, 因为延拓来的度规一定会满足它。之所以有这么一个奇妙的性质, 其原因在于两个因素的共同作用: 首要的因素是, 爱因斯坦场方程是广义协变的, 其成立不受坐标变换的影响。其次, 场方程解出来的度规的各分量均是实解析函数, 这意味着, 在一个开集上的度规足以决定整个更大的时空区域的度规。

基林视界

(9.41) 式描写的球对称时空显然有一个基林向量场 ξ ,

$$\xi = \partial_t. \quad (9.59)$$

很容易计算出这个基林向量场的模长平方, 为

$$\langle \xi, \xi \rangle = g_{tt} = -f(r). \quad (9.60)$$

一般来说, $r \rightarrow \infty$ 时 $f(r) \rightarrow 1$, 所以 ξ 在空间无穷远处是类时的。假设 $f(r)$ 仅在 $r = r_h$ 处有唯一一个零点, 由于经过零点 $f(r)$ 要变号, 所以很显然, $r > r_h$ 时均有 $f(r) > 0$, 因此这时 ξ 均类时。而 $r < r_h$ 时有 $f(r) < 0$, 所以这时 ξ 类空。当然, 在 $r = r_h$ 的视界面上 ξ 类光。从而视界面同时也是基林视界。

不妨将注意力限制于 (V, U) 坐标的右侧四分之一 (第四象限)。设想在基林向量 ξ 的作用之下, t 坐标平移一个无穷小量 δt , 则根据 (9.45) 式, 有

$$\delta v = \delta u = \delta t. \quad (9.61)$$

进而根据 $V = +e^{\kappa v}$, $U = -e^{-\kappa u}$, 有

$$\delta V = \kappa V \delta v = \kappa V \delta t, \quad \delta U = -\kappa U \delta u = -\kappa U \delta t. \quad (9.62)$$

从而即可以得到基林向量场 ξ 在克鲁斯卡坐标中的表达式, 为

$$\xi = \partial_t = \kappa \left(V \frac{\partial}{\partial V} - U \frac{\partial}{\partial U} \right). \quad (9.63)$$

不难验证，其它象限求出来的最终表达式是一样的。这个结论在形式上完全和近视界几何所得出来的一样。特别的， $V = U = 0$ 是一个分岔球。

9.2.2 施瓦西黑洞的克鲁斯卡延拓

奇点

对于施瓦西解， $f(r) = (1 - \frac{r_g}{r})$ ，所以度规在 $r = 0$ 处也是奇异的。前面提到过，对于施瓦西黑洞而言， $r = 0$ 对应的是时间的一个终点，注意，不是空间的终点，因为在 $r < r_g$ 区域，时间和径向空间互换了。因为 $r = 0$ 是时间终止的地方，所以此处的奇异性和视界面处的度规奇异性有根本性的不同，具体来说，视界面的度规奇异性是坐标奇异性，换一个坐标这种奇异性就可能被消除，比方说在克鲁斯卡坐标中视界所对应的 $U = 0$ 或 $V = 0$ 就没有什么奇异性。但是， $r = 0$ 处的奇异性是真正的施瓦西时空的奇异性，称作**时空奇异性**。

这种时空奇异性反映为无穷大的时空曲率，换言之，反映为黎曼曲率张量在 $r = 0$ 处的发散。计算表明，对于施瓦西解，有

$$R^{\rho\sigma\mu\nu}R_{\rho\sigma\mu\nu} = \frac{12r_g^2}{r^6}, \quad (9.64)$$

所以在 $r = 0$ 处黎曼曲率张量确实是发散的，这种发散不依赖于坐标系的选择！这个计算同时也表明，在 $r = r_g$ 的视界面处，时空曲率其实是正常的，并没有发散。所以黑洞视界并没有任何时空奇异性，它的特殊之处不在于奇异性，而在于，从时空整体的因果结构来说黑洞视界比较特殊。

施瓦西解的克鲁斯卡延拓

下面来考察具体的施瓦西时空的克鲁斯卡延拓。对于施瓦西解而言 $f(r) = (1 - \frac{r_g}{r})$ ，所以

$$2\frac{1}{f(r)} = \frac{2r}{r-r_g} = \frac{2r_g}{r-r_g} + 2. \quad (9.65)$$

与 (9.49) 式比较，即知 $r_h = r_g$ ， $\kappa = 1/(2r_g)$ ， $g(r) = 2$ ，从而

$$G(r) = 2r. \quad (9.66)$$

由此当然就可以定义克鲁斯卡坐标，根据 (9.58) 式，在这个坐标中施瓦西时空的度规就成为

$$ds^2 = -\frac{r_g}{\kappa^2} \frac{1}{r} e^{-2\kappa r} dV dU + r^2 d\Omega^2. \quad (9.67)$$

另外，根据 (9.53) 式，

$$VU = -\left(\frac{r-r_g}{r_g}\right)e^{2\kappa r}. \quad (9.68)$$

特别的，对应**时空奇异性**的 $r = 0$ 现在成为

$$r = 0 \Leftrightarrow VU = 1. \quad (9.69)$$

它对应 (V, U) 坐标系中位于一、三象限内的一条双曲线，其中在第三象限 (也即白洞区域) 内的部分描述的是时间的一个起始之处，而位于第一象限 (也即黑洞区域) 内的部分描述的是时间的一个终止之处。所以，在克鲁斯卡坐标系中， $VU = 1$ 描述的是时空在一、三象限内的**边界**，边界之外时空是不存在的！显然，施瓦西解的时空奇性是**类空的**。整个施瓦西黑洞的克鲁斯卡延拓如图 (9.3) 所示。注意，由于我们没有画出角度部分的坐标，所以

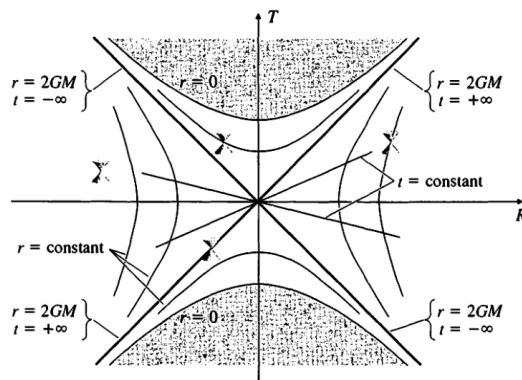


Figure 9.3: 施瓦西黑洞的克鲁斯卡延拓。图中 $V = T + R$, $U = T - R$, 请注意 $V - U$ 坐标系的象限划分和 $T - R$ 坐标系的象限划分的不同。

其实这个图中的每一个点都代表一个两维球面，面积为 $4\pi r^2$ 。

可以看到，在克鲁斯卡图的上部，任何一个类时或者类光的曲线 (切线与垂直轴的夹角小于等于 45 度) 只要过了视界那条直线就休息逃离上部四分之一的区域，只能撞在 $r = 0$ 的时空奇性上，所以上部四分之一的区域就是黑洞区域。

而在下部四分之一区域，注意到时间的单向性，可以看到任何一个类时或者类光曲线都会穿过视界逃离这个区域，这个区域和黑洞区域相对，就叫做白洞区域。

克鲁斯卡图可以直观地显示一些很微妙的事情，比如如图 (9.4) 所示的，某 A 掉进黑洞，某 B 在视界外面固定 r 距离处观察这一过程。 B 当然是通过接收 A 发出的光信号而观察他的，这些光信号就是图中所示的与 V 轴平行的那些线段，它们均与垂直轴成 45 度夹角。如图所示， A 在 P_A 点发出的信号将在 P_B 点到达 B ，这个时候的时间 t 是 $t = 1$ ，接着， A 在 Q_A 点发出的信号将在 Q_B 点到达 B ，这个时候的时间 t 是 $t = 100$ ，而 A 在 R_A 点发出的信号将在 R_B 点到达 B ，这个时候的时间 t 是 $t > 1000$ ，很显然，随着 A 越来越靠近视界， t 将趋于 ∞ ，即 B 永远也不能看到 A 真正掉进黑洞，只能看到他不断接近视界面。但是，在 A 本身看来，他掉进黑洞的事件当然发生了，而且发生在 A 的有限固有时之内。

有一个问题是，克鲁斯卡延拓以后的施瓦西时空还能被进一步延拓吗？回答是，不够了，克鲁斯卡延拓就是施瓦西时空的**最大延拓**。

这里有一个简单的方法来判断一个时空是否能被进一步延拓：那就是观察时空中的测地线，看它们在哪里终止。假如你能跟着一条测地线跑到其固有时参数或者仿射参数的无穷值处，那这时候你也就跑到时空的无穷远处去了。假如相反，测地线在一个固有时参数 (或者仿射参数) 的有限值处终止了，不能进一步继续延伸了，那这时候就一定有如下事情

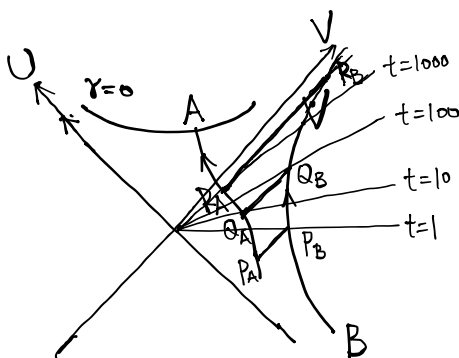


Figure 9.4: 某 A 掉进黑洞，某 B 在视界外固定 r 处观察他。

之一发生：要么这条测地线遇到了时空奇性，要么它遇到了坐标奇性。假设只是遇到了坐标奇性，那换一个更好的坐标系，测地线就还能继续延伸，这时候我们就说时空能被这个更好的坐标系延拓。而假如测地线遇到的是时空奇性，那我们就做不了什么了，它就真的无法进一步延伸了。**所谓的最大延拓**，指的是，任何在固有时参数（或者仿射参数）的有限值处终止的测地线所遇到的都是时空奇性的那种延拓。

不难检验，施瓦西解的克鲁斯卡延拓的确是最大延拓！

爱因斯坦-罗森桥

前面提到过，在克鲁斯卡延拓中，黑洞和白洞外面的区域一共有两片，分别对应 (V, U) 坐标系的第二象限和第四象限（即左侧的四分之一和右侧的四分之一）。这两个区域都在空间无穷远处渐近平坦，但是这两个渐近平坦的区域通过黑洞和白洞连接起来了。从图 (9.3) 中可以清楚的看到，第四象限（右侧）的观察者其实无法向第二象限（左侧）的观察者发射信号，因为任何这样的信号都会掉进黑洞，并不可避免地终止在时空奇性上。然而，我们不禁想知道，连接这两个区域的空间几何到底是怎么样的呢？

为此，我们不妨考察克鲁斯卡时空在 $t=0$ 时的纯空间部分，很显然，这就是一条穿过原点的水平轴，它在右侧四分之一的部分和在左侧四分之一的部分均对应黑洞的外部，空间度规均为

$$ds^2 = \left(1 - \frac{r_g}{r}\right)^{-1} dr^2 + r^2 d\Omega^2, \quad (9.70)$$

并且 $r > r_g$ ，这两个区域通过 $U = V = 0$ 的原点连接起来，但是 $U = V = 0$ 的原点其实代表一个半径 $r = r_g$ 的两维球面。所以完整的 $t=0$ 空间几何其实是两片渐近平坦的空间在 $r = r_g$ 的两维球面处粘合在一起，如图 (9.5) 所示。图中上面和下面的两片分别代表克鲁斯卡图中左右两侧的两个渐近平坦区。连接两片渐近平坦时空的粘合区（也就是图中的“喉咙”部分）就称作爱因斯坦-罗森桥，它就是所谓的时空虫洞的一种。

我们可以通过在近视界几何的探讨中引入的坐标 $\rho^2 = 4r_g(r - r_g)$ 来更清楚地考察爱因斯坦-罗森桥。不过具体的处理有所不同，为了避免混淆，这里我们将把 ρ 记为 z 。下面我们取 $\theta = \pi/2$ ，从而 $d\Omega^2 = d\phi^2$ ，而空间变成两维的，由极坐标 (r, ϕ) 表示。当这样处理以后，(9.70) 式描写的空间几何（现在是二维），就可以看作平坦的三维空间（取柱坐标

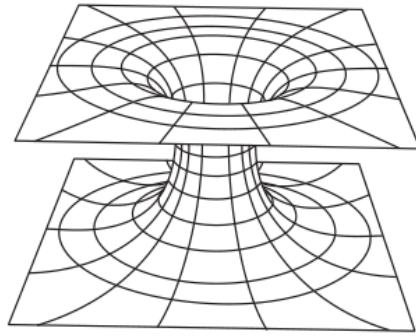


Figure 9.5: 爱因斯坦-罗森桥，图中我们已经取 $\theta = \pi/2$ ，因此两维球面变成一个圆周，而 $d\Omega^2 = d\phi^2$ 。

(r, ϕ, z) 中的由下式描述的三维超曲面

$$z^2 = 4r_g(r - r_g). \quad (9.71)$$

我们很容易验证度规的确是正确的，为此首先注意到平坦三维空间的度规为 $dr^2 + r^2 d\Omega^2 + dz^2$ ，其次，注意到 $z dz = 2r_g dr$ ，从而 $dr^2 + dz^2 = (1 + \frac{4r_g^2}{z^2}) dr^2 = (1 - \frac{r_g}{r})^{-1} dr^2$ ，所以这张三维超曲面上的诱导度规的确是 (9.70) 式。

下面只需在三维平坦空间中示意性地画出 (9.71) 式所描写的这张二维超曲面，显然，结果就是图 (9.5)。

进一步，在克鲁斯卡坐标中，如果定义 $V = T + R$, $U = T - R$ (从而 $VU = T^2 - R^2$)，然后取一些固定时间 T 的空间切片来考察 (如图 (9.6))，就可以看到虫洞的打开和关闭过程 (如图9.7)。在图 (9.6) 中，我们取定了 A, B, C, D, E 五个固定时间 T 的空间切片，与之相

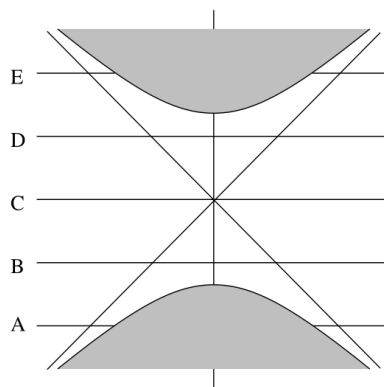


Figure 9.6: 取定 A, B, C, D, E 五个空间切片。

应的，在图 (9.7) 中，我们示意性地画出了随着时间演化，虫洞在黑洞视界后面打开最后又关闭的示意图。 $T = -1$ 时虫洞开始打开， $T = 1$ 时虫洞关闭。

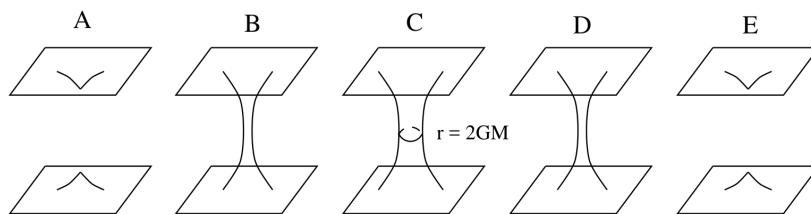


Figure 9.7: 与 A,B,C,D,E 五个空间切片相对应的虫洞打开和关闭过程示意图。

ER=EPR?

首先解释一下黑体字的小标题: ER 代表的就是 Einstein-Rosen bridge, 即刚刚讲述的爱因斯坦-罗森桥, 也即时空虫洞。而 EPR 代表的是 Einstein-Podolosky-Rosen paradox, 即 EPR 佯谬也就是通常所谓的量子纠缠。所以, 这个标题的字面含义就是, 时空虫洞等价于量子纠缠。

为了进一步理解这个小标题, 下面我们作一点解释说明。

前面提到过, 在施瓦西解的克鲁斯卡延拓中, 位于左侧四分之一区域的观察者并不能发送信号给位于右侧四分之一区域的观察者。但是, 这两个观察者的确有一种方式通信, 那就是两者都跳进黑洞里, 然后在各自终结于时空奇性之前在黑洞视界后面相互通信, 甚至在黑洞里彼此遇到。

但是, 以上这一段话真有物理含义吗? 毕竟, 我们知道位于克鲁斯卡图底部四分之一的白洞其实很可能只有数学含义, 因为并没有什么物理过程可以产生白洞。同样的, 假设右侧四分之一的那一片宇宙是物理的话, 那克鲁斯卡图左侧四分之一的那一片宇宙也可能只有数学含义, 它可能完全源于我们的数学延拓, 而不描写任何真正的物理现实。如果真是这样的话, 那上一段话当然就没什么真正的物理含义。

然而, 的确有一种可能的物理方式实现在黑洞视界后面相互通信。

为了说明这一点, 我们不妨设想克鲁斯卡图的左右两侧描写的不是两片相互独立的宇宙, 而是近似描写同一个宇宙的两个相距遥远的区域。从而, 克鲁斯卡图里的黑洞就可以近似成为分别位于这两个相距遥远区域的两个黑洞, 它们在视界后面通过爱因斯坦-罗森桥相互连接。近似来说, 爱因斯坦场方程当然允许这样的双黑洞解存在, 因为近似描写它们的克鲁斯卡延拓的确是爱因斯坦场方程的解。

现在, 位于这两个相距遥远区域的观察者就只要跳进各自区域的黑洞就能实现视界后面的通信了! 因为他们的黑洞在视界后面有爱因斯坦-罗森桥相连接。

什么情况下两个相距遥远的黑洞可以通过爱因斯坦-罗森桥相连接呢? 近些年来对量子引力的研究暗示了一个答案, 那就是, 如果这两个黑洞之间有很强的量子纠缠, 那么它们的视界后面就可能存在爱因斯坦-罗森桥, 这就是小标题所谓的 ER=EPR。

ER=EPR 在多大程度上成立是远未清楚的一个问题。但如果它能推向极端, 那也许意味着任何量子纠缠之间都有时空虫洞相连接, 只不过通常这种虫洞是微观上的微型虫洞。当然, 这种推论很可能已经走得太远了。

9.2.3 Reissner-Nordstrom 黑洞的克鲁斯卡延拓

除了施瓦西黑洞以外, 类似的处理也可以应用于 Reissner-Nordstrom 黑洞 (RN 黑洞), 也就是所谓的带电的球对称黑洞。它的度规也有如下形式

$$ds^2 = -f(r)dt^2 + \frac{1}{f(r)}dr^2 + r^2d\Omega^2, \quad (9.72)$$

式中 $f(r)$ 为

$$f(r) = 1 - \frac{2GM}{r} + \frac{e^2}{r^2}, \quad (9.73)$$

其中 M 为黑洞的质量, e^2 由黑洞所携带的电荷和磁荷决定, 并且这里我们假定 e^2 不太大。这是引力场与电磁场耦合系统的一个解, 不过具体的求解过程我们推迟到后面的章节中来处理, 本章我们只是简单地给出这个解。

很显然 (9.73) 式给出的函数 $f(r)$ 有两个零点, 分别记为 r_+, r_- (设 $r_+ > r_-$), 它们由下式给出

$$r_{\pm} = GM \pm \sqrt{(GM)^2 - e^2}. \quad (9.74)$$

进而可以将函数 $f(r)$ 重写成

$$f(r) = \frac{1}{r^2}(r - r_+)(r - r_-). \quad (9.75)$$

很显然, RN 解也有一个基林向量场 $\xi = \partial_t$, 并且 $r = r_{\pm}$ 的超曲面均为它的基林视界, 因为 $(\xi, \xi)|_{r_{\pm}} = 0$! 同样容易看清的是, $r_- < r < r_+$ 的区域发生了时空互换, 从而使得这个区域为单向膜区, $r = r_+$ 为单向膜区开始的地方, 称作外视界, 它也就是 RN 黑洞的未来事件视界, 而 $r = r_-$ 为单向膜区结束的地方, 称作内视界, 它其实是 RN 解的柯西视界。关于柯西视界的定义, 我们稍后会有一个简单的讨论, 更详细的讨论请参阅 E. Witten 的综述文章 Light Rays, Singularities, and All That, 中文译本叫《光线、奇点, 以及其它》。

根据上面的讨论, 容易得到

$$\begin{aligned} 2\frac{1}{f(r)} &= \frac{2r^2}{(r - r_+)(r - r_-)} \\ &= \frac{2r_+^2}{(r_+ - r_-)(r - r_+)} + \frac{2r_-^2}{(r_- - r_+)(r - r_-) + 2} \\ &= \frac{1}{\kappa_+(r - r_+)} + \frac{1}{\kappa_-(r - r_-)} + 2, \end{aligned} \quad (9.76)$$

式中 κ_{\pm} 分别为内外视界上的表面引力, 由下式给出

$$\kappa_{\pm} = \frac{r_{\pm} - r_{\mp}}{2r_{\pm}^2}. \quad (9.77)$$

由于现在有两个极点 (而不是一个), 所以与 (9.49) 式比较, 在每个极点控制的邻域内都有一个 $g(r)$ 函数和相应的 $G(r)$ 函数, 不妨分别记作 $g_{\pm}(r)$ 和 $G_{\pm}(r)$ 。而且, 在每个极点所控制的邻域内我们都要引入一组 (V, U) 光锥坐标, 不妨记作 (V_{\pm}, U_{\pm}) 。

为了给出 V_{\pm}, U_{\pm} 的具体定义, 我们将上一段的讨论与 (9.49) 式比较, 即可知, 在 $r_h = r_+$ 的邻域内 (从而 $r > r_-$), $g_+(r) = \frac{1}{\kappa_-(r-r_-)} + 2$, 积分得出

$$G_+(r) = \frac{1}{\kappa_-} \ln \left(\frac{r-r_-}{r_-} \right) + 2r. \quad (9.78)$$

相应的, 有

$$e^{\kappa_+(v-u)} = \left| \left(\frac{r-r_+}{r_+} \right) \right| e^{\kappa_+ G_+(r)}. \quad (9.79)$$

类似的, 在 $r_h = r_-$ 的邻域内 (从而 $r < r_+$), 有 $g_-(r) = \frac{1}{\kappa_+(r-r_+)} + 2$, 积分得出

$$G_-(r) = \frac{1}{\kappa_+} \ln \left| \left(\frac{r-r_+}{r_+} \right) \right| + 2r = \frac{1}{\kappa_+} \ln \left(\frac{r_+-r}{r_+} \right) + 2r. \quad (9.80)$$

从而有

$$e^{\kappa_-(v-u)} = \left| \left(\frac{r-r_-}{r_-} \right) \right| e^{\kappa_- G_-(r)}. \quad (9.81)$$

注意, 这两个邻域有一个交叠区, 即 $r_- < r < r_+$ 的单向膜区。

对于 $r_h = r_+$ 的邻域, 接下来的一切讨论都和 (9.2.1) 小节中给出的一样。比如我们可以定义 V_+, U_+ 以使得

$$U_+ V_+ = - \left(\frac{r-r_+}{r_+} \right) e^{\kappa_+ G_+(r)}. \quad (9.82)$$

与 (9.79) 比较可知, (V_+, U_+) 可以分成四片, 对于 $r_- < r < r_+$, 取

$$V_+ = \pm e^{\kappa_+ v}, \quad U_+ = \pm e^{-\kappa_+ u}, \quad (9.83)$$

它覆盖 (V_+, U_+) 坐标系的一、三象限 (即上部和下部)。而对于 $r > r_+$, 则取

$$V_+ = \pm e^{\kappa_+ v}, \quad U_+ = \mp e^{-\kappa_+ u}, \quad (9.84)$$

它覆盖 (V_+, U_+) 坐标系的二、四象限 (即左侧和右侧)。

但是, 对于 $r_h = r_-$ 的邻域, 接下来的讨论会和 (9.2.1) 小节在细节上有点出入。这是因为上一段已经把 $r_- < r < r_+$ 的交叠区定义在 (V_+, U_+) 坐标系的一、三象限了, 而 (V_-, U_-) 坐标当然是要和 (V_+, U_+) 坐标拼接成一个整体的, 为了能拼接起来, 则 $r_- < r < r_+$ 的交叠区在 (V_-, U_-) 坐标系中也必须定义在一、三象限。所以, 当 $r_- < r < r_+$ 时, 我们需要定义

$$V_- = \pm e^{\kappa_- v}, \quad U_- = \pm e^{-\kappa_- u}. \quad (9.85)$$

进而当然, $0 < r < r_-$ 的区域就只能在 (V_-, U_-) 坐标系的二、四象限。所以, 当 $0 < r < r_-$ 时, 需要定义

$$V_- = \pm e^{\kappa_- v}, \quad U_- = \mp e^{-\kappa_- u}. \quad (9.86)$$

与 (9.81) 式比较, 即有

$$U_-V_- = \left(\frac{r-r_-}{r_-}\right)e^{\kappa_-G_-(r)}, \quad (9.87)$$

注意, 前面没有负号!

代入 $\kappa_+/\kappa_- = -r_-^2/r_+^2$, $\kappa_-/\kappa_+ = -r_+^2/r_-^2$, 以及 $G_{\pm}(r)$ 的表达式, 即将 (9.82) 式和 (9.87) 式重新表达成

$$U_+V_+ = -\left(\frac{r-r_+}{r_+}\right)\left(\frac{r_-}{r-r_-}\right)^{r_-^2/r_+^2}e^{2\kappa_+r}$$

$$U_-V_- = \left(\frac{r-r_-}{r_-}\right)\left(\frac{r_+}{r_+-r}\right)^{r_+^2/r_-^2}e^{2\kappa_-r}.$$

从这两个式子不难看出, 在单向膜区的边界, 当 $r \rightarrow r_-$ 时, $U_+V_+ \rightarrow \infty$, 所以 $r = r_-$ 在 (V_+, U_+) 坐标系一、三象限的无穷远处。类似的, 当 $r \rightarrow r_+$ 时, $U_-V_- \rightarrow \infty$, 所以 $r = r_+$ 也在 (V_-, U_-) 坐标系一、三象限的无穷远处。

另外, 从上一段的第二个式子, 我们还能得到, $r = 0$ 处的时空奇性满足

$$r = 0 \Leftrightarrow U_-V_- = -1, \quad (9.88)$$

当然, 这是位于 (V_-, U_-) 坐标系二、四象限 (即左右两侧) 的一条双曲线。这说明, 与施瓦西黑洞不同, RN 黑洞的时空奇性是类时的。

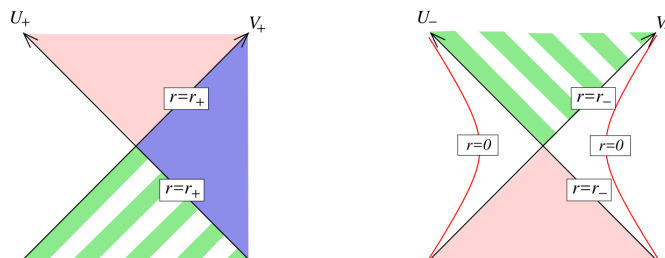


Figure 9.8: RN 黑洞的两种克鲁斯卡坐标卡。

RN 解的两种 (V, U) 坐标称作两种坐标卡, (V_+, U_+) 坐标卡如图 (9.8) 左图, 其中位于底部的第三象限是白洞区, 而位于上部的第一象限是黑洞区。而 (V_-, U_-) 坐标卡如图 (9.8) 右图, 其中位于底部第三象限的是黑洞区, 而位于上部第一象限的是白洞区。注意, (V_-, U_-) 坐标卡的第二、四象限 (即左右两侧) 是有限的, 时空奇性的 $U_-V_- = -1$ 是其边界, 所以和 (V_+, U_+) 坐标卡不同, (V_-, U_-) 坐标卡的第二、四象限不是两个渐近平坦的时空区域。

但是, 由于单向膜区是交叠区, 同时被两个坐标卡所覆盖, 所以这两个坐标卡其实是要拼接起来的, 比方说, 像图 (9.8) 所暗示的那样, 将右边的 (V_-, U_-) 坐标卡拼接在左边 (V_+, U_+) 坐标卡的下方。但是, (V_-, U_-) 坐标卡同样可以拼接在 (V_+, U_+) 坐标卡的上方。这导致一个结果, 即我们可以交错地拼接两种坐标卡, 进而将时空无穷地延拓下去。这是 RN 解与施瓦西解的又一个不同, 即它的克鲁斯卡延拓可以无穷无尽地永远进行下去。

9.2.4 Eddington-Finkelstein 坐标

实际上, 如果仅仅是想消除视界面位置的坐标奇异性, 那除了克鲁斯卡坐标以外, 还有一种更简单的坐标也能做到这一点, 那就是爱丁顿-芬克斯坦 (Eddington-Finkelstein) 坐标。下面以施瓦西时空为例来讲述这种坐标。

在通常的坐标系中, 施瓦西黑洞的时空度规为

$$ds^2 = -\left(1 - \frac{r_g}{r}\right) dt^2 + \left(1 - \frac{r_g}{r}\right)^{-1} dr^2 + r^2 d\Omega^2.$$

即 $f(r) = \left(1 - \frac{r_g}{r}\right)$, 从而

$$r_* = \int \frac{dr}{f(r)} = r + r_g \ln \left| \left(\frac{r - r_g}{r_g} \right) \right|. \quad (9.89)$$

这个坐标有时候也称作乌龟坐标。正如我们已经知道的, 有了 r_* 以后, 进而就能定义类光坐标 $u = t - r_*$, $v = t + r_*$, 但是正如前面提过的 (v, u) 坐标在视界面处依然是奇异的。

爱丁顿-芬克斯坦的想法是, 不同时引入 u, v , 而只引入一半, 即只引入 v 或者只引入 u , 只引入 v 就称作内行爱丁顿-芬克斯坦坐标, 只引入 u 就称作外行爱丁顿-芬克斯坦坐标。

以内行爱丁顿芬克斯坦坐标为例, 它即是将原来的 (t, r) 坐标变换成 (v, r) 坐标, 这时候

$$dt = dv - dr_* = dv - \left(1 - \frac{r_g}{r}\right)^{-1} dr, \quad (9.90)$$

消去施瓦西度规中的 dt^2 项, 即可以得到内行爱丁顿-芬克斯坦坐标下的度规, 为

$$ds^2 = -\left(1 - \frac{r_g}{r}\right) dv^2 + 2dvdr + r^2 d\Omega^2. \quad (9.91)$$

特别的, 在 $r = r_g$ 的视界面上, 这个度规变成

$$ds^2|_{r_g} = 2dvdr + r^2 d\Omega^2, \quad (9.92)$$

这是一个定义良好的度规, 没有任何病态, 只是相当于在视界面上 (v, r) 变成了类光坐标。所以, 内行爱丁顿-芬克斯坦坐标在 $r = r_g$ 处没有奇异性!

所以, (v, r) 坐标的好处是, r 可以从视界外一直延拓到视界内, 直到最终碰到 $r = 0$ 处的时空奇性。很显然, 原来的基林向量场 $\xi = \partial_t$ 在 (v, r) 坐标中就变成了

$$\xi = \partial_v. \quad (9.93)$$

不妨来分析一下 (v, r) 坐标中的径向类光测地线, 根据 (9.91), 这时候有

$$ds^2 = -\left(1 - \frac{r_g}{r}\right) dv^2 + 2dvdr = 0. \quad (9.94)$$

这个方程的一个解是

$$v = \text{constant}, \quad (9.95)$$

这描述的是内行类光测地线。另一个解由下式描写

$$\frac{dv}{dr} = 2\left(1 - \frac{r_g}{r}\right)^{-1} dr, \quad (9.96)$$

积分即得

$$v = 2r + 2r_g \ln\left|\left(\frac{r-r_g}{r_g}\right)\right| + \text{constant}. \quad (9.97)$$

这描述的是外行类光测地线，注意，由于绝对值的符号问题，它要分成视界外和视界内两部分来讨论。

那么在 $r = r_g$ 的视界面上，径向外行类光测地线是什么呢？很显然，由 (9.92) 式，这时候径向类光测地线方程变成了

$$ds^2|_{r_g} = 2dvdr = 0, \quad (9.98)$$

因此 $r = r_g$ 本身变成了径向外行类光测地线！这与前面所说的，视界面是一个类光超曲面是吻合的。

我们可以用芬克斯坦图来画出这些径向类光测地线。所谓芬克斯坦图就是以 (t_*, r) 为坐标的图 (9.9)，其中纵坐标

$$t_* = v - r. \quad (9.99)$$

在芬克斯坦图中，内行类光测地线由 $t_* + r = v = \text{constant}$ 描述，因此和竖直轴成 45 度夹

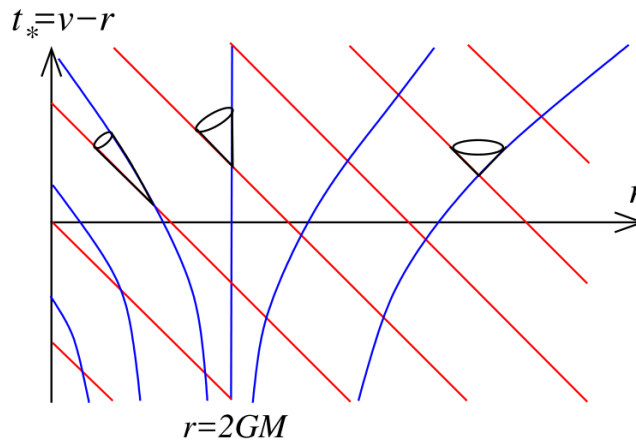


Figure 9.9: 内行坐标的芬克斯坦图。红线是内行类光测地线，蓝线是外行类光测地线。由于光锥的倾斜，在视界之内，外行测地线实际上并不指向外。

角，在图 (9.9) 中以红线表示。在图中，外行类光测地线以蓝线表示。很显然，外行类光测地线的行为在视界内外有很大不同。在视界外，外行测地线的确是向外行走的。

然而，在视界之内，光锥强烈地向着 $r = 0$ 的奇点方向倾斜了，这导致视界内的外行类光测地线实际上根本不向外行走，而是同样朝向 $r = 0$ 的奇点方向，并在有限时间 t_* 之

内撞上奇点。分界线就是 $r = r_g$ 的这条外行类光测地线，它完全沿着视界面，在芬克斯坦图中是竖直蓝线。

图 (9.9) 中内行类光测地线和外行类光测地线夹角处的小圆锥就是未来光锥，任何有质量粒子的世界线都只能从光锥之内穿过。现在，我们立即就能理解黑洞为什么叫黑洞了：因为从图 (9.9) 中很清楚地能看到，你只要进入黑洞，就再也出不来了。因为视界之内的未来光锥都是向着奇点倾斜的，你被俘获了！不过，从图 (9.9) 中也可以看出，在掉进黑洞的人本身看来，视界面并没有什么特殊的地方，这个人在穿过视界时不会感觉到任何异常，只是它向外发送的信号再也到不了无穷远处的观察者了。

当然，也可以引入外行爱丁顿-芬克斯坦坐标，即 (u, r) 坐标，这个坐标系中的度规是

$$ds^2 = -\left(1 - \frac{r_g}{r}\right)du^2 - 2dudr + r^2d\Omega^2. \quad (9.100)$$

除了 u, v 这两个记号的差别之外，更要注意的是，这个度规和前面内行坐标下度规的区别，即现在的 $-2dudr$ 项前面有个负号，而内行坐标下的相应项是 $2dvdr$ ，没有负号。

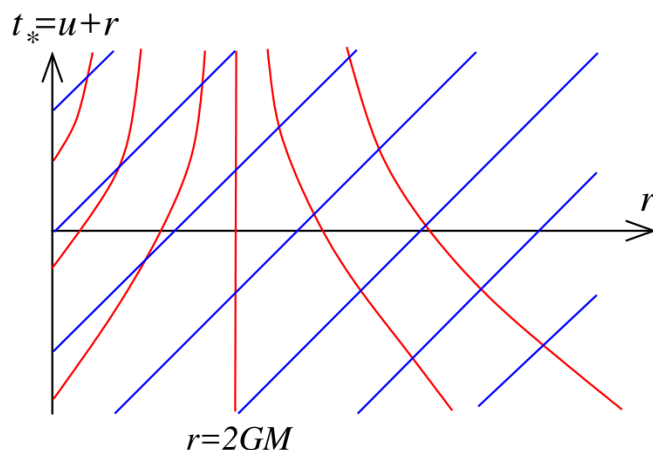


Figure 9.10: 外行坐标的芬克斯坦图。红线是内行类光测地线，蓝线是外行类光测地线。由于光锥向外倾斜，在视界之内，内行测地线实际上并不指向内。

外行坐标的芬克斯坦图如图 (9.10) 所示。很显然，这描写的是一个白洞，而不是黑洞，因为粒子只能从视界内出来，而不可能从视界外跑进视界之内。

为什么原来的施瓦西坐标 (t, r) 既可以描述黑洞也可以描述白洞，而爱丁顿-芬克斯坦坐标的 (v, r) 坐标就只能描写黑洞， (u, r) 坐标就只能描写白洞呢？原因在于，原来的施瓦西坐标 (t, r) 是时间反演对称的，即其度规在 $t \rightarrow -t$ 的变换之下是不变的，因此它描述白洞和黑洞都同样成立。而根据 v, u 的定义可知，在 $t \rightarrow -t$ 的时间反演之下，将有 $v \rightarrow -u$ 以及 $u \rightarrow -v$ 。很容易验证， (v, r) 坐标在时间反演之下，其度规并不保持不变，实际上度规 (9.91) 在时间反演之下就变成了度规 (9.100)，所以爱丁顿-芬克斯坦坐标破坏了时间反演对称性，因此只能分别描述黑洞和白洞，并且黑洞的时间反演就是白洞。

9.3 彭罗斯图

如何在保持一个时空因果结构的前提下将整个时空的“地图”画在一个有限的纸张区域上呢？显然，为此就要将时空的无穷远变成有限，结果表明在闵可夫斯基号差的流形上，无穷远比人们想象的要更为有趣一些。这是因为，人们可以沿着空间方向走向无穷，也可以沿着类时方向走向无穷，甚至还可以沿着类光方向走向无穷，而这三种无穷每一种都有不同的结构。如何直观地用“地图”显示这些无穷的结构呢？这就是本节要讨论的内容。

9.3.1 外尔变换

变无穷为有限的一个基本方法是通过所谓的外尔变换，其定义是：给定一个时空 M 和一个度规场 $g_{\mu\nu}(x)$ ，我们可以构建一个新的度规场 $\tilde{g}_{\mu\nu}(x)$ 如下

$$\tilde{g}_{\mu\nu}(x) = \Omega^2(x)g_{\mu\nu}(x), \quad (9.101)$$

式中 $\Omega(x)$ 是一个非零的光滑函数。上述变换就叫做时空度规的外尔变换，在广义相对论的文献中有时也称之为共形变换，它和共形场论中的共形变换有密切联系但并不完全是一回事。

很明显，外尔变换相当于把时空距离在每一个局部上都做了一个伸缩， $\Omega(x)$ 就是伸缩因子。不过，这种局部伸缩是保角的，具体来说即是，如果有两个切向量场 X^μ 和 Y^μ ，记 $\langle X, Y \rangle_g = g_{\mu\nu}X^\mu Y^\nu$ ，记 $\langle X, Y \rangle_{\tilde{g}} = \tilde{g}_{\mu\nu}X^\mu Y^\nu$ ，则

$$\frac{\langle X, Y \rangle_{\tilde{g}}}{\sqrt{\langle X, X \rangle_{\tilde{g}} \langle Y, Y \rangle_{\tilde{g}}}} = \frac{\langle X, Y \rangle_g}{\sqrt{\langle X, X \rangle_g \langle Y, Y \rangle_g}}, \quad (9.102)$$

这很容易直接验证。当然，这个结果默认了切向量场 X^μ 和 Y^μ 均不类光。对于类光向量场 Z^μ ，我们有

$$\langle Z, Z \rangle_g = 0 \Leftrightarrow \langle Z, Z \rangle_{\tilde{g}} = 0. \quad (9.103)$$

即是说，类光向量在外尔变换以后的度规下依然类光。类似的，一个类时向量在变换以后的度规下依然类时，而类空向量在变换以后的度规下也依然类空。所以，外尔变换保持时空的局部光锥结构。换言之，外尔变换保持时空的因果结构。

变无穷为有限的关键技巧就在于通过合适的伸缩因子 $\Omega(x)$ 将无穷收缩到有限。为了将无穷收缩到有限，伸缩因子 $\Omega(\mathbf{x}, t)$ 必须满足如下条件

$$\Omega(\mathbf{x}, t) \rightarrow 0, \quad \text{当 } |\mathbf{x}| \rightarrow \infty \text{ 和 (或) } |t| \rightarrow \infty. \quad (9.104)$$

因此，“无穷”可以等同于这样的时空点 $x = (\mathbf{x}, t)$ ，它使得 $\Omega(\mathbf{x}, t) = 0$ 。这些“无穷”点并不存在于原来的时空之中，但是可以额外添加到时空图上，作为时空的边界（这些边界本身不是时空的一部分），这样的操作称之为时空的共形紧化。

外尔变换和测地线

不仅如此，还可以证明，一条类光测地线在外尔变换后的新度规下依然是一条类光测地线。（由于外尔变换改变了时空距离，所以一条类时测地线在变换之后不一定依然是类

时测地线，当然，由于外尔变换保持时空的因果结构，类时曲线在变换之后将依然是类时曲线。)

为了证明上述结论，我们先来计算一下克里斯托夫联络在外尔变换之下的变换关系，

$$\begin{aligned}\Gamma_{\rho\sigma}^{\mu}[\tilde{g}] &= \frac{1}{2}\tilde{g}^{\mu\nu}(\partial_{\rho}\tilde{g}_{\nu\sigma} + \partial_{\sigma}\tilde{g}_{\nu\rho} - \partial_{\nu}\tilde{g}_{\rho\sigma}) \\ &= \frac{1}{2}\Omega^{-2}g^{\mu\nu}(\partial_{\rho}(\Omega^2g_{\nu\sigma}) + \partial_{\sigma}(\Omega^2g_{\nu\rho}) - \partial_{\nu}(\Omega^2g_{\rho\sigma})) \\ &= \Gamma_{\rho\sigma}^{\mu}[g] + \Omega^{-1}(\delta_{\sigma}^{\mu}D_{\rho}\Omega + \delta_{\rho}^{\mu}D_{\sigma}\Omega - g_{\rho\sigma}D^{\mu}\Omega)\end{aligned}\quad (9.105)$$

式中，在最后一行因为 Ω 是一个标量函数，所以我们可以将普通导数替换成协变导数。

现在，考察一条原度规之下的类光测地线，它满足如下两个方程

$$\begin{aligned}\frac{d^2x^{\mu}}{d\lambda^2} + \Gamma_{\rho\sigma}^{\mu}[g]\frac{dx^{\rho}}{d\lambda}\frac{dx^{\sigma}}{d\lambda} &= 0, \\ g_{\rho\sigma}\frac{dx^{\rho}}{d\lambda}\frac{dx^{\sigma}}{d\lambda} &= 0.\end{aligned}\quad (9.106)$$

利用第一个方程，我们有，在外尔变换之后的度规下

$$\frac{d^2x^{\mu}}{d\lambda^2} + \Gamma_{\rho\sigma}^{\mu}[\tilde{g}]\frac{dx^{\rho}}{d\lambda}\frac{dx^{\sigma}}{d\lambda} = \Omega^{-1}(\delta_{\sigma}^{\mu}D_{\rho}\Omega + \delta_{\rho}^{\mu}D_{\sigma}\Omega - g_{\rho\sigma}D^{\mu}\Omega)\frac{dx^{\rho}}{d\lambda}\frac{dx^{\sigma}}{d\lambda}\quad (9.107)$$

但是，对于类光测地线，由于 $g_{\rho\sigma}\frac{dx^{\rho}}{d\lambda}\frac{dx^{\sigma}}{d\lambda} = 0$ ，所以上式等号右边的最后一项自动为零。进而可以将上式重写成

$$\frac{d^2x^{\mu}}{d\lambda^2} + \Gamma_{\rho\sigma}^{\mu}[\tilde{g}]\frac{dx^{\rho}}{d\lambda}\frac{dx^{\sigma}}{d\lambda} = 2\frac{dx^{\mu}}{d\lambda}\frac{1}{\Omega}\frac{d\Omega}{d\lambda}.\quad (9.108)$$

但最后的这个方程其实同样是测地线方程，只不过参数 λ 不再是仿射参数，可以通过选择一个合适的参数变换 $\lambda(\tilde{\lambda})$ 来抵消掉最后这个方程右边的项，从而得到一个以合适的 $\tilde{\lambda}$ 为仿射参数的标准的测地线方程。

因此，外尔变换的确将一条类光测地线变成一条类光测地线！

外尔曲率张量

外尔变换会改变黎曼曲率张量和里奇曲率张量，但是可以证明（计算比较繁琐，这里从略），可以组合出一个在外尔变换之下保持不变的曲率张量，称作外尔曲率张量，记作 $C_{\rho\sigma\mu\nu}$ ，其定义为

$$C_{\rho\sigma\mu\nu} = R_{\rho\sigma\mu\nu} - \frac{2}{D-2}(g_{\rho[\mu}R_{\nu]\sigma} - g_{\sigma[\mu}R_{\nu]\rho}) + \frac{2}{(D-1)(D-2)}Rg_{\rho[\mu}g_{\nu]\sigma}.$$

式中 D 表示时空的维数。由定义，外尔曲率张量其实就是从黎曼曲率张量中去除所有可能的非零收缩，所以，外尔曲率张量所有可能的收缩都为零！另外，外尔曲率张量有和黎曼曲率张量同样的指标对称性！所谓的外尔曲率张量在外尔变换下保持不变，指的是

$$C^{\rho}{}_{\sigma\mu\nu}[\tilde{g}] = C^{\rho}{}_{\sigma\mu\nu}[g].\quad (9.109)$$

如果一个时空的外尔曲率张量等于零，则我们就称之为共形平坦的。

9.3.2 闵可夫斯基时空的彭罗斯图

下面来考察 3+1 维闵可夫斯基时空的共形紧化，并讲述其彭罗斯图。闵可夫斯基时空的度规是

$$ds^2 = -dt^2 + dr^2 + r^2 d\Omega^2. \quad (9.110)$$

下面引入类光坐标 u, v

$$u = t - r, \quad v = t + r. \quad (9.111)$$

当然由于 $r \geq 0$, 所以还要满足

$$v \geq u. \quad (9.112)$$

进而将时空度规写成

$$ds^2 = -dudv + \frac{(u-v)^2}{4} d\Omega^2. \quad (9.113)$$

下面来了关键的一步。注意到以上类光坐标的取值范围都是 $(-\infty, +\infty)$, 所以我们可以利用 \tan 函数, 将坐标的取值范围变成有限, 具体来说即是引入新的类光坐标 \tilde{U}, \tilde{V}

$$u = \tan(\tilde{U}), \quad v = \tan(\tilde{V}) \quad (9.114)$$

显然 $\tilde{U}, \tilde{V} \in (-\frac{\pi}{2}, \frac{\pi}{2})$, 并且满足

$$\tilde{V} \geq \tilde{U}. \quad (9.115)$$

在这个新的坐标中, 度规为

$$ds^2 = \frac{1}{(2 \cos \tilde{U} \cos \tilde{V})^2} [-4d\tilde{U}d\tilde{V} + \sin^2(\tilde{U} - \tilde{V})d\Omega^2]. \quad (9.116)$$

注意, 当我们趋于闵可夫斯基时空的边界时, 即 $\tilde{U}, \tilde{V} \rightarrow \pm\frac{\pi}{2}$ 时, 这个度规是发散的, 反映了时空距离在“无穷”处是发散的。

为了进行共形紧化, 我们选择伸缩因子为

$$\Omega(x) = 2 \cos \tilde{U} \cos \tilde{V}, \quad (9.117)$$

则可以得到外尔变换以后新的度规

$$d\tilde{s}^2 = \Omega^2(x)ds^2 = -4d\tilde{U}d\tilde{V} + \sin^2(\tilde{U} - \tilde{V})d\Omega^2. \quad (9.118)$$

这个新的度规和原来的度规有相同的因果结构, 但是现在, 一切时空距离都是有限的。

彭罗斯图就是这个外尔变换之后新的时空的“地图”(它和原来的时空有相同的因果结构), 图中我们只标示 \tilde{U}, \tilde{V} 坐标, 不标示角度部分坐标, 因此彭罗斯图中的每一个点都代表一个半径为 $\sin(\tilde{U} - \tilde{V})$ 的两维球面。

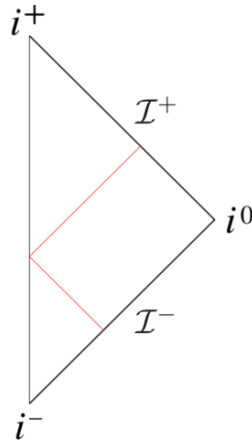


Figure 9.11: 闵可夫斯基时空的彭罗斯图。

为了画出彭罗斯图，我们取竖直方向为时间方向，水平方向为空间方向，当然，光线都是沿着与竖直方向成 45 度夹角的方向行进。这意味着 \tilde{U}, \tilde{V} 的光锥坐标轴都是与竖直方向成 45 度夹角的。如果定义 $\tilde{V} = T + R, \tilde{U} = T - R$, T 是竖直轴, R 是水平轴, 则 $\tilde{V} \geq \tilde{U}$ 意味着整个彭罗斯图只可能占据右侧 $R \geq 0$ 的那半边。

如图 (9.11) 所示, 现在我们将 $\Omega(x) = 0$ 的共形边界画到彭罗斯图上, 也就是将闵可夫斯基时空的“无穷”在彭罗斯图上画出来。闵可夫斯基时空有几种不同的无穷:

首先, **类空无穷, 通常记作 i^0** 。它即是取 t 有限, 取 $r \rightarrow \infty$ 。很显然, 这相当于 u, v 坐标的 $u \rightarrow -\infty, v \rightarrow \infty$, 换成 \tilde{U}, \tilde{V} 坐标即是

$$\tilde{U} = -\frac{\pi}{2}, \quad \tilde{V} = \frac{\pi}{2}, \quad (9.119)$$

显然, 这在彭罗斯图上是水平轴上的一个点。

其次, **过去类时无穷 i^- , 以及未来类时无穷 i^+** 。它们即是取 r 有限, 取 $t \rightarrow \pm\infty$ 。用 u, v 坐标即是 $u \rightarrow \pm\infty, v \rightarrow \pm\infty$ 。换成 \tilde{U}, \tilde{V} 坐标即是

$$\tilde{U} = \pm\frac{\pi}{2}, \quad \tilde{V} = \pm\frac{\pi}{2}, \quad (9.120)$$

显然, 这在彭罗斯图上是竖直轴上方的一个点 (i^+) 和下方的一个点 (i^-)。

第三, **过去类光无穷, 记作 \mathcal{I}^-** 。它即是取 $t \rightarrow -\infty, r \rightarrow \infty$, 并让 $t+r$ 有限。在 u, v 坐标中即是, v 有限, $u \rightarrow -\infty$ 。换成 \tilde{U}, \tilde{V} 坐标即是

$$\tilde{U} = -\frac{\pi}{2}, \quad |\tilde{V}| \neq \frac{\pi}{2}. \quad (9.121)$$

这是彭罗斯图下方的一条斜边。

第四, **未来类光无穷, 记作 \mathcal{I}^+** 。它即是取 $t \rightarrow \infty, r \rightarrow \infty$, 并让 $t-r$ 有限。在 u, v 坐标中即是, u 有限, $v \rightarrow \infty$ 。换成 \tilde{U}, \tilde{V} 坐标即是

$$|\tilde{U}| \neq \frac{\pi}{2}, \quad \tilde{V} = \frac{\pi}{2}. \quad (9.122)$$

这是彭罗斯图上方的一条斜边。

在彭罗斯图中，所有的类时测地线，如果一直向两端延伸下去，都会起于 i^- ，终止于 i^+ 。所有的类光测地线，如果一直延伸下去，都会起于 \mathcal{I}^- ，终止于 \mathcal{I}^+ 。注意彭罗斯图左边那条竖直的直线，它并不是时空的边界，而仅仅是 $r=0$ 的坐标原点，为了说明这一点，我们在图 (9.11) 中用红线画出了一条类光测地线，它开始于 \mathcal{I}^- ，然后到达 $r=0$ 的竖直线，之后继续前进，返回 $r>0$ 处，最终到达 \mathcal{I}^+ 。

彭罗斯图是将四维时空用一张二维图来表示，因此会忽略两个维度，人们通常让这未标示的两个维度为球坐标的角度部分，从而使得彭罗斯图上的每一个点代表一个二维球面。这意味着彭罗斯图通常是用来处理球对称时空，或者说有 $SO(3)$ 对称性的时空。这是一些最简单的时空，通常也是最重要的。

正如我们已经看到的，闵可夫斯基时空有两个位于未来类时无穷和过去类时无穷的边界点，尤其是同时还有两个类光边界。在量子场论中，这导致人们自然地问如下关于零质量粒子场（比如引力场）的散射问题：即在 \mathcal{I}^- 上给定渐近态，让它演化，然后在 \mathcal{I}^+ 上读出最终的渐近态。这类问题密切地牵涉到散射 S 矩阵的定义。

以上论点也可以推广到渐近平坦时空，所谓渐近平坦是指在类空和类光方向分别渐近趋于 i^0 和 \mathcal{I}^\pm 的时空。 i^0 是一个点因此无法利用，然而由于渐近平坦时空有类光边界 \mathcal{I}^\pm ，所以同样能考察零质量粒子场（比如引力场）的散射问题，同样能定义散射 S 矩阵。尤其是考虑到渐近平坦时空的量子引力，这时候微分同胚不变性将使得理论不存在局域的可观测量，任何可观测量都要用两个渐近边界 \mathcal{I}^\pm 上的渐近态来定义，这就使得渐近平坦时空量子引力的可观测量必定为散射 S 矩阵。

但是正如后面的章节中会讨论的，如果宇宙学常数为负，那时空将不是渐近平坦的而是渐近 AdS (Anti de-Sitter) 的，这种时空在类时的方向上没有共形边界，而在类光和类空方向上的边界结合在一起了，从而有唯一一个共形的渐近边界 \mathcal{S} (与 i^0 是一个点不同，也与 \mathcal{I}^\pm 分成两片不同)。 \mathcal{S} 本身是类时的 (它的法向量类空)，这时候人们就可以在这个渐近边界 \mathcal{S} 上定义共形场论，这个共形场论的关联函数 (而不是散射 S 矩阵) 才是原来渐近 AdS 时空的可观测量，这也就是所谓的 AdS/CFT 对应，CFT 就是共形场论 (conformal field theory)。

但是如果宇宙学常数为正，那时空将渐近 dS (de-Sitter)，这种时空在类空的方向上没有渐近边界，而类光和类时方向上的边界结合在一起了，在这种情况下，这些共形边界本身是类空的 (法向量类时)，与 AdS/CFT 不同，这时候如何定义原来渐近 dS 时空的可观测量是一件尚没有研究清楚的问题，甚至还能不能定义可观测量都不太清楚。

关于 AdS 时空和 dS 时空，以及关于它们的彭罗斯图，我们将在后面的章节中再具体讨论。

9.3.3 施瓦西时空的彭罗斯图

克鲁斯卡时空的彭罗斯图

正如前文解释的，最好的展示一个时空因果结构的方式就是画出其彭罗斯图。下面我们来讨论施瓦西时空的彭罗斯图。当然，施瓦西时空的完整结构要用克鲁斯卡延拓来展示，

所以我们将要讨论的其实是克鲁斯卡延拓以后的施瓦西时空的彭罗斯图。

首先，我们回忆一下克鲁斯卡坐标 V, U ，它们的取值范围都是 $(-\infty, +\infty)$ 。和前面讨论闵可夫斯基时空时所做的一样，我们可以利用 \tan 函数来将坐标的无穷转换成有限，具体来说即是引入新的坐标，

$$U = \tan \tilde{U}, \quad V = \tan \tilde{V}. \quad (9.123)$$

当然 $\tilde{U}, \tilde{V} \in (-\frac{\pi}{2}, \frac{\pi}{2})$ ，注意现在**并没有** $\tilde{V} \geq \tilde{U}$ 的限制，所以最终的彭罗斯图将同时占据左右两侧，左右对称，而不是只有右侧的部分！有了这个坐标以后，进而就能将克鲁斯卡度规 (9.67) 重写成

$$ds^2 = \frac{1}{(\cos \tilde{U} \cos \tilde{V})^2} \left[-\frac{r_g}{\kappa^2} \frac{1}{r} e^{-2\kappa r} d\tilde{V} d\tilde{U} + r^2 (\cos \tilde{U} \cos \tilde{V})^2 d\Omega^2 \right]. \quad (9.124)$$

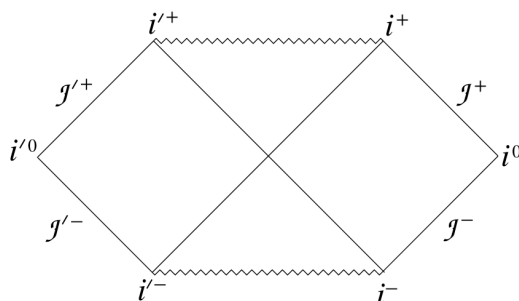


Figure 9.12: 施瓦西时空的彭罗斯图。上部代表黑洞区域，下部代表白洞区域，上下部两条水平的波浪线代表时空奇性。右侧是黑洞外部的一个渐近平坦区域，左侧是另一个渐近平坦区域。和克鲁斯卡图中一样，过原点的两条 45 度斜线代表视界。

取共形因子 $\Omega(x) = \cos \tilde{U} \cos \tilde{V}$ ，进而就能将上述度规外尔变换成如下形式

$$d\bar{s}^2 = \Omega^2(x) ds^2 = -\frac{r_g}{\kappa^2} \frac{1}{r} e^{-2\kappa r} d\tilde{V} d\tilde{U} + r^2 (\cos \tilde{U} \cos \tilde{V})^2 d\Omega^2. \quad (9.125)$$

完全类似于闵可夫斯基时空情形，我们也可以讨论这个时空在无穷远处满足 $\Omega(x) = 0$ 的共形边界。但是，除了共形边界以外，克鲁斯卡时空还有两个边界，那就是满足 $UV = 1$ 的时空齐性，很显然

$$UV = 1 \Leftrightarrow \tan \tilde{U} \tan \tilde{V} = 1 \Leftrightarrow \tilde{U} + \tilde{V} = \pm \frac{\pi}{2}. \quad (9.126)$$

这在以 \tilde{U}, \tilde{V} 为坐标的彭罗斯图上是上和下的两条水平线，通常画成波浪线以代表时空奇性。整个彭罗斯图如图 (9.12) 所示。

彭罗斯图 (9.12) 有两个渐近平坦区域，因此它有类光边界 \mathcal{I}^\pm 和类光边界 \mathcal{I}'^\pm 。从彭罗斯图中，我们可以给出黑洞的更好定义，所以黑洞区域，就是指无法发送信息到未来类光无穷 \mathcal{I}^+ 和 \mathcal{I}'^+ 的区域，也就是彭罗斯图的上部。黑洞区域的边界就称作未来事件视界，记为 \mathcal{H}^+ 。

类似的，可以定义白洞区域为无法接收到来自过去类光无穷 \mathcal{I}^- 和 \mathcal{I}'^- 的信息的区域，也就是彭罗斯图的下部。白洞区域的边界就是过去事件视界，记作 \mathcal{H}^- 。

重要的是，为了定义黑洞，我们需要知道整个时空的因果结构，尤其是需要利用到未来类光无穷。并没有直接利用时空在某给定时刻的空间截片 Σ 来定义黑洞区域的方式，至少任何这样的定义都不是一个良定义。这件事意味着，其实一个观察者无从真正地知道它是否在黑洞区域内，除非他知道时空的整个未来演化。

黑洞的形成：弱宇宙监督

卡鲁斯卡延拓以后的施瓦西解包含一些非物理的东西，比如白洞区域就是非物理的，真实的施瓦西时空不是先有一个白洞区域，然后接着一个黑洞区域。真实的黑洞通常是由非常大质量的恒星发生引力坍缩所形成的。所以真实黑洞的因果结构并不由彭罗斯图 (9.12) 所刻画。

为了得到一个真实黑洞的彭罗斯图，原则上就必须得研究恒星的引力坍缩，这当然比求解施瓦西解要复杂一些，所以我们放到后面的章节中再进行。实际上，为了推导出真实黑洞的彭罗斯图，我们可以考虑如下一个假想的引力坍缩过程：我们假想一个均匀填充满光子的薄层球壳，球壳里面是真空，假设所有的光子都正在以光速朝着 $r=0$ 的球心飞行，因此这个球壳正在以光速向内坍缩！很显然，在所有的光子都汇聚到 $r=0$ 点时（甚至在此之前），由于光子能量密度的高度集中，时空会形成一个黑洞。所以这就是一个物质发生引力坍缩进而形成黑洞的假想例子。

但这个假想例子很容易精确描述，首先，根据伯克霍夫定理的推论，虽然球壳本身在以光速运动，但球壳里面所包围的空腔一定保持为平坦的真空，所以空腔里面的时空由图 (9.13) 中最左边那幅平坦时空彭罗斯图的阴影部分描写（请读者想想为什么空腔里面对应阴影部分？）。其次，根据伯克霍夫定理，球壳外面的时空必定由施瓦西解所描述，所以球壳外部的时空必定由图 (9.13) 中间那幅施瓦西时空彭罗斯图的阴影部分描写。最后，将这两个阴影部分分别裁剪下来，并在以光速运动的球壳位置拼接起来，就得到了最终的彭罗斯图，也就是图 (9.13) 中最右边那幅图。

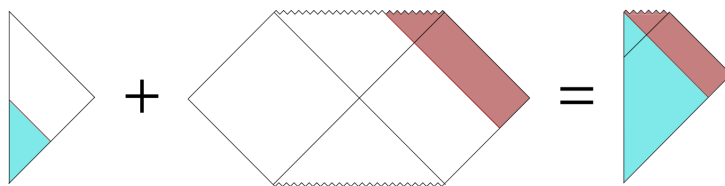


Figure 9.13: 裁剪并拼接两个彭罗斯图的阴影部分，以得到真实黑洞的彭罗斯图。

所以，球对称引力坍缩形成的黑洞必定具有图 (9.13) 中最右边那幅彭罗斯图，图中与竖直方向成 45° 角的那条斜线段同样代表黑洞视界。但是现在这个视界是从左边的 $r=0$ 处开始的，所以视界面积并不是恒定不变的 $4\pi r_g^2$ ，而是从零开始逐渐增加到这个稳定值。

当然，以上是一个假想的例子，对于真实的球对称引力坍缩，坍缩星体的表面必定走的是类时世界线，所以真实的球对称引力坍缩的彭罗斯图大体如图 (9.14) 所示。注意，从

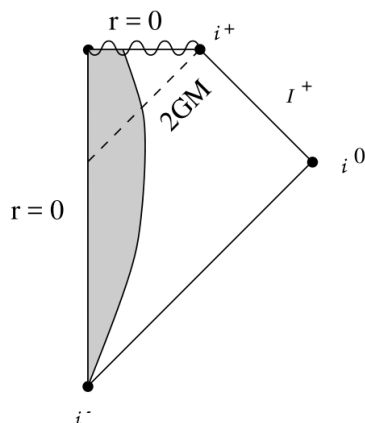


Figure 9.14: 球对称引力坍缩的彭罗斯图。

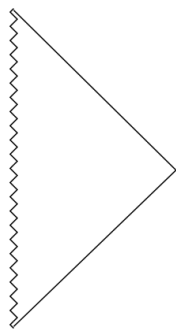
这个彭罗斯图可以看到，真实的黑洞并没有一个白洞区域。

从真实黑洞的彭罗斯图中可以看到，黑洞有一个重要的性质，即它的时空奇性必定被一个未来事件视界所包围。这意味着，黑洞的时空奇性对无穷远处的观察者完全不能产生影响。如果一个时空奇性不被未来事件视界包围，从而能够对无穷远处的观察者产生影响，比如克鲁斯卡时空中白洞里面的时空奇性，那这种时空奇性就称作裸奇性。

爱因斯坦场方程允许存在裸奇性的解，比如克鲁斯卡时空中的白洞，再比如宇宙大爆炸开始时的时空奇性。其实，如果施瓦西解中的质量 $M < 0$ ，那就会有裸奇性。这时候，施瓦西度规可以写成

$$ds^2 = -\left(1 + \frac{2G|M|}{r}\right)dt^2 + \left(1 + \frac{2G|M|}{r}\right)^{-1}dr^2 + r^2d\Omega^2. \quad (9.127)$$

很明显这时候根本不存在视界！但是， $r=0$ 处的时空奇性依然存在。可以按照讨论闵可夫斯基时空类似的方式讨论这个时空的彭罗斯图，即先引入类光坐标 v, u ，再引入 \tilde{V}, \tilde{U} ，最终得到的彭罗斯图如图 (9.15) 所示。很显然， $r=0$ 处的时空奇性 (图中的波浪线) 并没有

Figure 9.15: $M < 0$ 的施瓦西解的彭罗斯图。

被一个未来事件视界包围，从而在 \mathcal{I}^+ 上可以看到这个奇性，所以称之为裸奇性。

但 $M < 0$ 显然是非物理的！实际上，人们常常相信广义相对论中的裸奇性都是非物理的（宇宙大爆炸开始时的时空奇性是一个例外）。实际上，有一个所谓的弱宇宙监督猜测，说的就是这件事情：

弱宇宙监督猜测：假设物质满足所谓的主能量条件，并假定度规场和物质场都具有一般性的、光滑的初始条件，那么渐近平坦时空中将不会演化形成裸奇性。

这里所谓的主能量条件是对物质场能量动量张量的一个物理限制，要求物质场能动张量满足某种正定性，具体内容我们也许会在后面的章节中介绍。而所谓的初始条件，就是定义在时空的一个空间截片 Σ 上的初始数据，利用场方程我们可以让这些初始数据向着未来不断演化。

假如弱宇宙监督是正确的，那如图 (9.16) 所示的那种引力坍缩就是不可能的！实际

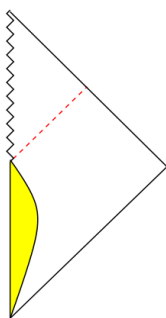


Figure 9.16: 自动形成裸奇性的引力坍缩。

上，这幅图本身也有误导人的地方，因为一旦裸奇性形成，你就不知道从奇性中将冒出什么东西，因此这个时候实际上不可能决定场位形在图中虚线所示的光锥内部的演化。这意味着，我们的动力学演化过程必须终止于虚线位置，不能再进一步朝着未来演化了。

弱宇宙监督猜测还没有得到证明，而只有一些数值模拟的支持。并且弱宇宙监督意味着我们没有机会直接观测到时空奇性处发生了什么，因为它们都被未来事件视界包围起来了。而时空奇性的地方正是我们预期量子引力会起重要作用的地方，因此弱宇宙监督也意味着我们很可能失去了一个直接观测量子引力效应的机会。

宇宙大爆炸开始时的时空奇性的确是裸奇性，但是由于它处于时间的开端处，而不是在时空动力学演化的未来自动形成的，所以它并不破坏弱宇宙监督猜测。很可能，我们最终对量子引力的证据要从对宇宙开端的观测中寻找。

9.3.4 RN 解的彭罗斯图

对于带电的 RN 解，我们有两种克鲁斯卡坐标卡 (V_+, U_+) 和 (V_-, U_-) ，应此可以分别引入 $(\tilde{V}_+, \tilde{U}_+)$ 和 $(\tilde{V}_-, \tilde{U}_-)$

$$V_{\pm} = \tan \tilde{V}_{\pm}, \quad U_{\pm} = \tan \tilde{U}_{\pm}. \quad (9.128)$$

$(\tilde{V}_+, \tilde{U}_+)$ 坐标卡的彭罗斯图如图 (9.17) 左图， $(\tilde{V}_-, \tilde{U}_-)$ 坐标卡如右图。图中字母 W 用于标示白洞区域，字母 B 用于标示黑洞区域。注意，图中的时空奇性是类时的。将这两种坐

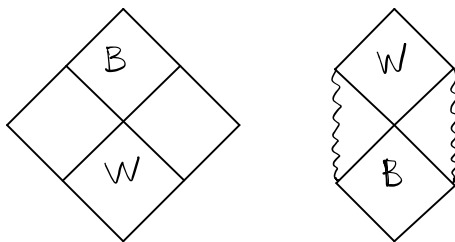


Figure 9.17: RN 解两种坐标卡的彭罗斯图。

标卡的彭罗斯图沿着重叠区 (也就是单向膜区) 交替地拼接起来, 就得到 RN 解完整的彭罗斯图, 如图 (9.18) 所示。

RN 解的彭罗斯图揭示了内视界 $r=r_-$ 的含义。为此, 考虑定义在空间截片 Σ 上的初始数据, 如图 (9.19) 所示。 Σ 称作一个柯西面。然后, 我们利用场方程将柯西面上的初始数据朝前演化。

然而, 一旦我们遇到图中的类时时空奇性, 这种演化就不再可能了, 因为还需要额外知道场在时空奇性处的信息。因此, 从图 (9.19) 中容易看到, Σ 上的初始数据不能演化到超出内视界 $r=r_-$ 之外, 进而 $r=r_-$ 处的类光超曲面就称作柯西视界。柯西视界之外的时空不能由柯西面上的初始数据完全决定。

柯西视界通常被认为是不稳定的。为了理解这一点, 让我们考察两个观察者 A 和 B, 如图 (9.19) 所示, 观察者 A 一直待在黑洞外面, 并以恒定的频率向黑洞里面发射信号, 观察者 B 掉进了黑洞里面, 在那里他接收 A 发出的信号。但是, 正如图中所示, 当 B 靠近柯西视界时, 信号将变得越来越密集。也即是说, 信号在柯西视界处无限蓝移了。这意味着, 黑洞外面渐近区域的一个小扰动就会在柯西视界上造成一个发散的扰动, 从而外面的任何一个小微扰都会使得柯西视界变得不稳定。

柯西视界的这种不稳定, 意味着上面 RN 解彭罗斯图的很多部分, 包括类时的时空奇性, 都是非物理的! 一种可能性是, 扰动会使得 $r=r_-$ 处的柯西视界变成时空奇性。实际上, 柯西视界的不稳定性正是如下强宇宙监督猜测所预言的:

强宇宙监督猜测: 对于满足适当能量条件的物质, 一般性的初始条件将不会导致出现柯西视界, 与之相关的, 类时的时空奇性也不会形成。也就是说, 整个时空都可以由柯西面上的初始数据演化决定! 这样的时空也称作整体双曲时空。所以强宇宙监督猜测相当于说, 物理的时空都是整体双曲的。

注意, 强宇宙监督猜测既不弱于但也不强于弱宇宙监督猜测 (不要被名字误导), 两个猜测虽然相关, 但逻辑上是相互独立的。

本章参考文献

1. 本章对克鲁斯卡延拓的处理是受知乎 @ 非对称虫洞的专栏文章的启发, 这篇文章叫《从 \tan 函数到彭罗斯图(Penrose diagram)》, 链接是 <https://zhuanlan.zhihu.com/p/188750243>

2. 本章部分内容参考了 David Tong 的在线讲义, General Relativity. 尤其是本章不少插图是来自于 David Tong 的这本在线讲义。

$$GM^2 > p^2 + q^2$$

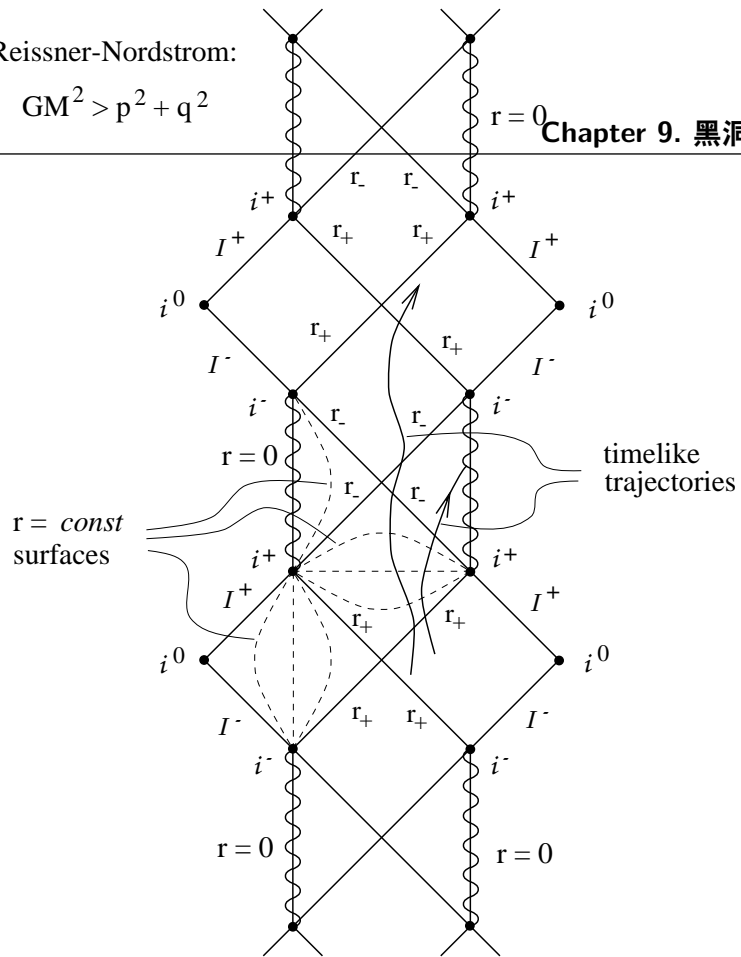


Figure 9.18: RN 解的彭罗斯图。

3. 关于虫洞打开和关闭的过程讨论来自于 Sean Carroll, Spacetime and Geometry.
4. 另外, 也推荐读者阅读 L. Susskind 理论最小值系列新出的 General Relativity.

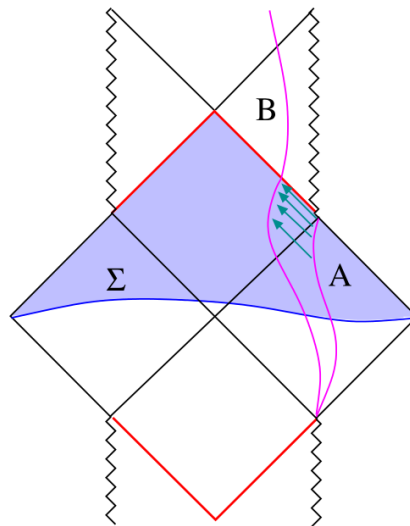


Figure 9.19: 初始数据定义在空间截片 Σ 上，但是这些数据不能演化超过 $r = r_-$ 位置的柯西视界 (图中用红线标示)。当然完整的 RN 解即有未来柯西视界也有过去柯西视界。

10. 带电黑洞与极端黑洞等等

10.1 RN 带电黑洞

10.1.1 Einstein-Maxwell 方程

本章主要是处理上一章遗留的一些问题。尤其是，我们要具体解出带电的球对称引力场解，也就是 RN 解。为此需要将电磁场耦合到引力场上，也就是研究弯曲时空中的电磁场方程。首先，由于克里斯托夫联络是一个对称联络，因此不难验证弯曲时空中的电磁场张量 $F_{\mu\nu}$ 可以写成

$$F_{\mu\nu} = D_\mu A_\nu - D_\nu A_\mu = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (10.1)$$

由最右边的表达形式不难验证如下恒等式

$$\partial_\rho F_{\mu\nu} + \partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} = 0. \quad (10.2)$$

在弯曲时空中，这个式子表面看似不协变，其实可以利用克里斯托夫联络是对称联络的性质改写成如下明显协变的形式

$$D_\rho F_{\mu\nu} + D_\mu F_{\nu\rho} + D_\nu F_{\rho\mu} = 0. \quad (10.3)$$

但后文我们用得更多的是前面那个看似不协变的表达。

进一步，弯曲时空中的电磁场作用量 S_{em} 可以写成

$$S_{\text{em}} = -\frac{1}{4} \int d^4x \sqrt{|g|} F_{\mu\nu} F^{\mu\nu}. \quad (10.4)$$

根据最小作用量原理对 A_μ 进行变分, 即可以得到弯曲时空中的电磁场方程

$$D_\mu F^{\mu\nu} = 0. \quad (10.5)$$

注意, 和 (10.2) 式不同, 这个方程必须用协变导数。

注意到 $D_\mu F^{\mu\nu} = \partial_\mu F^{\mu\nu} + \Gamma_{\mu\rho}^\mu F^{\rho\nu} + \Gamma_{\mu\rho}^\nu F^{\mu\rho}$, 由于 $\Gamma_{\mu\rho}^\nu$ 关于 μ, ρ 对称, 而 $F^{\mu\rho}$ 关于 μ, ρ 反对称, 所以等号右边最后一项自动为零。再根据第四章中得出的结果 $\Gamma_{\mu\nu}^\mu = \frac{1}{\sqrt{|g|}} \partial_\nu \sqrt{|g|}$, 进而不难有 $D_\mu F^{\mu\nu} = \partial_\mu F^{\mu\nu} + \Gamma_{\mu\rho}^\mu F^{\rho\nu} = \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} F^{\mu\nu})$ 。从而弯曲时空中的电磁场方程可以写作

$$\partial_\mu (\sqrt{|g|} F^{\mu\nu}) = 0. \quad (10.6)$$

方程 (10.2) 和方程 (10.6) 就是弯曲时空中的 Maxwell 方程。

为了代爱因斯坦场方程, 我们需要求出弯曲时空中的电磁场能量动量张量。首先回忆起物质场能动量张量的定义 $\delta S_{\text{em}} = \frac{1}{2} \int d^4x \sqrt{|g|} T^{\mu\nu} \delta g_{\mu\nu}$ 。为了将电磁场作用量对度规变分, 我们利用 $\delta(F_{\mu\nu} F^{\mu\nu}) = \delta(g^{\mu\rho} g^{\nu\sigma} F_{\mu\nu} F_{\rho\sigma}) = (\delta g^{\mu\rho}) F_{\mu\nu} F_{\rho}{}^\nu + (\delta g^{\nu\sigma}) F_{\mu\nu} F^{\mu\sigma}$, 再利用 $\delta g^{\mu\nu} = -g^{\mu\rho} g^{\nu\sigma} \delta g_{\rho\sigma}$, 不难进一步得到 $\delta(F_{\mu\nu} F^{\mu\nu}) = -2\delta g_{\mu\nu} F^{\mu\rho} F^{\nu\rho}$ (注意电磁场张量 F 两指标的反对称性)。进而不难算出 S_{em} 对度规场 $g_{\mu\nu}$ 的变分, 从而得到

$$T^{\mu\nu} = F^{\mu\rho} F^{\nu\rho} - \frac{1}{4} g^{\mu\nu} F_{\rho\sigma} F^{\rho\sigma}. \quad (10.7)$$

这和第二章中给出的平坦时空中的结果吻合。

第八章已经看到, 爱因斯坦场方程可以写成

$$R_{\mu\nu} = 8\pi G(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T). \quad (10.8)$$

注意到, 对于电磁场能动张量, 根据 (10.7) 式, 有

$$T = g_{\mu\nu} T^{\mu\nu} = 0. \quad (10.9)$$

从而电磁场与引力场耦合的爱因斯坦场方程可以写作

$$R_{\mu\nu} = 8\pi G T_{\mu\nu}. \quad (10.10)$$

10.1.2 RN 解

RN 解即是前面所述的爱因斯坦-麦克斯韦方程的球对称解。从第八章的讨论可以知道, 球对称时空的度规必定可以写成如下形式

$$ds^2 = -A(t, r) dt^2 + B(t, r) dr^2 + r^2 d\Omega^2. \quad (10.11)$$

容易算得 $\sqrt{|g|} = \sqrt{AB} r^2 \sin\theta$ 。

球对称性意味着电磁场张量的独立非零分量只可能为 F_{tr} 和 $F_{\theta\phi}$ 。由 $0 = \partial_\theta F_{tr} + \partial_t F_{r\theta} + \partial_r F_{\theta t} = \partial_\theta F_{tr}$, 即知 F_{tr} 独立于 θ , 同理它也独立于 ϕ 。即 F_{tr} 仅仅只能是 t, r 的函数。类似

的, 取 t, θ, ϕ 三个指标, 用恒等式 (10.2), 即知 $F_{\theta\phi}$ 独立于时间 t , 同理它也独立于 r 。所以 $F_{\theta\phi}$ 只能是 θ, ϕ 的函数。

进而 $F^{tr} = g^{tt}g^{rr}F_{tr} = -F_{tr}/(AB)$ 。进而由场方程 $\partial_t(\sqrt{|g|}F^{tr}) = 0$ 得

$$\partial_t\left(\frac{r^2}{\sqrt{AB}}F_{tr}\right) = 0, \quad (10.12)$$

再由场方程 $\partial_r(\sqrt{|g|}F^{tr}) = 0$ 又可得

$$\partial_r\left(\frac{r^2}{\sqrt{AB}}F_{tr}\right) = 0. \quad (10.13)$$

由这两个结果即有

$$F_{tr} = \frac{Q_e}{4\pi r^2} \sqrt{AB}, \quad (10.14)$$

式中 Q_e 为一个常数。

类似的, $F^{\theta\phi} = g^{\theta\theta}g^{\phi\phi}F_{\theta\phi} = F_{\theta\phi}/(r^4 \sin^2 \theta)$ 。进而由场方程 $\partial_\theta(\sqrt{|g|}F^{\theta\phi}) = 0$ 得

$$\partial_\theta\left(\frac{F_{\theta\phi}}{r^2 \sin \theta}\right) = 0, \quad (10.15)$$

又由场方程 $\partial_\phi(\sqrt{|g|}F^{\theta\phi}) = 0$ 得

$$\partial_\phi\left(\frac{F_{\theta\phi}}{r^2 \sin \theta}\right) = 0. \quad (10.16)$$

由这两个结果即有

$$F_{\theta\phi} = \frac{Q_m}{4\pi} \sin \theta, \quad (10.17)$$

式中 Q_m 为一个常数。

到此为止, 电磁场的场方程就已经解完了, 下面就要进一步求解爱因斯坦场方程。为此需要具体算出电磁场的能量动量张量。由 $T_{tt} = g^{rr}F_{tr}F_{tr} - \frac{1}{4}g_{tt}(2F_{tr}F_{tr}g^{tt}g^{rr} + 2F_{\theta\phi}F_{\theta\phi}g^{\theta\theta}g^{\phi\phi})$, 并代入上面的 (10.14) 式和 (10.17) 式, 可得

$$T_{tt} = \frac{1}{2} \left[\left(\frac{Q_e}{4\pi}\right)^2 + \left(\frac{Q_m}{4\pi}\right)^2 \right] \frac{A}{r^4}. \quad (10.18)$$

同理可以算得

$$\begin{aligned} T_{rr} &= -\frac{1}{2} \left[\left(\frac{Q_e}{4\pi}\right)^2 + \left(\frac{Q_m}{4\pi}\right)^2 \right] \frac{B}{r^4} \\ T_{\theta\theta} &= \frac{1}{2} \left[\left(\frac{Q_e}{4\pi}\right)^2 + \left(\frac{Q_m}{4\pi}\right)^2 \right] \frac{1}{r^2} \\ T_{\phi\phi} &= \sin^2 \theta T_{\theta\theta}. \end{aligned} \quad (10.19)$$

能动量张量的非对角分量都等于零, 尤其是 $T_{tr} = 0$ 。另外, 由 T_{tt} 和 T_{rr} 的表达式容易看出

$$T_{tt}/A + T_{rr}/B = 0. \quad (10.20)$$

第八章已经计算了度规 (10.11) 对应的里奇张量, 结果如下

$$\begin{aligned} R_{tt} &= \frac{A''}{2B} + \frac{A'}{rB} - \frac{A'}{4B} \left(\frac{A'}{A} + \frac{B'}{B} \right) - \frac{\ddot{B}}{2B} + \frac{\dot{B}}{4B} \left(\frac{\dot{A}}{A} + \frac{\dot{B}}{B} \right) \\ R_{rr} &= -\frac{A''}{2A} + \frac{B'}{rB} + \frac{A'}{4A} \left(\frac{A'}{A} + \frac{B'}{B} \right) + \frac{\ddot{B}}{2A} - \frac{\dot{B}}{4A} \left(\frac{\dot{A}}{A} + \frac{\dot{B}}{B} \right) \\ R_{\theta\theta} &= 1 - \left(\frac{r}{B} \right)' - \frac{r}{2B} \left(\frac{A'}{A} + \frac{B'}{B} \right) \\ R_{\phi\phi} &= \sin^2 \theta R_{\theta\theta}. \end{aligned} \quad (10.21)$$

以及

$$R_{tr} = R_{rt} = \frac{\dot{B}}{rB}. \quad (10.22)$$

里奇张量的其它分量都是零。另外, 由 (10.21) 式容易得到

$$R_{tt}/A + R_{rr}/B = \frac{1}{rB} \left(\frac{A'}{A} + \frac{B'}{B} \right). \quad (10.23)$$

下面的推导完全类似于第八章。首先, 由 $T_{tr} = 0$ 可得 $R_{tr} = 0$, 进而根据 (10.22) 式知函数 B 独立于时间 t , 即

$$\dot{B} = 0. \quad (10.24)$$

其次, 由场方程 $R_{\mu\nu} = 8\pi G T_{\mu\nu}$ 易有 $R_{tt}/A + R_{rr}/B = (8\pi G)(T_{tt}/A + T_{rr}/B)$, 进而由 (10.20) 式和 (10.23) 式, 可得

$$\left(\frac{A'}{A} + \frac{B'}{B} \right) = (\ln(AB))' = 0. \quad (10.25)$$

所以 $AB = f(t)$, $f(t)$ 为 t 的某函数, 即

$$A(t, r) = \frac{1}{B(r)} f(t), \quad (10.26)$$

所以球对称度规为

$$ds^2 = -\frac{1}{B(r)} f(t) dt^2 + B(r) dr^2 + r^2 d\Omega^2. \quad (10.27)$$

重新定义时间, 使得 $\sqrt{f(t)} dt \rightarrow dt$, 进而吸收掉 $f(t)$, 从而可知, 即使考虑到与电磁场的耦合, 球对称时空也有某种推广的伯克霍夫定理, 即它的度规必定为

$$ds^2 = -A(r) dt^2 + B(r) dr^2 + r^2 d\Omega^2. \quad (10.28)$$

注意式中函数 A, B 均只依赖于 r , 不依赖于时间 t , 而且满足

$$AB = 1. \quad (10.29)$$

考虑到上述结果, 再根据场方程 $R_{\theta\theta} = (8\pi G) T_{\theta\theta}$, 即可得到

$$R_{\theta\theta} = 1 - \left(\frac{r}{B} \right)' = (8\pi G) \frac{1}{2} \left[\left(\frac{Q_e}{4\pi} \right)^2 + \left(\frac{Q_m}{4\pi} \right)^2 \right] \frac{1}{r^2}. \quad (10.30)$$

定义

$$e^2 = \frac{G}{4\pi}(Q_e^2 + Q_m^2). \quad (10.31)$$

进而可将上面方程重写成

$$1 - \left(\frac{r}{B}\right)' = \frac{e^2}{r^2}. \quad (10.32)$$

积分即得 $B(r) = 1/(1 + \frac{c}{r} + \frac{e^2}{r^2})$, 进而 $A(r) = 1 + \frac{c}{r} + \frac{e^2}{r^2}$. 与牛顿近似比较, 可知常数 $c = -2GM$. 所以, 最终的解为

$$A(r) = 1/B(r) = 1 - \frac{2GM}{r} + \frac{e^2}{r^2}. \quad (10.33)$$

正是上一章给出的 RN 解。而电磁场张量的 (10.14) 式和 (10.17) 式就成为

$$F_{tr} = \frac{Q_e}{4\pi r^2}, \quad F_{\theta\phi} = \frac{Q_m}{4\pi} \sin\theta. \quad (10.34)$$

这描述的是一个静止的带电荷 Q_e 和磁荷 Q_m 的粒子。

以上推导过程说明了, Einstein-Maxwell 方程的球对称解是唯一的, 必定为 RN 解, 这就是 Einstein-Maxwell 方程的**伯克霍夫定理**。但这里其实有一点小小的漏洞, 因为有一个例外, 即除了 RN 解之外, Einstein-Maxwell 方程还允许一个球对称解, 即具有几何 $AdS_2 \times S^2$ 并带适当电通量的解。不过, 正如后文我们将要看到的, 这个解可以作为 RN 解的某种极限而出现。至于以上推导的小漏洞到底出在哪里, 后文我们讲到这个解的时候就可以看到了。

10.2 极端黑洞

超极端黑洞

正如我们已经知道的, RN 解有内外两个视界 r_{\pm}

$$r_{\pm} = GM \pm \sqrt{(GM)^2 - e^2}, \quad (10.35)$$

它们对应于函数 $f(r) = A(r) = 1 - \frac{2GM}{r} + \frac{e^2}{r^2}$ 的两个零点。但要让这两个零点存在, 显然要求 $|e| < GM$, 这也就是我们上一章具体分析的情况。

而如果 $|e| > GM$, 那函数 $f(r)$ 将不存在零点, 相应的 RN 解也就不存在视界。但是, $r = 0$ 处的时空奇性依然存在, 那这就变成一个裸奇性, 正如上一章所说, 这是非物理的。这种非物理的 RN 解有时候也称作超极端黑洞。

不妨设一个超极端黑洞仅仅带电荷, 即 $Q_m = 0$, 那这时候超极端的条件 $|e| > GM$ 相当于要求 $Q_e^2/(4\pi) > GM^2$. 但这意味着库伦排斥力超过了万有引力, 因此这种黑洞很可能无法通过恒星的引力坍缩等物理机制形成。所以, 我们通常认为这种超极端黑洞是不存在的。

极端黑洞

所谓的极端黑洞，就是满足如下条件的 RN 黑洞

$$|e| = GM. \quad (10.36)$$

很显然，这时候内外视界合并了， $r_+ = r_- = GM$ ，所以极端黑洞并没有一个单向膜区。相应 RN 度规变成

$$ds^2 = -\left(1 - \frac{GM}{r}\right)^2 dt^2 + \left(1 - \frac{GM}{r}\right)^{-2} dr^2 + r^2 d\Omega^2. \quad (10.37)$$

由于 $r_+ = r_-$ ，所以极端黑洞的 $\kappa_{\pm} = 0$ ，即它的表面引力为零。进而根据霍金温度公式知，极端黑洞的温度为零，所以极端黑洞并不会霍金辐射。这一点和前面关于超极端黑洞不存在的论证是一致的，否则如果极端黑洞辐射的话，由于它会同等地辐射正反粒子，所以其电荷会保持不变，但是因为辐射，其质量却会减少，那它就会成为一个超极端黑洞，而超极端黑洞我们相信是不存在的。

实际上，在超引力理论中，人们的确可以证明 $|e| \leq GM$ ，这通常被称作 Bogomolnyi 能限，Bogomolnyi 能限饱和的黑洞就是极端黑洞，它通常受一定的超对称性保护，从而不会霍金辐射。

回到极端黑洞。在克鲁斯卡延拓中，现在 $r = GM$ 变成了一个双极点 (double pole)，而不是单极点。但是稍微小心点的话，克鲁斯卡延拓还是可以类似地进行，这时候对于 $r > GM$ 和 $0 < r < GM$ 两个区域需要定义两种不同的坐标卡，最后要交错地将两种坐标卡在 $r = GM$ 处拼接起来。彭罗斯图也可以类似地讨论，最后得到的彭罗斯图如图 (10.1) 所示。

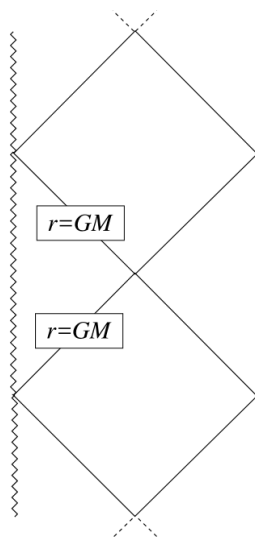


Figure 10.1: 极端黑洞的彭罗斯图。

相比于正常的 RN 黑洞，极端黑洞有一些不同的性质。比方说，考察视界外 $r = \rho$ 处

的某点到视界 $r = r_+$ 的固有空间距离。对于正常的 RN 黑洞，有

$$s = \int_{r_+}^{\rho} \sqrt{g_{rr}} dr = \int_{r_+}^{\rho} \frac{dr}{\sqrt{(1-r_+/r)(1-r_-/r)}} < \infty. \quad (10.38)$$

这是一个有限的距离。但是，对于极端黑洞

$$s = \int_{GM}^{\rho} \frac{dr}{1-GM/r} = \infty. \quad (10.39)$$

所以，视界外的点到极端黑洞视界的固有距离是无穷大。然而，类时或者类光测地线却可以在有限固有时或有限仿射参数值之内到达这个视界。

下面我们来考察极端黑洞的近视界几何，我们将会发现结果和正常 RN 黑洞近视界的 Rindler 几何有很大不同。为了进行这种考察，我们记

$$R = GM, \quad (10.40)$$

并设

$$r = R + \eta, \quad (10.41)$$

式中 $\eta \ll R$ 为近视界的小量。将这些假设代入极端黑洞的度规 (10.37)，即可以得到

$$ds^2 = -\frac{\eta^2}{R^2} dt^2 + R^2 \frac{d\eta^2}{\eta^2} + R^2 d\Omega^2. \quad (10.42)$$

式中前两项是两维反德西特 (Anti-de Sitter, AdS) 时空 Poincare patch 的度规，最后一项描写一个固定半径 R 的两维球面。这样的几何就记作 $AdS_2 \times S^2$ 。在这里，我们发现了 AdS 时空作为极端黑洞的近视界几何而出现了，实际上，马达西那 (Juan Maldacena) 就是通过考察弦论中一些极端黑膜的近视界极限发现 AdS/CFT 对应的。

以上近视界几何本身也就是 Einstein-Maxwell 方程伯克霍夫定理的那个例外解，很清楚，之所以这个解没有包含在原来对 RN 解的推导过程之中，是因为其两维球面部分和其余的时空坐标完全分离了，这一点与前面的球对称时空度规假设有根本性不同。而且，这个解当然不是渐近平坦的。

上述近视界几何的度规 (10.42) 可以写成另一种常见形式，为此引入新的坐标 z ,

$$z = R^2/\eta. \quad (10.43)$$

则可以将 (10.42) 式改写成

$$ds^2 = R^2 \left(\frac{-dt^2 + dz^2}{z^2} \right) + R^2 d\Omega^2. \quad (10.44)$$

这相当于将 AdS_2 表示成了 (t, z) 平面的上半平面 (取 t 轴为水平轴)，很明显， $z = 0$ (也就是 t 轴) 是 AdS_2 空间的共形边界，显然这个边界是类时的。

多黑洞解

下面考虑仅仅带电荷的极端黑洞，即假设磁荷 $Q_m = 0$ ，不妨将电荷简记为 Q 。从而电磁张量仅有 $F_{tr} = \frac{Q}{4\pi r^2}$ ，即电磁势

$$A_t = \frac{Q}{4\pi r}. \quad (10.45)$$

这样的极端黑洞满足条件 $Q^2/(4\pi) = GM^2$ 。假设我们考虑两个这样的极端黑洞，电荷分别为 $Q_1 > 0, Q_2 > 0$ ，质量分别为 M_1, M_2 ，那极端条件将意味着

$$Q_1 Q_2 / (4\pi) = GM_1 M_2. \quad (10.46)$$

这就意味着两个黑洞之间的库伦排斥力刚好与万有引力相平衡。因此，我们容易想到，也许多个这样的极端黑洞放在一起，最后的解也是稳定的。

但是，多个极端黑洞的解当然不会有球对称性，为了证明这样的解存在我们需要重新求解爱因斯坦-麦克斯韦方程。下面我们给出求解方法。

首先，我们要改写一下单个极端黑洞的解 (10.37)，我们将注意力集中在视界之外的几何，为此定义新的径向坐标 ρ

$$\rho = r - GM. \quad (10.47)$$

进而就能将极端 RN 解 (10.37) 重写成

$$ds^2 = -H(\rho)^{-2} dt^2 + H(\rho)^2 (d\rho^2 + \rho^2 d\Omega^2) \quad \text{其中 } H(\rho) = 1 + \frac{GM}{\rho}. \quad (10.48)$$

另外，电磁势现在显然可以写成 (注意到极端条件 $Q = M\sqrt{4\pi G}$)

$$A_t = \frac{1}{\sqrt{4\pi G}} (1 - H(\rho)^{-1}). \quad (10.49)$$

假如将三维空间坐标记作 \mathbf{x} ，则不难看出 $\rho = |\mathbf{x}|$ ，而 $d\rho^2 + \rho^2 d\Omega^2 = (d\mathbf{x})^2$ 。

受上一段的启发，我们假设多黑洞解具有如下形式，

$$ds^2 = -H(\mathbf{x})^{-2} dt^2 + H(\mathbf{x})^2 (d\mathbf{x})^2. \quad (10.50)$$

以及电磁势

$$A_t = \frac{1}{\sqrt{4\pi G}} (1 - H(\mathbf{x})^{-1}). \quad (10.51)$$

式中 $H(\mathbf{x})$ 是某个待定的函数。将这个假设代入爱因斯坦-麦克斯韦方程 (主要是里奇张量的计算比较冗长一些，但也不难)，即可以发现，非常凑巧，麦克斯韦方程和爱因斯坦场方程均简化为如下方程

$$\nabla^2 H(\mathbf{x}) = 0. \quad (10.52)$$

式中 ∇^2 为三维空间的拉普拉斯算符。如果我们要求 $|\mathbf{x}| \rightarrow \infty$ 时 $A_t \rightarrow 0$, 那么这个方程的一般解可以写成

$$H(\mathbf{x}) = 1 + \sum_{a=1}^N \frac{GM_a}{|\mathbf{x} - \mathbf{x}_a|}, \quad (10.53)$$

式中 \mathbf{x}_a 为一组任意给定的常矢量。

所以我们最终找到了爱因斯坦-麦克斯韦方程的一组解, 它描述 N 个满足极端条件 $Q_a = M_a \sqrt{4\pi G}$ 的极端黑洞, 它们分别位于任意的空间位置 \mathbf{x}_a 。这些极端黑洞并不相对运动, 因为它们之间的库伦排斥力和万有引力刚好平衡了。

10.3 带宇宙学常数的黑洞

完全类似于前面求解 RN 黑洞的过程, 我们也可以求解带宇宙学常数的球对称解。依然假定物质场是电磁场, 整个系统还是爱因斯坦-麦克斯韦系统, 不过现在假定宇宙学常数 Λ 非零。

现在的爱因斯坦场方程为 $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi GT_{\mu\nu}$, 将指标 μ, ν 进行缩并, 并利用 $T = 0$, 即有 $R = 4\Lambda$, 进而即可将爱因斯坦场方程重写成

$$R_{\mu\nu} = \Lambda g_{\mu\nu} + 8\pi GT_{\mu\nu}. \quad (10.54)$$

受 RN 解求解过程的启发, 现在直接假设球对称时空为

$$ds^2 = -f(r)dt^2 + f(r)^{-1}dr^2 + r^2d\Omega^2. \quad (10.55)$$

所以现在可以直接跳到求解过程的最后一步, 即求解 $R_{\theta\theta} = \Lambda g_{\theta\theta} + 8\pi GT_{\theta\theta}$ 。根据 (10.21) 式, 现在

$$R_{\theta\theta} = 1 - (rf)'. \quad (10.56)$$

而根据 (10.19) 式

$$T_{\theta\theta} = \frac{1}{2} \left[\left(\frac{Q_e}{4\pi} \right)^2 + \left(\frac{Q_m}{4\pi} \right)^2 \right] \frac{1}{r^2}. \quad (10.57)$$

所以我们要求解的方程变成

$$1 - (rf)' = \Lambda r^2 + \frac{e^2}{r^2}, \quad (10.58)$$

积分并利用牛顿近似, 即得

$$f(r) = 1 - \frac{2GM}{r} + \frac{e^2}{r^2} - \frac{1}{3}\Lambda r^2. \quad (10.59)$$

这就是带宇宙学常数的 RN 解。

假设 $e = 0$, 则 $f(r) = 1 - \frac{2GM}{r} - \frac{1}{3}\Lambda r^2$, 这当然就是带宇宙学常数的施瓦西解。记

$$L^2 = \frac{3}{|\Lambda|}. \quad (10.60)$$

L 大体可以理解为整个时空的半径。则当宇宙学常数 $\Lambda > 0$ 时, 解为

$$ds^2 = -\left(1 - \frac{2GM}{r} - \frac{r^2}{L^2}\right)dt^2 + \left(1 - \frac{2GM}{r} - \frac{r^2}{L^2}\right)^{-1}dr^2 + r^2d\Omega^2. \quad (10.61)$$

这就是所谓的德西特时空中的施瓦西黑洞。很显然, 度规各分量中宇宙学常数的那一项 (即 r^2/L^2 项) 在 $r \ll L$ 的小尺度上是可以忽略的, 但是在 $r \sim L$ 的宇宙学尺度上, 这一项将起主导作用, 并使得时空不再渐近平坦。特别的, 当 $r \gg 2GM$ 时, 这个时空的度规渐近于

$$ds^2 = -\left(1 - \frac{r^2}{L^2}\right)dt^2 + \left(1 - \frac{r^2}{L^2}\right)^{-1}dr^2 + r^2d\Omega^2. \quad (10.62)$$

这就是德西特时空。所以宇宙学常数大于零时, 解渐近于德西特时空, 而不是渐近平坦。另外, 根据以上求解过程也容易知道, 德西特时空本身当然也是宇宙学常数大于零时真空爱因斯坦场方程的解。

德西特时空本身在 $r = L$ 处有一个事件视界, 在这个视界附近 $f(r) \sim \frac{2}{L}(L - r)$, 与 $(2\kappa)(L - r)$ 比较可知德西特视界的表面引力 κ 为

$$\kappa = \frac{1}{L}, \quad (10.63)$$

利用霍金温度公式, 可知德西特时空的温度为

$$T_{dS} = \frac{1}{2\pi L} \frac{\hbar}{k_B}. \quad (10.64)$$

依然假设 $e = 0$ 。而当宇宙学常数 $\Lambda < 0$ 时, 球对称解为

$$ds^2 = -\left(1 - \frac{2GM}{r} + \frac{r^2}{L^2}\right)dt^2 + \left(1 - \frac{2GM}{r} + \frac{r^2}{L^2}\right)^{-1}dr^2 + r^2d\Omega^2. \quad (10.65)$$

这就是所谓的反德西特时空中的施瓦西黑洞。特别的, 当 $r \gg 2GM$ 时, 这个时空的度规渐近于

$$ds^2 = -\left(1 + \frac{r^2}{L^2}\right)dt^2 + \left(1 + \frac{r^2}{L^2}\right)^{-1}dr^2 + r^2d\Omega^2. \quad (10.66)$$

这就是反德西特时空。所以宇宙学常数小于零时, 解渐近于反德西特时空。反德西特时空本身也是爱因斯坦场方程在宇宙学常数小于零时的真空解。显然, 如果不考虑黑洞, 反德西特时空本身并没有视界。

实际上, 德西特时空和反德西特时空的对称性比球对称性还要大, 它们是所谓的最大对称空间, 关于这一点, 我们会在后面的章节中进一步研究, 在那里, 我们还将讨论这两个时空的一些常用坐标, 以及它们的彭罗斯图。

11. 时空对称性

11.1 基林向量场以及最大对称空间

11.1.1 基林向量场

前面在求解施瓦西解的时候，我们已经看到了对称性的力量，在那里，正因为假设了时空度规具有球对称性，所以对爱因斯坦场方程的求解可以大大简化。真空爱因斯坦场方程的平凡解，即平坦的闵可夫斯基时空，更是一个具有很高对称性的例子，实际上，它是一个所谓的具有最大对称性的例子，它具有洛伦兹变换下的对称性以及时空平移对称性，两者合起来也称作庞加莱对称性。也即是说，闵可夫斯基时空的度规 $\eta_{\mu\nu}$ 在庞加莱变换下保持不变。

任何一个给定的时空都有一个给定的度规场，在广义相对论中这个度规场当然是爱因斯坦场方程的解，因此人们常常将给定时空记作 (M, g) ，其中 M 就表示时空流形，而 g 就是其上的那个给定度规场（当然它在不同的局部坐标中有不同的分量形式）。不过，物理文献中也常常会省略符号 M ，而直接称给定的度规场 $g_{\mu\nu}$ 为给定时空。进而，更一般的时空对称性可以定义如下：一个给定时空度规 $g_{\mu\nu}(x)$ 在微分同胚变换 $\phi : x \rightarrow x'$ 之下称作不变的，如果变换以后的度规场 $g'_{\mu\nu}$ 对同一时空点 x 的函数依赖与变换之前的 $g_{\mu\nu}(x)$ 相同，即

$$g'_{\mu\nu}(x) = g_{\mu\nu}(x) \quad \text{对任意 } x. \quad (11.1)$$

这样的微分同胚变换 ϕ 就称作等度规变换。

一个给定时空 $g_{\mu\nu}(x)$ 所有的等度规变换构成一个群, 称作等度规群, 或者时空对称群。也即是说, 假设 $\phi: x \rightarrow x'$ 是一个等度规变换, $\phi': x' \rightarrow x''$ 也是一个等度规变换, 那么 ϕ 和 ϕ' 的函数复合 $\phi' \circ \phi: x \rightarrow x''$ 必然也是一个等度规变换。另外, $id: x \rightarrow x$ 的恒等变换显然是一个等度规变换。而且, 对于任何一个等度规变换 $\phi: x \rightarrow x'$, 其逆变换 $\phi^{-1}: x' \rightarrow x$ 显然也是一个等度规变换。等度规变换的这几条性质都可以很容易地根据定义验证。满足这几条性质的等度规变换的集合就称作构成了一个等度规群, 每一个等度规变换都是这个群里的一个元素, 函数的复合运算 \circ 就称作群的乘法, 而恒等变换 id 就称作群的单位元。

很多时候, 我们可以考虑无穷小等度规变换 $x^\mu \rightarrow x^\mu + \varepsilon^\mu(x)$, 式中 $\varepsilon^\mu(x)$ 为无穷小向量场。根据上面对等度规变换的定义, 可知, 无穷小等度规变换必然使得变换前后度规场的无穷小改变量为零, 即 $\delta g_{\mu\nu} = 0$ 。而根据第四章关于微分同胚变换和李导数的讨论 $\delta g_{\mu\nu} = -\mathcal{L}_\varepsilon g_{\mu\nu}$, 所以对于无穷小等度规变换, 必定有

$$\mathcal{L}_\varepsilon g_{\mu\nu} = 0. \quad (11.2)$$

可以记 $\varepsilon^\mu(x) = \varepsilon X^\mu(x)$, 式中 ε 为无穷小常数, $X^\mu(x)$ 为一个正常的向量场, 那这也即是说, 每一个无穷小变换都可以由一个正常的向量场生成。因此, 无穷小等度规变换所对应的向量场 X^μ 必然满足

$$\mathcal{L}_X g_{\mu\nu} = D_\mu X_\nu + D_\nu X_\mu = 0. \quad (11.3)$$

这样的生成等度规变换的向量场 X_μ 就称作基林向量场! 上述方程就称作基林方程, 基林向量场就是满足基林方程的向量场。

实际上, 根据第六章对微分同胚群的讨论, 可知每一个基林向量场 $X = X^\mu \partial_\mu$ 都会生成一个连续的单参微分同胚变换 ϕ_τ , 只要取 X 为单参变换曲线 (流线) 的切向量, 即满足 $X = \partial_\tau$ 。很显然, 这些单参微分同胚变换都是等度规变换!

对于任意基林矢量 X , 设 ϕ_τ 为它所生成的单参等度规变换, 则有 $X = \partial_\tau$ 。在时空流形上选取合适的局部坐标, 使得参数 τ 为此局部坐标的一个独立坐标分量。从而由李导数的原始定义, 在此局部坐标系中, 基林方程变成

$$0 = \mathcal{L}_X g_{\mu\nu} = \partial_\tau g_{\mu\nu}. \quad (11.4)$$

又由于黎曼曲率张量和里奇张量均由度规张量及其偏导决定, 进而有, 在此特殊的局部坐标系中, $\mathcal{L}_X R_{\rho\sigma\mu\nu} = \partial_\tau R_{\rho\sigma\mu\nu} = 0$ (里奇张量与此类似), 但 $\mathcal{L}_X R_{\rho\sigma\mu\nu}$ 是一个张量, 它在一个特殊的坐标系中为零, 就必然在任意坐标系中都为零。也即是说, 对于任意基林向量 X , 必定有

$$\mathcal{L}_X R_{\rho\sigma\mu\nu} = 0, \quad \mathcal{L}_X R_{\mu\nu} = 0, \quad \mathcal{L}_X R = 0. \quad (11.5)$$

假设 X 为基林向量场, Y 也为基林向量场, 则 X 与 Y 的任意线性组合 $c_1 X + c_2 Y$ 也必定为基林向量场 (式中 c_1, c_2 为任意的两个常数), 因为根据李导数的性质有

$$\mathcal{L}_{c_1 X + c_2 Y} g_{\mu\nu} = c_1 \mathcal{L}_X g_{\mu\nu} + c_2 \mathcal{L}_Y g_{\mu\nu} = 0. \quad (11.6)$$

也即是说, 一个时空所有基林向量场的集合构成一个线性空间。不仅如此, 正如后文将要证明的, 它还必定是一个有限维线性空间。

实际上, 不仅是构成线性空间, 而且基林向量场的集合构成了一个李代数。因为根据第六章证明的重要代数关系 $[\mathcal{L}_X, \mathcal{L}_Y] = \mathcal{L}_{[X, Y]}$ 可知, 若 X 为基林向量场, Y 也为基林向量场, 则 $[X, Y]$ 必定为基林向量场, 因为有

$$\mathcal{L}_{[X, Y]}g_{\mu\nu} = [\mathcal{L}_X, \mathcal{L}_Y]g_{\mu\nu} = \mathcal{L}_X\mathcal{L}_Yg_{\mu\nu} - \mathcal{L}_Y\mathcal{L}_Xg_{\mu\nu} = 0. \quad (11.7)$$

所以基林向量场的集合在李括号 $[X, Y]$ 下构成一个李代数。这说明, 基林向量场生成的等度规群为一个李群。与微分同胚群是一个无穷维李群不同, 由于基林向量场的李代数是有限维的, 所以这个等度规李群也必定为有限维的。

对于任意基林向量 ξ_ρ , 根据黎曼曲率张量的定义, 我们有

$$D_\mu D_\nu \xi_\rho - D_\nu D_\mu \xi_\rho = -\xi_\sigma R^\sigma_{\rho\mu\nu}. \quad (11.8)$$

利用基林方程 $D_\mu \xi_\rho = -D_\rho \xi_\mu$ 即有

$$D_\mu D_\nu \xi_\rho + D_\nu D_\rho \xi_\mu = -\xi_\sigma R^\sigma_{\rho\mu\nu}. \quad (11.9)$$

将指标 $(\mu\nu\rho)$ 进行轮换, 可得

$$\begin{aligned} D_\mu D_\nu \xi_\rho + D_\nu D_\rho \xi_\mu &= -\xi_\sigma R^\sigma_{\rho\mu\nu} \\ D_\nu D_\rho \xi_\mu + D_\rho D_\mu \xi_\nu &= -\xi_\sigma R^\sigma_{\mu\nu\rho} \\ D_\rho D_\mu \xi_\nu + D_\mu D_\nu \xi_\rho &= -\xi_\sigma R^\sigma_{\nu\rho\mu}. \end{aligned}$$

将这三式的第一式加上第二式再减去第三式, 即有

$$2D_\nu D_\rho \xi_\mu = -\xi_\sigma (R^\sigma_{\rho\mu\nu} + R^\sigma_{\mu\nu\rho} - R^\sigma_{\nu\rho\mu}). \quad (11.10)$$

利用黎曼曲率张量的代数性质 $R^\sigma_{\rho\mu\nu} + R^\sigma_{\mu\nu\rho} + R^\sigma_{\nu\rho\mu} = 0$, 即可得

$$D_\nu D_\rho \xi_\mu = \xi_\sigma R^\sigma_{\nu\rho\mu}. \quad (11.11)$$

方程 (11.11) 是一个二阶微分方程, 它的一个重要推论是, 给定 ξ_μ 和 $L_{\mu\nu} \equiv D_\mu \xi_\nu$ 在时空点 x_0 的值, 则就决定了 ξ_μ 在任意时空点 x 的值。为了看清楚这一点, 我们取一条任意的连接 x_0, x 两点的曲线, 设曲线的参数为 τ , 设 $v = \partial_\tau = v^\rho \partial_\rho$ 为这条曲线的切向量, 进而我们可以把二阶微分方程 (11.11) 改写成如下一阶常微分方程组

$$\begin{aligned} \frac{D\xi_\nu}{D\tau} &\equiv v^\mu D_\mu \xi_\nu = v^\mu L_{\mu\nu} \\ \frac{DL_{\nu\rho}}{D\tau} &\equiv v^\mu D_\mu L_{\nu\rho} = v^\mu \xi_\sigma R^\sigma_{\mu\nu\rho}. \end{aligned} \quad (11.12)$$

进而只要沿着这条连接 x_0, x 两点的曲线积分这一组常微分方程, 就能由 ξ_μ 和 $L_{\mu\nu}$ 在 x_0 点的初始值得到 ξ_μ 在点 x 处的值。

以上结果立即有如下推论：(1). 假如基林向量和它的一阶协变导数在某点均为零，则此基林向量场在任意时空点都恒等于零。(2). 在一个 $D = n$ 维的时空流形上，由于根据基林方程 $L_{\mu\nu}$ 的两指标反对称，从而 $L_{\mu\nu}$ 只有 $n(n-1)/2$ 个独立分量，加上 ξ_μ 有 n 个独立分量，所以基林向量场只有 $n + n(n-1)/2 = n(n+1)/2$ 个独立的初始值。由于某些初始值可能没有相应的满足基林方程的基林向量场，从而， n 维时空流形上最多只有 $n(n+1)/2$ 个线性独立的基林向量（从而证明了前面提到的基林向量场所构成的线性空间为有限维），换言之， n 维时空的等度规李群最多只有 $n(n+1)/2$ 维。

具有 $n(n+1)/2$ 个线性独立基林向量场的时空就称作**最大对称时空**，或者称作**最大对称空间**。这里所谓的空间是一个数学概念，指的当然就是时空。也就是说，在最大对称时空中，对于任意 $\xi_\mu(x_0) = a_\mu$ (a_μ 为任意向量) 和任意 $L_{\mu\nu}(x_0) = b_{\mu\nu}$ ($b_{\mu\nu}$ 为任意反对称矩阵)，均存在一个与之相应的基林向量场。

11.1.2 最大对称空间

最大对称空间的定义

最大对称空间的一个例子是闵可夫斯基时空，也就是黎曼曲率张量为零的 n 维时空。这时候，可以选择通常的 n 维笛卡尔直角坐标系，使得度规张量为 $\eta_{\mu\nu} = \text{diag}\{-1, 1, \dots, 1\}$ (diag 表示对角矩阵)，而联络系数均为零。在这种坐标系中，基林向量满足的方程 (11.11) 变成

$$\partial_\nu \partial_\rho \xi_\mu = 0. \quad (11.13)$$

很显然，这个方程的解为（取初始位置 $x_0 = 0$ ）

$$\xi_\mu(x) = a_\mu + b_{\mu\nu} x^\nu \Leftrightarrow \xi^\mu(x) = a^\mu + b^{\mu\nu} x_\nu, \quad (11.14)$$

式中 a_μ 和 $b_{\mu\nu}$ 为常数。代入基林方程，即有

$$b_{\mu\nu} = -b_{\nu\mu}. \quad (11.15)$$

从而 $b_{\mu\nu}$ 的独立分量个数为 $n(n-1)/2$ 。

若取系数 $b_{\mu\nu} = 0$ ，取无穷小量 $\varepsilon_\mu = \varepsilon a_\mu$ ，式中 ε 为无穷小量，则等度规变换 $x^\mu \rightarrow x^\mu + \varepsilon^\mu$ 为无穷小时空平移。很显然独立的无穷小时空平移共有 n 个。时空平移对称性意味着闵可夫斯基时空是均匀的。通过合适地选取一组 $a_{(\mu)}^\nu = \delta_\mu^\nu$ ，我们可以得到如下 n 个线性独立的时空平移基林向量场 $P_\mu = a_{(\mu)}^\nu \partial_\nu = \partial_\mu$ ，即

$$P_\mu = \partial_\mu. \quad (11.16)$$

若取系数 $a_\mu = 0$ ，取无穷小量 $\varepsilon_{\mu\nu} = \varepsilon b_{\mu\nu}$ ，则等度规变换 $x^\mu \rightarrow x^\mu + \varepsilon^\mu_\nu x^\nu$ 为无穷小洛伦兹变换，独立的无穷小洛伦兹变换有 $n(n-1)/2$ 个。洛伦兹对称性意味着闵可夫斯基时空关于 $x = 0$ 的原点是各向同性的，因为洛伦兹对称性意味着时空绕原点“旋转”对称。通过合适地选取一组反对称 $b_{(\mu,\nu)}^{\rho\sigma} = -\delta_\mu^\rho \delta_\nu^\sigma + \delta_\nu^\rho \delta_\mu^\sigma$ ，我们可以得到如下 $n(n-1)/2$ 个线性独立的“旋转”基林向量场 $M_{\mu\nu} = b_{(\mu,\nu)}^{\rho\sigma} x_\sigma \partial_\rho$ ，即

$$M_{\mu\nu} = x_\mu \partial_\nu - x_\nu \partial_\mu. \quad (11.17)$$

闵可夫斯基时空有 n 个独立的平移对称性, $n(n-1)/2$ 个独立的洛伦兹对称性, 从而一共有 $n+n(n-1)/2=n(n+1)/2$ 个线性独立的基林向量, 从而闵可夫斯基时空为最大对称时空。简单的计算表明, 闵可夫斯基时空 $n(n+1)/2$ 个线性独立基林向量场 $P_\mu, M_{\mu\nu}$ 满足如下代数关系

$$\begin{aligned} [P_\mu, P_\nu] &= 0, \quad [M_{\mu\nu}, P_\sigma] = -\eta_{\mu\sigma}P_\nu + \eta_{\nu\sigma}P_\mu, \\ [M_{\mu\nu}, M_{\rho\sigma}] &= -\eta_{\mu\rho}M_{\nu\sigma} + \eta_{\nu\rho}M_{\mu\sigma} + \eta_{\mu\sigma}M_{\nu\rho} - \eta_{\nu\sigma}M_{\mu\rho}. \end{aligned} \quad (11.18)$$

闵可夫斯基时空的等度规群称作庞加莱群, 上式就是庞加莱群的李代数。

将闵可夫斯基时空的例子涉及的一些概念推而广之, 就能搞清楚在一般的时空中, 基林向量场的初始数据 $\xi_\mu(x_0)$ 以及 $L_{\mu\nu}(x_0)$ 是如何联系到时空的均匀性以及各向同性的。具体来说:

- 我们定义**均匀时空**为这样的时空, 它具有如下无穷小等度规变换, 这些变换可以将**任何**给定点 x_0 变换到其邻域内的任何其它点。从而均匀时空必然允许如下基林向量场, 这些基林向量场可以在任意给定点 x_0 处取遍所有可能值。这也就是说, 对于任何初始值 $\xi_\mu(x_0) = a_\mu$ 都必然存在一个与之相应的基林向量场。显然, 这样的线性独立基林向量场有 n 个, 它们就称作 n 个时空平移基林向量场。
- 我们称一个时空关于 x_0 点是**各向同性的**, 假如它有如下等度规变换, 这些变换保持 x_0 点不动, 但却可以将 x_0 点的任何切向量旋转到 x_0 点的任何其它切向量。因此这样的时空必然具有如下基林向量场, 它们满足 $\xi_\mu(x_0) = 0$, 但是 $L_{\mu\nu}(x_0) = D_\mu\xi_\nu(x_0)$ 是任意反对称矩阵 (共有 $n(n-1)/2$ 个独立分量)。很显然, 这样的线性独立基林向量场有 $n(n-1)/2$ 个, 它们就称作 $n(n-1)/2$ 个绕 x_0 点的“旋转”基林向量场。
- 一个既均匀又各向同性的时空就称作一个**最大对称时空**, 很显然, 它也即是具有 $n+n(n-1)/2=n(n+1)/2$ 个线性独立基林向量场的 n 维时空。

这些定义有一些重要的推论, 比如: (1). 一个关于任意 x 点均各向同性的时空必定同时是一个均匀时空。(2). 因此, 一个关于每一点均各向同性的时空必定是一个最大对称时空。

显然, 只有推论 (1) 是需要证明的。证明的关键在于表明, 两个分别绕 x 点和 $x+dx$ 点的“旋转”基林向量的差必定是一个平移基林向量。为此我们取两个分别绕 x_0 点和 x_0+dx 点的“旋转”基林向量 X^μ 和 Y^μ , 因为是“旋转”基林向量, 所以它们分别满足

$$X^\mu(x_0) = 0, \quad Y^\mu(x_0+dx) = 0. \quad (11.19)$$

特别是

$$(D_\mu Y_\nu)(x_0+dx) \neq 0. \quad (11.20)$$

下面, 考虑这两个基林向量的差

$$Z^\mu(x) = X^\mu(x) - Y^\mu(x), \quad (11.21)$$

显然 Z^μ 依然是一个基林向量。

在 x_0 点, 我们有

$$Z^\mu(x_0) = X^\mu(x_0) - Y^\mu(x_0) = -Y^\mu(x_0 + dx - dx). \quad (11.22)$$

将 $Y^\mu(x)$ 在 $x_0 + dx$ 处展开, 即有 (注意 $Y^\mu(x_0 + dx) = 0$)

$$Z_\mu(x_0) = dx^\nu \partial_\nu Y_\mu(x_0 + dx) = dx^\nu D_\nu Y_\mu(x_0 + dx) \neq 0. \quad (11.23)$$

并且, 由于 $DDY \sim Y$, 即 (11.11) 式, 所以 $D_\mu Z_\nu(x_0) = 0$, 所以 $Z^\mu(x)$ 定义了 x_0 处的一个平移基林向量。因此推论 (1) 得证。

最大对称空间与常曲率空间

定理: n 维最大对称空间必然是常曲率空间, 也即是说, 它的黎曼曲率张量必然取如下形式

$$R_{\rho\sigma\mu\nu} = K(g_{\rho\mu}g_{\sigma\nu} - g_{\rho\nu}g_{\sigma\mu}), \quad (11.24)$$

式中 K 是一个常数

$$K = \frac{R}{n(n-1)}. \quad (11.25)$$

换言之, 对于任意两个线性独立的切向量 v_1, v_2 , 最大对称空间的截面曲率必定为

$$K(\Pi) = \frac{R_{\rho\sigma\mu\nu}v_1^\rho v_2^\sigma v_1^\mu v_2^\nu}{\langle v_1, v_1 \rangle \langle v_2, v_2 \rangle - \langle v_1, v_2 \rangle^2} = K. \quad (11.26)$$

也即是说, 最大对称空间的任意截面曲率都必定是同一个常数 K , 所以称作常曲率空间。

证明: 假设 ξ 是任意一个基林向量场, 则根据 (11.5) 式, 我们有 $\mathcal{L}_\xi R_{\rho\sigma\mu\nu} = 0$, 按照李导数的定义式将式子左边展开, 即有

$$\xi^\lambda D_\lambda R_{\rho\sigma\mu\nu} + R_{\lambda\sigma\mu\nu} D_\rho \xi^\lambda + R_{\rho\lambda\mu\nu} D_\sigma \xi^\lambda + R_{\rho\sigma\lambda\nu} D_\mu \xi^\lambda + R_{\rho\sigma\mu\lambda} D_\nu \xi^\lambda = 0.$$

将指标升降一下, 再利用一下黎曼曲率张量的代数性质适当交换一下指标, 并取定任意时空点 x_0 , 这个式子即可以重写成

$$(D^\lambda R_{\rho\sigma\mu\nu})\xi_\lambda(x_0) + (R^\lambda_{\sigma\mu\nu}\delta_\rho^\gamma - R^\lambda_{\rho\mu\nu}\delta_\sigma^\gamma + R^\lambda_{\nu\rho\sigma}\delta_\mu^\gamma - R^\lambda_{\mu\rho\sigma}\delta_\nu^\gamma)D_\gamma \xi_\lambda(x_0) = 0.$$

对于最大对称空间, 总有足够多的线性独立基林向量 ξ , 使得 $\xi_\lambda(x_0) = a_\lambda$ 可以为任意向量, 同时使得 $D_\gamma \xi_\lambda(x_0) = L_{\gamma\lambda}(x_0) = b_{\gamma\lambda}$ 为任意反对称矩阵。从而上面的式子意味着, 在最大对称空间中, 对于任意 a_μ 和任意反对称矩阵 $b_{\gamma\lambda}$, 总有

$$(D^\lambda R_{\rho\sigma\mu\nu})a_\lambda + (R^\lambda_{\sigma\mu\nu}\delta_\rho^\gamma - R^\lambda_{\rho\mu\nu}\delta_\sigma^\gamma + R^\lambda_{\nu\rho\sigma}\delta_\mu^\gamma - R^\lambda_{\mu\rho\sigma}\delta_\nu^\gamma)b_{\gamma\lambda} = 0.$$

而这又意味着, 式中 a_λ 项前面的系数, 以及 $b_{\gamma\lambda}$ 项前面的系数在将指标 γ, λ 反称化以后, 两者都得为零, 从而有

$$D^\lambda R_{\rho\sigma\mu\nu} = 0, \quad (11.27)$$

以及

$$\begin{aligned} & R^\lambda_{\sigma\mu\nu}\delta_\rho^\gamma - R^\lambda_{\rho\mu\nu}\delta_\sigma^\gamma + R^\lambda_{\nu\rho\sigma}\delta_\mu^\gamma - R^\lambda_{\mu\rho\sigma}\delta_\nu^\gamma \\ &= R^\gamma_{\sigma\mu\nu}\delta_\rho^\lambda - R^\gamma_{\rho\mu\nu}\delta_\sigma^\lambda + R^\gamma_{\nu\rho\sigma}\delta_\mu^\lambda - R^\gamma_{\mu\rho\sigma}\delta_\nu^\lambda. \end{aligned} \quad (11.28)$$

将 (11.28) 式的指标 γ 和 ρ 进行缩并, 即可得

$$(n-1)R^\lambda_{\sigma\mu\nu} = R_{\nu\sigma}\delta_\mu^\lambda - R_{\mu\sigma}\delta_\nu^\lambda. \quad (11.29)$$

重命名一下指标, 并将指标适当升降, 即可以写成

$$(n-1)R_{\rho\sigma\mu\nu} = g_{\rho\mu}R_{\sigma\nu} - g_{\rho\nu}R_{\sigma\mu}. \quad (11.30)$$

但根据黎曼曲率张量的代数性质, 这个式子关于指标 ρ, σ 也必须反对称, 从而有

$$g_{\rho\mu}R_{\sigma\nu} - g_{\rho\nu}R_{\sigma\mu} = -(g_{\sigma\mu}R_{\rho\nu} - g_{\sigma\nu}R_{\rho\mu}). \quad (11.31)$$

缩并指标 ρ 和 μ , 即有

$$nR_{\sigma\nu} = g_{\sigma\nu}R. \quad (11.32)$$

将这个式子代入 (11.30) 式, 即有

$$R_{\rho\sigma\mu\nu} = \frac{R}{n(n-1)}(g_{\rho\mu}g_{\sigma\nu} - g_{\rho\nu}g_{\sigma\mu}). \quad (11.33)$$

再代入 (11.27) 式, 即有 R 是一个常数! 从而定理得证。

还可以证明, **最大对称空间具有唯一性**, 即唯一由截面曲率 K 决定, 具体来说即是:

唯一性定理: 假定给定了两个度规 $g_{\mu\nu}(x)$ 和 $g'_{\mu\nu}(x')$, 它们有相同个数的正负本征值 (比方说两者都为闵可夫斯基号差) 且都满足最大对称空间的条件 (11.24), 即

$$R_{\rho\sigma\mu\nu} = K(g_{\rho\mu}g_{\sigma\nu} - g_{\rho\nu}g_{\sigma\mu}) \quad (11.34)$$

$$R'_{\rho\sigma\mu\nu} = K(g'_{\rho\mu}g'_{\sigma\nu} - g'_{\rho\nu}g'_{\sigma\mu}). \quad (11.35)$$

两式中的截面曲率 K 是相同的。则, $g_{\mu\nu}(x)$ 和 $g'_{\mu\nu}(x')$ 必定微分同胚等价, 即必定存在微分同胚变换 $x \rightarrow x'$, 使得

$$g'_{\mu\nu}(x') = g_{\rho\sigma}(x) \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu}. \quad (11.36)$$

证明: 证明过程过于技术化, 感兴趣的读者可以参见温伯格《引力和宇宙学》第十三章对称空间第 13.2 节。

最大对称空间的度规

将最大对称空间满足的条件 (11.24) 式代入外尔曲率张量的定义

$$C_{\rho\sigma\mu\nu} = R_{\rho\sigma\mu\nu} - \frac{2}{n-2}(g_{\rho[\mu}R_{\nu]\sigma} - g_{\sigma[\mu}R_{\nu]\rho}) + \frac{2}{(n-1)(n-2)}Rg_{\rho[\mu}g_{\nu]\sigma},$$

不难得到, 最大对称空间的外尔曲率张量必定为零, 即

$$C_{\rho\sigma\mu\nu} = 0, \quad (11.37)$$

换言之, 最大对称空间必定是共形平坦的, 即其度规张量必定可以写成

$$ds^2 = \Omega^2(x)\eta_{\mu\nu}dx^\mu dx^\nu, \quad (11.38)$$

式中 $\Omega(x)$ 为一个合适的共形因子, $\eta_{\mu\nu} = \text{diag}\{-1, 1, \dots, 1\}$ 为平坦的闵可夫斯基时空的度规。

为了确定式 (11.45) 中的共形因子 $\Omega(x)$, 方法之一是利用如下引理:

关于外尔变换的引理: 在如下外尔变换之下

$$\tilde{g}_{\mu\nu}(x) = \Omega^2(x)g_{\mu\nu}(x), \quad (11.39)$$

克里斯托夫联络的变换关系为 (推导见第九章关于外尔变换的相关讲述)

$$\Gamma_{\rho\sigma}^\mu[\tilde{g}] = \Gamma_{\rho\sigma}^\mu[g] + \gamma_{\rho\sigma}^\mu \quad (11.40)$$

式中

$$\begin{aligned} \gamma_{\rho\sigma}^\mu &= \Omega^{-1}(\delta_\sigma^\mu D_\rho \Omega + \delta_\rho^\mu D_\sigma \Omega - g_{\rho\sigma} D^\mu \Omega) \\ &= \delta_\sigma^\mu D_\rho \ln \Omega + \delta_\rho^\mu D_\sigma \ln \Omega - g_{\rho\sigma} D^\mu \ln \Omega. \end{aligned} \quad (11.41)$$

可以直接算得, 变换以后的黎曼曲率张量为

$$R^\rho_{\sigma\mu\nu}[\tilde{g}] = R^\rho_{\sigma\mu\nu}[g] + D_\mu \gamma_{\nu\sigma}^\rho - D_\nu \gamma_{\mu\sigma}^\rho + \gamma_{\mu\lambda}^\rho \gamma_{\nu\sigma}^\lambda - \gamma_{\nu\lambda}^\rho \gamma_{\mu\sigma}^\lambda. \quad (11.42)$$

而里奇张量为

$$\begin{aligned} R_{\mu\nu}[\tilde{g}] &= R_{\mu\nu}[g] - (n-2)(D_\mu D_\nu \ln \Omega) - g_{\mu\nu}(D^\rho D_\rho \ln \Omega) \\ &\quad + (n-2)(D_\mu \ln \Omega)(D_\nu \ln \Omega) - (n-2)g_{\mu\nu}(D^\rho \ln \Omega)(D_\rho \ln \Omega). \end{aligned}$$

进一步, 可以得到

$$\begin{aligned} R[\tilde{g}] &= \Omega^{-2} \left[R[g] - 2(n-1)(D^\mu D_\mu \ln \Omega) \right. \\ &\quad \left. - (n-2)(n-1)(D^\mu \ln \Omega)(D_\mu \ln \Omega) \right]. \end{aligned} \quad (11.43)$$

进而可以证明, 外尔曲率张量在外尔变换前后保持不变, 即

$$C^\rho_{\sigma\mu\nu}[\tilde{g}] = C^\rho_{\sigma\mu\nu}[g]. \quad (11.44)$$

上面所有式子中的协变导数以及指标升降都是用度规 g 来定义的。以上所有式子的推导都是直接的，只是推导过程可能有些繁琐，读者如果没有时间自己验证，把它们记下来备查也就可以了。

下面将上述引理应用到 (11.45) 式假设的度规，并且为了接下来的操作，我们进一步取 $\Omega(x) = 1/\omega(x)$ ，即假设最大对称空间的度规为

$$ds^2 = \omega^{-2} \eta_{\mu\nu} dx^\mu dx^\nu. \quad (11.45)$$

代入上面的引理，并注意到对于平坦度规 $\eta_{\mu\nu}$ ，相应的联络系数以及曲率均为零，进而即有 (记 $\partial^2 = \partial^\mu \partial_\mu = \eta^{\mu\nu} \partial_\mu \partial_\nu$ ，以及 $(\partial\omega)^2 = \partial^\mu \omega \partial_\mu \omega = \eta^{\mu\nu} \partial_\mu \omega \partial_\nu \omega$)，

$$R = (n-1)[2\omega\partial^2\omega - n(\partial\omega)^2] \quad (11.46)$$

$$R_{\mu\nu} = (n-2)\omega^{-1}\partial_\mu\partial_\nu\omega + \eta_{\mu\nu}[\omega^{-1}\partial^2\omega - (n-1)\omega^{-2}(\partial\omega)^2]. \quad (11.47)$$

另一方面，将最大对称空间的条件 (11.24) 进行适当缩并，即有

$$R_{\mu\nu} = (n-1)Kg_{\mu\nu} = (n-1)K\omega^{-2}\eta_{\mu\nu}, \quad R = n(n-1)K. \quad (11.48)$$

代入上面的 (11.46) 式和 (11.47) 式，即有

$$2\omega\partial^2\omega - n(\partial\omega)^2 = nK \quad (11.49)$$

$$\begin{aligned} & (n-2)\omega^{-1}\partial_\mu\partial_\nu\omega + \eta_{\mu\nu}[\omega^{-1}\partial^2\omega - (n-1)\omega^{-2}(\partial\omega)^2] \\ & = (n-1)K\omega^{-2}\eta_{\mu\nu}. \end{aligned} \quad (11.50)$$

将 (11.49) 式代入 (11.50) 式，即有

$$\partial_\mu\partial_\nu\omega = \frac{1}{n}\eta_{\mu\nu}\partial^2\omega. \quad (11.51)$$

另外，将 (11.49) 式整理一下，即有

$$\frac{2}{n}\omega\partial^2\omega - (\partial\omega)^2 = K. \quad (11.52)$$

从 (11.51) 式可以看出，当 $\mu \neq \nu$ 时， $\partial_\mu\partial_\nu\omega = 0$ ，所以 $\omega(x^0, \dots, x^{n-1})$ 可以表示为 n 个待定一元函数之和，即

$$\omega = f_0(x^0) + f_1(x^1) + \dots + f_{n-1}(x^{n-1}). \quad (11.53)$$

又，在 (11.51) 式中取 $\mu = \nu = \alpha$ (注意下面对 α 没有求和)，并代入上式，即有 $\partial^2\omega = \eta_{\alpha\alpha}n\partial_\alpha\partial_\alpha f_\alpha(x^\alpha)$ ，由于这个式子对任意 $\alpha = 0, 1, \dots, n-1$ 变量均成立，所以必有 $\partial^2\omega$ 为常数，不妨记为

$$\partial^2\omega = na, \quad (11.54)$$

式中 a 为某常数。将这个式子代入 (11.51) 式，即有

$$\partial_\mu\partial_\nu\omega = a\eta_{\mu\nu}. \quad (11.55)$$

积分两次，即得

$$\omega = a \frac{1}{2} \eta_{\mu\nu} x^\mu x^\nu + b_\mu x^\mu + c. \quad (11.56)$$

式中 b_μ 和 c 均为积分常数。将 (11.56) 式代入 (11.52) 式，即有

$$K = 2ac - \eta^{\mu\nu} b_\mu b_\nu. \quad (11.57)$$

根据上面关于最大对称空间的唯一性定理，只要 K 相同，无论 a, b_μ, c 取什么值，所得到的最大对称空间都必定微分同胚等价。为此不妨取 $b_\mu = 0, c = 1$ ，则 (11.57) 式要求 $K = 2a$ ，代入 (11.56) 式，因此最终我们可以将最大对称空间的 ω 取作

$$\omega = 1 + \frac{K}{4} \eta_{\mu\nu} x^\mu x^\nu. \quad (11.58)$$

从而最大对称空间的度规必定可以取为如下形式

$$ds^2 = \frac{\eta_{\mu\nu} dx^\mu dx^\nu}{(1 + \frac{K}{4} \eta_{\rho\sigma} x^\rho x^\sigma)^2}. \quad (11.59)$$

这个度规最早是黎曼在他那篇伟大的演讲《论几何学之基础假设》中给出的（当然黎曼处理的是欧几里德号差的情形），只不过黎曼当时没有给出推导过程。

从最大对称空间的度规表达式 (11.59) 容易看出，1. 当 $K = 0$ 时，对应的度规即是平坦的闵可夫斯基时空度规。2. 当 $K > 0$ 时，对应的最大对称时空就称作 de Sitter 时空。这时候我们可以记

$$K = \frac{1}{L^2}, \quad (11.60)$$

L 称作 de Sitter 时空半径。这时候度规 (11.59) 与上一章给出的 de Sitter 度规的等价关系我们会在后面的章节中进一步研究。3. 当 $K < 0$ 时，对应的最大对称时空称作 Anti de Sitter 时空，这时候可以记

$$K = -\frac{1}{L^2}, \quad (11.61)$$

L 为 Anti de Sitter 时空半径。同样，后面的章节会进一步证明这时候度规 (11.59) 与上一章给出的 Anti de Sitter 度规等价。

特别的，对于 $K = -1/L^2$ ，并且假设我们研究的是二维欧几里德号差的情形（即将度规 $\eta_{\mu\nu}$ 替换成 $\delta_{\mu\nu}$ ），记相应的坐标为 $x^\mu = (x, y)$ ，则这时候 (11.59) 式成为

$$ds^2 = \frac{dx^2 + dy^2}{(1 - \frac{x^2 + y^2}{4L^2})^2}. \quad (11.62)$$

这正是第一章中给出来的庞加莱圆盘世界的度规。

最大对称空间与爱因斯坦场方程

最大对称空间可以作为带宇宙学常数的爱因斯坦场方程的真空解。为了看清楚这一点，我们首先写出带宇宙学常数的真空爱因斯坦场方程

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 0. \quad (11.63)$$

将指标 μ, ν 进行缩并，即有 $R = \frac{2n}{(n-2)}\Lambda$ ，代入上式，即有

$$R_{\mu\nu} = \frac{2}{(n-2)}\Lambda g_{\mu\nu}. \quad (11.64)$$

与最大对称空间的 $R_{\mu\nu} = (n-1)Kg_{\mu\nu}$ 进行比较，可知，只要取

$$\Lambda = \frac{(n-1)(n-2)}{2}K, \quad (11.65)$$

则最大对称空间就是相应真空爱因斯坦场方程的解。(11.65) 式告诉我们，当宇宙学常数 $\Lambda > 0$ 时，de Sitter 时空是真空爱因斯坦场方程的解，而当 $\Lambda < 0$ 时，Anti de Sitter 时空是真空爱因斯坦场方程的解。

11.2 对称性与守恒定律

场的守恒量

在广义相对论中，如果有一个守恒流 J^μ ，它满足如下协变守恒方程

$$D_\mu J^\mu = 0, \quad (11.66)$$

则必定有一个相应的守恒荷。

为了看清楚这一点，我们取时空的一个区域 V ，如图 (11.1) 所示，假设 Σ_1 和 Σ_2 分别为 V 的过去和未来两个类空边界， B 为 V 的类时边界，从而 V 的完整边界 ∂V 可以记为

$$\partial V = \Sigma_1 \cup \Sigma_2 \cup B. \quad (11.67)$$

假设 J^μ 为满足 $D_\mu J^\mu = 0$ 的守恒流。则根据第四章中讨论的弯曲时空中的高斯定理，我们有

$$\begin{aligned} 0 &= \int_V d^n x \sqrt{|g|} D_\mu J^\mu = \int_{\partial V} d^{n-1} x \sqrt{|h|} n_\mu J^\mu \\ &= \int_{\Sigma_1} d^{n-1} x \sqrt{|h|} n_\mu J^\mu + \int_{\Sigma_2} d^{n-1} x \sqrt{|h|} n_\mu J^\mu + \int_B d^{n-1} x \sqrt{|h|} n_\mu J^\mu, \end{aligned}$$

式中 h 为边界 ∂V 上的诱导度规， n_μ 为 ∂V 的单位法向量，规定为从 V 的内部指向外侧。

现在，假设没有流穿过 V 的类时边界 B ，即在 B 上有， $n_\mu J^\mu = 0$ ，则上式的结果可以表达成

$$\begin{aligned} 0 &= \int_{\Sigma_1} d^{n-1} x \sqrt{|h|} n_\mu J^\mu + \int_{\Sigma_2} d^{n-1} x \sqrt{|h|} n_\mu J^\mu \\ &= - \int_{\Sigma_1} d^{n-1} x \sqrt{|h|} n_\mu J^\mu + \int_{\Sigma_2} d^{n-1} x \sqrt{|h|} n_\mu J^\mu, \end{aligned} \quad (11.68)$$

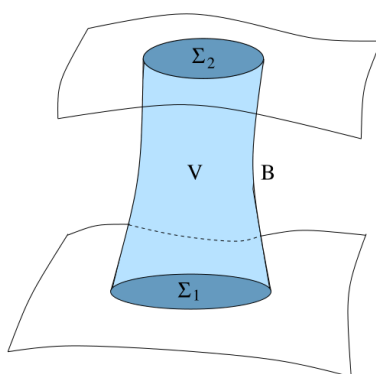


Figure 11.1: 弯曲时空中的荷守恒。

式中第二行我们已经将 Σ_1 的法向反向使得它顺着 (而不是逆着) 时间方向了, 虽然依然是记作 n_μ 。假设我们记类空超曲面 Σ 上的荷 $Q(\Sigma)$ 为

$$Q(\Sigma) = \int_{\Sigma} d^{n-1}x \sqrt{|h|} n_\mu J^\mu, \quad (11.69)$$

式中 Σ 的法向量 n_μ 为顺着时间方向。则 (11.68) 式就可以重新写作

$$Q(\Sigma_1) = Q(\Sigma_2). \quad (11.70)$$

特别的, 我们可以取 Σ 为任意时刻 t 的整个完整的空间截片, 假设在空间无穷远处 $J^\mu \rightarrow 0$, 从而使得在无穷远的 B 处 $n_\mu J^\mu = 0$ 自动满足, 则 (11.70) 式就意味着 $Q(\Sigma)$ 是一个不依赖于时间的守恒荷! 所以, 协变守恒流的存在必然意味着相应守恒荷的存在!

你可能会想, 能量动量张量 $T^{\mu\nu}$ 也满足协变守恒方程 $D_\mu T^{\mu\nu} = 0$, 所以类似的也必定会有相应的守恒荷。但事实上, 情况并非如此, 因为对于两个或两个以上指标的张量, 我们有

$$D_\mu T^{\mu\nu} = \partial_\mu T^{\mu\nu} + \Gamma_{\mu\rho}^\mu T^{\rho\nu} + \Gamma_{\mu\rho}^\nu T^{\mu\rho} = \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} T^{\mu\nu}) + \Gamma_{\mu\rho}^\nu T^{\mu\rho}. \quad (11.71)$$

很显然, 相比于单个指标的流 J^μ 的协变散度, 上式最右边多出来了一项。这多出来的一项就使得无法将协变散度 $D_\mu T^{\mu\nu}$ 的积分转化为边界上的积分, 也即, **不再有相应的高斯定理了!**

所以, 上面关于守恒荷的推导这时候就不再成立了, 因此即使有协变守恒方程 $D_\mu T^{\mu\nu} = 0$, 也并不意味着存在相应的守恒荷! 也即是说, 对于类似于 $T^{\mu\nu}$ 这样的高阶张量, 协变守恒和实际存在守恒荷是两回事!

基林向量可以帮助我们克服能量动量张量的上述困难。因为假设时空中存在一个基林向量场 ξ_μ , 则我们就可以利用能量动量张量构造如下单个指标的流⁴

$$J_\xi^\nu = -\xi_\mu T^{\mu\nu}. \quad (11.72)$$

不难验证这是一个守恒流，因为

$$D_\nu J_\xi^\nu = -(D_\nu \xi_\mu) T^{\mu\nu} - \xi_\mu D_\nu T^{\mu\nu}, \quad (11.73)$$

利用能量动量张量的协变守恒，并注意到 $T^{\mu\nu}$ 是一个对称张量，即有

$$D_\nu J_\xi^\nu = -(D_\nu \xi_\mu) T^{\mu\nu} = -\frac{1}{2}(D_\nu \xi_\mu + D_\mu \xi_\nu) T^{\mu\nu} = 0. \quad (11.74)$$

上式的最后我们利用了基林方程。

进而，利用守恒流 J_ξ^ν ，我们就可以构造相应的守恒荷

$$Q_\xi(\Sigma) = \int_\Sigma d^{n-1}x \sqrt{|h|} n_\mu J_\xi^\mu. \quad (11.75)$$

所以，与每一个线性独立的基林向量相对应的，必定有一个守恒荷！特别的，假设基林向量 ξ 处处类时，即恒有 $g_{\mu\nu} \xi^\mu \xi^\nu < 0$ ，则相应的守恒荷可以等同为物质场的总能量

$$E = Q_\xi(\Sigma). \quad (11.76)$$

对于最大对称空间，与它的 n 个平移基林向量相对应的守恒荷可以等同为物质场的能量和动量，与 $n(n-1)/2$ 个“旋转”基林向量相对应的守恒荷可以等同为物质场的角动量。

粒子的守恒量

以上考察的是物质场的守恒量，如果与引力场耦合的物质是一个粒子，而不是场，则对守恒量的考察又会有些不同。

考察一个在引力场中自由运动的粒子，其速度矢量为 $u^\mu = \frac{dx^\mu}{d\tau}$ ， τ 为此粒子的固有时参数或者仿射参数。用坐标无关的数学记号，这个速度矢量即是

$$u = \frac{dx^\mu}{d\tau} \partial_\mu = \partial_\tau. \quad (11.77)$$

自由粒子走的是测地线，所以满足如下测地线方程

$$D_u u^\mu = u^\nu D_\nu u^\mu = 0. \quad (11.78)$$

假设时空有基林向量场 ξ ，则可以证明如下量必定为守恒量

$$Q \equiv \xi \cdot u \equiv \langle \xi, u \rangle = \xi_\mu u^\mu. \quad (11.79)$$

证明如下

$$\begin{aligned} \frac{d}{d\tau}(\xi \cdot u) &= \frac{dx^\nu}{d\tau} D_\nu(\xi \cdot u) = u^\nu D_\nu(\xi \cdot u) \\ &= u^\nu u^\mu (D_\nu \xi_\mu) + \xi_\mu u^\nu D_\nu u^\mu \\ &= \frac{1}{2} u^\nu u^\mu (D_\nu \xi_\mu + D_\mu \xi_\nu) = 0. \end{aligned} \quad (11.80)$$

式中第二行我们利用了测地线方程，最后一个等于号利用了基林方程。

尤其是, 当 ξ 为类时基林向量场时, 相应的守恒量可以等同于单位质量的粒子在引力场中的守恒能量, 可以记为 ε

$$\varepsilon \equiv -\xi \cdot u. \quad (11.81)$$

举例来说, 对于如下球对称度规

$$ds^2 = -A(r)dt^2 + B(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \quad (11.82)$$

显然有时间平移不变性, 以及沿着 ϕ 方向的旋转不变性, 相应的基林向量场分别为

$$\xi = \partial_t, \quad k_\phi = \partial_\phi. \quad (11.83)$$

根据上面的结论, 必有如下守恒量

$$\varepsilon = -\xi \cdot u = -\langle \partial_t, u \rangle = -u_t = -g_{tt}u^t = A(r)\frac{dt}{d\tau}. \quad (11.84)$$

以及

$$l = k_\phi \cdot u = \langle \partial_\phi, u \rangle = u_\phi = g_{\phi\phi}u^\phi = r^2 \sin^2\theta \frac{d\phi}{d\tau}. \quad (11.85)$$

显然, 这两个守恒量正是第八章在研究粒子在球对称时空中运动时所得到的守恒量。

11.3 共形基林向量场

我们知道, 等度规变换是一种微分同胚变换, 它要求变换前后的度规场保持不变, 有时候我们会放松这个要求, 仅仅要求变换前后的度规场相差一个外尔尺度因子。具体来说, 即考虑微分同胚变换 $x \rightarrow x'$, 记变换之后的度规场为 $g'_{\mu\nu}(x)$, 我们要求

$$g'_{\mu\nu}(x) = \Omega^2(x)g_{\mu\nu}(x). \quad (11.86)$$

注意, 这和直接进行外尔变换有所不同, 这里的 $g'_{\mu\nu}(x)$ 是微分同胚变换之后的度规场, 它满足

$$g'_{\mu\nu}(x') = g_{\rho\sigma}(x) \frac{\partial x^\rho}{\partial x'^\mu} \frac{\partial x^\sigma}{\partial x'^\nu}. \quad (11.87)$$

上面这样的微分同胚变换就称作共形变换 (注意和外尔变换区分开来, 虽然文献中也常常称外尔变换为共形变换, 这个地方的术语有点混乱), 它是等度规变换的一个推广, 因为根据上面的定义, 等度规变换当然可以看作是一种特殊的共形变换 (即 $\Omega(x) = 1$ 的特殊情况)。一个时空流形所有共形变换的集合就构成了时空的共形变换群, 等度规群是共形变换群的一个子群。

假设一个上面这样的无穷小共形变换由某向量场 X^μ 生成, 即 $x'^\mu = x^\mu + \varepsilon X^\mu(x)$, 则由无穷小微分同胚的定义, 有

$$g'_{\mu\nu}(x) = g_{\mu\nu}(x) - \mathcal{L}_{\varepsilon X} g_{\mu\nu} = g_{\mu\nu}(x) - \varepsilon (D_\mu X_\nu + D_\nu X_\mu). \quad (11.88)$$

另一方面, 假设将 (11.86) 式中的共形因子 Ω 按照无穷小量 ε 展开成 (注意对于 $\varepsilon = 0$ 的恒等变换必定有 $\Omega = 1$) $\Omega(x) = 1 + \varepsilon\sigma_X(x)$, 则由 (11.86) 式, 有

$$g'_{\mu\nu}(x) = g_{\mu\nu}(x) + 2\varepsilon\sigma_X(x)g_{\mu\nu}(x). \quad (11.89)$$

比较这个式子和上面的 (11.88) 式, 即有

$$D_\mu X_\nu + D_\nu X_\mu = -2\sigma_X(x)g_{\mu\nu}(x). \quad (11.90)$$

或者写成

$$\mathcal{L}_X g_{\mu\nu} = -2\sigma_X(x)g_{\mu\nu}(x). \quad (11.91)$$

这个方程就称作共形基林方程, 满足这个方程的向量场 X^μ 就称作**共形基林向量场**。

不难证明, 所有的共形基林向量场在向量场的李括号运算下构成了一个李代数, 它也就是时空共形变换群的李代数。为了证明这一点, 只需注意到, 若 X 为共形基林向量场, Y 也为共形基林向量场, 则必有

$$\mathcal{L}_{[X,Y]}g_{\mu\nu} = \mathcal{L}_X\mathcal{L}_Yg_{\mu\nu} - \mathcal{L}_Y\mathcal{L}_Xg_{\mu\nu} = -2(X\sigma_Y - Y\sigma_X)g_{\mu\nu}, \quad (11.92)$$

只需令 $\sigma_{[X,Y]} \equiv X\sigma_Y - Y\sigma_X$, 从而 $[X,Y]$ 也必定是共形基林向量场。

为了进一步改写共形基林方程, 我们将式 (11.90) 中的指标 μ, ν 进行缩并, 即有 $D_\mu X^\mu = -n\sigma_X(x)$ (n 为时空的维数), 反代入 (11.90) 式, 即得

$$D_\mu X_\nu + D_\nu X_\mu = \frac{2}{n}(D_\rho X^\rho)g_{\mu\nu}(x). \quad (11.93)$$

不难发现, 基林方程是共形基林方程在 $D_\rho X^\rho = 0$ 时的特殊情况, 因为对基林方程的指标进行缩并可知满足基林方程就自动满足 $D_\rho X^\rho = 0$, 从而, 任何基林向量场也自动是共形基林向量场。

以平坦的 n 维闵可夫斯基时空为例子, 除了前面讨论过的基林向量 $P_\mu = \partial_\mu$ 和 $M_{\mu\nu} = x_\mu\partial_\nu - x_\nu\partial_\mu$ 自动也是共形基林向量以外, 它还有如下共形基林向量。

- 首先, 由于 x^μ 满足共形基林方程 $\partial_\mu x_\nu + \partial_\nu x_\mu = 2\eta_{\mu\nu}$, 所以有如下共形基林向量

$$D = x^\mu\partial_\mu. \quad (11.94)$$

它生成的无穷小微分同胚变换为 $x^\mu \rightarrow x^\mu + \varepsilon x^\mu = e^\varepsilon x^\mu$, 这相当于将坐标进行放缩, 所以 D 称作伸缩变换 (dilatation) 的生成元。

- 此外, 还有 n 个共形基林向量 $C^{(m)} = (2x^m x^\mu - \eta^{m\mu} x_\rho x^\rho)\partial_\mu$ 。这是因为向量 $C^{(m)}$ 满足如下共形基林方程

$$\partial_\mu C_\nu^{(m)} + \partial_\nu C_\mu^{(m)} = 4x^m \eta_{\mu\nu}. \quad (11.95)$$

$C^{(m)}$ 称作特殊共形变换的生成元, 通常记 $C^\mu \equiv C^{(\mu)}$ 。

生成元 D 和 C_μ 扩大了时空平移和洛伦兹变换的庞加莱代数。在最终的共形代数中，除了原来的庞加莱代数关系以外，还得包括如下代数关系

$$\begin{aligned}
 [D, P_\mu] &= -P_\mu, & [D, C_\mu] &= C_\mu \\
 [M_{\mu\nu}, D] &= 0, & [C_\mu, C_\nu] &= 0 \\
 [M_{\mu\nu}, C_\sigma] &= -\eta_{\mu\sigma}C_\nu + \eta_{\nu\sigma}C_\mu \\
 [P_\mu, C_\nu] &= 2(\eta_{\mu\nu}D - M_{\mu\nu}).
 \end{aligned} \tag{11.96}$$

注意算符 $D = x^\mu \partial_\mu$ 是一个数齐次度的算符， $P_\mu = \frac{\partial}{\partial x^\mu}$ 当然齐次度是 -1 ，因此就有上面的第一个式子，类似的， C^μ 的齐次度是 $+1$ ，因此有第二个式子。第三个式子无非是说 D 本身是一个洛伦兹标量，类似的，第五个式子说的无非是 C_μ 为一个洛伦兹矢量（和 P_μ 一样）。真正需要验证的是第四个式子和最后一个式子，而这不难通过直接计算来验证。共形李代数在共形量子场论（简称共形场论）中有关键的重要性。

共形基林向量场与守恒量

对于讨论共形对称性而言，有一类物质场是特殊的，那就是其作用量在度规场的外尔变换下保持不变的物质场。这种物质场有一个特殊的性质，即其能动张量的指标缩并等于零，或者说其能动张量的迹为零，即 $T^\mu_\mu = 0$ 。为了看清楚这一点，我们考虑一个尺度因子为 $\Omega(x) = (1 + \varepsilon\sigma(x))$ 的外尔变换，因此变换前后度规场的改变量为 $\delta g_{\mu\nu} = 2\varepsilon\sigma(x)g_{\mu\nu}$ ，进而变换前后物质场作用量 S_m 的改变量为

$$\begin{aligned}
 \delta S_m &= \frac{1}{2} \int d^n x \sqrt{|g|} T^{\mu\nu} \delta g_{\mu\nu} = \frac{1}{2} \int d^n x \sqrt{|g|} T^{\mu\nu} 2\varepsilon\sigma(x)g_{\mu\nu} \\
 &= \int d^n x \sqrt{|g|} T^\mu_\mu \varepsilon\sigma(x).
 \end{aligned} \tag{11.97}$$

如果物质场作用量在任意外尔变换之下都保持不变，即有 $\delta S_m = 0$ ，那由上式即有

$$T^\mu_\mu = 0. \tag{11.98}$$

在第十章中，我们看到了这种物质场的一个典型例子，那就是电磁场。

可以证明，如果存在一个满足 $T^\mu_\mu = 0$ 的守恒能动张量，那么对于每一个共形基林向量 X^μ ，都必定存在一个相应的守恒流 J_X^μ

$$J_X^\mu \equiv T^{\mu\nu} X_\nu. \tag{11.99}$$

证明如下（利用了 $D_\mu T^{\mu\nu} = 0$ ），

$$D_\mu J_X^\mu = T^{\mu\nu} D_\mu X_\nu = \frac{1}{2} T^{\mu\nu} (D_\mu X_\nu + D_\nu X_\mu). \tag{11.100}$$

代入共形基林方程 (11.90)，即有

$$D_\mu J_X^\mu = -\sigma_X g_{\mu\nu} T^{\mu\nu} = -\sigma_X T^\mu_\mu = 0, \tag{11.101}$$

最后一个等号利用了 $T^\mu_\mu = 0$ 的假定。

正如前面讲过的, 每一个守恒流都有一个相应的守恒量。因此, 如果物质场满足 $T^\mu_\mu = 0$, 那么每一个共形基林向量都有一个相应的守恒量。特别的, 在平坦闵可夫斯基时空中, 具有这种性质的物质场就称作共形场。所以, 平坦时空的电磁场是一个共形场。

不仅电磁场有这种特殊性, 而且作为光子世界线的光线也有相应的特殊性。可以证明, 这时候对于每一个共形基林向量 X , $Q = \langle X, u \rangle$ 均沿着光线守恒, 这里 u 为光线的切向量, 满足测地线方程以及类光条件 $g_{\mu\nu}u^\mu u^\nu = 0$ 。证明过程完全类似于前面关于粒子守恒量中的相关讨论, 只是最后需要额外用一下类光条件。

11.4 静态时空与引力红移

这一节我们利用基林向量场来刻画稳态时空和静态时空, 并重新回顾静态时空的引力红移公式。

11.4.1 稳态时空与静态时空

定义: 所谓的**稳态时空**, 即存在类时基林向量场 ξ 的时空。稍后我们将解释为什么称这样的时空为稳态时空。

为了接下来的讨论, 下面我们引入一个数学概念, 叫做**向量场的积分曲线**。虽然相关的内容我们其实已经见过多次了, 但是我们还一直没有给出这个数学上的专有名称。我们知道, 对于时空中任何向量场 $V = V^\mu \partial_\mu$, 均可以把它看成是某种时空“流体”的流速场, 进而通过积分下面的微分方程就可以得到相应的时空“流线”

$$\frac{dx^\mu}{d\tau} = V^\mu(x(\tau)). \quad (11.102)$$

这些“流线”就称作向量场 V^μ 的**积分曲线**, 积分曲线就是那个专门的数学概念。显然, 参数 τ 就是积分曲线的参数, 而积分曲线的切向量就是 V , 因为

$$V = V^\mu(x(\tau))\partial_\mu = \frac{dx^\mu}{d\tau}\partial_\mu = \partial_\tau. \quad (11.103)$$

假设 ξ 为稳态时空的类时基林向量场, 设 ξ 积分曲线的参数为 t , 从而有

$$\xi = \partial_t. \quad (11.104)$$

又由于 ξ 类时, 所以我们总可以在时空上取到以参数 t 为 x^0 的局部坐标, 在此局部坐标中 ξ 所满足的基林方程变成

$$0 = \mathcal{L}_\xi g_{\mu\nu} = \partial_t g_{\mu\nu}. \quad (11.105)$$

换言之, 对于稳态时空, 我们总可以取到如下局部坐标, 使得在此局部坐标中度规场不依赖于坐标的 0 分量, 即 $g_{\mu\nu}$ 不依赖于 $x^0 = t$ 。也即是说, 稳态时空总可以使得度规场与时间坐标无关, 这也就是我们称之为稳态时空的原因!

根据上面所述, 对于稳态时空, 我们总可以取局部坐标 $(t, \mathbf{x}) = (t, x^1, x^2, x^3)$, 使得度规为 (式中 $i, j = 1, 2, 3$)

$$ds^2 = g_{tt}(\mathbf{x})dt^2 + 2g_{ti}(\mathbf{x})dtdx^i + g_{ij}(\mathbf{x})dx^i dx^j. \quad (11.106)$$

注意, 上式中的时间空间交叉项 $g_{ti}(\mathbf{x})$ 一般来说不等于零。

如果能够进一步取合适的局部坐标, 使得上式中的时间空间交叉项等于零, 即使得

$$g_{ti}(\mathbf{x}) = 0. \quad (11.107)$$

那这样的稳态时空就称作**静态时空**。换言之, 静态时空必定可以取合适的局部坐标以使得度规场为如下形式

$$ds^2 = g_{tt}(\mathbf{x})dt^2 + g_{ij}(\mathbf{x})dx^i dx^j. \quad (11.108)$$

实际上, 这是对类时基林向量场 ξ 的一个要求, 具有上述性质的类时基林向量场就称作**超曲面正交的**。所谓超曲面正交, 即是指必定存在空间超曲面与这样的类时基林向量场的积分曲线正交, 从而使得度规的时间空间交叉项消失。因此, 我们有如下定义:

定义: 所谓的**静态时空**, 即存在超曲面正交的类时基林向量场的时空。

施瓦西时空是不是一个静态时空呢? 实际上不是的, 的确, 它有一个基林向量场 $\xi = \partial_t$, 但这个基林向量场只在黑洞视界之外类时, 在视界之内是类空的。所以整体来看, 施瓦西时空并非静态的, 但它在视界之外的部分的确是静态的!

注意, 稳态时空和静态时空都是时空本身内禀的性质, 和具体的局部坐标无关。比方说, 假设有某二维时空, 在某局部坐标中其度规可以表达成 $ds^2 = -t^{-4}dt^2 + dx^2$, $t > 0$, 这个度规的 g_{tt} 分量依赖于时间 t , 看起来这似乎不是一个稳态时空, 但实际上, 只要取坐标变换 $t' = t^{-1}$, $x' = x$, 就能把度规表达式变换成 $ds^2 = -dt'^2 + dx'^2$, 这当然是一个稳态时空, 而且是静态时空。

11.4.2 静态时空的引力红移

下面我们讨论如何利用类时基林向量场 ξ 来处理静态时空中的引力红移效应。

为此我们首先要讨论一个问题, 即时空中一个协变速度为 u^μ (满足归一化条件 $u^\mu u_\mu = -1$) 的观察者测量某个物理量 J^μ 的结果是什么。

定义: 速度为 u^μ 的观察者 O 测量物理量 J^μ , 测到的时间分量为

$$J^\mu \equiv -J_u \equiv -J_\mu u^\mu. \quad (11.109)$$

如果是测量能量密度, 那测到的结果即是

$$T^{uu} \equiv T_{uu} \equiv T_{\mu\nu} u^\mu u^\nu. \quad (11.110)$$

所谓 O 测到的时间分量, 即在 O 自己的参考系中观测 J^μ 所得到的时间分量。很容易看到上面的定义是合理的, 因为协变速度 u^μ 的方向当然就是观察者 O 自身参考系的时间轴方向, 单位向量 u^μ 就是这个时间轴的单位向量, 所以观察到的 J^μ 的时间分量当然就是 J^μ 向着观察者时间轴的投影, 即 J^μ 和单位向量 u^μ 的内积。能量密度的定义与此类似。

下面来讨论静态时空中的引力红移。假设静态时空中有两个静态观察者 O_1 和 O_2 , 他们的速度分别为 u_1^μ 和 u_2^μ 。假定 O_1 在时空点 p_1 发射一个光子, 并被 O_2 在时空点 p_2 接收到。光子应该沿一条切向量为 k^μ 的类光测地线传播, 由于 k^μ 可以看作是由光子角频率

和空间波矢量所构成的协变向量, 所以根据上面的定义, O_1 和 O_2 测到的光子角频率应该分别为

$$\omega_1 = -(k_\mu u_1^\mu)_{p_1}, \quad \omega_2 = -(k_\mu u_2^\mu)_{p_2}. \quad (11.111)$$

静态时空的时间方向当然就是类时基林向量场 ξ^μ 的方向, 而由于两个观察者都为静态, 所以他们的协变速度应该都只有时间分量非零, 从而 u_1^μ 和 u_2^μ 应该分别与 ξ^μ 成比例, 再注意到协变速度是归一化的向量, 即有

$$u_1^\mu = \frac{\xi^\mu}{\sqrt{-\xi^\rho \xi_\rho}}|_{p_1}, \quad u_2^\mu = \frac{\xi^\mu}{\sqrt{-\xi^\rho \xi_\rho}}|_{p_2}. \quad (11.112)$$

根据作类光测地线运动的光子的能量守恒 (因为存在类时基林向量场), 必有

$$(\xi_\mu k^\mu)_{p_1} = (\xi_\mu k^\mu)_{p_2}. \quad (11.113)$$

因此我们有

$$\frac{\omega_1}{\omega_2} = \frac{\sqrt{-\xi^\rho \xi_\rho}|_{p_2}}{\sqrt{-\xi^\rho \xi_\rho}|_{p_1}}. \quad (11.114)$$

这就是利用类时基林向量场表达的引力红移公式。

如果取合适的坐标系, 进而将静态时空的度规写成 $ds^2 = g_{tt}(\mathbf{x})dt^2 + g_{ij}(\mathbf{x})dx^i dx^j$, 也就是说, 类时基林向量场为 $\xi = \partial_t$, 从而

$$-\xi^\rho \xi_\rho|_{\mathbf{x}} = -\langle \xi, \xi \rangle|_{\mathbf{x}} = -\langle \partial_t, \partial_t \rangle|_{\mathbf{x}} = -g_{tt}(\mathbf{x}). \quad (11.115)$$

代入上面的红移公式 (11.114), 即有

$$\frac{\omega_1}{\omega_2} = \left(\frac{g_{tt}(\mathbf{x}_{p_1})}{g_{tt}(\mathbf{x}_{p_2})} \right)^{-\frac{1}{2}}. \quad (11.116)$$

这正是第三章中导出的引力红移公式。但现在, 公式 (11.114) 的好处是, 它是坐标无关的! 而且, 很显然, 公式 (11.114) 也可以推广到稳态时空中的稳态观察者。

11.5 基林视界

第九章引入了基林视界的概念, 但是只进行了初步的研究, 本节我们将进行更系统的研究。所谓的基林视界, 其定义是, 首先它是一个类光超曲面, 其次, 其类光法向量 (也是一个切向量) 刚好是一个基林向量。当一个超曲面同时满足这两点时, 就称之为一个基林视界。

设有一个基林视界 \mathcal{N} , 设与之相应的基林向量为 K^μ 。第九章我们证明了 K^μ 在 \mathcal{N} 上必定满足如下方程

$$K^\mu D_\mu K^\nu = w(x)K^\nu. \quad (11.117)$$

如果我们定义标量函数 $S(x) = K^\mu K_\mu$, 进而即有 $D_\nu S = 2K^\mu D_\nu K_\mu$ 。利用基林方程 $D_\nu K_\mu = -D_\mu K_\nu$, 即有 $D_\nu S = -2K^\mu D_\mu K_\nu$, 将这个结果限制在超曲面 \mathcal{N} 上, 并代入上面的 (11.117) 式, 即有 $D_\nu S|_{\mathcal{N}} = -2w(x)K_\nu|_{\mathcal{N}}$ 。换言之, 在超曲面 \mathcal{N} 上必有

$$K_\mu = f(x)D_\mu S = f(x)\partial_\mu S, \quad (11.118)$$

式中 $f(x) = -1/(2w(x))$ 。另外, 由于 K^μ 是 \mathcal{N} 的类光法向量, 所以超曲面 \mathcal{N} 必定满足如下方程

$$S(x)|_{\mathcal{N}} = (K^\mu K_\mu)|_{\mathcal{N}} = 0. \quad (11.119)$$

利用 (11.118) 式, 不难验证超曲面 \mathcal{N} 的类光法向量 K^μ 必定满足所谓的 Frobenius 定理, 即满足

$$K_{[\mu} D_\nu K_{\rho]} = 0. \quad (11.120)$$

式中 $[\mu\nu\rho]$ 表示三个指标的全反对称化。验证上式的一个办法是, 首先根据克里斯托夫联络是一个对称联络验证 $K_{[\mu} D_\nu K_{\rho]} = K_{[\mu} \partial_\nu K_{\rho]}$, 然后再根据 (11.118) 式有 $K_{[\mu} \partial_\nu K_{\rho]} = 0$ 。

结合基林方程, 也可以将上面这个结果写成 (不难验证下式中的三个指标是全反对称的)

$$(K_\rho D_\mu K_\nu + K_\mu D_\nu K_\rho - K_\nu D_\mu K_\rho)|_{\mathcal{N}} = 0. \quad (11.121)$$

以 $D^\mu K^\nu$ 去缩并上式, 即得在 \mathcal{N} 上有 (利用了基林方程)

$$\begin{aligned} K_\rho D^\mu K^\nu D_\mu K_\nu &= -(D^\mu K^\nu)K_\mu D_\nu K_\rho + (D^\mu K^\nu)K_\nu D_\mu K_\rho \\ &= -(D^\mu K^\nu)K_\mu D_\nu K_\rho - (D^\nu K^\mu)K_\nu D_\mu K_\rho = -2(D^\mu K^\nu)K_\mu D_\nu K_\rho, \end{aligned}$$

代入 (11.117) 式, 即得

$$K_\rho D^\mu K^\nu D_\mu K_\nu|_{\mathcal{N}} = -2w(x)K^\nu D_\nu K_\rho|_{\mathcal{N}} = -2w^2(x)K_\rho|_{\mathcal{N}}. \quad (11.122)$$

所以只要 $K_\rho \neq 0$, 就有

$$w^2(x) = -\frac{1}{2}(D^\mu K^\nu)(D_\mu K_\nu)|_{\mathcal{N}}. \quad (11.123)$$

对于基林视界 \mathcal{N} , 由于基林向量场满足方程 (11.11), 所以其基林向量场 K^μ 满足

$$D_\nu D_\rho K_\mu = K_\sigma R^\sigma_{\nu\rho\mu}. \quad (11.124)$$

另外, K^μ 既是 \mathcal{N} 的类光法向量, 也是它上面的一个切向量, 在 \mathcal{N} 上 K^μ 的积分曲线就是一些扫出整个 \mathcal{N} 的类光测地线, \mathcal{N} 由这些类光测地线铺成。进而我们有如下引理

引理: 沿着 \mathcal{N} 上每一条 K^μ 的积分曲线, $w^2(x)$ 都是常数。

证明：任取 \mathcal{N} 的一个切向量场 t^μ ，利用 (11.123) 式和 (11.124) 式，我们有

$$\begin{aligned} t^\rho \partial_\rho w^2 &= t^\rho D_\rho w^2 = -t^\rho (D^\mu K^\nu)(D_\rho D_\mu K_\nu)|_{\mathcal{N}} \\ &= -t^\rho (D^\mu K^\nu) K_\sigma R^\sigma_{\rho\mu\nu}|_{\mathcal{N}} \\ &= -(D^\mu K^\nu) R_{\sigma\rho\mu\nu} K^\sigma t^\rho|_{\mathcal{N}}. \end{aligned} \quad (11.125)$$

沿着 K^ρ 的积分曲线，因此用 K^ρ 代替式中的 t^ρ ，即有

$$K^\rho \partial_\rho w^2 = -(D^\mu K^\nu) R_{\sigma\rho\mu\nu} K^\sigma K^\rho|_{\mathcal{N}} = 0. \quad (11.126)$$

式中利用了 $R_{\sigma\rho\mu\nu}$ 关于指标 σ, ρ 的反对称性。所以，引理得证。

下面我们再引入一个概念，叫分岔基林视界，它是一个基林视界，并且其上所有类光测地线都是从一个满足 $K^\mu = 0$ 的分岔球 (记为 B) 发出的。对于分岔基林视界，我们可以证明如下定理：

定理：如果 \mathcal{N} 是一个分岔基林视界，则 $w^2(x)$ 在整个 \mathcal{N} 上都是常数。

证明：由于 $w^2(x)$ 沿着 \mathcal{N} 的每一条类光测地线都是常数，而所有这些类光测地线都从分岔球 B 发出，因此只要证明 $w^2(x)$ 在 B 上是一个常数，就必定有 $w^2(x)$ 在整个 \mathcal{N} 上都是常数。另一方面，由于 $K^\sigma|_B = 0$ ，因此根据 (11.125) 式，沿着 B 的任意切向量 t^ρ ，我们有

$$t^\rho \partial_\rho w^2 = -(D^\mu K^\nu) R_{\sigma\rho\mu\nu} t^\rho K^\sigma|_B = 0. \quad (11.127)$$

因此这就证明了 w^2 在 B 上是一个常数，从而定理得证。

由于上述定理，所以 $w(x)$ 在整个分岔基林视界 \mathcal{N} 上必定只能和某个常数 κ 相差正负号，即有 $w(x) = \pm\kappa$ ，代入 (11.117) 式，即有，在分岔基林视界上

$$K^\mu D_\mu K^\nu = \pm\kappa K^\nu. \quad (11.128)$$

常数 κ 就称作表面引力。当然，这个结论我们在第九章中提到过，这里只是给出了具体的证明。

上述结论的意义在于，由于表面引力决定了霍金温度，所以上述结论就说明了，在整个分岔基林视界上，**霍金温度是一个常数**！这可以看作是分岔基林视界的热力学第零定律，它和黑洞热力学第零定律密切相关，因为很多黑洞 (比如克鲁斯卡延拓下的施瓦西黑洞) 的事件视界本身是分岔基林视界的一部分，所以分岔基林视界的热力学第零定律也意味着，在这些黑洞的事件视界上，霍金温度是一个常数，这也就是黑洞的热力学第零定律。

12. 德西特空间与反德西特空间

本章将承接前面的章节，系统地介绍一下德西特以及反德西特空间，包括它们的一些常用坐标，以及它们的彭罗斯图。

12.1 德西特和反德西特空间的定义

欧几里德号差的最大对称空间

n 维欧几里德空间 \mathbb{R}^n 是一个曲率为零的最大对称空间。除此之外， n 维欧几里德号差的最大对称空间还有截面曲率为正的 n 维球面 S^n ，以及截面曲率为负的 n 维双曲空间 H^n 。

其中 S^n 可以嵌入到 $n+1$ 维欧几里德空间 \mathbb{R}^{n+1} 中作为它的一张超曲面。具体来说，取 $z^A = (z^1, z^2, \dots, z^{n+1})$ 为 \mathbb{R}^{n+1} 的笛卡尔坐标， \mathbb{R}^{n+1} 上有如下平坦的黎曼度规

$$ds^2 = (dz^1)^2 + (dz^2)^2 + \dots + (dz^{n+1})^2 = \delta_{AB} dz^A dz^B. \quad (12.1)$$

则 S^n 可以由下面方程定义

$$S^n: (z^1)^2 + \dots + (z^{n+1})^2 = L^2 \Leftrightarrow \delta_{AB} z^A z^B = L^2. \quad (12.2)$$

式中 L 为 S^n 的半径。 S^n 上的黎曼度规可以通过将 \mathbb{R}^{n+1} 的平坦度规限制在超曲面 S^n 上而诱导得出。

为了证明上面定义的 S^n 的确是一个最大对称空间，我们指出它有如下 $(n+1)(n+1-1)/2 = n(n+1)/2$ 个基林向量场

$$M_{AB} = z_A \partial_B - z_B \partial_A, \quad (12.3)$$

式中 $z_A \equiv \delta_{AB} z^B = z^A$ 。即 S^n 有最大可能数目的线性独立基林向量，从而是最大对称空间。为了验证 M_{AB} 的确是基林向量场，我们不妨以 $M_{12} = z^1 \partial_2 - z^2 \partial_1$ 为例，它生成的无穷小微分同胚为

$$z^1 \rightarrow z'^1 = z^1 - \varepsilon z^2, \quad z^2 \rightarrow z'^2 = z^2 + \varepsilon z^1. \quad (12.4)$$

不难看出，在此微分同胚下，有

$$(z'^1)^2 + (z'^2)^2 = (z^1 - \varepsilon z^2)^2 + (z^2 + \varepsilon z^1)^2 = (z^1)^2 + (z^2)^2. \quad (12.5)$$

从而 S^n 的超曲面方程 (12.2) 在此微分同胚下保持不变，这说明 M_{12} 是 S^n 的一个切向量。类似的，可以验证度规 $ds^2 = \delta_{AB} dz^A dz^B$ 在此微分同胚下也保持不变。从而 M_{12} 是 S^n 上的一个基林向量。完全类似的， M_{AB} 也是 S^n 的基林向量，所以 S^n 是最大对称空间。实际上， S^n 是截面曲率为 $K = 1/L^2$ 的最大对称空间。

容易算得 S^n 的基林向量场满足如下李代数

$$[M_{AB}, M_{CD}] = \delta_{AD} M_{BC} + \delta_{BC} M_{AD} - \delta_{AC} M_{BD} - \delta_{BD} M_{AC}. \quad (12.6)$$

另一方面，很显然 n 维球面在 \mathbb{R}^{n+1} 的绕坐标原点的任意旋转之下均保持不变，而且欧氏度规 (12.1) 也在这个旋转之下保持不变。通常记这样的旋转的集合为 $SO(n+1)$ ，当然，它就是 S^n 的等度规群。进而即知，(12.6) 就是 $SO(n+1)$ 的李代数。

为了定义 n 维双曲空间 H^n ，我们可以引入 $n+1$ 维闵可夫斯基时空 $\mathbb{R}^{1,n}$ ，其坐标为 $z^A = (z^0, z^1, \dots, z^n)$ ，相应的度规张量为

$$ds^2 = -(dz^0)^2 + (dz^1)^2 + (dz^2)^2 + \dots + (dz^n)^2 = \eta_{AB} dz^A dz^B. \quad (12.7)$$

式中 $\eta_{AB} = \text{diag}\{-1, 1, \dots, 1\}$ 。 H^n 为 $\mathbb{R}^{1,n}$ 中的双曲面，由下式定义

$$H^n: -(z^0)^2 + (z^1)^2 + \dots + (z^n)^2 = -L^2 \Leftrightarrow \eta_{AB} z^A z^B = -L^2. \quad (12.8)$$

H^n 上的黎曼度规可以由 (12.7) 式的度规限制在超曲面 H^n 上而诱导得出。值得注意的是，虽然 $\mathbb{R}^{1,n}$ 是闵可夫斯基号差的，但 H^n 并不是， H^n 的度规其实是欧几里德号差的，即度规张量的本征值全为正。这是因为，在 H^n 上可以利用 (12.8) 式消去号差为负的变量 z^0 。

类似于 S^n 在 \mathbb{R}^{n+1} 的空间旋转之下保持不变， H^n 在闵可夫斯基时空 $\mathbb{R}^{1,n}$ 的洛伦兹变换下是保持不变的，换言之， $\mathbb{R}^{1,n}$ 的洛伦兹群也是 H^n 的等度规群，通常记作 $SO(1, n)$ (是四维闵可夫斯基时空的洛伦兹群 $SO(1, 3)$ 的推广)。同样类似于 S^n 的情况，不难验证， H^n 有如下 $n(n+1)/2$ 个线性独立基林向量场 (记 $z_A \equiv \eta_{AB} z^B$)

$$M_{AB} = z_A \partial_B - z_B \partial_A. \quad (12.9)$$

它们满足的李代数是

$$[M_{AB}, M_{CD}] = \eta_{AD} M_{BC} + \eta_{BC} M_{AD} - \eta_{AC} M_{BD} - \eta_{BD} M_{AC}. \quad (12.10)$$

因此， H^n 也为最大对称空间，实际上， H^n 的截面曲率为 $K = -1/L^2$ 。

德西特 (de Sitter) 空间和反德西特 (Anti-de Sitter) 空间

下面我们可以引入 n 维德西特空间 (dS_n) 和反德西特空间 AdS_n 了。

dS_n 其实是 S^n 的推广, 是将 S^n 某一个分量的号差由正改成负而得来 (因此 dS_n 是闵可夫斯基号差的)。因此, dS_n 可以嵌入 $n+1$ 维闵可夫斯基时空 $\mathbb{R}^{1,n}$ 中 (即把 \mathbb{R}^{n+1} 某一个分量的号差改成负的)。具体来说, 设 $\mathbb{R}^{1,n}$ 的坐标为 $z^A = (z^0, z^1, \dots, z^n)$, 相应的度规张量为

$$ds^2 = -(dz^0)^2 + (dz^1)^2 + (dz^2)^2 + \dots + (dz^n)^2 = \eta_{AB} dz^A dz^B. \quad (12.11)$$

则 dS_n 为 $\mathbb{R}^{1,n}$ 中的“球面”, 由下式定义

$$dS_n: -(z^0)^2 + (z^1)^2 + \dots + (z^n)^2 = L^2 \Leftrightarrow \eta_{AB} z^A z^B = L^2. \quad (12.12)$$

dS_n 上的度规可以由 (12.11) 式诱导得出。不难看出, dS_n 有基林向量场 $M_{AB} = z_A \partial_B - z_B \partial_A$ 。 dS_n 的等度规群为 $SO(1, n)$ 。作为 $\mathbb{R}^{1,n}$ 中的“球面”, dS_n 的截面曲率同样为 $K = 1/L^2$ 。

类似的, AdS_n 其实是 H^n 的推广, 是将 H^n 的 z^n 分量的号差由正改成负而得来 (因此 AdS_n 是闵可夫斯基号差的)。因此, AdS_n 可以嵌入 $n+1$ 维时空 $\mathbb{R}^{2, n-1}$ 中 (即把 $\mathbb{R}^{1, n}$ 中 z^n 分量的号差改成负的)。具体来说, 设 $\mathbb{R}^{2, n-1}$ 的坐标为 $z^A = (z^0, z^1, \dots, z^n)$, 相应的度规张量为

$$ds^2 = -(dz^0)^2 + (dz^1)^2 + \dots + (dz^{n-1})^2 - (dz^n)^2 = \eta_{AB} dz^A dz^B. \quad (12.13)$$

式中 $\eta_{AB} = \text{diag}\{-1, 1, \dots, 1, -1\}$ 。则 AdS_n 为 $\mathbb{R}^{2, n-1}$ 中的“双曲面”, 由下式定义

$$AdS_n: -(z^0)^2 + \sum_{i=1}^{n-1} (z^i)^2 - (z^n)^2 = -L^2 \Leftrightarrow \eta_{AB} z^A z^B = -L^2. \quad (12.14)$$

作为 $\mathbb{R}^{2, n-1}$ 中的“双曲面”, AdS_n 的截面曲率同样为 $K = -1/L^2$ 。

类似的, AdS_n 上有基林向量场 $M_{AB} = z_A \partial_B - z_B \partial_A$, AdS_n 的等度规群为 $\mathbb{R}^{2, n-1}$ 的“洛伦兹群”, 记为 $SO(2, n-1)$ 。值得一提的是, $SO(2, n-1)$ 同样是 $n-1$ 维闵可夫斯基时空 $\mathbb{R}^{1, n-2}$ 的共形变换群, 这其实也是 AdS_n/CFT_{n-1} 对应成立的一个必要条件。比方说, 对于 $n=5$ 的 AdS_5 情形, 其等度规群为 $SO(2, 4)$, 而 $SO(2, 4)$ 同样是 4 维闵可夫斯基时空 $\mathbb{R}^{1, 3}$ 的共形变换群, 这两者的等同是 AdS_5/CFT_4 对应之所以成立的保证之一。

德西特空间和反德西特空间作为常曲率空间

上一章中, 作为一种常曲率空间, 我们已经引入过 n 维德西特空间和 n 维反德西特空间了, 而且我们还给出了它们的一种度规, 也就是黎曼在其演讲《论几何学之基础假设》中给出的度规。为了看清楚上面对德西特空间和反德西特空间的定义其实与上一章一致, 下面我们根据本章的定义重新来推导上一章的那个度规。

先看德西特空间。为此, 我们需要作一个投影 (本质上是球极投影的一种推广), 即将 $n+1$ 维的坐标 $z^A = (z^0, z^1, \dots, z^n)$ 投影到 n 维的 $x^\mu = (x^0, x^1, \dots, x^{n-1})$, 这个投影的定义如下

$$z^\mu = \frac{1}{1 + \frac{x^2}{4L^2}} x^\mu, \quad \mu = 0, 1, \dots, n-1 \quad (12.15)$$

以及

$$z^n = L \left(\frac{1 - \frac{x^2}{4L^2}}{1 + \frac{x^2}{4L^2}} \right) \quad (12.16)$$

这里 $x^2 \equiv \eta_{\mu\nu} x^\mu x^\nu$, 其中 n 维的 $\eta_{\mu\nu}$ 为 $\eta_{\mu\nu} = \text{diag}\{-1, 1, \dots, 1\}$ 。

要看清上面定义的投影的确满足方程 (12.12), 只需注意到

$$\eta_{\mu\nu} z^\mu z^\nu = \frac{x^2}{(1 + \frac{x^2}{4L^2})^2} \Rightarrow \eta_{\mu\nu} z^\mu z^\nu + (z^n)^2 = L^2. \quad (12.17)$$

进而, 可以根据上面的投影定义直接验证 dS_n 上的度规为

$$ds^2 = \eta_{AB} dz^A dz^B = \eta_{\mu\nu} dz^\mu dz^\nu + (dz^n)^2 = \frac{\eta_{\mu\nu} dx^\mu dx^\nu}{(1 + \frac{x^2}{4L^2})^2}. \quad (12.18)$$

这正好是上一章给出的度规! 这也验证了, 德西特时空的截面曲率的确为 $K = 1/L^2$ 。

反德西特空间是类似的。不过, 这时候 $n+1$ 维坐标 $z^A = (z^0, z^1, \dots, z^n)$ 到 n 维坐标 $x^\mu = (x^0, x^1, \dots, x^{n-1})$ 的投影应该定义为

$$z^\mu = \frac{1}{1 - \frac{x^2}{4L^2}} x^\mu, \quad \mu = 0, 1, \dots, n-1 \quad (12.19)$$

以及

$$z^n = L \left(\frac{1 + \frac{x^2}{4L^2}}{1 - \frac{x^2}{4L^2}} \right). \quad (12.20)$$

不难验证, 这的确满足方程 (12.14)。进而可以算得 AdS_n 上的度规为

$$ds^2 = \eta_{AB} dz^A dz^B = \eta_{\mu\nu} dz^\mu dz^\nu - (dz^n)^2 = \frac{\eta_{\mu\nu} dx^\mu dx^\nu}{(1 - \frac{x^2}{4L^2})^2}. \quad (12.21)$$

这正好是上一章给出的度规! 这也验证了, 反德西特时空的截面曲率的确为 $K = -1/L^2$ 。

12.2 德西特空间的一些常用坐标以及彭罗斯图

下面我们介绍 dS_n 的一些常用坐标, 并讨论 dS_n 的彭罗斯图。

12.2.1 整体坐标与彭罗斯图

整体坐标

先从 dS_n 的所谓整体坐标入手。为了引入整体坐标, 让我们先观察定义德西特超曲面的方程 (12.12), 重写如下

$$dS_n: \quad -(z^0)^2 + (z^1)^2 + \dots + (z^n)^2 = L^2. \quad (12.22)$$

从这个方程可以看出 dS_n 其实是一个拓扑为 $\mathbb{R} \times S^{n-1}$ 的双曲面, 其中 \mathbb{R} 为时间方向, 由坐标 z^0 参数化。给定时间 z^0 , 相应的空间 $n-1$ 维球面 S^{n-1} 由下式给出

$$(z^1)^2 + \dots + (z^n)^2 = L^2 + (z^0)^2 = \text{常数} > 0. \quad (12.23)$$

超曲面方程 (12.22) 显然可以通过如下参数化方式解出来, 即定义

$$z^0 = L \sinh \tau, \quad z^a = L n^a \cosh \tau, \quad (12.24)$$

式中 $a = 1, 2, \dots, n$, n^a 是定义 $n-1$ 维单位球面的 n 维单位向量, n^a 满足

$$\delta_{ab} n^a n^b = 1, \quad \delta_{ab} n^a dn^b = 0, \quad \delta_{ab} dn^a dn^b = d\Omega_{n-1}^2, \quad (12.25)$$

式中 $d\Omega_{n-1}^2$ 表示 $n-1$ 维单位球面上的标准度规。进而不难得到, dS_n 的度规可以表示为

$$\begin{aligned} ds^2 &= [-(dz^0)^2 + \sum_{a=1}^n (dz^a)^2] |_{dS_n} \\ &= L^2 [-d\tau^2 + \cosh^2 \tau d\Omega_{n-1}^2]. \end{aligned} \quad (12.26)$$

在上述整体坐标中, 明显的对称性是 S^{n-1} 的旋转对称性, 也即是完整的等度规群 $SO(1, n)$ 中的 $SO(n)$ 子群。

与 (12.26) 式相应的, 欧几里德号差的最大对称空间 S^n 上的相应度规可以由 (12.26) 式通过将时间 τ “维克旋转” 到虚时间而得到, 即

$$\tau \equiv i\theta \Rightarrow -d\tau^2 + \cosh^2 \tau d\Omega_{n-1}^2 \rightarrow d\theta^2 + \cos^2 \theta d\Omega_{n-1}^2 = d\Omega_n^2. \quad (12.27)$$

这告诉我们闵氏号差的 dS_n 和欧氏号差的 S^n 其实只相差一个“维克旋转”。

彭罗斯图和视界

为了讨论 dS_n 的彭罗斯图, 我们引入共形时间 η , 其定义为

$$d\eta = \frac{d\tau}{\cosh \tau}, \quad (12.28)$$

不难验证这相当于

$$\cos \eta = \frac{1}{\cosh \tau}, \quad \tau \in (-\infty, +\infty) \Rightarrow \eta \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right). \quad (12.29)$$

引入共形时间后, 即可将度规 (12.26) 式重写成

$$\begin{aligned} ds^2 &= L^2 \cosh^2 \tau \left(-\frac{d\tau^2}{\cosh^2 \tau} + d\Omega_{n-1}^2 \right) \\ &= \frac{L^2}{\cos^2 \eta} (-d\eta^2 + d\Omega_{n-1}^2). \end{aligned} \quad (12.30)$$

因此, 如果我们取共形因子 $\Omega = \frac{1}{L} \cos \eta$, 则由 (12.30) 式可见, dS_n 外尔等价于具有如下度规的所谓爱因斯坦静态宇宙 $(-\frac{\pi}{2}, \frac{\pi}{2}) \times S^{n-1}$

$$d\tilde{s}^2 = \Omega^2 ds^2 = -d\eta^2 + d\Omega_{n-1}^2. \quad (12.31)$$

由共形紧化的定义, dS_n 的无穷远对应 $\Omega = 0$ 的地方, 也就是对应

$$\eta = \pm \frac{\pi}{2}. \quad (12.32)$$

通常称 $\eta = \frac{\pi}{2}$ 为未来无穷, 记作 \mathcal{I}^+, i^+ , 因为在现在的情形中类时未来无穷和类光未来无穷合并在一起了; 称 $\eta = -\frac{\pi}{2}$ 为过去无穷, 记作 \mathcal{I}^-, i^- , 因为类时过去无穷和类光过去无穷合并了。

为了画出 dS_n 的彭罗斯图, 通常进一步将空间部分的 $d\Omega_{n-1}^2$ 写成 $d\Omega_{n-1}^2 = d\chi^2 + \sin^2 \chi d\Omega_{n-2}^2$, 这里 $\chi \in [0, \pi]$, $\chi = 0$ 对应 S^{n-1} 的北极, $\chi = \pi$ 对应 S^{n-1} 的南极。从而 dS_n 外尔等价于

$$d\tilde{s}^2 = -d\eta^2 + d\chi^2 + \sin^2 \chi d\Omega_{n-2}^2. \quad (12.33)$$

因此, 我们可以在 (η, χ) 平面上画出 dS_n 的彭罗斯图, 如图 (12.1)(图中每一个点代表一个 $n-2$ 维球面)。

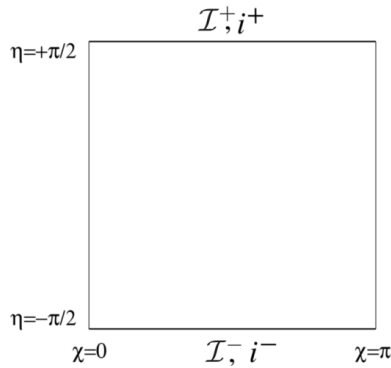


Figure 12.1: dS_n 的彭罗斯图, 图中每一个点代表一个 $n-2$ 维球面。

注意, 彭罗斯图中左侧的直线 $\chi = 0$ 和右侧的直线 $\chi = \pi$ 并非时空的边界, 而只是空间部分 S^{n-1} 的北极和南极。

现在, 考虑一个位于北极 $\chi = 0$ 的观察者, 这个观察者将有一个事件视界 (event horizon), 它也即是此观察者能够从中接收到信号的时空区域的边界。同时, 北极观察者还有一个粒子视界 (particle horizon), 它也即是此观察者能够通过发射信号影响的区域的边界。如图 (12.2) 所示。北极观察者能够接收信号的区域和能够影响的区域的交集称作北极的因果菱形 (causal diamond), 即图 (12.2) 中右侧的图的红色部分。

不难看出, 北极观察者的事件视界和粒子视界分别是彭罗斯图上的两条对角线, 它们满足方程

$$\chi = \pm\left(\eta - \frac{\pi}{2}\right). \quad (12.34)$$

12.2.2 静态坐标以及其它

在 dS_n 的整体坐标中, 度规场显然是依赖于时间 τ 的, 然而, dS_n 上还可以引入一组适当的局部坐标 (而不是整体坐标), 其中度规场不依赖于时间, 甚至是静态的, 这就是所谓的静态坐标。

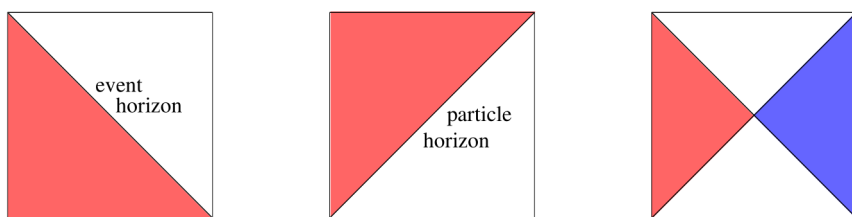


Figure 12.2: 左边的图表示北极观察者能够看到的时空区域，其边界即就是这个观察者的事件视界。中间的图表示北极观察者能够通过发射信号影响的时空区域，其边界即就是这个观察者的粒子视界。右边这幅图的红色部分表示北极观察者的因果菱形，蓝色部分表示南极观察者的因果菱形。

为了引入静态坐标，我们将 (12.22) 式重写为

$$\sum_{k=1}^{n-1} (z^k)^2 = (z^1)^2 + \dots + (z^{n-1})^2 = L^2 + (z^0)^2 - (z^n)^2. \quad (12.35)$$

然后，通过下式定义空间径向坐标 r

$$r^2 \equiv \sum_{k=1}^{n-1} (z^k)^2 \Rightarrow (z^n)^2 - (z^0)^2 = L^2 \left(1 - \frac{r^2}{L^2}\right). \quad (12.36)$$

假如 $r \leq L$ ，则根据上式可以自然地定义

$$\begin{aligned} z^k &= rn^k, \quad (k = 1, 2, \dots, n-1) \\ z^0 &= L \left(1 - \frac{r^2}{L^2}\right)^{\frac{1}{2}} \sinh\left(\frac{t}{L}\right), \quad z^n = L \left(1 - \frac{r^2}{L^2}\right)^{\frac{1}{2}} \cosh\left(\frac{t}{L}\right), \end{aligned} \quad (12.37)$$

式中 n^k 为 $n-1$ 维单位向量，它定义了一个 $n-2$ 维单位球面。根据上面的定义，即可以得到 dS_n 的度规为

$$\begin{aligned} ds^2 &= -(dz^0)^2 + (dz^n)^2 + \sum_{k=1}^{n-1} (dz^k)^2 \\ &= -\left(1 - \frac{r^2}{L^2}\right) dt^2 + \left(1 - \frac{r^2}{L^2}\right)^{-1} dr^2 + r^2 d\Omega_{n-2}^2. \end{aligned} \quad (12.38)$$

这就是 dS_n 的静态度规，它也正是第 10 章中解出来的德西特度规。

显然， $r = L$ 是一个德西特视界，同时也是基林向量场 $\xi = \partial_t$ 的基林视界。根据 (12.37) 式，也有

$$\begin{aligned} \xi = \partial_t &= \frac{\partial z^A}{\partial t} \partial_A = \frac{1}{L} (z^n \partial_0 + z^0 \partial_n) \\ &= \frac{1}{L} (z_n \partial_0 - z_0 \partial_n) = \frac{1}{L} M_{n0}. \end{aligned} \quad (12.39)$$

所以， ξ 正比于 M_{n0} ，是 dS_n 的基林向量场之一。

为了看清楚静态坐标所覆盖的区域, 我们由 (12.37) 式, 有

$$z^n + z^0 = L\left(1 - \frac{r^2}{L^2}\right)^{\frac{1}{2}} e^{t/L} \geq 0, \quad z^n - z^0 = L\left(1 - \frac{r^2}{L^2}\right)^{\frac{1}{2}} e^{-t/L} \geq 0. \quad (12.40)$$

所以静态坐标仅仅覆盖 $z^n \geq |z^0|$ 的四分之一时空。 $r = L$ 的德西特视界就是这四分之一时空的边界。

下面考察 $r \rightarrow L$ 时, 坐标如何趋于德西特视界。为此令 $r = L(1 - \varepsilon^2/2)$ (ε 为无穷小量), 代入 (12.40) 式, 即有

$$z^n + z^0 = \varepsilon e^{t/L}, \quad z^n - z^0 = \varepsilon e^{-t/L}. \quad (12.41)$$

当 $\varepsilon \rightarrow 0$ 时, 如果同时令 $\varepsilon e^{t/L} \rightarrow$ 有限(即同时取 $t \rightarrow \infty$), 则显然有 $\varepsilon e^{-t/L} \rightarrow 0$, 这时的德西特视界为

$$z^n = z^0, \quad z^n + z^0 = \text{有限}. \quad (12.42)$$

反过来, 若当 $\varepsilon \rightarrow 0$ 时, 同时令 $\varepsilon e^{-t/L} \rightarrow$ 有限(即同时取 $t \rightarrow -\infty$), 则必然有 $\varepsilon e^{t/L} \rightarrow 0$, 这时的德西特视界为

$$z^n = -z^0, \quad z^n - z^0 = \text{有限}. \quad (12.43)$$

总之, 静态坐标的德西特视界必然满足 $z^n = \pm z^0$ 。

翻译到整体坐标中, 根据整体坐标的定义 (12.24), 这时候有

$$z^0 = L \sinh \tau, \quad z^n = L \cosh \tau \cos \chi, \quad (12.44)$$

式中 χ 为 $n-1$ 维单位球面上的极角。因此 $z^n = \pm z^0$ 相当于

$$\begin{aligned} \cosh \tau \cos \chi = \pm \sinh \tau &\Rightarrow \cos \chi = \pm \frac{\sinh \tau}{\cosh \tau} = \pm \sin \eta \\ &\Rightarrow \chi = \pm \left(\eta - \frac{\pi}{2}\right). \end{aligned} \quad (12.45)$$

这正好是上一小节中北极的事件视界和粒子视界所满足的方程! 所以, 北极的事件视界和粒子视界正好就是静态坐标的德西特视界。

实际上, 容易进一步验证, 静态坐标中的 $r = 0$ 就相应于整体坐标中的北极 $\chi = 0$, 进而沿着 $r = 0$ 的静止世界线有 $t/L = \tau$, 这也正好就是上一小节中北极观察者的世界线。综上所述可知, 静态坐标其实仅仅覆盖了北极的因果菱形! 它正好是德西特彭罗斯图的四分之一。

庞加莱坐标

还有一种坐标也很常用, 即所谓的庞加莱坐标。为了定义庞加莱坐标, 我们引入空间坐标 $\mathbf{x} = (x^1, x^2, \dots, x^{n-1})$, 并定义

$$\begin{aligned} z^n - z^0 &= L e^{-t} \geq 0, \quad z^n + z^0 = L(e^t - \mathbf{x}^2 e^{-t}) \\ z^k &= L e^{-t} x^k, \quad (k = 1, 2, \dots, n-1). \end{aligned} \quad (12.46)$$

注意到 $(z^n)^2 - (z^0)^2 = (z^n - z^0)(z^n + z^0) = L^2 - L^2 \mathbf{x}^2 e^{-2t}$, 由此不难验证上面的定义满足 (12.22) 式。进而可以求出相应的 dS_n 度规, 为

$$ds^2 = L^2(-dt^2 + e^{-2t} d\mathbf{x}^2). \quad (12.47)$$

由于有 $z^n - z^0 \geq 0$, 所以上面的坐标仅仅覆盖了整个 dS_n 的一半。但是上面的坐标还不是庞加莱坐标, 为了进一步引入庞加莱坐标, 我们将 (12.47) 重写为

$$ds^2 = L^2 e^{-2t}(-e^{2t} dt^2 + d\mathbf{x}^2). \quad (12.48)$$

进而只要定义 $\tau \equiv e^t \geq 0$, 则有

$$ds^2 = L^2 \frac{1}{\tau^2}(-d\tau^2 + d\mathbf{x}^2). \quad (12.49)$$

最终的这个度规就是 dS_n 的庞加莱度规, 相应的坐标即是庞加莱坐标。很显然, 在庞加莱坐标中, $\tau = 0$ 是时空的共形边界, 相应于过去无穷 \mathcal{I}^-, i^- , 未来无穷所在的那一半区域没有被庞加莱坐标覆盖。

12.3 反德西特空间的一些常用坐标

AdS_n 空间的坐标可以类似于前面 dS_n 情形处理。为此我们从方程 (12.14) 开始, 将之重写如下

$$AdS_n: -(z^0)^2 + \sum_{k=1}^{n-1} (z^k)^2 - (z^n)^2 = -L^2. \quad (12.50)$$

稍微调整一下, 也可以写成

$$(z^0)^2 + (z^n)^2 = \sum_{k=1}^{n-1} (z^k)^2 + L^2. \quad (12.51)$$

不难看出这个空间具有拓扑 $S^1 \times \mathbb{R}^{n-1}$, 因为对于 \mathbb{R}^{n-1} 上给定 z^k 的点, 上面方程描述了 (z^0, z^n) 平面中的一个圆周 S^1 。由于度规在 (z^0, z^n) 平面是负定的, 这表明这个 S^1 是闭合类时曲线。具有闭合类时曲线的时空, 其因果性是有问题的。为了避免这个问题, 我们可以取上述时空的万有覆盖空间, 也就是将时间圆周 S^1 替换成 \mathbb{R} , 即 $S^1 \rightarrow \mathbb{R}$ 。实际上, 我们通常所谓的 AdS 时空都是指这个覆盖空间。

整体坐标

为了定义 AdS_n 的整体坐标, 我们取 (12.51) 式的如下解

$$(z^0)^2 + (z^n)^2 = L^2 \cosh^2 \rho, \quad \sum_{k=1}^{n-1} (z^k)^2 = L^2 \sinh^2 \rho. \quad (12.52)$$

进一步, 可以取如下参数化

$$z^0 = L \cosh \rho \sin \tau, \quad z^n = L \cosh \rho \cos \tau, \quad z^k = L n^k \sinh \rho, \quad (12.53)$$

式中 n^k 为 $(n-1)$ 维单位矢量, 它定义了一个 $n-2$ 维单位球面。根据上面的参数化不难得到如下 AdS_n 的度规

$$ds^2 = L^2 [-\cosh^2 \rho d\tau^2 + d\rho^2 + \sinh^2 \rho d\Omega_{n-2}^2]. \quad (12.54)$$

这就是所谓的 AdS_n 的整体坐标, 其中 ρ 的取值范围是 $[0, +\infty)$, τ 的取值范围本来是 $[0, 2\pi]$, 忽略 2π 周期等同进而过渡到万有覆盖空间以后就变成了 $\tau \in (-\infty, +\infty)$ 。

假设定义 $r \equiv L \sinh \rho$, 并定义 $t \equiv L\tau$, 则度规 (12.54) 就变成了

$$ds^2 = -(1 + \frac{r^2}{L^2})dt^2 + (1 + \frac{r^2}{L^2})^{-1}dr^2 + r^2 d\Omega_{n-2}^2. \quad (12.55)$$

这正是我们在第十章中得出爱因斯坦场方程的 AdS 解时解出来的度规, 它是静态 dS 度规的对应物。

彭罗斯图以及共形边界

我们可以将 AdS_n 的整体度规重写作

$$ds^2 = L^2 \cosh^2 \rho [-d\tau^2 + \frac{d\rho^2}{\cosh^2 \rho} + \tanh^2 \rho d\Omega_{n-2}^2]. \quad (12.56)$$

通过下式引入一个新的径向坐标 ψ

$$d\psi = \frac{d\rho}{\cosh \rho}. \quad (12.57)$$

这个方程的解是

$$\cos \psi = \frac{1}{\cosh \rho}. \quad (12.58)$$

由于 $\rho \in [0, +\infty)$, 所以

$$\cosh \rho \in [1, +\infty) \Rightarrow \psi \in [0, \frac{\pi}{2}). \quad (12.59)$$

由

$$\cosh \rho = \frac{1}{\cos \psi} \Rightarrow \sinh \rho = \tan \psi, \tanh \rho = \sin \psi, \quad (12.60)$$

从而可以将度规变成

$$\begin{aligned} ds^2 &= \frac{L^2}{\cos^2 \psi} [-d\tau^2 + d\psi^2 + \sin^2 \psi d\Omega_{n-2}^2] \\ &= \frac{L^2}{\cos^2 \psi} [-d\tau^2 + d\Omega_{n-1}^2]. \end{aligned} \quad (12.61)$$

从上面的结果可以看出, AdS_n 时空外尔等价于

$$d\tilde{s}^2 = -d\tau^2 + d\Omega_{n-1}^2, \quad 0 \leq \psi < \frac{\pi}{2}. \quad (12.62)$$

也即是说, AdS_n 外尔等价于爱因斯坦静态宇宙 (它有 $0 \leq \psi \leq \pi$) 的一半。给定时间 τ , 空间超曲面是一个 $n-1$ 维的半球, 边界在 $\psi = \frac{\pi}{2}$ 处 (它显然不在时空本身的定义域内), 这个空间边界当然是一个 $n-2$ 维球面 S^{n-2} 。 $\psi = \frac{\pi}{2}$ 也是共形因子 $\Omega = \cos \psi / L = 0$ 的地方, 因此对应 AdS_n 的共形边界, 记作 \mathcal{I} 。

从 (12.61) 式可以看出, 这个共形边界 \mathcal{I} 是类时的, 具有度规

$$d\tilde{s}^2|_{\psi=\pi/2} = -d\tau^2 + d\Omega_{n-2}^2. \quad (12.63)$$

所以 \mathcal{I} 的拓扑是

$$\mathcal{I} \simeq \mathbb{R} \times S^{n-2}. \quad (12.64)$$

它可以看作是对 $n-1$ 维闵可夫斯基时空 $\mathbb{R}^{1,n-2}$ 的空间部分作了一个共形紧化

$$\mathbb{R}^{1,n-2} \rightarrow \mathbb{R} \times S^{n-2}. \quad (12.65)$$

\mathcal{I} 其实是未来类光无穷、过去类光无穷、以及类空无穷结合起来的产物。

可以在 (τ, ψ) 平面上画出 AdS_n 的彭罗斯图, 如图 (12.3) 左图。图中与竖直方向成 45

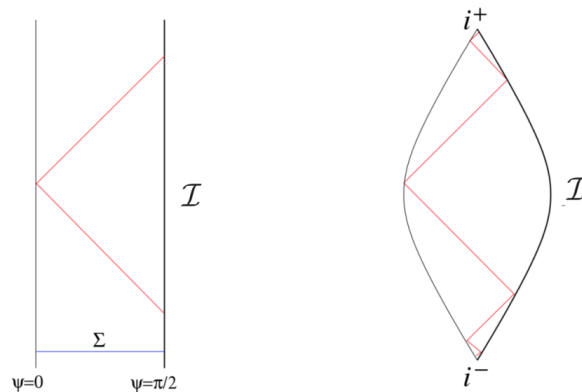


Figure 12.3: 反德西特时空的彭罗斯图, 图中每一个点代表一个 $n-2$ 维球面。

度角的那些直线代表类光测地线。由于 τ 无限延伸, 而 ψ 有限, 所以很难在保持光线沿着 45 度角方向的前提下进一步将整幅彭罗斯图压缩到一个有限区域内, 因为一旦我们将 τ 压缩到有限了, 那就一定会同时把整个 ψ 轴压缩成一个点, 最后的结果将如图 (12.3) 右图。我们通常不会这么做, 通常我们就将图 (12.3) 的左图当作 AdS_n 的彭罗斯图, 虽然它的 τ 方向向两端无限延伸。注意, 图 (12.3) 中左边 $\psi = 0$ 的那条直线并不是 AdS_n 的边界, 而只是 $\psi = 0$ 的径向坐标的坐标原点。

从彭罗斯图中可以清楚地看到, 光线可以在有限的坐标时内抵达共形边界 \mathcal{I} 。这意味着, 取 AdS_n 的任意一个空间截片 Σ (如图 (12.3) 所示), 在 Σ 的未来一定存在一些点, 它的某一些过去指向的因果线 (类光测地线) 在过去并不会与 Σ 相交, 而是跑到共形边界 \mathcal{I} 上去了。这意味着, Σ 上的初始数据并不足以决定场位形在这些未来点处的值, 换言之,

AdS_n 不是整体双曲时空! (关于整体双曲的定义, 请参见第九章在介绍强宇宙监督猜测时的讲述。)

为了决定场位形在整个 AdS 时空的演化, 我们不仅需要初始数据, 还需要在 \mathcal{S} 上指定合适的边界条件, 比方说让 \mathcal{S} 完美地反射光线等等。由于需要在共形边界 \mathcal{S} 上给定边界条件, 这也就意味着, 可以有一些场本身就生存在 \mathcal{S} 上。特别的, 可以在 \mathcal{S} 上定义共形场论。这些事实, 结合 AdS_5 的等度规群 (即 $SO(2,4)$) 与边界四维闵氏时空的共形群相同的事实, 对于 AdS/CFT 对应的引入起到了重要作用。 AdS/CFT 对应说的就是 AdS 时空的量子引力理论等价于共形边界 \mathcal{S} 上的一个没有引力的共形场论。

庞加莱坐标

和 dS 时空一样, AdS 时空也有一组庞加莱坐标。为了定义它, 我们注意到下式可以满足定义了 AdS 时空的 (12.50) 式

$$\begin{aligned} z^\alpha &= Lr x^\alpha, & z^n - z^{n-1} &= Lr \\ z^n + z^{n-1} &= L(r^{-1} + r\eta_{\alpha\beta}x^\alpha x^\beta). \end{aligned} \quad (12.66)$$

式中 $\alpha, \beta = 0, 1, \dots, n-2$ 。利用上式, 可以将 AdS 的度规改写成

$$\begin{aligned} ds^2 &= L^2 \left(\frac{dr^2}{r^2} + r^2 \eta_{\alpha\beta} dx^\alpha dx^\beta \right) \\ &= L^2 (d\rho^2 + e^{2\rho} \eta_{\alpha\beta} dx^\alpha dx^\beta) \quad (r = e^\rho) \\ &= L^2 \frac{1}{z^2} (\eta_{\alpha\beta} dx^\alpha dx^\beta + dz^2) \quad (r = z^{-1}). \end{aligned} \quad (12.67)$$

最后一行的度规就是庞加莱度规, 相应的坐标就是庞加莱坐标。

$z=0$ 就对应 AdS 的共形边界, 很显然, 共形边界具有标准的闵氏度规

$$ds^2|_{z=0} = \eta_{\alpha\beta} dx^\alpha dx^\beta. \quad (12.68)$$

这个结果表面看起来与 (12.63) 式有点出入, 因为 (12.68) 式并不与 (12.63) 式微分同胚等价, 但是两者共形等价, 也就是说, 在合适的微分同胚变换下, 两者只相差一个外尔共形因子。实际上, 共形紧化手续并不能唯一决定共形边界上的度规, 因为共形紧化时所乘的共形因子 Ω 并不是唯一的, 这导致共形紧化只能将边界上的度规确定到一个共形等价类, 而 (12.63) 式与 (12.68) 式正是属于同一个共形等价类, 所以两者并不矛盾。同时, 这也是为什么共形边界上只能定义共形场论的原因, 因为只有共形场论才不依赖于确定的背景度规, 而只依赖于背景度规的一个共形等价类。

13. 克尔黑洞，以及黑洞热力学定律

自施瓦西解发现以后，直到 1963 年，克尔才发现真空爱因斯坦场方程的另一个重要性可与施瓦西解相提并论的精确解，也就是著名的克尔解，或者说克尔黑洞，因为它描写一个旋转的黑洞。克尔黑洞的具体求解过程比施瓦西解复杂太多，因此下文中我们将略过这个求解过程，直接给出最后的克尔度规，并从这个度规出发对克尔黑洞的一些物理性质进行讨论。

除了关于克尔黑洞的相关讨论以外，本章的另一个重点是对所谓的黑洞热力学定律进行初步探讨。这个探讨与对克尔黑洞的一些讨论密切相关，因此我们将这两个主题放在同一章节中。

13.1 克尔黑洞的度规与视界

13.1.1 克尔黑洞的度规

在所谓的 Boyer-Lindquist 坐标下，克尔黑洞的度规可以写为

$$ds^2 = -\frac{\Delta}{\rho^2}(dt - a\sin^2\theta d\phi)^2 + \frac{\sin^2\theta}{\rho^2}[(r^2 + a^2)d\phi - a dt]^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2. \quad (13.1)$$

式中

$$\rho^2 = r^2 + a^2 \cos^2\theta, \quad \Delta = r^2 - 2GMr + a^2. \quad (13.2)$$

通常也将上面的度规展开, 写成更便于直接读出各度规分量的形式

$$\begin{aligned} ds^2 &= -\left(1 - \frac{2GMr}{\rho^2}\right)dt^2 + \frac{\Sigma}{\rho^2} \sin^2 \theta d\phi^2 \\ &\quad - \frac{4GMar}{\rho^2} \sin^2 \theta dt d\phi + \frac{\rho^2}{\Delta} dr^2 + \rho^2 d\theta^2 \\ &= g_{tt} dt^2 + g_{\phi\phi} d\phi^2 + 2g_{t\phi} dt d\phi + g_{rr} dr^2 + g_{\theta\theta} d\theta^2. \end{aligned} \quad (13.3)$$

式中

$$\Sigma = (r^2 + a^2)^2 - a^2 \Delta \sin^2 \theta. \quad (13.4)$$

其中 a 称作克尔参数, 不难看出, 当 $a=0$ 时, 克尔解退回到球对称的施瓦西解。另外, 显然可以从式中读出 $g_{tt}, g_{\phi\phi}, g_{t\phi}, g_{rr}, g_{\theta\theta}$ 。而且由于 $g_{t\phi} \neq 0$, 所以即使在 r 很大处, 克尔解也不是一个静态解, 而只是一个稳态解。

由于各度规分量都不依赖于 t , 所以这个解显然具有时间平移不变性, 相应的基林向量场可以记为 $\xi = \partial_t$ 。但是, 由于克尔参数 a 的引入, 这个解不再具有空间旋转不变性, 而仅仅只具有沿着坐标 ϕ 的旋转不变性, 即在 $\phi \rightarrow \phi + c$ (c 为常数) 的变换下解保持不变, 相应的基林向量场可以记作 $\eta = \partial_\phi$, 换言之, 克尔解只具有轴对称性。一般地, 克尔解的任意基林向量场可以写作 ξ 和 η 的线性组合, 即

$$K = a\xi + b\eta. \quad (13.5)$$

假设按照 (t, ϕ, r, θ) 的顺序排列坐标分量, 则易见克尔解的度规张量具有如下形式

$$g_{\mu\nu} = \begin{pmatrix} g_{tt} & g_{t\phi} & 0 & 0 \\ g_{t\phi} & g_{\phi\phi} & 0 & 0 \\ 0 & 0 & g_{rr} & 0 \\ 0 & 0 & 0 & g_{\theta\theta} \end{pmatrix}. \quad (13.6)$$

通过一些代数运算, 可以得到如下有用结果

$$D \equiv -\det \begin{pmatrix} g_{tt} & g_{t\phi} \\ g_{t\phi} & g_{\phi\phi} \end{pmatrix} = g_{t\phi}^2 - g_{tt}g_{\phi\phi} = \Delta \sin^2 \theta. \quad (13.7)$$

度规张量的逆矩阵 $g^{\mu\nu}$ 由下式给出

$$g^{\mu\nu} = \begin{pmatrix} -D^{-1}g_{\phi\phi} & D^{-1}g_{t\phi} & 0 & 0 \\ D^{-1}g_{t\phi} & -D^{-1}g_{tt} & 0 & 0 \\ 0 & 0 & 1/g_{rr} & 0 \\ 0 & 0 & 0 & 1/g_{\theta\theta} \end{pmatrix}, \quad (13.8)$$

所以, $g^{rr} = 1/g_{rr} = \Delta/\rho^2$, $g^{\theta\theta} = 1/g_{\theta\theta} = 1/\rho^2$ 。值得注意的是, 当 $D=0$, $g^{\mu\nu}$ 不存在, 从而 $D=0$ 对应某种坐标奇异性。

我们可以将克尔度规 (13.3) 重写作,

$$\begin{aligned} ds^2 &= \left(g_{tt} - \frac{g_{t\phi}^2}{g_{\phi\phi}}\right) dt^2 + g_{\phi\phi} (d\phi - \omega dt)^2 + g_{rr} dr^2 + g_{\theta\theta} d\theta^2 \\ &= -\frac{\rho^2 \Delta}{\Sigma} dt^2 + \frac{\Sigma}{\rho^2} \sin^2 \theta (d\phi - \omega dt)^2 + \frac{\rho^2}{\Delta} dr^2 + \rho^2 d\theta^2. \end{aligned} \quad (13.9)$$

式中

$$\omega \equiv -\frac{g_{t\phi}}{g_{\phi\phi}} = \frac{2GMa r}{\Sigma}. \quad (13.10)$$

正如我们即将看到的, 这实际上描述的是一个绕轴旋转的时空。

ω 的物理含义

由于克尔时空有基林向量场 η , 进而根据第十一章第二节的知识可知, 对于一个协变速度为 $u^\mu = \left(\frac{dt}{d\tau}, \frac{d\phi}{d\tau}, \frac{dr}{d\tau}, \frac{d\theta}{d\tau}\right)$ (τ 为固有时) 的粒子, 其单位质量的角动量 l 为守恒量, 这里 l 为

$$l = \eta \cdot u = g_{t\phi} \frac{dt}{d\tau} + g_{\phi\phi} \frac{d\phi}{d\tau}. \quad (13.11)$$

设想从空间无穷远处丢一个粒子到克尔时空里, 其协变速度为 u , 则它的 l 是守恒的。另外, 对于克尔解 (13.3), 在空间无穷远处 $g_{t\phi} \rightarrow 0$, $g_{\phi\phi} \rightarrow r^2 \sin^2 \theta$ 。设想这个粒子在空间无穷远处是静止的, 即 $\frac{d\phi}{d\tau} \rightarrow 0$, 则很显然, 在空间无穷远处这个粒子的角动量 $l = 0$ 。由于 l 守恒, 所以在粒子的下落过程中, 将保持 $l = 0$, 这是一个**零角动量**粒子。进而根据 (13.11) 式, 在下落过程中这个粒子将获得一个位置依赖的角速度 $\omega(r, \theta)$ (即使它的角动量始终为零!),

$$\omega(r, \theta) \equiv \frac{d\phi}{dt} = \frac{d\phi}{d\tau} / \frac{dt}{d\tau} = -\frac{g_{t\phi}}{g_{\phi\phi}}. \quad (13.12)$$

很显然, 这个 $\omega(r, \theta)$ 正是前面 (13.10) 式定义的 ω 。

对粒子在克尔时空中获得零角动量的角速度现象的解释是, 克尔时空本身在旋转 (如图 (13.1) 所示), 这种时空旋转使得其中的粒子跟着旋转! 通常也称这种效应为参考系拖拽效应, 指在广义相对论中, 一个旋转的物体 (比如旋转的黑洞) 会拖拽着周围的时空一起旋转。

13.1.2 克尔黑洞的视界和无限红移面

根据第十一章第四节关于引力红移的讨论, 对于稳态时空中的一个稳态观察者而言, 满足 $\langle \xi, \xi \rangle = 0$ 的面, 就是这个观察者的无限红移面。对于克尔解, 由于 $\xi = \partial_t$, 所以这个条件相当于 $g_{tt} = 0$, 代入克尔度规, 即有

$$1 - \frac{2GMr}{\rho^2} = 0 \Leftrightarrow r^2 - 2GMr + a^2 \cos^2 \theta = 0. \quad (13.13)$$

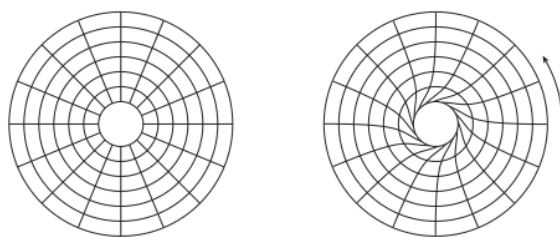


Figure 13.1: 旋转的物体带动周围时空一起旋转。

解得两个根, 为

$$r_{S\pm} = GM \pm \sqrt{(GM)^2 - a^2 \cos^2 \theta}. \quad (13.14)$$

但是, 与施瓦西黑洞不同, 克尔黑洞关于稳态观察者的无限红移面, 并不是它的未来事件视界!

为了考察克尔黑洞的内外视界, 我们按照第九章第一节的讨论, 考察其单向膜开始和结束的地方。设单向膜由超曲面 $f(r, \theta) = 0$ 给出, 则单向膜开始和结束的地方满足的方程是

$$g^{\mu\nu} \partial_\mu f \partial_\nu f = 0. \quad (13.15)$$

按照假设, 这也即是

$$\begin{aligned} g^{rr} (\partial_r f)^2 + g^{\theta\theta} (\partial_\theta f)^2 &= 0 \\ \Rightarrow \frac{\Delta}{\rho^2} (\partial_r f)^2 + \frac{1}{\rho^2} (\partial_\theta f)^2 &= 0. \end{aligned} \quad (13.16)$$

由于 $\partial_r f$ 与 $\partial_\theta f$ 不同时为零, 不难看出这个方程的解为

$$\partial_\theta f = 0, \quad \Delta = 0. \quad (13.17)$$

由前一个方程得 $f(r, \theta)$ 实际上与 θ 无关, 而由 $\Delta = 0$ 的两个根可以得到内外视界 r_\pm ,

$$r_\pm = GM \pm \sqrt{(GM)^2 - a^2}, \quad (13.18)$$

的确与 θ 无关! 注意, $r_+ \leq r_{S+}$ 。

值得注意的是, 虽然 r_\pm 与 θ 无关, 但这并不意味着给定时刻的视界面是一个球面, 因为克尔度规在 $r = r_\pm$ 的超曲面上诱导的度规并不是球面度规! 当然, 由于视界面处 $D = \Delta \sin^2 \theta = 0$, 所以它对应 Boyer-Lindquist 坐标的一种奇异性, 这种奇异性不是时空本身的奇性, 而是坐标奇性, 可以通过选取新的坐标来去除。视界面的特殊性不在于这种坐标奇性, 而在于从时空整体的因果结构而言, 视界面是特殊的, 类似于施瓦西时空的视界。稍后我们会进一步讨论克尔时空的整体因果结构。

从事件视界 r_{\pm} 的表达式 (13.18) 也可以看出, 符合物理要求的克尔解必须满足如下能限

$$GM \geq a. \quad (13.19)$$

因为如果 $a > GM$, 那克尔解将不存在事件视界, 进而出现裸奇性 (正如后文将要进一步讲述的, $\rho^2 = 0$ 对应时空奇性)。

克尔黑洞的视界和无限红移面如图 (13.2) 所示, 其中位于 r_{S+} 与 r_+ 之间的阴影区域就是所谓的能层, 至于为什么称之为能层, 我们稍后再讨论。值得注意的是, 在能层区, 人们还是有可能逃离黑洞的, 因为单向膜还没开始。

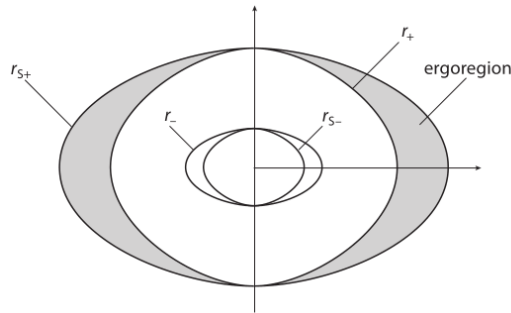


Figure 13.2: 克尔黑洞的视界和无限红移面, r_{S+} 和 r_+ 之间的阴影区就是所谓的能层。注意, 在能层区, 人们还是有可能逃离黑洞的, 因为单向膜还没开始。

13.1.3 稳态观察者

在一个稳态时空中 (比如克尔解在空间足够远处), 所谓的稳态观察者指的即是其世界线 (通常并不是测地线) 与类时基林向量场 ξ 的积分曲线相重合的那些观察者, 换言之, 稳态观察者的协变速度 $u = \gamma\xi$, 式中 γ 为归一化系数。但是, 对于像克尔解这样的存在两个基林向量场 (ξ 和 η) 的稳态时空, 由于 $\tilde{\xi} = a\xi + b\eta$ (a, b 为常数) 也为基林向量场, 因此只要 $\tilde{\xi}$ 类时, 则由它的积分曲线给出来的观察者也是稳态观察者。不过, 我们这里谈到稳态观察者的时候仅仅指由 $\xi = \partial_t$ 的积分曲线给出来的观察者。很多书也称这样的观察者为静态观察者, 不过, 按照本书的定义, 静态观察者仅存在于静态时空中, 因此对于稳态时空中 ξ 的积分曲线, 我们还是称之为稳态观察者。

对于克尔解, 现在我们考察一个协变速度 $u = \gamma(\xi + \Omega\eta)$ 的观察者, 这里 γ 为归一化系数, Ω 不必为常数, 而是可以依赖于 r, θ 。则我们说这样的观察者为具有角速度 Ω 的观察者。为了看出这一点, 我们可以积分协变速度的方程 $\frac{dt}{d\tau} = \gamma, \frac{d\phi}{d\tau} = \gamma\Omega$, 进而有观察者的角速度

$$\frac{d\phi}{dt} = \frac{d\phi}{d\tau} / \frac{dt}{d\tau} = \Omega(r, \theta). \quad (13.20)$$

我们将看到, 仅在无限红移面 r_{S+} 之外才存在角速度 $\Omega = 0$ 的观察者, 因此人们又把 r_{S+} 称作**稳态极限面**, 意思是, 在这个面之内 $g_{tt} > 0$ 区域的观察者无法具有固定的 (r, θ, ϕ) 位置, 而必须跟着时空一起转动, 具有非零的角速度 Ω 。

为了看清楚这一点, 我们注意到作为协变速度要求 u 类时, 即 $\langle u, u \rangle < 0$, 根据 u 的定义即有

$$\begin{aligned} \langle \xi, \xi \rangle + 2\Omega \langle \xi, \eta \rangle + \Omega^2 \langle \eta, \eta \rangle &< 0 \\ \Leftrightarrow g_{tt} + 2g_{t\phi}\Omega + g_{\phi\phi}\Omega^2 &< 0. \end{aligned} \quad (13.21)$$

进而有

$$\Omega_-(r, \theta) < \Omega(r, \theta) < \Omega_+(r, \theta), \quad (13.22)$$

其中 Ω_{\pm} 是 (13.21) 关于 Ω 的一元二次多项式的两个根, 由下式给出

$$\Omega_{\pm} = \frac{-g_{t\phi} \pm \sqrt{g_{t\phi}^2 - g_{tt}g_{\phi\phi}}}{g_{\phi\phi}} = \omega \pm \frac{\sqrt{\Delta}\rho^2}{\Sigma \sin \theta}, \quad (13.23)$$

上式最后我们代入了 (13.10) 式关于 ω 的定义, 也代入了 (13.7) 式的结果。

在稳态极限面 r_{S+} 上, $g_{tt} = 0$, 从而有 $\Omega_- = 0$ (注意 $g_{t\phi} < 0$), 并且不难知道

$$\Omega_- = \begin{cases} < 0, & r > r_{S+} \\ = 0, & r = r_{S+} \\ > 0, & r < r_{S+} \end{cases}. \quad (13.24)$$

另外, 再注意到 $\Omega_+ > 0$ 恒成立, 而且在观察者从稳态极限面 r_{S+} 进入能层, 并逐渐向视界 r_+ 靠近的过程中, Δ 在减小, 从而 Ω_+ 逐渐减小, 而 Ω_- 逐渐增加。在 $r = r_+$ 的视界面上, $\Delta = 0$, 从而

$$\Omega_+ = \Omega_- = \omega(r_+) = \frac{a}{r_+^2 + a^2} > 0, \quad (13.25)$$

式中用到 $\Delta(r_+) = 0$, 即 $r_+^2 + a^2 = 2GM r_+$ 。

以上讨论也意味着: 在 $r > r_{S+}$ 的区域, 观察者的角速度 Ω 可正可负, 尤其是, 可以取 $\Omega = 0$! 而在 $r_+ < r < r_{S+}$ 的能层区, 必定有 $\Omega > 0$, 观察者必须跟着时空一起旋转。尤其是在 $r = r_+$ 的视界面上, 这时候 Ω 只能取 $\omega(r_+) = \frac{a}{r_+^2 + a^2} > 0$! 因此, 通常称这个角速度为克尔黑洞的旋转角速度, 记为 Ω_H ,

$$\Omega_H \equiv \omega(r_+) = \frac{a}{r_+^2 + a^2}. \quad (13.26)$$

很显然, Ω_H 是一个常数, 所以 $\xi_H = \xi + \Omega_H \eta$ 必定为一个基林向量场。而且由于在视界面 $r = r_+$ 上 $\Omega_+ = \Omega_- = \Omega_H$, 进而由上面的讨论过程可以知道, 在视界面 $r = r_+$ 上, 必有 $\langle \xi_H, \xi_H \rangle|_{r_+} = 0$, 换言之, 视界面 $r = r_+$ 同时是基林向量场 ξ_H 的基林视界!

13.2 彭罗斯过程

由于能层的特殊性, 彭罗斯想到一个聪明的想法, 使得人们在原则上可以从一个旋转黑洞中提取能量, 这就是所谓的彭罗斯过程。这个过程可以将克尔黑洞的总能量 (也就是

质量 M) 部分提取出来, 转换成可资利用的能量, 而且, 彭罗斯过程的能量转换率非常高, 反复利用彭罗斯过程最高大约可以将黑洞总能量的百分之 29 转换成可利用的能量。在把质量转换成其它能量形式的各种机制中, 这是一个异常巨大的转换百分数。比方说, 实验室里可实现的转化率最高的过程, 就是将氢聚变成氦的核聚变反应, 但这个反应的转换百分数还不到百分之 1。不妨科幻地想象一下, 宇宙中也许有足够发达的文明, 它可以建立一个巨大的旋转黑洞旁边, 并利用彭罗斯过程从旋转黑洞中提取能量, 那这样的文明就将有用不完的能量!

为了说清楚彭罗斯的想法, 我们考察基林向量 $\xi = \partial_t$, 根据第 11 章第 2 节, 我们知道有相应的守恒能量 (在 11.2 节我们是记单位质量的守恒能量 $\varepsilon = -\xi \cdot u$)

$$E = -\xi \cdot P = -\xi_\mu P^\mu. \quad (13.27)$$

式中 P 为粒子的协变动量, 对于质量为 m 的粒子, $P = mu$ 。粒子的协变动量 P 当然是类时的! 如果 ξ 也类时, 则由于两个类时向量的内积一定是负的 (请读者想一想为什么?), 因此以上定义的守恒能量 E 就一定为正。但是, 如果 ξ 类空, 那相应的能量 E 就可以为负! 而在克尔解中, ξ 正是既可以类时也可以类空, 具体来说, 在 r_{S+} 之外, $\langle \xi, \xi \rangle = g_{tt} < 0$, 这时 ξ 类时; 但是, 在能层 $r_+ < r < r_{S+}$ 之内, $g_{tt} > 0$, 这时 ξ 就类空。因此, 在能层之内, 允许粒子的能量为负!

以上就是彭罗斯的观察, 而彭罗斯的想法是: 我们让一个协变动量为 P 的物体沿着测地线进入能层, 因为它沿着测地线运动, 所以其 $E = -\xi \cdot P$ 守恒, 从而, 这个 E 在能层之内也和能层之外一样保持为正。现在, 一旦这个物体进入了能层, 我们可以设法让它分裂成两半 (比方说如果这个物体是一个粒子, 我们可以让它衰变成两个粒子), 每一半都沿着自己的测地线运动, 能量和动量守恒要求

$$P = P_1 + P_2, \quad E = E_1 + E_2. \quad (13.28)$$

但是, 在能层之内, 我们有可能让 $E_1 < 0$, 进而使得 $E_2 = E - E_1 > E$, 也即是说, 使得另一半获得比初始时更大的能量! 当然, 第一半由于 $E_1 < 0$, 因此它是无法逃出能层之外的 (因为能层外要求能量为正), 这一半一般来说会进一步掉进黑洞视界之内。但是对于 $E_2 > E$ 的这另一半, 它完全可以逃离到能层的外面而被外面的我们所利用 (因为 r_{S+} 不是视界, 视界之内才是无法逃离的, 能层之内则可以逃离)。这样一来, 我们就通过花费一份较小的能量 E , 而获得了一份更大的能量 $E_2 > E$, 额外获得的能量当然是从旋转黑洞中提取出来的, 因此, 这样我们就从克尔黑洞中提取出了能量。这, 就是彭罗斯过程!

从克尔黑洞中提取能量会使得黑洞的质量 M 减少。假如从克尔黑洞中提取了过多的能量, 那会不会最终破坏 $GM \geq a$ 的能限要求, 从而制造出裸奇性呢? 下文的分析表明, 这种情况永远不会发生, 因为黑洞质量 M 的减少必然会同时伴随着克尔参数 a 的减少, 从而使得能限永远得以满足。

为了看清楚这一点, 不妨考虑基林向量场 $\xi_H = \xi + \Omega_H \eta$ 。我们知道, 视界 r_+ 也是 ξ_H 的基林视界, 即 ξ_H 在视界上类光, 不难进一步看到, 在视界之外 (包括能层内), ξ_H 都是类时的。这意味着在视界外恒有

$$\xi_H \cdot P \leq 0 \Leftrightarrow \xi \cdot P + \Omega_H \eta \cdot P \leq 0 \Rightarrow -E + \Omega_H j \leq 0. \quad (13.29)$$

式中 $j = \eta \cdot P$ 为粒子的守恒角动量。将这个条件运用到两半中掉进黑洞的那一半, 也就是满足 $E_1 < 0$ 的那一半, 即有

$$0 > E_1 \geq \Omega_H j_1. \quad (13.30)$$

这意味着 $j_1 < 0$, 从而这一半会减少克尔黑洞的总角动量, 而且角动量减少得“更多”, 这样, 我们在从克尔黑洞提取能量的同时, 必然也提取了它的“更多”角动量。

事实上, 正如后文将会解释的, 克尔黑洞的总角动量 J 为

$$J = Ma. \quad (13.31)$$

(请注意克尔黑洞的旋转角速度 Ω_H 取决于克尔参数 a) 这样, 从克尔黑洞中提取角动量就意味着 a 也会同时减少, 而且减少得“更多”, 从而保证能限 $GM \geq a$ 始终得以满足。换言之, 彭罗斯过程提取的黑洞能量实际上是黑洞的转动能量, 反复利用彭罗斯过程耗尽一个克尔黑洞转动能量的最终结果就是, 使它变成一个施瓦西黑洞。

为了进一步看清提取能量的过程不会制造裸奇性, 我们考察克尔黑洞视界面的表面积 A (给定一个时刻, 给定 $r = r_+$, 注意在 $r = r_+$ 上 $\Delta = 0$)

$$\begin{aligned} A &= \int_0^\pi d\theta \int_0^{2\pi} d\phi \sqrt{g_{\theta\theta}g_{\phi\phi}}|_{r=r_+} = 4\pi(r_+^2 + a^2) \\ &= 8\pi GM [GM + \sqrt{(GM)^2 - a^2}] \\ &= 8\pi [G^2 M^2 + G\sqrt{G^2 M^4 - J^2}]. \end{aligned} \quad (13.32)$$

式中我们代入了 $r_+ = GM + \sqrt{G^2 M^2 - a^2}$ 。假如克尔黑洞同时丢失 δM 的质量和 δJ 的角动量, 则其面积的改变量为

$$\begin{aligned} \frac{\delta A}{4G} &= 2\pi \left[2GM\delta M + \frac{2G^2 M^3 \delta M - J\delta J}{\sqrt{G^2 M^4 - J^2}} \right] \\ &= \frac{2\pi J}{\sqrt{G^2 M^4 - J^2}} \left[\frac{\delta M}{\Omega_H} - \delta J \right]. \end{aligned} \quad (13.33)$$

这个式子从最后一行代入 Ω_H 的定义式并倒推上去比较容易证明。

掉进黑洞的那一半物体会使得黑洞质量减少 $\delta M = E_1$, 同时使得角动量减少 $\delta J = j_1$ 。进而, 在式 (13.33) 中代入与 $E_1 \geq \Omega_H j_1$ 对应的 $\delta M \geq \Omega_H \delta J$, 即有

$$\delta A \geq 0. \quad (13.34)$$

这意味着, 在彭罗斯过程中, 虽然黑洞的质量减少了, 但是其视界面积是单调递增的! 特别的, 这保证了, 彭罗斯过程永远不会使得克尔黑洞的视界消失进而出现裸奇性!

下面我们解释为什么说克尔黑洞的总角动量为 $J = Ma$ 。这里的关键依据是, 在广义相对论中, 一个时空的总能量动量以及总角动量完全取决于时空度规在空间无穷远处的渐近形式, 关于这一点后面的章节会进一步展开讲, 这里暂且把它接受为基本原理吧。

因此,为了看清楚为什么克尔黑洞的角动量为 $J = Ma$, 我们需要考察克尔度规在空间无穷远 ($r \rightarrow +\infty$) 处的渐近形式。根据 (13.3) 式可知, 当 $r \rightarrow +\infty$ 时有

$$ds^2 \rightarrow -\left(1 - \frac{2GM}{r}\right)dt^2 + \left(1 + \frac{2GM}{r}\right)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) - \frac{4GMa}{r}\sin^2\theta dt d\phi. \quad (13.35)$$

式中第一行所描述的正是一个质量为 M 的星体在空间无穷远处的渐近度规, 而第二行 $2g_{t\phi}dt d\phi$ 的渐近形式 $-\frac{4GMa}{r}\sin^2\theta dt d\phi$ 正是取决于这个星体的旋转角动量。为了看得更清楚, 不妨引入直角坐标 $x = r\sin\theta\cos\phi, y = r\sin\theta\sin\phi$, 注意到 $xdy - ydx = r^2\sin^2\theta d\phi$, 进而即有

$$2g_{t\phi}dt d\phi \rightarrow -\frac{4GMa}{r}\sin^2\theta dt d\phi = -\frac{4GMa}{r^3}(xdy - ydx)dt. \quad (13.36)$$

正如后面章节中将会看到的, 这一渐近形式正是一个在 $x-y$ 平面内旋转角动量为 $J = Ma$ 的星体的渐近度规。所以, **克尔黑洞的总角动量为 $J = Ma$!**

13.3 黑洞热力学定律

本节将从计算克尔黑洞的表面引力出发, 进而对黑洞热力学定律进行一个探讨。

13.3.1 克尔黑洞的表面引力

假设记克尔黑洞的视界面 ($r = r_+$ 处) 为 \mathcal{H} , 则由于 \mathcal{H} 同时是基林向量场 $\xi_H = \xi + \Omega_H\eta$ 的基林视界, 所以在 \mathcal{H} 上有

$$(\xi_H^\mu D_\mu \xi_H^\nu = \kappa \xi_H^\nu)|_{\mathcal{H}}, \quad (13.37)$$

式中常数 κ 为克尔黑洞的表面引力。原则上, 我们可以通过这个式子计算出克尔黑洞的 κ 。

不过, 用下面的等效方法计算 κ 将会更快。为此我们注意到 $\partial_\mu \langle \xi_H, \xi_H \rangle = 2\langle \xi_H, D_\mu \xi_H \rangle = 2\xi_H^\nu D_\mu \xi_{H\nu}$, 利用基林方程 $D_\mu \xi_{H\nu} + D_\nu \xi_{H\mu} = 0$, 即有 $\partial_\mu \langle \xi_H, \xi_H \rangle = -2\xi_H^\nu D_\nu \xi_{H\mu}$, 限制在视界面上又有 $-2\xi_H^\nu D_\nu \xi_{H\mu}|_{\mathcal{H}} = -2\kappa \xi_{H\mu}|_{\mathcal{H}}$, 从而即有

$$\partial_\mu \langle \xi_H, \xi_H \rangle|_{\mathcal{H}} = -2\kappa \xi_{H\mu}|_{\mathcal{H}}. \quad (13.38)$$

用这个式子计算 κ 只需用到普通导数, 无需计算协变导数, 因此要更方便一些。

不过, 上面的式子要求坐标系在视界面上非奇异, 而我们知道 Boyer-Lindquist 坐标在视界面上是奇异的。解决这个问题的办法之一是, 引入相应的爱丁顿-芬克斯坦坐标。

为此, 我们首先回顾 (13.1) 式, 重写如下

$$ds^2 = -\frac{\Delta}{\rho^2}(dt - a\sin^2\theta d\phi)^2 + \frac{\sin^2\theta}{\rho^2}[(r^2 + a^2)d\phi - a dt]^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2. \quad (13.39)$$

定义 r_* 坐标如下

$$dr_* = \frac{r^2 + a^2}{\Delta} dr. \quad (13.40)$$

观察上面的 (13.39) 式, 进而引入爱丁顿-芬克斯坦坐标 (v, χ, r, θ) , 定义如下

$$dv = dt + dr_* = dt + \frac{r^2 + a^2}{\Delta} dr, \quad d\chi = d\phi + \frac{a}{\Delta} dr. \quad (13.41)$$

利用这个式子消去 (13.39) 式中的 dt 和 $d\phi$, 即可以得到

$$ds^2 = -\frac{\Delta}{\rho^2} (dv - a \sin^2 \theta d\chi)^2 + \frac{\sin^2 \theta}{\rho^2} [(r^2 + a^2) d\chi - adv]^2 + 2(dv - a \sin^2 \theta d\chi) dr + \rho^2 d\theta^2. \quad (13.42)$$

为了方便读出各度规分量, 我们展开上式, 得

$$ds^2 = -\left(1 - \frac{2GMr}{\rho^2}\right) dv^2 + \frac{\Sigma}{\rho^2} \sin^2 \theta d\chi^2 + \rho^2 d\theta^2 + 2dvdr - 2a \sin^2 \theta d\chi dr - \frac{4GMar}{\rho^2} \sin^2 \theta dvd\chi. \quad (13.43)$$

注意到 $g_{vv} = g_{tt}$, $g_{\chi\chi} = g_{\phi\phi}$, $g_{v\chi} = g_{t\phi}$ 。特别的, 这个度规中的 g_{rr} 分量为零! 并且 Δ 不再出现在度规分量的分母上, 这使得即使在视界面上, 度规分量也不会发散, 这就消除了视界面处的坐标奇异性!

根据 (13.41) 式, 不难得到爱丁顿-芬克斯坦坐标中的基林向量 ξ_H

$$\xi_H = \xi + \Omega_H \eta = \partial_t + \Omega_H \partial_\phi = \partial_v + \Omega_H \partial_\chi. \quad (13.44)$$

换言之, 在 (v, χ, r, θ) 坐标中 $\xi_H^\mu = (1, \Omega_H, 0, 0)$ 。为了应用 (13.38) 式, 我们需要计算 $\langle \xi_H, \xi_H \rangle$, 有

$$\begin{aligned} \langle \xi_H, \xi_H \rangle &= g_{vv} + 2g_{v\chi} \Omega_H + g_{\chi\chi} \Omega_H^2 \\ &= g_{tt} + 2g_{t\phi} \Omega_H + g_{\phi\phi} \Omega_H^2. \end{aligned} \quad (13.45)$$

(注意, $g_{vv} = g_{tt}$, $g_{\chi\chi} = g_{\phi\phi}$, $g_{v\chi} = g_{t\phi}$) 从而不难得到 (利用定义 $\omega = -\frac{g_{t\phi}}{g_{\phi\phi}}$, 再利用 $g_{t\phi}^2 - g_{tt}g_{\phi\phi} = \Delta \sin^2 \theta$)

$$\begin{aligned} \langle \xi_H, \xi_H \rangle &= g_{tt} + 2g_{t\phi} \Omega_H + g_{\phi\phi} \Omega_H^2 \\ &= g_{\phi\phi} (\Omega_H - \omega)^2 + g_{tt} - \frac{g_{t\phi}^2}{g_{\phi\phi}} \\ &= g_{\phi\phi} (\Omega_H - \omega)^2 - \frac{\rho^2 \Delta}{\Sigma}. \end{aligned} \quad (13.46)$$

利用在视界上 $\Omega_H = \omega|_{\mathcal{H}}$, 以及 $\Delta|_{\mathcal{H}} = 0$, 即可以得到

$$\partial_\mu \langle \xi_H, \xi_H \rangle|_{\mathcal{H}} = -\frac{\rho^2}{\Sigma} (\partial_\mu \Delta)|_{\mathcal{H}}. \quad (13.47)$$

为了应用 (13.38) 式, 我们还需要知道 (v, χ, r, θ) 坐标中的 $\xi_{H\mu}$, 为此, 将 $\xi_H^\mu = (1, \Omega_H, 0, 0)$ 的指标降下去, 即有

$$\begin{aligned}\xi_{H\mu} &= (g_{vv} + \Omega_H g_{v\chi}, \Omega_H g_{\chi\chi} + g_{v\chi}, g_{vr} + \Omega_H g_{\chi r}, 0) \\ &= (g_{tt} + \Omega_H g_{t\phi}, \Omega_H g_{\phi\phi} + g_{t\phi}, g_{vr} + \Omega_H g_{\chi r}, 0).\end{aligned}\quad (13.48)$$

在视界面上, $(g_{t\phi}^2 - g_{tt}g_{\phi\phi})|_{\mathcal{H}} = \Delta \sin^2 \theta|_{\mathcal{H}} = 0$, 所以 $-\omega|_{\mathcal{H}} = \frac{g_{t\phi}}{g_{\phi\phi}}|_{\mathcal{H}} = \frac{g_{t\phi}}{g_{\phi\phi}}|_{\mathcal{H}}$, 所以,

$$\begin{aligned}\xi_{H\mu}|_{\mathcal{H}} &= (g_{tt} + \Omega_H g_{t\phi}, \Omega_H g_{\phi\phi} + g_{t\phi}, g_{vr} + \Omega_H g_{\chi r}, 0)|_{\mathcal{H}} \\ &= (g_{t\phi}(\Omega_H - \omega|_{\mathcal{H}}), g_{\phi\phi}(\Omega_H - \omega|_{\mathcal{H}}), g_{vr} + \Omega_H g_{\chi r}, 0) \\ &= (0, 0, 1 - \Omega_H a \sin^2 \theta, 0).\end{aligned}\quad (13.49)$$

上式最后一行利用了视界面上 $\Omega_H = \omega|_{\mathcal{H}}$, 并代入了 g_{vr} 和 $g_{\chi r}$ 的表达式。

另一方面, 注意 $\Delta = (r - r_+)(r - r_-)$, 在 (v, χ, r, θ) 坐标中不难得到

$$\begin{aligned}\partial_\mu \langle \xi_H, \xi_H \rangle|_{\mathcal{H}} &= -\frac{\rho^2}{\Sigma} (\partial_\mu \Delta)|_{\mathcal{H}} = -\frac{\rho^2}{\Sigma}|_{\mathcal{H}} (0, 0, r_+ - r_-, 0) \\ &= -\frac{(r_+^2 + a^2 - a^2 \sin^2 \theta)}{(r_+^2 + a^2)^2} (0, 0, r_+ - r_-, 0) \\ &= -\frac{1 - \Omega_H a \sin^2 \theta}{r_+^2 + a^2} (0, 0, r_+ - r_-, 0).\end{aligned}\quad (13.50)$$

将结果 (13.49) 和结果 (13.50) 代入 (13.38) 式, 即可以读出

$$\kappa = \frac{r_+ - r_-}{2(r_+^2 + a^2)}.\quad (13.51)$$

13.3.2 再谈霍金辐射

从表面引力的计算结果 (13.51), 再根据公式 $T_{BH} = \hbar\kappa/(2\pi)$, 就能得到克尔黑洞的霍金温度。特别的, 我们注意到, 当克尔参数刚好使能限饱和时, 也就是 $GM = a$ 时, 有 $r_+ = r_-$, 因此这时候表面引力 $\kappa = 0$, 从而霍金温度也等于零, 这种克尔黑洞称作极端克尔黑洞。

霍金温度的存在意味着非极端的克尔黑洞会向外发出热辐射, 也就是霍金辐射。霍金辐射当然是一种量子现象, 对它的详细讨论要用到弯曲时空的量子场论, 这当然超出了本书的范围。不过, 定性地分析一下霍金辐射的产生机制倒是可以的。首先, 平坦时空的量子场论告诉我们, 量子场会在平坦时空中不断涨落, 有各种正反虚粒子对不断产生并湮灭, 一般来说, 能量守恒会阻止这些虚粒子对变成真实的粒子, 因为真实的粒子和真实的反粒子都具有大于零的能量, 因此从真空涨落中凭空产生真实粒子对会破坏能量守恒。但是, 如果有一个外加的场 (比方说电场), 那虚粒子对就有可能从这个外场中吸收能量, 并进而转化成真实的正反粒子对。实际上, 所谓的施温格效应指的就是这种外场作用下产生粒子对的现象。类似的, 在黑洞这样的弯曲时空中, 引力场的存在也有可能将真空涨落的虚粒子对转化为真实粒子对。

以克尔黑洞为例, 这时我们有一个基林向量场 ξ_H , 它在视界外是类时的, 而在视界之内是类空的, 这个区别就是霍金辐射的根源。比方说考虑一对正反粒子, 记正粒子的协变动量为 P , 记反粒子的协变动量为 \bar{P} , 则由于 ξ_H 在视界外类时, 所以在视界外显然有 $-\xi_H \cdot P > 0$, $-\xi_H \cdot \bar{P} > 0$, 即如果两粒子都在视界之外, 那么它们的能量都为正, 这时当然不能从零能的真空中凭空产生这一对粒子。但是, 场的真空涨落允许产生虚粒子对, 如果这些虚粒子对中的一个, 比方说反粒子的那个, 掉进了视界之内, 而另一个保持在视界之外, 那它们就有可能转化成真实的粒子。因为, 由于在视界之内 ξ_H 类空, 所以视界内的那个反粒子完全可能具有负的能量, 也即使得 $-\xi_H \cdot \bar{P} < 0$, 视界外的那个粒子当然能量为正, 这样一来, 两粒子的总能量就可以为零, 那零能真空产生这样的粒子对当然就不违反能量守恒! 黑洞之外的那个粒子跑到无穷远处就成了黑洞的辐射, 而掉进黑洞的粒子由于具有负能量, 所以会使得黑洞的质量减少, 所以总的效果就相当于黑洞的能量转化成了辐射出去的粒子的能量。这就是霍金辐射的基本机制!

霍金辐射特殊的地方在于, 它的辐射谱是一个黑体谱, 对应的黑体温度就是霍金温度, 换言之, 霍金辐射是一个热辐射, 从这个意义来说, 出来的辐射似乎是不携带任何信息的。这导致霍金提出黑洞信息损失问题, 霍金的想法是: 假设在量子引力中考虑一团物质坍缩成黑洞, 然后这黑洞再通过霍金辐射蒸发掉, 整个过程似乎是信息不守恒的。因为在黑洞形成之前, 物质和引力是处于一个量子的纯态, 而在坍缩成黑洞以后, 辐射出来的却是一个热谱 (所谓的混态), 不再是一个纯态, 这样原来的纯态的信息就凭空损失掉了! 但是, 在量子力学中, 一个纯态只会随着时间么正演化, 它永远也不会自发变成一个混态, 换言之, 在量子力学中, 孤立系统的信息永远是守恒的! 基于以上想法, 霍金提出, 黑洞原则上不可能满足量子力学规律! 因此为了能在量子的层次上描述黑洞, 我们的量子力学原理需要修改。

但, 与霍金的观点相对的, 也有不少物理学家相信, 霍金辐射的热谱只是一个半经典的结果, 如果对黑洞蒸发过程进行真正量子引力的研究就会发现, 辐射可以用某种微妙的方式携带信息, 整个黑洞的演化过程是信息守恒的, 和当前的量子力学原理兼容!

在某种意义上, 这两派观点的争论由于 AdS/CFT 的发现而得以了结, 因为 AdS/CFT 对应告诉我们, AdS 时空中的一切过程, 包括黑洞的形成和蒸发, 都可以用边界上的普通量子场论系统来描述, 因此当然是信息守恒的! 后来霍金也承认了他原来的观点可能不正确, 但是, 问题是, 霍金辐射是如何携带信息从而使得信息守恒的, 这就是所谓的信息黑洞问题, 这个问题目前依然远远没有解决。这方面的最新进展可以参看综述文章 *The entropy of Hawking radiation*, Ahmed Almheiri, Thomas Hartman, Juan Maldacena, Edgar Shaghoulian, and Amirhossein Tajdini.

下面我们来具体估算一下霍金温度和一颗黑洞需要多长时间才能蒸发完。不妨以施瓦西黑洞为例, 它的表面引力是 $1/(4GM)$, 恢复带量纲常数 c, k_B 以后, 其霍金温度为

$$k_B T_{BH} = \frac{\hbar c^3}{8\pi GM} \sim \frac{\hbar c^3}{GM}, \quad (13.52)$$

注意, 霍金温度反比于黑洞质量, 质量越大的黑洞温度越低。代入各物理常数的值以后, 可以得到 $T_{BH} \sim 6.2 \times 10^{-8} (M_\odot/M) \text{K}$, 其中 M_\odot 是太阳质量, 可以看到, 除了质量非常微小

的原初黑洞，通常黑洞的温度都非常低，很难从宇宙背景中分辨出这么微小的温度。

下面估算黑洞蒸发完大约需要多长时间。首先从黑洞辐射的斯特藩 (Stefan) 定律, 有

$$\frac{dE}{dt} \simeq -\sigma AT_{BH}^4, \quad (13.53)$$

式中 $\sigma = \frac{\pi^2 k_B^4}{60\hbar^3 c^2}$ 为斯特藩-玻尔兹曼常数, A 为黑洞表面积分, 对施瓦西黑洞 $A = 4\pi r_g^2$, 注意到

$$E = Mc^2, \quad A \sim \left(\frac{GM}{c^2}\right)^2, \quad (13.54)$$

由此可以得到

$$\frac{dM}{dt} \sim -\frac{\hbar c^4}{G^2 M^2}. \quad (13.55)$$

积分即可以得到黑洞的寿命

$$\tau_{BH} \sim \left(\frac{G^2}{\hbar c^4}\right) M^3. \quad (13.56)$$

以上估算过程忽略了霍金辐射所产生的反作用, 在 $dM/dt \ll M$ 时这是一个不错的近似, 但是, 在黑洞即将蒸发完的最后阶段, 反作用对黑洞质量的影响其实是不可忽略的。从上面的估算结果可以得到, 对于一个太阳质量大小的黑洞, 即 $M \sim M_\odot$, 有 $\tau_{BH} \gg 140$ 亿年, 即远远大于宇宙年龄。而那些在宇宙极早期产生的原初黑洞, 质量约为 $10^{14} \text{g} \sim 10^{-19} M_\odot$, 大约是地球上一座山的质量, 不难算出, 把它辐射完需要 $\tau_{BH} = 8.3 \times 10^{16} \text{s}$ 。也就是说, 即使是宇宙早期所产生的原初黑洞, 至今也还在辐射。当然, 随着黑洞质量的减少, 霍金辐射是越来越快的, 在一颗黑洞寿命的最后阶段, 它可能会终结于一场剧烈的“小爆炸”。

13.3.3 黑洞热力学定律

根据表面引力的式子 (13.51) 不难得到

$$\kappa = \Omega_H \frac{\sqrt{G^2 M^4 - J^2}}{J}. \quad (13.57)$$

进而即可以将 (13.33) 式写成

$$\begin{aligned} \frac{\delta A}{4G} &= \frac{2\pi}{\kappa} [\delta M - \Omega_H \delta J] \\ \Rightarrow \delta M &= \frac{\kappa}{2\pi} \delta \left(\frac{A}{4G}\right) + \Omega_H \delta J. \end{aligned} \quad (13.58)$$

与热力学第一定律 $\delta M = T \delta S + \Omega \delta J$ 比较, 并注意到 $\frac{\hbar \kappa}{2\pi}$ 就是黑洞的温度, 进而可知, 黑洞具有一个正比于其视界面积的熵

$$S_{BH} = \frac{A}{4G\hbar}, \quad (13.59)$$

称作贝肯斯坦-霍金熵!

而 (13.58) 式即是所谓的**黑洞热力学第一定律**, 我们上面给的是一个简化版的推导, 这个推导并不具备一般性, 它依赖于克尔解的假定, 更具一般性的推导可以参见 Eric Poisson, *A Relativist's Toolkit—The Mathematics of Black-Hole Mechanics*, 第 5.5.4 节, 或者参见原始论文, *The Four Laws of Black Hole Mechanics*, J. M. Bardeen, B. Carter and S. W. Hawking.

贝肯斯坦-霍金熵的存在和霍金温度的存在一样, 是一件非常不可思议的事情, 因为从经典物理的角度来看, 克尔黑洞只依赖于两个参数 M 和 a , 换言之, 经典的克尔黑洞只有两个自由度。而热力学熵却是系统自由度数目的衡量, 一个正比于面积的熵就意味着系统存在正比于面积的自由度数目的 (对于一个真实的黑洞, 这个自由度数目是巨大的, 比如即使对于一个质子那么大的小黑洞, 其 $S_{BH} \sim 10^{40}$, 一个巨大的天文数字), 这些自由度当然不可能仅仅是经典解的 M 和 a , 那么与黑洞熵所对应的这些微观自由度是什么呢? 这是一个远没有完全解决的问题, 不过, 注意到贝肯斯坦-霍金熵的表达式依赖于普朗克常数 \hbar , 而且它和引力常数 G 一起出现在表达式的分母上, 所以这意味着两件事情: 第一, 与黑洞熵所对应的微观自由度一定是量子的, 只有用量子引力才能解释; 第二, 由于 $G\hbar$ 出现在分母上, 所以黑洞熵必定不是一个微扰量子引力效应 (微扰指按照 $G\hbar$ 的幂级数展开, 因此 $G\hbar$ 的负幂次一定是非微扰的), 它必定有一个非微扰的起源。

关于黑洞熵的微观起源, 目前只有超弦理论对于一些具有超对称性的极端黑洞给出了正确的解释, 对于像施瓦西黑洞这样的黑洞, 其熵的微观起源是什么, 人们还远远没有了解, 这是一个非常重要非常困难的问题, 在这个问题上的进展将意味着我们对量子引力的理解取得了进展。

另外, 贝肯斯坦-霍金熵正比于黑洞表面积也是一个奇异的特征。因为通常的量子场论体系在给定能标或者给定温度之下, 其熵 (也就是自由度数目) 一定是正比于体积的, 但是黑洞熵却并非正比于体积而是正比于表面积, 另外还可以论证, 给定一个总能量等于黑洞质量的区域, 其总的熵 (自由度数) 一定小于相应黑洞的熵。基于这两点观察, 't Hooft 提出, 在量子引力中, 一个给定时空区域完全可以由分布在这个区域表面上的某些场论自由度来描述, 特别的, 黑洞区域可以由黑洞表面的某些场论自由度描述, 所以黑洞的熵才正比于表面积, 这也就是所谓的量子引力的全息原理。全息的意思就是区域表面全息性地编码了整个区域的信息, 就好像一张全息图一样!

目前来说, 全息原理已经不仅仅是一个猜测了, 因为的确有量子引力的例子精确地实现了全息原理, 这例子就是 Maldacena 提出来的 AdS/CFT 对应, 也就是 AdS 时空的某种量子引力完全由其边界上的某个共形场论描述。但是, 对于我们生活的这个真实的宇宙, 其量子引力怎样符合全息原理, 则目前人们还完全不知道。

由于黑洞视界的表面积正比于其熵, 而我们又知道, 在彭罗斯过程中, 视界面积 A 是单调递增的, 这不禁让我们想到热力学第二定律, 也就是所谓的熵增定律。实际上, 在黑洞力学中, 的确有一条严格证明的定理, 它说, 渐近平坦时空中任何黑洞的视界面积之总和在任何经典过程 (不限于彭罗斯过程) 中都只能增加, 不能减少。这就是霍金的黑洞面积不减定理, 具体来说:

霍金的面积不减定理: 如果物质场的能动张量满足所谓的类光能量条件 (后面的章节

会介绍), 并假定宇宙监督成立, 则一个渐近平坦时空中未来事件视界的面积之总和一定是时间的非减函数。

关于这个定理的证明, 可以参见 E. Witten 的经典综述文章 Light Rays, Singularities, and All That。中文译本叫《光线、奇点, 以及其它》, 是由我们一些人翻译的, 网上能搜到。

在历史上, 霍金温度的发现其实晚于黑洞熵概念的提出, 最早是贝肯斯坦根据黑洞力学第一定律 (13.58) 与热力学第一定律的相似性, 以及面积不减定理与热力学第二定律的相似性, 进而初步论证黑洞具有一个正比于其表面积的熵, 至于精确的比例系数, 则是后来霍金通过精确推导出霍金温度公式而确定下来的。

贝肯斯坦的基本推理如下: 设想一个里面包裹着黑洞的孤立系统, 设想这个系统的物质正在不断掉进黑洞里面, 因此这些物质的熵在减少, 如果黑洞没有熵或者在物质掉进去的时候其熵不增加, 那么整个系统总的熵就会自发减少, 从而破坏了热力学第二定律。所以, 贝肯斯坦说, 黑洞一定有一个熵, 而且在物质掉进去的时候, 其熵一定会增加, 贝肯斯坦进一步估算了黑洞熵要增加多少才能挽救热力学第二定律, 他发现, 大体上, 黑洞熵的增量应该正比于其视界面积的增量, 进而贝肯斯坦才提出, 黑洞有熵, 且其熵正比于视界面积。

下面讨论两个可以从面积不减定理中推导出来的结果。一个是黑洞合并过程中可能辐射 (不是霍金辐射, 而是经典的引力波辐射) 出来的能量, 另一个是对所谓的黑洞不能分裂的简单论证。

考虑两个质量分别为 M_1 和 M_2 的黑洞相互合并, 最终形成一个质量为 M_3 的黑洞, 在合并过程中通过引力波辐射出去的能量是 $M_1 + M_2 - M_3$, 因此这个过程的效率 η (注意不是上文的基林向量 η) 为

$$\eta = \frac{M_1 + M_2 - M_3}{M_1 + M_2} = 1 - \frac{M_3}{M_1 + M_2}. \quad (13.60)$$

假定初始的两个黑洞均是施瓦西黑洞, 因此视界面积 A_1, A_2 分别为, $A_1 = 16\pi G^2 M_1^2$, $A_2 = 16\pi G^2 M_2^2$ 。假定合并以后的黑洞面积为 A_3 , 则根据面积不减定理

$$A_3 \geq 16\pi G^2 (M_1^2 + M_2^2). \quad (13.61)$$

另外, 因为两个黑洞在合并过程中互相绕对方旋转, 因此一般来说角动量不为零, 最终合并以后的黑洞一般来说是一个克尔黑洞, 根据克尔黑洞的面积公式 (13.33), 有

$$A_3 = 8\pi [G^2 M_3^2 + G \sqrt{G^2 M_3^4 - J_3^2}] \leq 16\pi G^2 M_3^2. \quad (13.62)$$

结合以上两个不等式, 即有

$$M_3 \geq \sqrt{M_1^2 + M_2^2}, \quad (13.63)$$

因此辐射效率将受到如下限制

$$\eta \leq 1 - \frac{\sqrt{M_1^2 + M_2^2}}{M_1 + M_2} \leq 1 - \frac{1}{\sqrt{2}}. \quad (13.64)$$

辐射能是可以利用来做功的能量, 面积不减定理的这个限制有点类似于热力学第二定律对热机效率的限制。

下面我们来简单论证为什么黑洞不能分裂。假设一个施瓦西黑洞可以分裂成两个黑洞, $M_3 \rightarrow M_1 + M_2$. 则面积不减定理告诉我们

$$M_3 \leq \sqrt{M_1^2 + M_2^2} \leq M_1 + M_2. \quad (13.65)$$

而另一方面分裂过程只可能通过引力波辐射损失能量, 这要求 $M_3 \geq M_1 + M_2$, 这就导出了矛盾。所以黑洞不能分裂。

值得注意的是, 面积不减定理是一个单纯的经典结果, 考虑到量子的霍金辐射, 则黑洞能量可以减少, 同时保持角动量不变, 因此黑洞面积当然就可能减少。由于黑洞面积正比于热力学熵, 那这是否意味着霍金辐射过程会破坏热力学第二定律呢? 研究表明, 情况并非如此, 原因在于, 一个正在蒸发的黑洞本身并不是一个孤立系统, 只有再加上黑洞外面描写霍金辐射的量子场以后整个系统才可以看成孤立系统。因此这时候, 整个系统的熵除了黑洞熵以外, 还要加上外面量子场的熵 S_{outside} , 两者加起来的结果称作广义熵 S_{gen} ,

$$S_{\text{gen}} = \frac{A}{4G\hbar} + S_{\text{outside}}. \quad (13.66)$$

可以证明, 广义熵 S_{gen} 依然满足热力学第二定律, 即

$$\delta S_{\text{gen}} \geq 0. \quad (13.67)$$

具体的证明可以参见论文 A. C. Wall, A proof of the generalized second law for rapidly changing fields and arbitrary horizon slices, Phys. Rev. D 85, 104049, 2012, [arXiv:1105.3445 [gr-qc]]. [Erratum: Phys.Rev.D 87, 069904 (2013)].

以上我们讲述了黑洞热力学的第一定律和第二定律, 其实黑洞力学还满足所谓的热力学第零定律和第三定律。

具体来说, 前面第十一章中, 我们证明过具有分岔基林视界的黑洞满足热力学第零定律, 也即是, 在整个分岔基林视界上, 表面引力的平方是一个常数, 因此霍金温度是一个常数。但是对于一个由引力坍缩形成的真实黑洞, 其视界只是一个“不完整”的基林视界, 并没有一个分岔球面, 因此这时候第十一章中的证明就失效了。好在我们有一个更加一般性的第零定律:

黑洞第零定律: 假如物质场的能动张量满足所谓的主能量条件 (后面的章节中会介绍), 则表面引力 κ 在整个未来事件视界上是一个常数, 从而霍金温度是一个常数。

这条定律完全类似于热力学平衡系统的温度是一个常数这条热力学第零定律, 所以称作黑洞热力学第零定律。其证明可以参见 P.K. Townsend, Black Holes, Lecture notes, arXiv: gr-qc/9707012, 第 6.2.1 节。或者参见原始论文 The Four Laws of Black Hole Mechanics, J. M. Bardeen, B. Carter and S. W. Hawking.

而所谓的**黑洞热力学第三定律**, 是指: 不可能通过一个有限系列的操作 (无论这些操作多么理想), 将一个黑洞的表面引力 κ 减少至零 (也即是将霍金温度降低至绝对零度)。

很显然，这条定律类似于不可能将一个物体冷却到绝对零度这条热力学第三定律，所以称作黑洞热力学第三定律。由于极端黑洞的表面引力等于零，所以在这种情况下，第三定律的意思就是不可能通过有限的过程将一个非极端黑洞转变成极端黑洞，因为如果这可能性的话，那我们就可能进一步进行这样的过程，从而创造出裸奇性，这是宇宙监督假设所不允许的。第三定律的证明技术上相当复杂，我们推荐读者阅读 Eric Poisson, *A Relativist's Toolkit—The Mathematics of Black-Hole Mechanics*, 第 5.5.6 节的一个相关讨论（不是证明）。

13.4 克尔黑洞的彭罗斯图和时空奇性

彭罗斯图

前面讲述克尔黑洞的爱丁顿-芬克斯坦坐标时，我们引入了 r_* 坐标

$$r_* = \int \frac{r^2 + a^2}{\Delta} dr. \quad (13.68)$$

与第九章关于克鲁斯卡坐标的相关讲述比较，我们发现， $\Delta/(r^2 + a^2)$ 相当于那里的函数 $f(r)$ 。因此，类似于第九章，我们可以通过分析 $\Delta/(r^2 + a^2)$ 的零点，进而引入克鲁斯卡坐标，显然 Δ 有两个零点 r_{\pm} ，与带电 RN 黑洞有点类似，因此相应有两组克鲁斯卡坐标 (U_{\pm}, V_{\pm}) 。相应的彭罗斯图也有点类似于 RN 黑洞的彭罗斯图，如图 (13.3) 所示。

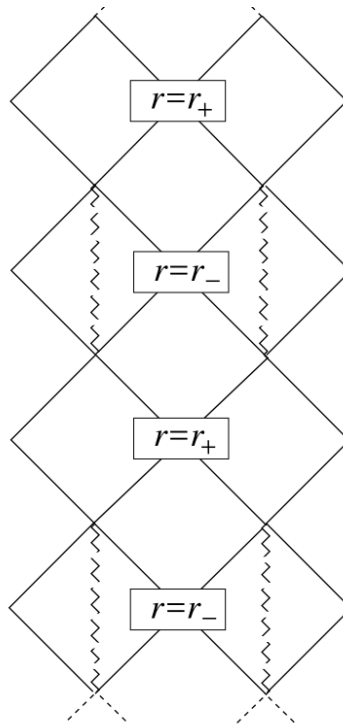


Figure 13.3: 克尔黑洞的彭罗斯图。

然而, 严格来说, 彭罗斯图仅仅只对球对称时空有定义, 因为彭罗斯图中的每一个点都代表一个两维球面 S^2 , 但克尔黑洞却不是球对称的, 而仅仅只有轴对称, 所以上面的彭罗斯图其实是取定了 θ 坐标的前提下画出来的, 图中每一点是代表一个由 $\phi \in [0, 2\pi)$ 坐标所描述的圆周 S^1 . 在彭罗斯图 (13.3) 中, 我们用断续的小折线表示时空奇性, 这与前面的章节中对时空奇性的图示方法略有不同, 这是因为, 正如稍后将会看到的, 克尔时空的奇性仅仅出现在 $\theta = \pi/2$ 处的圆周上, 而且这个圆周的半径不是零, 而在其它 θ 角位置是没有时空奇性的。正因为时空奇性只对应 $\theta = \pi/2$ 处的圆周, 而不对应由整个 (θ, ϕ) 角坐标所描述的半径为零的退化两维球面, 因此, 经过时空奇性以后, 时空依然可以继续, 反映在彭罗斯图 (13.3) 中即是, 在时空奇性那根线的两侧都各有一个时空“三角块”, 而不是像 RN 时空那样, 时空在奇性处就完全终止了。这也是克尔黑洞彭罗斯图与 RN 黑洞彭罗斯图的不同之处。

时空奇性

正如前面提到过的, 克尔时空的奇性与 RN 黑洞的奇性有一些根本性的不同, 虽然从彭罗斯图上看, 这两种时空奇性都是类时的。具体来说, 通过计算黎曼曲率张量 (或者计算 Kretschmann 标量), 研究它发散的地方, 可以确定, 克尔解的时空奇性出现在如下地方

$$\rho^2 = 0 \Rightarrow r = 0 \text{ 且 } \theta = \frac{\pi}{2}. \quad (13.69)$$

在 RN 黑洞以及施瓦西黑洞情形, $r = 0$ 的时空奇性处所对应的两维球面, 其半径也退化成了零, 但是, 克尔黑洞情形与此不同。为了看清楚这一点, 我们考察克尔度规 (13.1) 在 $r = 0$ 处的诱导度规 (给定时间 t), 不难得到

$$ds^2 = a^2 [\cos^2 \theta d\theta^2 + \sin^2 \theta d\phi^2]. \quad (13.70)$$

可见, (θ, ϕ) 坐标所描述的两维闭合曲面现在并没有退化成一个点, 虽然它上面的度规也并不是标准的球面度规。之所以不是球面度规, 当然是因为克尔解并没有空间旋转对称性。进一步将上面的度规 (13.70) 限制在 $\theta = \pi/2$ 处的时空奇性处, 即有

$$ds^2 = a^2 d\phi^2, \quad (13.71)$$

这描写的是一个半径为 a 的圆环! 这个圆环的角坐标为 $\phi \in [0, 2\pi)$ 。因此, 给定时刻, 克尔黑洞的时空奇性并不是一个点, 而是有一个圆环结构, 称作**奇环**!

闭合类时曲线 (时间机器)

前面说过, 克尔黑洞的类时奇性左右两侧都有时空存在 (与 RN 黑洞不同), 分别对应 $r > 0$ 和 $r < 0$ 的区域。实际上, 这个 $r < 0$ 的区域可以被一个在 $\theta \neq \pi/2$ 处经过 $r = 0$ 的类时观察者所访问, 这个类时观察者走的是类时测地线。总之, 克尔时空中存在 $r < 0$ 的区域。

这个新的区域有一个奇异的性质，它允许存在时间机器！换言之，它允许存在闭合的类时曲线，沿着这样的曲线，一个观察者可以回到过去。为了看清楚这一点，考察一根在给定 t, r 坐标同时给定 $\theta = \pi/2$ 处的闭合曲线，它当然由 $\phi \in [0, 2\pi)$ 所参数化。从克尔度规 (13.1)，不难得到这条闭合曲线上的诱导度规，为

$$ds^2 = \left(-\frac{a^2 \Delta}{r^2} + \frac{(r^2 + a^2)^2}{r^2} \right) d\phi^2 = \left(r^2 + a^2 + \frac{2GMa^2}{r} \right) d\phi^2. \quad (13.72)$$

很显然，只要 $r < 0$ 且绝对值足够小，则这条闭合曲线前面的度规系数就将为负！从而这条闭合曲线就成了一条类时曲线， ϕ 就成为这条闭合类时线的时间。换言之，在足够接近时空奇性且 $r < 0$ 的地方，周期变量 ϕ 成了时间坐标，因此随着 ϕ 演化的粒子可以自动会到自己的过去，这就是一台时间机器！

可惜的是，这台在克尔黑洞内部的时间机器很可能是非物理的！因为它藏在柯西视界的后面，而正如我们在第九章中讨论过的，柯西视界是不稳定的。

闭合类时线还出现在爱因斯坦场方程的其它一些解中，不过，就像裸奇性一样，人们通常也认为闭合类时线都是非物理的！也就是说，给定时空的一个合理初始条件，人们相信闭合类时线无法自动形成。

13.5 超辐射以及黑洞无毛定理

13.5.1 超辐射

彭罗斯过程还有一个场论版本，其中一个场（而不是一个粒子）被克尔黑洞所散射，但是散射波相比于入射波不但没有衰减，反而加强了，即散射波从克尔黑洞中提取出了能量，这种效应就叫做超辐射，它是彭罗斯过程的场论版本。

为了说清楚超辐射的机理，我们考察一个相互作用势能密度 $\mathcal{U}(\Phi) = 0$ 的标量场 Φ ，其能量动量张量为

$$T_{\mu\nu} = D_\mu \Phi D_\nu \Phi - \frac{1}{2} g_{\mu\nu} D^\rho \Phi D_\rho \Phi. \quad (13.73)$$

根据第十一章的知识，对于类时基林向量 $\xi = \partial_t$ ，相应的有守恒能量流

$$J^\nu = -\xi_\mu T^{\mu\nu} = -(\xi^\mu D_\mu \Phi) D^\nu \Phi + \frac{1}{2} D^\rho \Phi D_\rho \Phi \xi^\nu. \quad (13.74)$$

它满足协变守恒方程 $D_\mu J^\mu = 0$ 。

进而，根据第十一章第二节的讨论，这个标量场在黑洞外的空间截片 Σ 上的能量 $E(\Sigma)$ 为

$$E(\Sigma) = \int_\Sigma d^3x \sqrt{|h|} n_\mu J^\mu, \quad (13.75)$$

式中 h_{ij} 为 Σ 上的诱导度规， n^μ 是 Σ 的未来指向的法向量。现在，将 $D_\mu J^\mu$ 在如图 (13.4) 所示的阴影区域 V 上积分，假设在类空无穷远 i^0 处有 $D_\mu \Phi = 0$ ，从而运用弯曲时空中的高斯定理，即有

$$\begin{aligned} 0 &= \int_V d^4x \sqrt{|g|} D_\mu J^\mu \\ &= \int_{\Sigma_2} d^3x \sqrt{|h_2|} n_\mu^2 J^\mu - \int_{\Sigma_1} d^3x \sqrt{|h_1|} n_\mu^1 J^\mu + \int_{\mathcal{N}} dS_\mu J^\mu. \end{aligned} \quad (13.76)$$

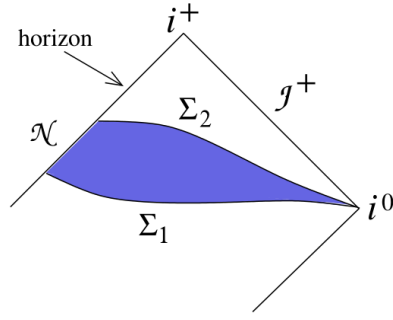


Figure 13.4: 在黑洞外的一个空间区域上积分。

其中 \mathcal{N} 为黑洞视界中夹在 Σ_1 和 Σ_2 之间的部分。重新组合一下以上结果，即有

$$E(\Sigma_1) - E(\Sigma_2) = - \int_{\mathcal{N}} dA dv \xi_H \cdot J. \quad (13.77)$$

式中 v 是前面引入的爱丁顿坐标，它可以看作是视界面上的类光时间， dA 是视界的空间面积元，并注意到底林向量 $-\xi_H$ 正好是类光超曲面 \mathcal{N} 的沿着黑洞“内向”的类光法向量。

由以上结果即知，标量场经黑洞散射以后，每单位类光时间黑洞的吸收功率为

$$\mathcal{P} = - \int_{\mathcal{N}} dA \xi_H \cdot J. \quad (13.78)$$

很明显

$$\xi_H \cdot J = -(\xi^\mu D_\mu \Phi)(\xi_{H\nu} D^\nu \Phi) + \frac{1}{2} D^\rho \Phi D_\rho \Phi (\xi \cdot \xi_H). \quad (13.79)$$

当限制在视界面上时这个式子的最后一项其实为零，因为这时候有 $\xi \cdot \xi_H = 0$ 。为了看清楚这一点，我们注意到

$$\xi \cdot \xi_H = g_{tt} + \Omega_H g_{t\phi} \Rightarrow \xi \cdot \xi_H|_{r_+} = \left(g_{tt} - \frac{g_{t\phi}^2}{g_{\phi\phi}} \right)_{r=r_+} = 0, \quad (13.80)$$

最后的等于零是因为，在视界面上 $g_{t\phi}^2 - g_{tt}g_{\phi\phi} = \Delta \sin^2 \theta = 0$ 。这意味着，我们可以把吸收功率表达成

$$\mathcal{P} = \int_{\mathcal{N}} dA (\xi^\mu D_\mu \Phi)(\xi_{H\nu} D^\nu \Phi) = \int_{\mathcal{N}} dA \frac{\partial \Phi}{\partial v} \left(\frac{\partial \Phi}{\partial v} + \Omega_H \frac{\partial \Phi}{\partial \chi} \right). \quad (13.81)$$

上式最后利用了 $\xi = \partial_t = \partial_v$, $\xi_H = \partial_v + \Omega_H \partial_\chi$ 。

我们可以将标量场 Φ 按照角动量模式展开，在爱丁顿-芬克斯坦坐标中，即是

$$\Phi = \Phi_0(r, \theta) \cos(\omega v - m\chi), \quad (13.82)$$

式中 ω 为标量场的频率，取 $\omega > 0$ ，另外，由于 χ 为一个角度坐标，具有 2π 的周期性，这就要求角动量模式 m 只能取整数，即 $m \in \mathbf{Z}$ 。将这个展开代入 (13.81) 式，并对时间 v 进行周期平均，就可以得到黑洞的平均吸收功率 $\overline{\mathcal{P}}$ ，为

$$\overline{\mathcal{P}} = \frac{1}{2} \left[\int dA \Phi_0^2(r_+, \theta) \right] \omega(\omega - m\Omega_H). \quad (13.83)$$

很显然，对于足够高频的波，这个吸收功率大于零，这时候正如我们期望的，黑洞吸收能量。但是，假设波的频率相比于角动量来说较小，满足

$$\omega < m\Omega_H \quad (13.84)$$

那这个吸收功率就是负的！从而这时候散射波实际上从黑洞中提取了能量！这就是场论版本的彭罗斯过程。

13.5.2 黑洞无毛定理

克尔黑洞是一种稳态黑洞，所谓稳态，指的是存在一个基林向量场 ξ ，并且在空间无穷远处， ξ 是类时向量。值得指出的是，这并不意味着时空全局都是稳态时空，就比如克尔黑洞，显然在视界之内 ξ 不是类时的，因此按照我们前面章节中对稳态时空的定义，克尔黑洞并不是全局的稳态时空，虽然在视界之外它的确是稳态时空，因此克尔黑洞依然称作稳态黑洞。

霍金在 1972 年证明，如果一个黑洞是稳态的，那么它必定是静态的（即在视界之外为静态）或者轴对称的。因此，如果一个稳态黑洞在旋转，那它就必须是轴对称的。这一观察结果可能令人费解，毕竟，应该有可能在黑洞外放置一些非对称的物质分布，并让它以非对称的方式潮汐扭曲视界。按理来说，这将产生一个仍然稳态和旋转的黑洞，但不是轴对称的。然而，霍金和哈特尔 (Hartle)(1972) 已经证明，事实上，这是错误的！原因是这样的物质分布会给黑洞施加扭矩，迫使它的旋转速度降下来直至成为静态，这个过程当然是非稳态的。因此，这样的情况会破坏稳态的条件。

在外部没有任何物质的情况下，稳态黑洞有一个极其简单的描述：首先，如果黑洞是静态的，那么它必须是球对称的，并且只能用施瓦西解来描述。这个美丽的唯一性定理，是第一个此类定理，由 Werner Israel 于 1967 年确立。这意味着在没有角动量的情况下，完全的引力坍缩必然会导致施瓦西黑洞。这似乎令人费解，因为最初的坍缩恒星可能是非常非球形的。1972 年，理查德·普莱斯 (Richard Price) 阐明了非球形恒星在引力坍缩过程中摆脱高阶多极矩的机制：这些多极矩辐射出去了，要么辐射到无穷远，要么辐射进黑洞。在辐射逐渐消失后，这个黑洞就稳定下来，变成了最终的球形。

其次，如果黑洞是轴对称的，那么它一定是克尔黑洞。Israel 唯一性定理的这种扩展是由 Brandon Carter(1971) 和 D.C.Robinson(1975) 确立的。

最后，黑洞唯一性定理可以推广到黑洞携带电荷的情况。这时，如果黑洞是静态的，那么它一定是 RN 黑洞 (Israel, 1968 年)。而如果它是轴对称的，那么它一定是稍后即将介绍的克尔-纽曼黑洞 (Mazur, 1982; Bunting, 未发表)。

我们看到，孤立的黑洞可以通过三个参数来唯一而完整地表征：质量、角动量和电荷，不需要其他参数。而这一非凡的性质正是约翰·惠勒著名短语“黑洞没有头发”的由来。惠勒的这句话后来被人们称作黑洞无毛定理。当然，实际上应该说黑洞三毛定理才准确。

基于此，钱德拉塞卡 (1987) 才有充分的理由写道：黑洞是质量从几个太阳质量到数百万太阳质量不等的宏观物体。在某种程度上，它们可以被认为是稳态的和孤立的，在这一程度上，他们都是，每一个都是，由克尔解精确地描述的。这是我们唯一一个对宏观物

体进行精确描述的例子。宏观物体, 正如我们在周围看到的那样, 受到各种力的支配, 这些力源于各种物理理论的各种近似。相比之下, 构建黑洞的唯一元素是我们对空间和时间的基本概念。因此, 几乎根据定义, 它们就是宇宙中最完美的宏观物体。由于广义相对论为它们的描述提供了一个唯一的双参数解族, 因此它们也是最简单的对象。

克尔-纽曼黑洞

根据上面所谈到的黑洞无毛定理可知, 在渐近平坦的四维时空中, 最具一般性的黑洞必定为克尔-纽曼黑洞, 它是爱因斯坦-麦克斯韦方程的解。这个解是纽曼在 1965 年发现的, 其度规为

$$ds^2 = -\frac{\rho^2 \Delta}{\Sigma} dt^2 + \frac{\Sigma}{\rho^2} \sin^2 \theta (d\phi - \omega dt)^2 + \frac{\rho^2}{\Delta} dr^2 + \rho^2 d\theta^2. \quad (13.85)$$

其中 $\rho^2 = r^2 + a^2 \cos^2 \theta$, $\Delta = r^2 - 2GMr + a^2 + Q^2$, $\Sigma = (r^2 + a^2)^2 - a^2 \Delta \sin^2 \theta$, $\omega = a(r^2 + a^2 - \Delta)/\Sigma$ 。与这个度规伴随的电磁场为

$$A_\mu dx^\mu = -\frac{Qr}{\rho^2} (dt - a \sin^2 \theta d\phi). \quad (13.86)$$

这里 Q 为黑洞所带的电荷。很显然, 当 $Q = 0$ 时, 克尔-纽曼解退化为克尔解。

类似于前面关于克尔解的讨论, 不难得到克尔-纽曼黑洞的内外视界 r_\pm 对应函数 Δ 的零点, 即有

$$r_\pm = GM \pm \sqrt{(GM)^2 - a^2 - Q^2}. \quad (13.87)$$

类似的, 可以得到克尔-纽曼黑洞的旋转角速度为 $\Omega_H = a/(r_+^2 + a^2)$, 表面引力为 $\kappa = \frac{r_+ - r_-}{2(r_+^2 + a^2)}$, 视界面积为 $A = 4\pi(r_+^2 + a^2)$, 这些结果均与克尔黑洞情形完全类似。同样, 存在基林向量场 $\xi_H = \partial_t + \Omega_H \partial_\phi$, 视界面为它的基林视界。进而可以定义一个静电势 Φ

$$\Phi \equiv -A_\mu \xi_H^\mu. \quad (13.88)$$

不难发现, 在视界面上的静电势为

$$\Phi_H \equiv \Phi|_{r=r_+} = \frac{r_+ Q}{r_+^2 + a^2}. \quad (13.89)$$

很显然, Φ_H 为一个常数。

进而可以将黑洞热力学第一定律推广为

$$\delta M = \frac{\kappa}{8\pi G} \delta A + \Omega_H \delta J + \Phi_H \delta Q. \quad (13.90)$$

14. 再谈能量动量张量

所有物质的共同特性是，有一个守恒的能量动量张量，通过这个张量，物质才能发出引力场并弯曲时空。物质的能动张量并非任意的，而是要满足一些合理的物理要求，比方说，任何观察者测到的能量密度都应该大于等于零。这些物理要求就是所谓的能量条件，它们就是本章的核心话题。另外，本章还要讨论一个物理模型，即所谓的理想流体模型。

14.1 能量条件

14.1.1 能动张量的本征值分解

压强

我们知道 $T^{ij}, i, j = 1, 2, 3$ 为动量流密度，但其实还有另外一个看待它们的角度。为了说清楚这一点，不妨假设我们是在平坦的闵可夫斯基时空中考察问题。这时候，能动张量所满足的守恒方程为 $\partial_\mu T^{\mu\nu} = 0$ 。取一个空间区域 V ，则这个区域内的总协变动量 p_V^ν 为 $p_V^\nu = \int_V d^3x T^{0\nu}(x)$ 。根据守恒方程 $\partial_\mu T^{\mu\nu} = 0$ 有， $\frac{d}{dt} p_V^\nu = \int_V d^3x \partial_t T^{0\nu}(x) = - \int_V d^3x \partial_i T^{i\nu} = - \int_{\partial V} dS_i T^{i\nu}$ ，式中 ∂V 表示区域 V 的边界， dS_i 为三维空间的面积元矢量。取 $\nu = j$ ，则有

$$\frac{d}{dt} p_V^j = - \int_{\partial V} dS_i T^{ij}. \quad (14.1)$$

根据牛顿第二运动定律可知， T^{ij} 为单位面积上的力，也就是压强！具体来说， T^{ij} 为作用在 i 方向的面积元 dS_i 上，并向着 j 方向推的压强。根据 $T^{ij} = T^{ji}$ 可知，这个压强也等于作用在 j 方向的面积元 dS_j 上，并向着 i 方向推的压强。

能动张量的本征值分解

将能量动量张量 $T^{\mu\nu}$ 看作是一个 4×4 的对称矩阵, 可以求解本征方程 $g_{\nu\sigma}T^{\mu\nu}e^\sigma = \lambda e^\mu$ 得到它的四个本征向量。最常见的情形是, 有一个本征向量类时 (记作 e_0^μ), 三个本征向量类空 (记作 $e_i^\mu, i = 1, 2, 3$), 其它情形 (比方说有本征向量类光的情形) 比较罕见, 这里不作讨论。作为对称矩阵的本征向量, 这四个向量 $e_a^\mu, a = 0, 1, 2, 3$ 当然两两正交, 因此可以正交归一化为

$$g_{\mu\nu}e_a^\mu e_b^\nu = \eta_{ab}, \quad (14.2)$$

式中 $\eta_{ab} = \text{diag}\{-1, 1, 1, 1\}$ 是标准的闵可夫斯基度规。将方程 (14.2) 看作矩阵方程, 然后两边取逆, 即可以由正交归一关系得到如下完备性关系,

$$\eta^{ab}e_a^\mu e_b^\nu = g^{\mu\nu}. \quad (14.3)$$

假设记 $T^{\mu\nu}$ 的四个相应本征值为 ρ, p_1, p_2, p_3 , 则线性代数的知识告诉我们, $T^{\mu\nu}$ 必定可以分解成

$$T^{\mu\nu} = \rho e_0^\mu e_0^\nu + p_1 e_1^\mu e_1^\nu + p_2 e_2^\mu e_2^\nu + p_3 e_3^\mu e_3^\nu. \quad (14.4)$$

根据上面关于压强的讨论可知, $p_i, i = 1, 2, 3$ 作为三个类空方向的本征值, 它的物理含义当然也是三个类空方向的压强 (注意这里的 p_i 不是表示粒子动量), 而 ρ 作为 $a = 0$ 方向 (也就是类时方向) 的本征值, 它的含义当然就是能量密度。

特别的, 如果某种物质满足 $p_1 = p_2 = p_3 \equiv p$, 则它就可以看作是一种理想流体, e_0^μ 即可以看作是这种理想流体的流速场, 记作 $e_0^\mu = u^\mu$ 。这时候, 能动张量可以写成

$$\begin{aligned} T^{\mu\nu} &= \rho e_0^\mu e_0^\nu + p(e_1^\mu e_1^\nu + e_2^\mu e_2^\nu + e_3^\mu e_3^\nu) \\ &= \rho e_0^\mu e_0^\nu + p(g^{\mu\nu} + e_0^\mu e_0^\nu) \\ &= (\rho + p)u^\mu u^\nu + pg^{\mu\nu}, \end{aligned} \quad (14.5)$$

式中第二行的等于号是利用了完备性关系。最终的表达式 $T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}$ 就称作理想流体的能动张量。

上述理想流体定义的解释在于: 在跟着流体一起以速度 u^μ 运动的参考系中, 也就是在四个正交归一矢量场 $\{e_a^\mu\}$ 所构成的活动标架中 (以加 ' 来表示), 流体的能动张量 T'^{ab} 具有球对称的形式, 即 $T'^{00} = \rho, T'^{0i} = T'^{i0} = 0, T'^{ij} = p\delta_{ij}$, 变换到实验室系中, 即有 $T^{\mu\nu} = T'^{ab}e_a^\mu e_b^\nu = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}$ 。不难看出, 对于理想流体, 有

$$T = T^\mu{}_\mu = 3p - \rho. \quad (14.6)$$

也可以利用 e_a^μ 将任何一个四维速度 v^μ 分解成

$$v^\mu = v^a e_a^\mu, \quad (14.7)$$

注意, 记号 v^μ 和 v^a 不止是换了一下指标, 两者根本就是不同的。则很显然 (利用正交归一关系)

$$g_{\mu\nu}v^\mu v^\nu = \eta_{ab}v^a v^b. \quad (14.8)$$

从而对于类时的速度场, 可以归一为

$$v^\mu = \gamma(e_0^\mu + ae_1^\mu + be_2^\mu + ce_3^\mu), \quad \gamma = (1 - a^2 - b^2 - c^2)^{-\frac{1}{2}}, \quad (14.9)$$

式中 $a^2 + b^2 + c^2 < 1$ 。而对于一个未来指向的类光向量 k^μ , 此时它的归一化可以相差一个任意的常数, 因此可以取

$$k^\mu = e_0^\mu + a'e_1^\mu + b'e_2^\mu + c'e_3^\mu, \quad a'^2 + b'^2 + c'^2 = 1. \quad (14.10)$$

很显然, 在相差一个无穷大的正归一化常数的前提下, 类光向量可以看成是类时向量的极限, 即将式 $a^2 + b^2 + c^2 < 1$ 中的小于号改成小于等于号 $a^2 + b^2 + c^2 \leq 1$ 。

14.1.2 弱能量条件和类光能量条件

弱能量条件

弱能量条件说的是, **时空中任何观察者所测到的能量密度必须非负**。假设以 v^μ 表示观察者的四维速度 (它当然是一个类时向量), 则弱能量条件说的是, 对于任意这样的类时 v^μ , 必有

$$T_{\mu\nu}v^\mu v^\nu \geq 0. \quad (14.11)$$

由于类时向量可以任意接近于类光向量, 所以根据连续性, 上述条件对于类光向量也成立。代入 (14.4) 式和 (14.9) 式, 即有

$$\rho + p_1 a^2 + p_2 b^2 + p_3 c^2 \geq 0. \quad (14.12)$$

由于 a, b, c 任意, 所以可以取 $a = b = c = 0$, 即有 $\rho \geq 0$ 。或者取 $b = c = 0, a \rightarrow 1$ (即 a 趋于 1, 注意到 $a^2 + b^2 + c^2 \leq 1$), 即有 $\rho + p_1 a^2 \geq 0 \rightarrow \rho + p_1 \geq 0$ 。因此 $\rho + p_1 \geq 0$ 。对 p_2, p_3 也有类似的结论。综上可知, 弱能量条件意味着

$$\rho \geq 0, \quad \rho + p_i \geq 0. \quad (14.13)$$

特别的, 负压强是允许的, 只要它别负得太多, 即满足 $p_i \geq -\rho$ 。

弱能量条件要求能量密度大于等于零。但是, 在有些情况下, 负能量密度也是有物理意义的, 比如负的宇宙学常数。我们知道, 宇宙学常数 Λ 对应于能动张量 $T_{\mu\nu}^\Lambda = -\frac{\Lambda}{8\pi G}g_{\mu\nu}$, 即对应能量密度 ρ_Λ 为

$$\rho_\Lambda = \frac{\Lambda}{8\pi G}. \quad (14.14)$$

因此负的宇宙学常数就对应负能量密度。但是，负宇宙学常数的确是有物理意义的，比如它对应反德西特时空。因此，弱能量条件是可以破坏的，反德西特时空就破坏了弱能量条件。

我们也可以用标量场来构造破坏弱能量条件的例子。我们知道，标量场 ϕ 的能动张量为

$$T_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - g_{\mu\nu}\left(\frac{1}{2}\partial^\sigma\phi\partial_\sigma\phi + \mathcal{U}(\phi)\right). \quad (14.15)$$

因此弱能量条件相当于

$$(v^\mu\partial_\mu\phi)^2 + \frac{1}{2}\partial^\sigma\phi\partial_\sigma\phi + \mathcal{U}(\phi) \geq 0. \quad (14.16)$$

式中第一项总是正的，但是第二项可正可负。但事实上，前两项结合起来总是正的，为了看清楚这一点，我们引入如下与 v^μ 正交的矢量 X_μ

$$X_\mu = \partial_\mu\phi + v_\mu(v^\nu\partial_\nu\phi), \quad (14.17)$$

满足 $v^\mu X_\mu = 0$ 。则由于 v^μ 类时，所以 X_μ 必定类空（或类光），因此满足 $X_\mu X^\mu \geq 0$ 。不难验证，弱能量条件 (14.16) 可以重写为

$$\frac{1}{2}(v^\mu\partial_\mu\phi)^2 + \frac{1}{2}X_\mu X^\mu + \mathcal{U}(\phi) \geq 0. \quad (14.18)$$

现在，前两项都是正的了！很明显，如果势能 $\mathcal{U}(\phi) \geq 0$ ，那么弱能量条件就是满足的。但是，如果 $\mathcal{U}(\phi) < 0$ ，那弱能量条件就会被破坏（只要取一个 $\partial_\mu\phi$ 足够小的平缓场位形，以使得前两项足够小）。然而，物理上并没有什么强烈的理由禁止 $\mathcal{U}(\phi) < 0$ 的势函数出现，所以弱能量条件是可以破坏的！

类光能量条件

所谓的类光能量条件，和弱能量条件类似，只是将向量 v^μ 改成任意未来指向的类光向量 k^μ ，即

$$T_{\mu\nu}k^\mu k^\nu \geq 0. \quad (14.19)$$

由于弱能量条件中的 v^μ 本身可以为类光向量，所以很显然，类光能量条件比弱能量条件更弱，换言之，弱能量条件蕴含着类光能量条件。代入 (14.4) 式和 (14.10) 式，即有

$$\rho + p_1 a'^2 + p_2 b'^2 + p_3 c'^2 \geq 0, \quad (14.20)$$

取 $b' = c' = 0, a' = 1$ ，则有 $\rho + p_1 \geq 0$ 。类似的也有 $\rho + p_i \geq 0$ 。因此，类光能量条件意味着

$$\rho + p_i \geq 0. \quad (14.21)$$

14.1.3 强能量条件和主能量条件等等

强能量条件

所谓的强能量条件，即是对任意类时向量场 v^μ ，要求

$$R_{\mu\nu}v^\mu v^\nu \geq 0. \quad (14.22)$$

实际上，这个条件是为了确保初始时相互平行的类时测地线簇经过足够长时间以后总是会开始汇聚，这可以看作是要求引力总是作为一种吸引力出现。利用爱因斯坦场方程的等价形式 $R_{\mu\nu} = 8\pi G(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu})$ ，可知，强能量条件等价于

$$(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu})v^\mu v^\nu \geq 0. \quad (14.23)$$

或者 $T_{\mu\nu}v^\mu v^\nu \geq -\frac{1}{2}T$ 。

值得注意的是，强能量条件并不一定比弱能量条件更强，当然也不一定比弱能量条件更弱，它只是一个不同的能量条件，之所以称之为强能量条件，完全是由于历史的原因。

代入 (14.4) 式和 (14.9) 式，即有

$$\gamma^2(\rho + p_1a^2 + p_2b^2 + p_3c^2) \geq \frac{1}{2}(\rho - p_1 - p_2 - p_3). \quad (14.24)$$

选取 $a = b = c = 0$ (从而 $\gamma = 1$)，即有 $\rho + p_1 + p_2 + p_3 \geq 0$ 。另外，将上面不等式写作

$$\gamma^2(\rho + p_1a^2 + p_2b^2 + p_3c^2) + \frac{1}{2}(p_1 + p_2 + p_3 - \rho) \geq 0, \quad (14.25)$$

然后取 $b = c = 0, a \rightarrow 1$ (从而 $\gamma \rightarrow +\infty$)，即知上式趋于 $\rho + p_1 \geq 0$ 。类似的，当然也有 $\rho + p_i \geq 0$ 。综上可知，强能量条件意味着

$$\rho + p_1 + p_2 + p_3 \geq 0, \quad \rho + p_i \geq 0. \quad (14.26)$$

很显然，强能量条件并没有蕴含弱能量条件。对于理想流体，这个条件即是

$$\rho + 3p \geq 0, \quad \rho + p \geq 0. \quad (14.27)$$

同样，不难找到强能量条件被破坏的物理情形。特别的，宇宙学常数 $\Lambda > 0$ 破坏了强能量条件。这是因为， $T_{\mu\nu}^\Lambda = -\frac{\Lambda}{8\pi G}g_{\mu\nu}$ ，与理想流体能动量张量比较可知，宇宙学常数对应

$$\rho_\Lambda = \frac{\Lambda}{8\pi G}, \quad p_\Lambda = -\rho_\Lambda, \quad (14.28)$$

式中 p_Λ 为宇宙学常数对应的压强，很显然，当 $\Lambda > 0$ 时，这是一种负压强！而且很显然破坏了 $\rho + 3p \geq 0$ 的要求。所以大于零的宇宙学常数破坏了强能量条件。实际上，这一点很容易理解，因为在 $\Lambda > 0$ 的德西特时空中，类时测地线簇并不会汇聚，而是会被宇宙膨胀效应拉开，所以当然破坏强能量条件。

主能量条件

所谓的主能量条件，其物理含义即是物质应该沿着类时或者类光世界线流动。更精确地说，这个要求相当于，对于任意协变速度为 v^μ 的观察者 (v^μ 为未来指向类时向量)，其观察到的能量流 J^μ 必定是一个未来指向的类时或者类光向量，这里

$$J^\mu = -T^{\mu\nu}v_\nu. \quad (14.29)$$

由此可知

$$J_\mu J^\mu \leq 0. \quad (14.30)$$

代入 (14.4) 式和 (14.9) 式可知，上式相当于要求

$$-\rho^2 + p_1^2 a^2 + p_2^2 b^2 + p_3^2 c^2 \leq 0. \quad (14.31)$$

取 $a = b = c = 0$ ，即有 $\rho^2 \geq 0$ ，进一步要求相应 J^μ 为未来指向的，则可推出 $\rho \geq 0$ 。反过来，如果取 $b = c = 0, a \rightarrow 1$ ，则有 $-\rho^2 + p_1^2 \leq 0$ ，进一步要求相应 J^μ 为未来指向的，则可以推出 $\rho \geq |p_1|$ 。类似的，当然也有 $\rho \geq |p_i|$ 。因此，主能量条件意味着

$$\rho \geq 0, \quad \rho \geq |p_i|. \quad (14.32)$$

平均类光能量条件

但是，以上能量条件均是对经典物质的要求，如果考虑到量子力学，那么以上所有能量条件都可能变得不再成立。具体来说，我们可以将量子物质通过下面的方程耦合到时空动力学，

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G \langle T_{\mu\nu} \rangle, \quad (14.33)$$

式中 $\langle T_{\mu\nu} \rangle$ 表示物质场能量动量张量在特定量子态上的算符期望值。量子场论的计算表明，以上所有能量条件在相当标准的量子场论中均不成立。然而，有一个更弱的能量条件在量子场论中却总是成立的，那就是所谓的平均类光能量条件。具体来说，可以证明：沿着任意一条无穷长的非时序类光测地线，任何合理的量子场论均满足如下条件

$$\int_{-\infty}^{+\infty} d\lambda \langle T_{\mu\nu} \rangle k^\mu k^\nu \geq 0. \quad (14.34)$$

式中 λ 是这条类光测地线的仿射参数， k^μ 是它的切向量，满足 $k^\mu = dx^\mu/d\lambda$ 。而所谓的非时序，指的是，这条类光测地线上任意两点都不可能用类时曲线连接起来。

14.2 理想流体

上一节讲到理想流体模型，其能动张量为

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}. \quad (14.35)$$

如果是在平坦时空中，那理想流体的能动张量即是

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + p\eta^{\mu\nu}. \quad (14.36)$$

那么，什么物理系统可以看作理想流体呢？这就是本节首先要讨论的问题。

14.2.1 能作为理想流体的两类物理系统

下面我们讨论两类能看作理想流体的物理系统。

无结构的质点气

我们要讨论的第一类系统是 (为简化起见, 限制在平坦时空中进行讨论): 由许多质点组成的理想气体, 这些质点的内部结构可以忽略, 并且它们仅仅只在相互碰撞的一瞬间有相互作用, 其余时间都是自由运动的。

根据本书第二章的相关讨论可知, 以上质点气的能动张量为

$$T^{\mu\nu} = \sum_n \frac{p_n^\mu p_n^\nu}{p_n^0} \delta^3(\mathbf{x} - \mathbf{x}_n). \quad (14.37)$$

在跟着流体一起运动的参考系中对上式进行统计平均, 注意到 $\langle p_n^i \rangle = 0$, $\langle \frac{p_n^i p_n^j}{p_n^0} \rangle = \frac{1}{3} \langle \frac{\mathbf{p}_n^2}{p_n^0} \rangle \delta_{ij}$, $i, j = 1, 2, 3$ (式中 \mathbf{p}_n 为粒子的三维动量, $\langle \cdot \rangle$ 表示统计平均), 易知, 平均以后的能动张量的确有球对称的形式 (因此满足理想流体的定义), 其中压强 p 和能量密度 ρ 分别为

$$\begin{aligned} p &= \frac{1}{3} \sum_{i=1}^3 \langle T^{ii} \rangle = \frac{1}{3} \sum_n \langle \frac{\mathbf{p}_n^2}{E_n} \delta^3(\mathbf{x} - \mathbf{x}_n) \rangle, \\ \rho &= \langle T^{00} \rangle = \sum_n \langle E_n \delta^3(\mathbf{x} - \mathbf{x}_n) \rangle, \end{aligned} \quad (14.38)$$

式中 $E_n = p_n^0$ 为粒子的能量。由相对论关系式 $\eta_{\mu\nu} p_n^\mu p_n^\nu = -m_n^2 \Rightarrow \mathbf{p}_n^2 + m_n^2 = E_n^2$ (m_n 为粒子质量) 可知 $\mathbf{p}_n^2 \leq E_n^2$, 进而可以推出

$$0 \leq p \leq \frac{1}{3} \rho. \quad (14.39)$$

对于冷的非相对论性气体, 可以作如下近似

$$E_n \simeq m_n + \frac{\mathbf{p}_n^2}{2m_n}. \quad (14.40)$$

故由 (14.38) 式, 可得

$$\rho \simeq \rho_0 + \frac{3}{2} p. \quad (14.41)$$

式中 ρ_0 为质量密度 (静能密度) 的统计平均值, 由下式给出

$$\rho_0 = \sum_n \langle m_n \delta^3(\mathbf{x} - \mathbf{x}_n) \rangle. \quad (14.42)$$

而对于热的极端相对论性气体, 我们有

$$E_n \simeq |\mathbf{p}_n| \gg m_n. \quad (14.43)$$

故由 (14.38) 式知

$$\rho \simeq 3p \gg \rho_0. \quad (14.44)$$

特别的, 光子的质量为零, 所以光子气当然是极端相对论性的, 从而对于光子气有 $\rho = 3p$ 。

一般来说 ρ 和 p 之间的函数关系就称作物态方程。很显然, 对于极端相对论性气体, 其物态方程为

$$p = w\rho, \quad w = 1/3. \quad (14.45)$$

当然, 这时候有 $T = T^\mu{}_\mu = 3p - \rho = 0$ 。而对于非相对论性理想气体, 其物态方程为 $p = \frac{2}{3}(\rho - \rho_0)$, 式中 $\rho - \rho_0$ 是理想气体动能密度的平均值。进一步, 根据非相对论性统计物理的能量均分定理, 可知 $\langle \frac{\mathbf{p}_n^2}{2m_n} \rangle = \frac{3}{2}k_B T$ (T 为温度), 进而可以算得

$$\begin{aligned} \rho - \rho_0 &= \sum_n \langle \frac{\mathbf{p}_n^2}{2m_n} \delta^3(\mathbf{x} - \mathbf{x}_n) \rangle = \sum_n \langle \frac{\mathbf{p}_n^2}{2m_n} \rangle \langle \delta^3(\mathbf{x} - \mathbf{x}_n) \rangle \\ &= \frac{3}{2}k_B T \sum_n \langle \delta^3(\mathbf{x} - \mathbf{x}_n) \rangle = \frac{3}{2}k_B T n(\mathbf{x}). \end{aligned} \quad (14.46)$$

式中 $n(\mathbf{x}) = \sum_n \langle \delta^3(\mathbf{x} - \mathbf{x}_n) \rangle$ 为粒子数密度的统计平均值。从而有, 非相对论理想气体的物态方程为

$$p = n(\mathbf{x})k_B T. \quad (14.47)$$

这当然是非相对论统计物理中的标准结论, 其中温度 T 不能太高, 否则粒子的平均动能就会大到需要考虑相对论效应。特别的, 在很多应用中, 可以近似认为 $T = 0$, 从而将这种极端冷的非相对论气体的物态方程写作

$$p = w\rho, \quad w = 0. \quad (14.48)$$

弯曲时空中的标量场

实际上, 弯曲时空中的标量场也能看作理想流体。为了看清楚这一点, 我们首先写出标量场的能动张量

$$T^{\mu\nu} = \partial^\mu \phi \partial^\nu \phi - g^{\mu\nu} \left[\frac{1}{2}(\partial\phi)^2 + \mathcal{U}(\phi) \right], \quad (14.49)$$

式中 $(\partial\phi)^2 \equiv \partial_\mu \phi \partial^\mu \phi$ 。为了将这个能动张量写成标准的理想流体形式, 我们令

$$\partial^\mu \phi \propto u^\mu \Rightarrow \partial^\mu \phi = K u^\mu. \quad (14.50)$$

为了确定其中的比例系数 K , 我们注意到 $(\partial\phi)^2 = K^2 u^2$, 利用 $u^2 = -1$, 即有 $K^2 = -(\partial\phi)^2$ 。进而可以将上面的标量场能动张量写成如下形式

$$T^{\mu\nu} = -(\partial\phi)^2 u^\mu u^\nu - g^{\mu\nu} \left[\frac{1}{2}(\partial\phi)^2 + \mathcal{U}(\phi) \right]. \quad (14.51)$$

与标准的 $T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}$ 比较, 即知

$$\begin{aligned} \rho + p &= -(\partial\phi)^2, \quad p = -\left[\frac{1}{2}(\partial\phi)^2 + \mathcal{U}(\phi) \right] \\ \Rightarrow p &= -\frac{1}{2}(\partial\phi)^2 - \mathcal{U}(\phi), \quad \rho = -\frac{1}{2}(\partial\phi)^2 + \mathcal{U}(\phi). \end{aligned} \quad (14.52)$$

这就是我们要讲的第二类能看作理想流体的物理系统。

14.2.2 理想流体的运动方程

平坦时空中质元组成的理想流体

在平坦时空中，理想流体的能动张量为

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + p\eta^{\mu\nu}. \quad (14.53)$$

这时，式中的归一化四维速度不妨取作

$$u^0 = \frac{1}{\sqrt{1-\mathbf{v}^2}}, \quad u^i = \frac{v^i}{\sqrt{1-\mathbf{v}^2}}, \quad i = 1, 2, 3. \quad (14.54)$$

在平坦时空中，能动量守恒方程为 $\partial_\mu T^{\mu\nu} = 0$ ，即

$$\partial^\nu p + (\rho + p)u^\mu \partial_\mu u^\nu + \{\partial_\mu [(\rho + p)u^\mu]\}u^\nu = 0. \quad (14.55)$$

将这个方程分开写作三维矢量方程和标量方程是方便的，其中的矢量方程就是通常流体力学中著名的欧拉方程的相对论版本，为了得到它，我们可以在方程 (14.55) 中取 $\mathbf{v} = i$ ，然后再利用 $\mathbf{v} = 0$ 时的方程 (14.55)，就可以得到

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{(1-\mathbf{v}^2)}{\rho + p} \left[\nabla p + \mathbf{v} \frac{\partial p}{\partial t} \right]. \quad (14.56)$$

至于标量方程，我们限制于在理想流体微观上由许多质点组成（也就是上面的第一类理想流体）的情形下研究它，因为这种情形下它可以写成特别简单的形式。在这种情形下，可以定义一个粒子数四维矢量 $n^\mu = n(x)u^\mu(x)$ ，式中 $n(x)$ 就是所谓的粒子数密度。粒子数守恒告诉我们， n^μ 满足如下连续性方程

$$0 = \partial_\mu n^\mu = \partial_\mu (nu^\mu). \quad (14.57)$$

或者写成

$$\partial_t \left(\frac{n}{\sqrt{1-\mathbf{v}^2}} \right) + \nabla \cdot \left(\frac{n\mathbf{v}}{\sqrt{1-\mathbf{v}^2}} \right) = 0. \quad (14.58)$$

另外，由于 $u^\nu u_\nu = -1$ ，所以还有 $0 = \partial_\mu (u^\nu u_\nu) = 2(\partial_\mu u^\nu)u_\nu$ ，即

$$(\partial_\mu u^\nu)u_\nu = 0. \quad (14.59)$$

为了得到标量方程，我们用 u_ν 去和 (14.55) 式收缩，再利用刚才的 (14.59) 式，即有

$$\partial_\mu [(\rho + p)u^\mu] = u^\mu \partial_\mu p. \quad (14.60)$$

下一步，我们利用连续性方程 (14.57) 来将 (14.60) 式改写作 $\partial_\mu [(\rho + p)u^\mu] = \partial_\mu \left[\left(\frac{\rho + p}{n} \right) nu^\mu \right] = nu^\mu \partial_\mu \left(\frac{\rho + p}{n} \right) = nu^\mu \partial_\mu (\rho/n) + npu^\mu \partial_\mu (1/n) + u^\mu \partial_\mu p$ ，进而有

$$u^\mu \partial_\mu (\rho/n) + pu^\mu \partial_\mu (1/n) = 0. \quad (14.61)$$

注意到 ρ/n 为每粒子能量, 而 $1/n$ 为每粒子占据的体积。而根据热力学第一定律, 我们有 $dE + pdV = TdS$, S 为热力学熵。定义 s 为每粒子的熵 (比熵), 则热力学第一定律可以写成

$$d(\rho/n) + pd(1/n) = Tds. \quad (14.62)$$

沿着理想流体的流线, 我们有微分 $d = u^\mu \partial_\mu$, 从而由上述热力学第一定律即有

$$u^\mu \partial_\mu (\rho/n) + pu^\mu \partial_\mu (1/n) = Tu^\mu \partial_\mu s. \quad (14.63)$$

代入 (14.61) 式, 即可以得到最终的标量方程为

$$u^\mu \partial_\mu s = 0 \Leftrightarrow \left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) s = 0. \quad (14.64)$$

即比熵 s 在任何随流体运动的点上不随时间变化! 这个结果有时候也称作能量方程或绝热方程。

所以, 对于平坦时空中由质元组成的理想流体, 其满足的基本微分方程有三个, 第一, 连续性方程 (14.57), 第二, 欧拉方程 (14.56), 第三, 绝热方程 (14.64)。除此之外, 为了研究这样的理想流体, 我们还需要合适的物态方程。

在常见的应用中绝热方程可以进一步简化。就像通常遇到的那样, 如果比熵 s 在某个初始时刻处处相同, 则根据绝热方程, 在此后的流动中 s 将保持处处相同并且不随时间变化。因此, 这时候就可以将绝热方程简单地写作

$$s = \text{常数}. \quad (14.65)$$

这时候根据热力学第一定律, 即有

$$d(\rho/n) + pd(1/n) = 0. \quad (14.66)$$

相对论性流体中的声速

作为一个例子, 让我们来计算在均匀静态的相对论性流体中的声速。假设在未扰动的状态中, $n = \bar{n}, \rho = \bar{\rho}, p = \bar{p}$ 均是常数, 而且 $\mathbf{v} = \bar{\mathbf{v}} = 0$ 。现在, 考虑声波引起的小扰动, 从而 $n = \bar{n} + \delta n, \rho = \bar{\rho} + \delta \rho, p = \bar{p} + \delta p, \mathbf{v} = 0 + \delta \mathbf{v}$ 。准确到扰动的一阶小量, 连续性方程 (14.58) 和欧拉方程 (14.56) 分别成为

$$\frac{\partial \delta n}{\partial t} + \bar{n} \nabla \cdot \delta \mathbf{v} = 0 \quad (14.67)$$

$$\frac{\partial \delta \mathbf{v}}{\partial t} = -\frac{\nabla \delta p}{\bar{p} + \bar{\rho}}. \quad (14.68)$$

另外, 由绝热方程 (14.66), 有

$$\begin{aligned} & -\frac{1}{\bar{n}^2} (\bar{p} + \bar{\rho}) d(\delta n) + \frac{1}{\bar{n}} d(\delta \rho) = 0 \\ \Rightarrow & -\frac{\bar{p} + \bar{\rho}}{\bar{n}} \delta n + \delta \rho = 0. \end{aligned} \quad (14.69)$$

定义

$$c_s^2 \equiv \left(\frac{\delta p}{\delta \rho} \right) = \left(\frac{\partial p}{\partial \rho} \right)_s, \quad (14.70)$$

式中的下标 s 表示绝热。则可以将方程 (14.69) 重写作

$$-\frac{\bar{p} + \bar{\rho}}{\bar{n}} \delta n + \frac{1}{c_s^2} \delta p = 0 \Rightarrow \frac{\delta p}{\bar{p} + \bar{\rho}} = \frac{c_s^2}{\bar{n}} \delta n. \quad (14.71)$$

代入 (14.68) 式, 即有

$$\frac{\partial \delta \mathbf{v}}{\partial t} = -\frac{c_s^2}{\bar{n}} \nabla \delta n. \quad (14.72)$$

联立此式和上面的 (14.67) 式, 即有

$$\left[\frac{\partial^2}{\partial t^2} - c_s^2 \nabla^2 \right] \delta n = 0. \quad (14.73)$$

这是一个标准的波动方程, 它表明声波的传播速度为 c_s , 恰如在非相对论性流体中那样。

在非相对论流体中声速远小于光速 1, 但是声速随着温度增高而增高, 因而值得检验一下是否在由高度相对论性质点 (例如高于 10^{13}K 的氢) 所组成的流体中, 声速 c_s 能否超过 1。此时, 由于 $p = \frac{1}{3}\rho$, 所以根据 c_s 的定义, 即有

$$c_s = \frac{1}{\sqrt{3}} < 1. \quad (14.74)$$

弯曲时空中的理想流体

对于弯曲时空中的理想流体, 其能动张量为

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + p g^{\mu\nu}. \quad (14.75)$$

相应的能动量守恒方程为

$$D_\mu T^{\mu\nu} = 0. \quad (14.76)$$

由此不难得到

$$D^\nu p + (\rho + p)u^\mu D_\mu u^\nu + \{D_\mu [(\rho + p)u^\mu]\}u^\nu = 0. \quad (14.77)$$

用 u_ν 去收缩这个方程, 并利用 $(D_\mu u^\nu)u_\nu = 0$ (成立的理由和平坦时空中的相应方程一样), 即有 $u^\nu D_\nu p = D_\mu [(\rho + p)u^\mu]$, 或者也可以写成

$$u^\mu D_\mu \rho + (\rho + p)D_\mu u^\mu = 0. \quad (14.78)$$

另外, 将 $u^\nu D_\nu p = D_\mu [(\rho + p)u^\mu]$ 反代入 (14.77) 式, 即有

$$(\rho + p)u^\mu D_\mu u^\nu = -(g^{\mu\nu} + u^\mu u^\nu)D_\mu p. \quad (14.79)$$

当然, 这就是弯曲时空中的欧拉方程。欧拉方程再加上上面的方程 (14.78) 就构成了弯曲时空中理想流动的基本运动微分方程。

15. 球对称恒星及其演化

前面的章节中我们系统讲述了黑洞这种最为奇异的天体，当然今天我们知道黑洞是存在的，然而由于广义相对论的黑洞解中总是存在时空奇性（即时空曲率发散的地方），而奇性的存在本身意味着广义相对论在此处失效，所以关于黑洞解是否描写了真实存在的天体（还是仅仅只有数学意义）就曾经是一个很难确定的难题（比如白洞通常就被认为是不存在的）。尤其是，如果黑洞真实存在，那它是如何形成的呢？这一章的目的之一就是要讨论这个问题。

结果表明，对黑洞如何形成的一个回答是，它可以作为恒星演化的可能归宿之一而出现。因此这就把我们带到一个同样重要而且有意思的课题，即利用广义相对论来探讨恒星的演化。当然，动态演化过程研究起来比较困难，但是考虑到恒星的演化通常是在天文学时间尺度上进行的，所以我们不妨从研究恒星的平衡态开始，具体来说，我们将从研究一颗静态球对称恒星开始。

15.1 星体结构微分方程

对于一颗静态球对称恒星，其外部的时空当然由施瓦西解所描述，所以关键是要研究星体的内部。而星体内部是有物质分布的，因此内部时空满足的显然不是真空爱因斯坦方程，而是有物质分布的爱因斯坦方程。

星体内部的物质分布可以看作是微观上由大量粒子所构成的理想流体，因此具有如下能量动量张量

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + pg_{\mu\nu}. \quad (15.1)$$

很显然 $T = g^{\mu\nu}T_{\mu\nu} = 3p - \rho$ 。由于星体被假定为静态的，从而描述物质分布的理想流体也是静态的，这意味着协变流速场 $u = u^\mu \partial_\mu$ 将正比于静态基林向量场 $\xi = \partial_t$ 。

根据第八章中的分析，静态球对称度规必定可以取成如下形式

$$ds^2 = -A(r)dt^2 + B(r)dr^2 + r^2d\Omega^2. \quad (15.2)$$

u 正比于 ∂_t 就意味着 u^μ 只有时间分量，进一步根据归一化条件 $g_{\mu\nu}u^\mu u^\nu = -1$ ，即可以定出在 (t, r, θ, ϕ) 坐标中，

$$u^\mu = A^{-\frac{1}{2}}(1, 0, 0, 0), \quad u_\mu = A^{\frac{1}{2}}(-1, 0, 0, 0). \quad (15.3)$$

代入 (15.1) 式，即得能动张量的非零分量为

$$\begin{aligned} T_{tt} &= \rho(r)A(r), & T_{rr} &= p(r)B(r) \\ T_{\theta\theta} &= p(r)r^2, & T_{\phi\phi} &= \sin^2\theta T_{\theta\theta}. \end{aligned} \quad (15.4)$$

另一方面，根据第八章的计算，静态球对称度规的里奇张量为

$$\begin{aligned} R_{tt} &= \frac{A''}{2B} + \frac{A'}{rB} - \frac{A'}{4B} \left(\frac{A'}{A} + \frac{B'}{B} \right) \\ R_{rr} &= -\frac{A''}{2A} + \frac{B'}{rB} + \frac{A'}{4A} \left(\frac{A'}{A} + \frac{B'}{B} \right) \\ R_{\theta\theta} &= 1 - \left(\frac{r}{B} \right)' - \frac{r}{2B} \left(\frac{A'}{A} + \frac{B'}{B} \right) \\ R_{\phi\phi} &= \sin^2\theta R_{\theta\theta}. \end{aligned} \quad (15.5)$$

进而由爱因斯坦场方程 $R_{\mu\nu} = \kappa(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T)$ 即可以得到

$$R_{tt} = \frac{A''}{2B} + \frac{A'}{rB} - \frac{A'}{4B} \left(\frac{A'}{A} + \frac{B'}{B} \right) = \kappa \frac{1}{2}(\rho + 3p)A \quad (15.6)$$

$$R_{rr} = -\frac{A''}{2A} + \frac{B'}{rB} + \frac{A'}{4A} \left(\frac{A'}{A} + \frac{B'}{B} \right) = \kappa \frac{1}{2}(\rho - p)B \quad (15.7)$$

$$R_{\theta\theta} = 1 - \left(\frac{r}{B} \right)' - \frac{r}{2B} \left(\frac{A'}{A} + \frac{B'}{B} \right) = \kappa \frac{1}{2}(\rho - p)r^2. \quad (15.8)$$

注意，只有三个方程，因为 $\theta\theta$ 分量和 $\phi\phi$ 分量所给出的方程是同一个。

考虑以上三个方程的组合 $\frac{R_{tt}}{A} + \frac{R_{rr}}{B} + \frac{2R_{\theta\theta}}{r^2}$ ，不难得到

$$\frac{1}{r^2} \left(1 - \left(\frac{r}{B} \right)' \right) = \kappa\rho. \quad (15.9)$$

积分这个式子，即可以得到 $\frac{r}{B} = r - 2Gm(r) + c$ ，式中 c 为积分常数， $m(r)$ 为

$$m(r) = 4\pi \int_0^r \rho(r')r'^2 dr'. \quad (15.10)$$

为了定出积分常数 c ，我们注意到度规在 $r=0$ 处必定是有限的，从而 $B(0)$ 有限，进而在 $\frac{r}{B} = r - 2Gm(r) + c$ 中取 $r=0$ ，即可以得到 $c=0$ ，即有

$$\frac{1}{B} = 1 - \frac{2Gm(r)}{r}. \quad (15.11)$$

假设恒星的半径为 R , 恒星外面施瓦西解的质量参数为 M , 即恒星外部的度规满足 $B = (1 - \frac{2GM}{r})^{-1}$, 则很显然, 为了衔接恒星内部和外部的解, 必然有

$$M = m(R) = 4\pi \int_0^R \rho(r)r^2 dr. \quad (15.12)$$

注意, 上面的 M 并不是星体内部所有物质分布的总能量 (虽然表达式看起来很像), 由于径向空间的弯曲, 星体内部所有物质分布的总能量应该由下式给出

$$4\pi \int_0^R \sqrt{B(r)}\rho(r)r^2 dr. \quad (15.13)$$

很显然, 这和上面的 M 并不同。实际上, 正如后面的章节中将会证明的, 上面的 M 是整个时空所有东西 (包括星体内部的物质分布和整个时空的引力场) 的总能量, 它当然和物质分布的能量是不同的。

一共三个独立方程, 我们才处理完一个, 为了继续进行下去, 我们把 (15.9) 式代入 (15.8) 式以消去其中的 ρ , 进而得到

$$1 - \left(\frac{r}{B}\right)' - \frac{r}{B} \left(\frac{A'}{A} + \frac{B'}{B}\right) = -\kappa p(r)r^2. \quad (15.14)$$

整理一下, 可以得到

$$\frac{r}{B} \frac{d \ln A}{dr} = \left(1 - \frac{1}{B}\right) + \kappa p r^2. \quad (15.15)$$

代入 (15.11) 式, 即可以得到

$$\begin{aligned} \frac{d \ln A}{dr} &= \frac{2G(m(r) + 4\pi p(r)r^3)}{r(r - 2Gm(r))} \\ &= \frac{2Gm(r)}{r^2} \left(1 + \frac{4\pi r^3 p(r)}{m(r)}\right) \left(1 - \frac{2Gm(r)}{r}\right)^{-1}. \end{aligned} \quad (15.16)$$

现在只剩下最后一个独立方程要处理了, 但是直接从前面那三个独立方程组合出最后这个方程的话有些繁琐, 一个简洁的办法是利用能动量守恒方程 $D_\mu T^{\mu\nu} = 0$ (由于爱因斯坦张量满足比安基恒等式, 所以这个式子必定成立) 来得到最后这个独立方程。第四章我们推导过矢量场的协变散度必定可以写成 $D_\mu A^\mu = \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} A^\mu)$, 完全类似的推导可以得到

$$D_\mu T^{\mu\nu} = \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} T^{\mu\nu}) + \Gamma_{\mu\rho}^\nu T^{\mu\rho}. \quad (15.17)$$

将 $D_\mu T^{\mu\nu} = 0$ 用到理想流体的能动张量, 并注意到 $D_\mu g^{\mu\nu} = 0$, 即有

$$0 = \frac{1}{\sqrt{|g|}} \partial_\mu (\sqrt{|g|} (\rho + p) u^\mu u^\nu) + \Gamma_{\mu\rho}^\nu (\rho + p) u^\mu u^\rho + g^{\mu\nu} \partial_\mu p. \quad (15.18)$$

注意到 u^μ 仅仅只有时间分量非零, 同时注意到球对称性的假设意味着 ρ 和 p 都仅仅是 r 的函数, 进而即可以从上式得到

$$0 = \frac{\rho + p}{A} \Gamma_{tt}^\nu + g^{r\nu} \frac{dp}{dr}. \quad (15.19)$$

根据第八章对克里斯托夫联络的计算可知, Γ_{μ}^{ν} 中只有 $\Gamma_{\mu}^r = \frac{1}{2} \frac{A'}{B}$ 非零, 因此在上式中取 $\nu = r$, 即可以得到

$$\frac{dp}{dr} = -\frac{1}{2}(\rho + p) \frac{d \ln A}{dr}. \quad (15.20)$$

代入 (15.16) 式, 即可以得到

$$\frac{dp}{dr} = -\frac{Gm(r)\rho(r)}{r^2} \left(1 + \frac{p(r)}{\rho(r)}\right) \left(1 + \frac{4\pi r^3 p(r)}{m(r)}\right) \left(1 - \frac{2Gm(r)}{r}\right)^{-1}. \quad (15.21)$$

这就是著名的, 相对论星体结构的 Tolman-Oppenheimer-Volkoff 方程。

Tolman-Oppenheimer-Volkoff 方程就是最主要的一个星体结构微分方程。另一个关键的结构方程是对 (15.10) 式的重新改写, 改写为

$$\frac{dm(r)}{dr} = 4\pi\rho(r)r^2. \quad (15.22)$$

很显然, 这个方程的初始条件是

$$m(0) = 0. \quad (15.23)$$

但是, 方程 (15.21) 和方程 (15.22) 还不够, 因为这只有两个方程, 但是未知的变量有 $\rho(r)$, $p(r)$ 和 $m(r)$ 三个, 所以还缺了一个方程。这个所缺的方程就是星体物质分布的物态方程。一般性的物态方程当然很复杂, 因为它可能和物质的比熵分布以及化学成分都有关系。但在最简单和最有用的情况下, 我们可以假定所谓的物态方程, 就是压强 p 作为能量密度 ρ 的一个给定函数关系, 即我们假定

$$p(r) = p(\rho(r)). \quad (15.24)$$

通常还要求物态满足下面的合理要求

$$\rho \geq 0, \quad p \geq 0. \quad (15.25)$$

补上了这样的物态方程, 方程的数目就够了。

为了联立求解星体结构方程 (15.21), (15.22) 以及物态方程 (15.24), 我们还需要知道初始条件, $m(r)$ 的初始条件为 $m(0) = 0$, 另一个初始条件就是 $\rho(0)$, 不妨假定

$$\rho(0) = \rho_0, \quad (15.26)$$

并记相应的 $p(0)$ 为

$$p(0) = p(\rho(0)) = p_0. \quad (15.27)$$

利用初始条件求解微分方程, 就可以得到函数 $\rho(r)$, $p(r)$ 以及 $m(r)$ 。有了这三个函数, 再代入方程 (15.11) 和方程 (15.16) 就能求出时空度中待定的 $A(r)$ 和 $B(r)$ 。

注意, 星体不是黑洞, 它内部不可能含有视界, 所以通常来说

$$A(r) > 0, \quad B(r) > 0, \quad (15.28)$$

换言之，必定有

$$2Gm(r) < r. \quad (15.29)$$

进一步，对于 $\rho \geq 0, p \geq 0$ 的合理物态，由 Tolman-Oppenheimer-Volkoff 方程可以知道，星体从内向外，其压强 $p(r)$ 必定是单调下降的，到星体的表面上，这个压强将下降为零。所以，星体的半径 R 可以定义为，使得下式成立的最小 R

$$p(R) = 0. \quad (15.30)$$

15.2 牛顿近似以及均匀密度星

牛顿近似

下面考虑星体结构微分方程的牛顿近似。这是指：1. 首先，时空弯曲可以忽略，比方说 $B(r)$ 近似等于 1，从而

$$2Gm(r) \ll r. \quad (15.31)$$

2. 其次，物质粒子的动能远远小于静止能量，由于压强是由动能导致的，从而这意味着

$$p \ll \rho \Rightarrow pr^3 \ll \rho r^3 \sim m(r). \quad (15.32)$$

不难看出，在牛顿近似下，Tolman-Oppenheimer-Volkoff 方程可以近似为

$$\frac{dp}{dr} = -\frac{Gm(r)\rho(r)}{r^2}. \quad (15.33)$$

这个方程当然也可以在牛顿引力的框架下推导出来，为此你只需要在星体内部考虑一个厚度为 dr 的薄球层，由球层内外压力差（内部压力高）与球层所包围物质对它的万有引力之间的平衡，也就是所谓压强梯度与星体自引力之间的平衡，就可以导出方程 (15.33)，如图 (15.1).

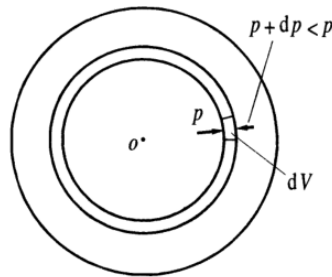


Figure 15.1: 由一个薄球层内外压力差与球层内部物质对它的万有引力的平衡可以导出方程 (15.33).

另外，同样不难看出，在牛顿近似下，方程 (15.16) 将近似为

$$\frac{d \ln A}{dr} = \frac{2Gm(r)}{r^2}. \quad (15.34)$$

实际上, 这个结果说明, 在牛顿近似下

$$\ln A = 2\varphi(r), \quad (15.35)$$

式中 $\varphi(r)$ 表示牛顿引力势。

为了看出这一点, 我们只需要注意到牛顿引力势 $\varphi(r)$ 满足泊松方程 $\nabla^2\varphi = 4\pi G\rho$ 。在球对称的情况下, 这个方程将成为

$$\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{d\varphi}{dr} \right) = 4\pi G\rho(r). \quad (15.36)$$

将这个方程积分一次, 即有 $r^2 \frac{d\varphi}{dr} = 4\pi G \int_0^r \rho(r') r'^2 dr' = Gm(r)$, 进而即有

$$\frac{d\varphi}{dr} = \frac{Gm(r)}{r^2}. \quad (15.37)$$

比较这个结果和 (15.34) 式, 即知 $\ln A = 2\varphi(r)$ 。

施瓦西内解

星体结构微分方程通常无法解析求解, 只能求数值解。不过, 对于一种最简单的假想物态, 最早在 1916 年施瓦西就已经求得了其解析解, 这就是所谓的施瓦西内解。这种最简单的情况就是所谓的均匀密度星, 即其物态方程是

$$\rho(r) = \rho_0 = \text{常数}. \quad (15.38)$$

很显然, 这时候

$$m(r) = \frac{4\pi}{3} \rho_0 r^3. \quad (15.39)$$

特别的, 总能量 M 为

$$M = \frac{4\pi}{3} \rho_0 R^3. \quad (15.40)$$

由于 ρ 为常数, 这时候 Tolman-Oppenheimer-Volkoff 方程可以写成

$$-\frac{p'(r)}{[\rho_0 + p(r)][\frac{\rho_0}{3} + p(r)]} = 4\pi Gr \left[1 - \frac{8\pi G\rho_0 r^2}{3} \right]^{-1}. \quad (15.41)$$

从星体表面向内积分 (因为星体表面的压强已知为 0), 即可以得到

$$\frac{p(r) + \rho_0}{3p(r) + \rho_0} = \left[\frac{1 - 8\pi G\rho_0 R^2/3}{1 - 8\pi G\rho_0 r^2/3} \right]^{1/2}. \quad (15.42)$$

代入 $\rho_0 = 3M/(4\pi R^3)$, 即可以得到 Tolman-Oppenheimer-Volkoff 方程的一种解析解

$$p(r) = \rho_0 \frac{(1 - 2GM/R)^{1/2} - (1 - 2GMr^2/R^3)^{1/2}}{(1 - 2GMr^2/R^3)^{1/2} - 3(1 - 2GM/R)^{1/2}}. \quad (15.43)$$

这时候, 星体中心的压强 p_0 为,

$$p_0 = \rho_0 \frac{1 - Y}{3Y - 1}, \quad (15.44)$$

式中

$$Y = (1 - 2GM/R)^{1/2}. \quad (15.45)$$

由于物质压强要大于零，所以从上面 p_0 的表达式可知，必然要求 $1/3 < Y < 1$ 。不难验证 $dp_0/dY < 0$ ，进而可知，中心压强 p_0 随着 GM/R 的增加而增加。这不难理解，因为在 R 一定时，若 M 越大，则星体的自引力就越强，为了抗衡自引力，星体的压强梯度就越大，由于 R 一定，则中心压强 p_0 就必然越大。反之，若 M 一定，则 R 变小，则为了产生所需的压强梯度也需要更大的 p_0 。

当 GM/R 大到使得 $Y = 1/3$ 时，显然就有 $p_0 \rightarrow +\infty$ 。因此对于均匀密度星，其 GM/R 有上限，由 $Y = 1/3$ 可知，这个上限为

$$(GM/R)_{\max} = 4/9. \quad (15.46)$$

在广义相对论中，星体质量上限的存在并非均匀密度星特有的结论，可以证明，只要假定 $\rho(r) \geq 0$ ，并且 $d\rho/dr \leq 0$ ，则任何球对称星体的 GM/R 都得以 $4/9$ 为上限！这个结论的证明可以参见温伯格《引力和宇宙学》第 11 章第 11.6 节。

实际上，普通星体的 GM/R 远小于 $4/9$ 。为做数值估算，恢复光速 c ，即考察 $\frac{GM/c^2}{R}$ 。以太阳为例，其 $GM_{\odot}/c^2 \approx 1.5\text{km}$ ，其 $R_{\odot} \approx 7 \times 10^5\text{km}$ ，所以，对于太阳

$$\frac{GM_{\odot}/c^2}{R_{\odot}} \approx 2 \times 10^{-6} \ll 4/9. \quad (15.47)$$

15.3 恒星演化

下面我们来简单概述恒星的演化过程，并弄清楚为何说黑洞是大质量恒星演化的可能归宿之一。我们将看到，恒星演化大体有三种可能的归宿，其一是最终成为白矮星，其二是成为中子星，最后一种可能归宿就是成为黑洞。对于演化形成白矮星的这种归宿，由于密度并不十分高，引力场也不足够强，所以实际上用牛顿引力就足够了。但是对于中子星和黑洞，广义相对论是不可或缺的。

在恒星形成之前，起初是一团密度不均匀的气体，以氢为主。在自引力的作用下，密度较大的地方就开始吸聚周围的气体，逐渐形成一个球对称的气团。同样是由于自引力，这个气团会不断收缩，气团收缩的过程会使得引力势能不断转化成热能，因此气团的温度 T 就不断升高。根据理想气体压强公式 $p = nk_B T$ (n 为粒子数密度)，随着温度的升高，气团的压强也不断增大，压强的梯度也不断增大，看来好像有可能和自引力达成平衡。但是，如果气团本身没有能源，一切热能都来自于引力势能转化的话，那这种平衡是不可能的。原因在于，即使某个时刻暂时平衡下来，但由于气团会通过热辐射损失能量，而它自身又没有能源，那它的温度就会降低，压强梯度就会减小，那就会继续收缩。总体来看就是，气团必须不断收缩，以使得引力势能不断转化成热辐射出去的能量。因此没有能源的话，平衡是不可能的。

经过一段时间的缓慢收缩，气团中心的温度和密度终于高到足以点燃热核反应（也就是氢聚合成氦的核聚变反应）的程度了。这个温度大约只要 1500 万度左右（以太阳为例）

就够了，当然，本来这个温度是不够高的，但是考虑到量子隧穿效应，它已经有不可忽视的概率使得两个氢核克服库伦排斥力而聚合成氦核了。气团中心附近（一个称作星核的中心球体）的这种核聚变反应能释放巨大的能量，从而提供能量来平衡热辐射损失的能量，这样，气团就能不再收缩，达成一种平衡态。这时候，气团就成为一颗恒星，其内的压强梯度满足平衡条件 (15.33)。

太阳目前就处在这种普通恒星阶段，据估计，太阳的星核部分氢变氦的聚变反应已经在稳定状态中持续了约 45 亿年，而且大约还能保持这种状态 50 亿年。

总有一天，星核内部的氢会全部聚变成氦，只有星核周围的一薄层氢还在聚变燃烧，如图 (15.2) 所示。当星核的温度尚未达到进一步点燃氦（使之再次聚变成更重元素）的核

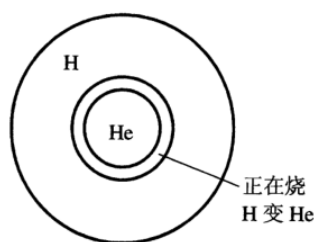


Figure 15.2: 星核内部氢燃烧后的情况.

聚变反应时，这时候星核就没有进一步的核聚变能源可用了，那就如氢尚未被点燃时类似，核心的氦球将在自引力作用下再次开始收缩，同时变得更热。这会使得星核周围薄层的氢燃烧加剧，从而导致星体更外层的部分膨胀和冷却，变成一个**红巨星**！“红”是由于星体表面的温度降低了，“巨”则是因为星体膨胀变得很大。红巨星阶段的星体会大量抛出其外层物质。

当然，氦核收缩导致的高温可能会进一步点燃氦的聚变反应，也就是烧氦变碳或氧的聚变反应，所释放的能量可以再次使得星核达到平衡。不过，这种靠氦燃烧维持的平衡所持续的时间远小于氢燃烧的情形。当星核的氦全部变成碳或氧时，星核又会再次收缩！

恒星晚年的命运因其抛出物质以后剩余的质量而定，对于质量较小的恒星，比如说太阳，星核的收缩已经不能提供足够的温度使碳和氧发生核聚变，因此靠核聚变来维持平衡就变得不再可能。那么，这时候还有什么力量提供足够的压强梯度来抗衡星体的自引力呢？

在经典物理的范畴的确是找不到抗衡力量了，不过，量子物理为我们提供了新的可能。这是因为，量子物理为我们找到了一种新的即使在零温下都能存在的压强产生机制，那就是所谓的电子简并压。具体来说，星体由氢、氦以及其它元素组成，之前的高温会使得原子都处于电离状态，因此星体可以看成是离子和电子的理想气体。以电子气为例，根据量子物理，电子气中每个电子的能量都会量子化成许多密集的能级，而由泡利不相容原理，每一个能级最多被两个电子占据，一个自旋向上、一个自旋向下。 $T = 0$ 时，电子会优先占据能量为零的最低能级，一个能级占满了就会接着往上占据更高的能级，最终，电子气中的所有电子将占满从最低能级到某个能量为 E_F 的能级的所有状态。 E_F 就叫费米能，其

值随密度的增大而增大，因为星核的密度很高，所以 E_F 非常大，以致于相比于 $k_B T$ 我们可以忽略温度的效应 ($k_B T \ll E_F$)，即可以看作零温 (即使从通常的观点来看，这时候星核的温度还是很高)。

也就是说，和经典物理的结论不同， $T = 0$ 时，电子不能全部集中在最低能态 (也就是能量为 0 的能级)，而是也有正的能量，因此也有运动，而且这个运动的能量 (虽然不是热运动的动能) 还可能很高，由此产生的压强就叫**电子简并压**，相应的电子气就叫**简并电子气**。在普通密度下， E_F 很小 (常见金属中自由电子气的 E_F 只有几个电子伏特)，因此电子简并压也很小。但是，前面说了，星核内部的 E_F 可以非常大，导致电子简并压也非常大，进而可能抗衡星体的自引力，使星体保持平衡，永不收缩 (因为这种抗衡不需要能源)。这种靠电子简并压维持的稳定星体就是**白矮星**。

一个孤立星体一旦演化成白矮星就不再会有重要的进一步演化，因此白矮星是恒星的可能归宿之一。因为温度比外界高，白矮星会不断辐射出能量，而又由于没有能源支持，所以热辐射会导致它的温度不断降低，直至温度接近外界温度而再也看不见，成了一颗所谓的“黑矮星”。

人类早已观测到白矮星的存在，天狼星 B 就是人类发现的第一颗白矮星。至于白矮星的“矮”字，则是指这些星的半径比普通恒星小得多 (部分原因是由于其密度高)，白矮星的半径约在 3 千到 2 万公里之间。据估计，初始质量小于 $6 \sim 8 M_\odot$ 的恒星都将经过红巨星阶段并抛出大量物质，最终成为剩余质量约为 $0.5 \sim 0.6 M_\odot$ 的白矮星。

当然，剩余质量越大的恒星其自引力就越强，因此，只有质量足够小的恒星其自引力才能被电子简并压所平衡。因此可以想见，白矮星的质量应该有一个上限，钱德拉塞卡最早求得了这个质量上限，大约为 $1.4 M_\odot$ ，称作钱德拉塞卡极限。

如果剩余质量大于钱德拉塞卡极限，那么电子简并压将不足以维持星体平衡，星体会在自引力作用下继续收缩，并一级一级地点燃星核内部的核聚变反应，直至将元素都聚合成铁和镍。这两者是结合得最紧密 (也就是核子的平均结合能最大) 的原子核，它们要再进行核聚变成为更重的原子核的话，那就不是释放能量了，而是要吸收能量，因此就不能成为星体平衡的能源了。于是，星核就会在自引力作用下急剧收缩，密度和温度都急剧升高。这时，自引力已经强到使得牛顿近似不再适用了，我们需要使用完整的 Tolman-Oppenheimer-Volkoff 方程 (15.21)。注意到 (15.21) 式右边的绝对值总是大于牛顿近似下 (15.33) 式右边的绝对值，因此这时候为达到平衡所需要的中心压强将更大，平衡将更难以实现。在这样的高温高密下，高能光子会将铁、镍原子核打碎成中子、质子和轻核，而且电子也将通过逆 β 衰变同质子反应，生成中子和中微子。当然，中微子会逸出星体之外。于是，中子在星核内将占据主导地位，注意，中子和电子一样，也是费米子，也服从泡利不相容原理，因此也会有一个费米能 E_F ，也会产生**中子简并压**！在星核达到原子核密度 ($\sim 10^{17} \text{kg} \cdot \text{m}^{-3}$) 时，中子的费米能也会远大于热能 $k_B T$ ，因此这时候的中子可以看成是零温的简并中子气。简并中子气的中子简并压也有可能抗衡星体自引力，使得星体达到平衡，这种靠中子简并压支撑的星体就称作**中子星**。

和白矮星类似，中子星当然也会有一个质量上限，超过这个质量上限星体自引力就会压倒中子简并压，使得星体进一步坍缩。不过，由于中子星的密度甚至超过了原子核密度，

人们对这种密度下的物态了解得不够确切，这给中子星质量上限的计算带来了一些不确定性。总之，按照今天人们的了解，中子星的质量上限大约为 $2 \sim 3M_{\odot}$ 。

中子星比白矮星还要小很多，典型中子星的半径只有 10km 。除了密度甚至超过原子核之外，中子星还有很多非常特别的表现：比如异常强的磁场、比如高速的自转（频率从 1Hz 到 1000Hz ）、比如接近光速的声速、甚至比如超流的内部等等。

中子星的理论模型最早是由 Oppenheimer 和 Volkoff 在 1939 年提出的。但是中子星的存在一直要到 1967 年发现脉冲星才得到观测证实。脉冲星是一种在地球上观测到的周期性电磁波信号的信号源，周期约为 1 秒甚至更短。对脉冲星的可靠解释就是：这是一颗旋转着的中子星，其表面的强磁场导致磁偶极辐射，辐射的方向性和旋转效应的结合就使得地球上接收到周期性信号。只有中子星由于半径足够小，而且自引力足够强，才能在如此高速的旋转中免于“散架”。

星核在形成中子星之前的收缩非常急剧，可以称之为引力坍缩，这种急剧的引力坍缩一旦被中子简并压所遏止，其强大的能量将表现为向外的冲击波，并使得星体的外层物质向四周飞出，形成能量急剧释放的**超新星爆发**。著名的超新星爆发遗址蟹状星云中就发现了中子星，这是对上述理论的支持。世界上最早记录下对超新星爆发观测的国家就是中国，其中宋史志卷九关于公元 1054 年（北宋期间）观测到的超新星爆发（SN1054）的记录特别受到国际上的重视，而蟹状星云也正是这次超新星爆发的遗址。

超新星爆发抛出的物质会形成星云（比如蟹状星云），这些星云中有大量的氢、也有一些比较重的元素（比如铁）。实际上，地球上的重元素就是这么来的，宇宙大爆炸的原初核合成中并没有这些重的元素，它们是来自于超新星爆发。爆发后的气团遗址当然又可以开始集聚，形成新一代的恒星，从而开始新一轮的恒星演化。据估计，我们的太阳就是这样的第三代恒星。

如果在超新星爆发抛出物质之后，星体的剩余质量依然大于中子星质量上限，那中子简并压也将无法抗衡星体自引力，实际上，这时候就没有任何力量可以抗衡星体自引力了，星体会急剧地坍缩下去，直至形成一个密度无穷大的奇点，外面再包裹一个事件视界，这也就是形成黑洞。所以黑洞是大质量恒星演化的可能归宿之一。但是，由于它涉及到无穷大的物质密度，所以早期人们一直很抗拒这种可能性。当然，如果密度大到（或者说尺寸小到）这个程度，那就需要考虑量子引力了，人们一般相信，考虑到量子引力以后奇点可能就不再是奇点了，不过，由于完整的量子引力理论今天还没有，所以奇点处到底会发生什么，今天依然是不知道的。

估算钱德拉塞卡极限

对于钱德拉塞卡极限的一个大致估计可以通过研究引力能和费米能之间的竞争来进行。为此，不妨假定半径为 R 的球对称星核包含 N 个电子和 N 个质子（所以整体是电中性的）。质子当然比电子重得多，因此承载了星核的绝大部分质量，而更轻的电子气则承载了大部分的压强。

既然电子之间相互排斥，我们不妨假定每个电子占据一个特征尺度为 λ 的体积， λ 就是电子的波长。由于半径 R 之内共有 N 个电子，所以 $\lambda \sim R/N^{1/3}$ 。利用德布罗意关系式

$p = 2\pi\hbar/\lambda$, 相应的电子特征动量 (也就是费米动量) p_F 为,

$$p_F \sim \hbar/\lambda \sim N^{1/3}\hbar/R. \quad (15.48)$$

假如星核被压缩, R 缩小, 从而 p_F 就会增大, 相应的费米能也会增大, 因此为了完成这种压缩就必须有一个外力做功, 这个外力当然就是星核的自引力。为了简单起见, 不妨假设这种压缩已经进行到如此地步, 以致于每个电子都在做高速的极端相对论运动, 换言之, 每个电子的能量为 $E = [(p_{FC})^2 + (m_e c)^2]^{1/2} \approx p_{FC}$ 。实际上, 最重的白矮星正好满足这个假定。因此, 星核中的总费米能 E_F (注意, 和上面不同, 这里不是每个电子的费米能) 就近似为

$$E_F \sim N(p_{FC}) \sim N^{4/3}\hbar c/R. \quad (15.49)$$

另一方面, 质子提供了绝大部分质量, 因此提供了绝大部分引力能 E_G , 它大约为

$$E_G \sim -G(m_p N)^2/R. \quad (15.50)$$

其中 m_p 是一个质子的质量, $m_p N$ 就是总质量。注意, 引力能是负的。


费米能和引力能都正比于 $1/R$, 但是两者对 N 的依赖关系不同。如果 N 足够大, 则负的引力能将占上风, 星核就会坍缩, 反之, 则是费米能占上风。不难得到, 临界的 N 值为

$$N_{\text{crit}} \sim (\hbar c/Gm_p^2)^{3/2} \sim 10^{57}. \quad (15.51)$$

相应的星核质量为

$$M_{\text{crit}} \sim N_{\text{crit}} m_p \sim M_{\odot}. \quad (15.52)$$

当然, 更精确的计算表明, 这个质量约为 $1.4M_{\odot}$, 这就是钱德拉塞卡极限。精确细致的计算参见温伯格《引力和宇宙学》第 11 章第 11.3 节。



16. 极简宇宙学

公元 2 世纪，托勒密创立地心说，告诉世人宇宙是一个嵌套 27 层的同心水晶球，地球是这个水晶球的中心。

一千多年后的 1543 年，哥白尼的《天体运行论》出版，它告诉世人，地球只是围绕太阳运动的行星之一。伽利略竖起天文望远镜证实了哥白尼的学说，他冲着望远镜笑道，这才是观星利器。

又过了 200 多年，赫歇尔架起更大号的望远镜，确定了圆盘状的银河系，并记录下了 5000 个星云。

时间来到 20 世纪 20 年代，哈勃叼着烟斗，瞄着胡克望远镜，测定了 10 万光年的银河系，看到了 90 万光年远的河外星系，他发现了星系间彼此远离的现象，宇宙在膨胀！

现在，请想象一下，你站在一片宁静的海滩上，脚下的沙粒细软无比。每一粒沙子都代表着宇宙中的一个星系。当你俯身拾起一粒沙子，这便是我们的银河系；而在你的指尖之间，还有无数其他的“沙粒”，它们是散布于浩瀚空间中的其他星系。这就是我们所处宇宙的一个缩影——广阔、神秘且充满了未知。

随着广义相对论的提出，以及随后哈勃发现宇宙膨胀的现象，人类对于宇宙的认识发生了革命性的变化。从大爆炸理论到暗物质与暗能量的研究，每一步进展都在帮助我们绘制出一幅更加精确的宇宙图景。

16.1 从热大爆炸到宇宙学原理

16.1.1 热大爆炸的概念

大爆炸理论的基本观察是遥远星系的红移。它们的光谱向较长波长方向偏移，离我们越远，偏移越大。这表明它们正在远离我们；它们与我们的距离在增加。根据广义相对论，这种现象不是因为星系本身在移动，而是空间本身在膨胀。随着空间的膨胀，通过空间传播的光线的波长也“膨胀”了。

在把星系看作“沙子”的大尺度上，宇宙的这种膨胀似乎是均匀的，整个宇宙以相同的速率膨胀，可以通过一个随时间变化的尺度因子 $a(t)$ 来描述， $a(t)$ 是宇宙尺寸的膨胀倍率，在宇宙的当前取 1，宇宙的膨胀速率由哈勃参数 $H \equiv \dot{a}/a$ 表征。通过测量当前的膨胀速率并结合对宇宙能量内含的知识，然后应用广义相对论方程，可以推算出 $a(t)$ 作为时间 t 的函数。

大爆炸理论是伽莫夫在 1948 年提出的，它的基础就是刚才说的宇宙膨胀模型，这是弗里德曼等人建立的。从今天的观点来看，大爆炸理论是膨胀宇宙模型的自然延伸，但在伽莫夫提出大爆炸理论的当时，弗里德曼模型预言的宇宙年龄明显与事实不符，在这样的基础上，伽莫夫的理论就完全不被学术界所接受，这种局面持续了近 20 年，直到这理论预言的宇宙微波背景被发现和证实。下面我们定性地谈一谈大爆炸理论背后的物理思想。

远古的宇宙中不可能有星系

在今天的宇宙中，物质是结团的，但是从这个现状出发，根据膨胀宇宙模型往前追溯，就不可避免地得到一个非常重大的推论：星系等物质集团不是亘古恒存的，而是在宇宙演化过程中产生的！

今天，星系内的平均物质密度比全宇宙的平均密度大 5 个数量级，也即是说，物质是局地地结团的。根据膨胀宇宙模型往前追溯到宇宙尺度因子 $a(t) < 10^{-2}$ 时，宇宙体积是今天的 $a^3 < 10^{-6}$ 倍，从而宇宙的平均密度将比今天大 6 个数量级以上，比今天的星系密度还大。这表明，那时候是不能有星系存在的，今天如星系这样的物质团块只能是宇宙演化的产物。

这结论深刻地指出早期宇宙并不是今天宇宙的缩影，在其演化过程中，宇宙的面貌发生过质的变化。那么，结团前的宇宙物质以什么状态存在？

星系是均匀宇宙气体碎裂的结果

一个自然的猜测是，一切星系结构形成之前的宇宙是有微小密度差别的均匀气体，宇宙的自引力不稳定性，也就是密度偏大的那些地方对周围的引力也会更大，因此会不断吸引周围的气体，使得密度进一步增大，这就会导致原来的均匀宇宙气体碎裂成团，进而演化出今天看到的一切物质集团结构。这是伽莫夫理论的一个要点。当然，在伽莫夫提出大爆炸理论的当时，这只能是一个猜测。

膨胀的宇宙来自大爆炸

如果结构形成之前的宇宙内含物是密度和温度都均匀的气体，那随着宇宙的膨胀，不仅是它的密度会逐步降低，而且其温度也会逐步降低，相应的推论就是，越早的时候宇宙的温度越高，如果追溯到非常早的时期，宇宙内含物的密度和温度将比今天高出几十个数量级。温度的大幅度变化会导致物质状态的相变，因此，早期宇宙的演化会因温度的巨大差别而有实质性的变化。

如果往前推到极端，宇宙的膨胀应该是从密度和温度都为无穷的状态开始的，如果宇宙是有限的，它起初的体积又很小，而温度又非常高，那么宇宙的行为就很像一次爆炸，因此在历史上，伽莫夫理论的反对者讥讽地称之为“宇宙大爆炸”，这就是这个名称的来源。

关于大爆炸理论，一个最常见的误解就是认为宇宙最早开始于一个点。实际上，我们今天的可观测宇宙在刚开始的时候的确很小，在大爆炸理论框架能处理的前提下，它很可能直径小于 1 毫米。但是，相比于早期极端短的时间，1 毫米其实并不是一个很小的尺度，至少远非一个点。另一方面，在可观测宇宙之外，我们的宇宙很可能还在延伸，甚至无限延伸，而如果今天的宇宙是无限的，那它很可能一开始就是无限的，那就更不是从一个点开始的了。当然，今天的理论还不能处理宇宙是如何开始的，因为最早的那个时刻需要量子引力才能描述，而目前还不存在完备的量子引力理论。

原子和分子是宇宙演化中产生的

宇宙大爆炸理论的另一个深刻结论是，原子分子也不是天然存在的，它们也是宇宙演化到一定阶段的产物。这里最奇妙的就是，这么小的原子和这么大的星系，两者竟然同样都是宇宙演化过程的产物，而不是永恒的天然存在。

为了理解这一点，需要追溯到星系形成前宇宙气体的构成，它是由中性原子构成的，其中氢原子的丰度约为 75%，是这气体中的主要成分。氢原子是由氢核与电子所构成的束缚系统，它的结合能为 13.6eV。因此这些中性原子气体的温度不能太高，当温度 T 高于 10^4K 时，粒子的平均热动能是 $k_B T \approx 1\text{eV}$ ，这时候能量超过 13.6eV 的光子大量存在，它们与氢原子的热碰撞将使原子电离。因此在这个温度以及更高的温度下，宇宙气体必定处于电离状态，即是等离子气体，其组分为原子核、自由电子、以及光子。原子和分子是随着宇宙膨胀降温的结果，之前是不存在的。

伽莫夫发现，在宇宙膨胀降温使等离子气体变成中性原子之后，光子与电子就几乎不散射了，因此这时候的光子将会自由地在宇宙中飞行，成为宇宙的背景，这种背景光子应该在今天依然存在而且可以被观测。这是伽莫夫理论的一个关键性预言。这个预言在 1965 年被观测证实，这是宇宙大爆炸理论从被否定到被重视和接受的转机。

化学元素也是演化的产物

在等离子状态的宇宙中，原子核是气体的主要成分，而它是由若干核子（也就是质子和中子）组成的，每一个核子在原子核中的平均结合能为 1MeV 的量级，所以，当气体的温度高于 1MeV（即 10^{10}K ）时，组分粒子间的热碰撞会使得原子核解离，原子核在这样的

高温下将不复存在。没有原子核就没有化学元素，因此，就连化学元素也是在宇宙膨胀降温到一定程度以后的产物。

原始汤

在宇宙温度高于 1MeV 的阶段，物质成分是比较原子核更基本的粒子，即质子、中子、电子以及光子等。当然，按照今天的认识，质子和中子也并不是物质最基本的组元，它们是由夸克和胶子所组成的复合粒子。按照粒子物理的标准模型，如果密度超过核子，而且温度也足够高，那么核子作为复合粒子也不能存在，这时候的宇宙内含物将是由夸克、轻子以及规范玻色子所组成的粒子气体。

因此大爆炸不是指一个瞬间发生的导致宇宙开始的爆炸事件，而是描述宇宙从一个高温、高密度状态开始迅速膨胀的过程。早期宇宙充满了几乎均匀的基本粒子“原始汤”，这些粒子处于长时间的热平衡状态，因此我们可以用少量的热力学变量来描述早期宇宙的状态，这使得宇宙的时间演化可以计算出来。

事实上，应用描述高温下宇宙能量内含的粒子物理标准模型，可以得到 $a(t)$ 在约 140 亿年前接近于零。这标志着宇宙开始于 140 亿年前的这个时刻，即大爆炸开始的时刻。在这个时间点，宇宙的密度趋向于无穷大，称作大爆炸奇点，我们定义这个时刻为 $t = 0$ 时刻。但是实际上，当密度达到普朗克密度 $\rho_{Pl} \approx 10^{96} \text{kg}/\text{m}^3$ 时，广义相对论本身就失效了，需要量子引力理论（目前还不存在）来描述这一非常短的时期。因此，严格来说，包括 $t = 0$ 在内的这些最早的时间其实必须从大爆炸理论中排除。

16.1.2 早期宇宙简史

下面我们概要性地简述一下早期宇宙的简史。因为热大爆炸，早期宇宙的温度很高，从而粒子就有很高的能量。因此为了描述这个时期的物质，我们需要粒子物理的知识。粒子物理的标准模型具有 $SU(3)_c \otimes SU(2)_w \otimes U(1)_y$ 的规范对称性。在我们今天的低能世界中，标准模型的某些对称性是自发破缺了的。当宇宙的温度超过 100GeV(约 10^{15} 开尔文)时，就达到了标准模型的自然能标，这大约就是宇宙的年龄小于 10^{-11}s 时的情况。这时候粒子的原始汤由自由的无质量费米子（夸克和轻子）以及传递相互作用的无质量规范玻色子构成。当然，标准模型中还有一种粒子，即希格斯玻色子。

正是希格斯玻色子使得电弱对称性 ($SU(2)_w \otimes U(1)_y$) 发生了自发破缺，这也正是早期宇宙中发生的一种相变。电弱相变之后，电弱相互作用就变成了两种相互独立的相互作用：其一是由有质量的规范玻色子 W^\pm 和 Z^0 传递的弱相互作用，其二就是由无质量 γ 光子传递的电磁相互作用。通过电弱相变，费米子也获得了质量，这些质量来源于它们与希格斯玻色子的相互作用。电弱相变是当宇宙冷却到临界温度 $T_c \sim 160\text{GeV}$ 时发生的，大约发生在 $t \sim 10^{-11}\text{s}$ 时。

早期宇宙的另一种相变是 QCD(量子色动力学) 相变，也就是夸克-强子相变，大约发生在 $t \sim 10^{-5}\text{s}$ ，其临界温度约为 $T_c \sim 150\text{MeV}$ 。相变之前，夸克是自由的，相变之后它们就形成了强子，包括重子（例如核子中的中子 n 和质子 p ）和介子（例如 π 介子和 K 介子）。QCD 相变之后，宇宙的物质内容就从夸克胶子等离子体转变成了强子气。

根据量子场论，每一种粒子都有相应的反粒子，它的质量与自旋等等性质都和粒子相同，除了荷相反，这里的荷包括电荷和色荷等。如果某种粒子不带有荷，那它就是自己的反粒子，典型的比方光子。当温度 $T \gg m$ 时，这里 m 是粒子的质量，这时粒子和反粒子不断地相互湮灭，同时也不断地产生，宇宙中粒子和反粒子的数目大致是一样的。但是，当温度 $T \ll m$ ，粒子和反粒子依然相互湮灭，但是热能已经不足以产生粒子-反粒子对了。因此随着宇宙的不断冷却（由于宇宙膨胀），重的粒子与反粒子就趋向于相互湮灭，同时额外产生出更多更轻的粒子-反粒子。如果宇宙中粒子和反粒子的数目严格相当，那么冷却到今天将只有光子和中微子¹留下来。但是今天的宇宙中显然还存有重子物质（主要由质子 p 和中子 n 构成，比如我们的银河系），这就表明，在早期宇宙中，核子和电子必然比它们相应的反粒子要稍微多一些，这多出来的一些就存留到了今天。

如果再追问早期宇宙中重子与反重子这种不对称的来源，则最自然的回答可能是：宇宙开始时重子和反重子是一样多的，只是随之出现了破坏重子数守恒的物理过程，导致重子数略微超过了反重子数。当然，在今天的粒子物理标准模型中，重子数始终是守恒的，但是 1970 年代人们也提出了电、弱、强大统一理论 (GUT)，其中重子数并不守恒。研究表明，重子数不守恒加上宇宙极早期膨胀中对热平衡的短暂偏离就可以从原本正反粒子对称的宇宙中产生出过剩的质子和中子。当然，大统一理论目前还只是理论探讨，远没有得到实验的证实。

已知的，有质量但是质量最轻的，同时又参与强或者电弱相互作用的粒子，是电子，因此宇宙中最后的正反物质湮灭事件是电子-正电子湮灭，这大约发生在 $t \sim 1\text{s}$ 时，这时的温度约为 $T \sim m_e \sim 0.5\text{MeV} \sim 10^{10}\text{K}$ 。在这之前宇宙中仍然存在大量的正反电子对，然而由于温度和密度相比更早的时候已经明显降低，中微子以及反中微子与其它粒子的相互作用率远小于宇宙膨胀率，它们的平均自由飞行时间显著变长，近似成为与其它粒子没有相互作用的自由粒子，从而不再与其它粒子共处于热平衡中，这称之为中微子退耦，这些退耦的中微子一直存留到今天。常常把中微子退耦的时刻记作 t_{vd} ，相应的温度记作 T_{vd} 。因此，在最后的电子-正电子湮灭事件之后，宇宙中的反物质就只有反中微子了，因此之后的宇宙原始汤将由大量的光子、中微子、反中微子，以及少量剩余下来的质子、中子以及电子组成。

值得说明的是，除了粒子物理标准模型中的粒子以外，宇宙中很可能还含有暗物质粒子，它们的性质目前人们还不太清楚。我们暂且先忽略暗物质粒子。

宇宙诞生之后几分钟，约在温度 $T \sim 100\text{keV}$ 时，质子和中子就结合成了轻元素的原子核。这也就是所谓的大爆炸核合成 (Big Bang Nucleosynthesis, BBN)，以占通常物质的质量百分比衡量，它大约产生了 75% ^1H ，25% ^4He ， 10^{-42}H ， 10^{-43}He ，以及 10^{-97}Li （其它化学元素是后来主要在恒星中形成的）。这时候这些物质是完全离子化的，所有的电子都还是自由电子。在这种等离子体中，光子不断地被电子所散射，相邻两次散射之间的平均自由程是短的。这就意味着，这时候的宇宙对于光来说是不透明的。

直到诞生之后的 37 万年，宇宙才变得透明。这时候的温度大约为 $T \sim 3000\text{K}$ （~

¹对中微子振荡的观测表明，中微子也有很小的质量，但是，中微子的相互作用是如此微弱，以至于即使在低于中微子质量的温度下，中微子与反中微子也很少相互湮灭。

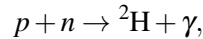
0.25eV), 这时电子和原子核结合形成了电中性的原子, 从而使得光子的平均自由程变得比可观测宇宙的半径更大, 宇宙就透明了。这个事件就称作宇宙复合过程 (recombination)。复合之后, 原初光子就可以自由地在空间飞行而几乎不被散射了, 因此也就从与其它粒子的热平衡中退耦出来了。今天我们可以观察到这些原初光子, 也就是所谓的宇宙微波背景 (cosmic microwave background, CMB), 它们是从早期宇宙来的光, 通过对它的观测, 我们可以“看见”大爆炸。

复合过程之后, 宇宙充满氢原子和氦原子气, 同时还有微量的锂。大约在宇宙诞生之后的几亿年, 从这个气体中形成了第一代恒星, 但是大多数气体形成了星际气体。大约在宇宙诞生之后的 7 亿年, 恒星发出的辐射重新离子化了这些星际气体。

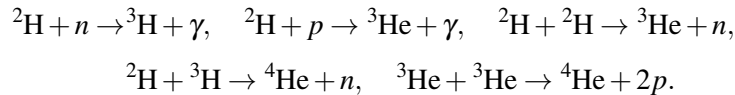
16.1.3 原初核合成

观测表明当前宇宙中氦约占宇宙总质量的 25%, 正如前面提到过的, 这主要是由早期宇宙的原初核合成 (大爆炸核合成) 所决定的, 恒星内部的核反应虽然也在生成氦, 但其值只占上述丰度的很小一部分。原初核合成涉及的宇宙温度大约在 10^{10}K 到稍低于 10^9K 之间。温度高于 10^{10}K 时, 即使质子中子结合成氦核也会被高能光子所击碎, 这也就是所谓的光分裂。而在 10^{10}K 到稍低于 10^9K 这个温度区间所涉及的物理知识早已在地球实验室内得到了确证, 因此人们对原初核合成的理论是很有信心的。

由于核合成时核子数密度不是很高, 同时宇宙的快速膨胀又导致合成反应时间很短 (核合成大约发生在 $t \sim 100\text{s}$ 时), 因此在原初核合成中只有快速的两粒子反应可以发生。首先是质子和中子结合成氘核 ${}^2\text{H}$, 多余的能量和动量由一个光子带走, 反应式就是



继而是一系列生成 ${}^3\text{H}$ 、 ${}^3\text{He}$ 、 ${}^4\text{He}$ 的后续反应, 反应式分别为



由于不存在质量数为 5 的稳定核素, 反应链至此中断。 ${}^4\text{He}$ 作为主要产物逐渐累积, 直到数量足够时核反应才得以继续, 结果就是在 ${}^4\text{He}$ 的基础上生成了很少一点 ${}^7\text{Li}$ 。由于不存在质量数为 8 的稳定核素, 反应链就终结了! 至于今天宇宙中各种比 ${}^7\text{Li}$ 更重的元素, 则是后来在恒星内部的核反应以及超星系爆发中形成的, 具体过程上一章已经讲过了。

原初核合成的整个反应链必须经过第一步质子和中子结合为氘核, 这就是所谓的“氘关卡”。氘核的结合能比氦核小很多, 氘核在温度降低到 $3 \times 10^9\text{K}$ 时就可以保持稳定, 但氘核在这一温度下刚形成便会碎裂, 因此真正有意义的核合成过程只在温度略低于 10^9K 时才能由于通过了“氘关卡”而开始发生。最后的产物是大量的 ${}^4\text{He}$ 和微量的 ${}^2\text{H}$ 、 ${}^3\text{He}$ 以及 ${}^7\text{Li}$ (${}^3\text{H}$ 不稳定, 很快会转变成 ${}^3\text{He}$)。若以 ${}^4\text{He}$ 的产量为单位, 则 ${}^2\text{H}$ 、 ${}^3\text{He}$ 的产量约为 10^{-5} , 而 ${}^7\text{Li}$ 约为 10^{-10} 。

氦的丰度约为 25% 在理论上是如何确定的呢? 为此假设以 N_n 和 N_p 分别代表宇宙中的总中子数与总质子数, 以 n_n 和 n_p 分别代表原初核合成的时候中子和质子的数密度, 记

这两个数密度的比值为 σ ，很显然

$$\sigma = \frac{n_n}{n_p} = \frac{N_n}{N_p}. \quad (16.1)$$

由于核合成之后宇宙的主要成分就是氦元素与氢元素 (含一个质子, 不含中子), 所以很显然 $\sigma < 1$, 具体是多少我们稍后再来确定。则, 宇宙中的总核子数 $N = N_n + N_p = (\sigma + 1)N_p$, 其中全部中子都与同数量的质子结合成氦元素, 所以氦元素中所含的核子数 $N_{\text{He}} = 2N_n = 2\sigma N_p$, 因此, 以质量百分比计, 原初核合成所产生的氦丰度为

$$Y = \frac{N_{\text{He}}}{N} = \frac{2\sigma N_p}{(\sigma + 1)N_p} = \frac{2\sigma}{\sigma + 1} = 2\left(\frac{n_n}{n_p}\right)\left(1 + \frac{n_n}{n_p}\right)^{-1}. \quad (16.2)$$

所以, 宇宙中氦的丰度完全取决于核合成时 n_n 与 n_p 的比值。

这个比值可以这样来确定。在中微子退耦前 (注意 $t_{\text{vd}} \approx 1\text{s}$, 远在核合成的时间之前), 中子和质子可以通过以下弱作用过程相互转化:

$$p + e \leftrightarrow n + \nu_e, \quad p + \bar{\nu}_e \leftrightarrow n + e^+, \quad (16.3)$$

式中 e, e^+ 分别代表电子和正电子, ν_e 和 $\bar{\nu}_e$ 分别代表相应的电子中微子及其反中微子。因为中子质量略大于质子质量 ($m_n \simeq m_p + 2.5m_e$), 所以质子变中子的过程相比于其逆过程较难发生。比如, 由于 $m_p + m_e \simeq m_n - 1.5m_e < m_n$, 所以一个静止的质子和一个静止的电子不能变成一个哪怕是静止的中子 (能量守恒不允许), 但是反过来却没有这个问题。当然, 在温度高达 10^{10}K 以上时, 电子远远不是静止的, 所以 $p + e \rightarrow n + \nu_e$ 可以发生, 但无论如何其发生的概率总小于逆过程的发生概率, 根据玻尔兹曼统计可知, 当正反过程达到统计平衡时, 必定有

$$\sigma = \frac{n_n}{n_p} = e^{-\frac{\Delta m}{k_B T}}, \quad (16.4)$$

式中 $\Delta m \equiv m_n - m_p$. 当宇宙温度降至 $T_{\text{vd}} \simeq 10^{10}\text{K}$ 时, 中微子退耦, 此后上述 n, p 相互转变的弱作用过程就基本停止了, 因此 $\frac{n_n}{n_p}$ 的值便大致冻结在了 $e^{-\frac{\Delta m}{k_B T_{\text{vd}}}}$. 当然, 以上讨论略去了自由中子的自发衰变 $n \rightarrow p + e + \bar{\nu}_e$, 因为其半衰期约为 10 分钟, 远大于中微子退耦时的宇宙年龄 $t_{\text{vd}} \sim 1\text{s}$. 不过, 等到氦合成时宇宙年龄 ($t \simeq 100\text{s}$) 相比中子半衰期来说已经可观了。因此, 等到氦合成时, $\sigma = \frac{n_n}{n_p}$ 的值将比之前的冻结值 $e^{-\frac{\Delta m}{k_B T_{\text{vd}}}}$ 略低, 大约为 1/7 弱一点。

将 $\sigma = \frac{n_n}{n_p} = 1/7$ 代入 (16.2) 式, 就得到氦丰度约为 $Y \simeq 0.25$, 与观测相当吻合。不过, 考虑到恒星内部的核反应也会产生 ${}^4\text{He}$, 虽然比原初核合成的产量少很多, 但也会使得原初核合成时的氦丰度略低于今天观测到的氦丰度, 考虑这个因素以后由今天观测的氦丰度推断出来的原初氦丰度约为 0.23, 与上述理论值 0.25 也基本吻合。

原初核合成的其它产物 ${}^2\text{H}$ 、 ${}^3\text{He}$ 以及 ${}^7\text{Li}$, 它们虽然丰度很小, 但对于验证理论也很重要, 定量计算原初核合成产物的丰度还需要用到一个重要的物理参数 η , 它就是宇宙中重子与光子数密度之比, 即 $\eta = n_b/n_\gamma$, η^{-1} 就代表每个重子周围的光子数, 它通过影响光分裂的难易程度而影响核合成开始的时刻, 进而影响产物的丰度。 ${}^4\text{He}$ 的丰度只微弱地依赖于 η , 然而 ${}^2\text{H}$ 、 ${}^3\text{He}$ 以及 ${}^7\text{Li}$ 的丰度却对 η 值相当敏感。计算表明, 只要假定 η 值在 $3.4 \times 10^{-10} \sim 5 \times 10^{-10}$ 的范围之内, 则上述四种产物的理论丰度都能与观测丰度符合。

16.1.4 宇宙微波背景 (CMB) 和宇宙学原理

伽莫夫的热大爆炸理论预言了有弥漫整个宇宙的微波背景存在。1960 年代美国普林斯顿大学的 Dicke 和 Peebles 认识到背景辐射问题对宇宙学的重要性, 因此就开始组织团队准备观测仪器, 以探寻这个辐射。然而, 当他们还在做准备时, 贝尔实验室的彭齐亚斯和威尔逊意外地发现了宇宙微波背景辐射。1965 年这两组人同时在一期《天体物理杂志》上分别发表论文, 彭齐亚斯和威尔逊的论文很简短, 只客观地指出, 在 4080MHz 发现有一个 3.5K 的过剩天线温度, 在测量限度内它是各向同性的, 并与季节的变化无关。Dicke 和 Peebles 等人的论文则详细解释了这信号的宇宙学意义, 也就是指出它很可能是宇宙微波背景。1978 年彭齐亚斯和威尔逊因为他们的发现获得了诺贝尔物理学奖。

各向同性的背景辐射的发现, 其首要意义是确证了星系形成之前的早期宇宙的确充满了均匀并且各向同性的热气体。当然, 其中均匀性并不直接来自于对微波背景的观测, 而是来自于我们对所谓哥白尼原理的信念, 即地球并不是宇宙中心, 同样的, 宇宙中任何一处都不是宇宙中心, 而是处处平权的。结合这个均匀性以及观测到的微波背景相对于地球的各向同性, 就能确立早期宇宙是处处均匀并且处处都各向同性的。由此而来的结果就是, 星系形成以后的宇宙在宇观大尺度上也是处处均匀且各向同性的。这也就是所谓的**宇宙学原理**。

根据 1989 年 11 月升空的宇宙背景探测器 (COBE, Cosmic Background Explorer) 测量到的结果, 宇宙微波背景辐射谱非常精确地符合温度为 $2.726 \pm 0.010\text{K}$ 的黑体辐射谱, 证实了银河系相对于背景辐射有一个相对的运动速度, 并且还验证, 扣除掉这个速度对测量结果带来的影响, 以及银河系内物质辐射的干扰, 宇宙背景辐射具有高度各向同性, 温度涨落的幅度只有大约百万分之五。目前公认的理论认为, 这个温度涨落起源于宇宙在形成初期极小尺度上的量子涨落, 它随着宇宙的暴胀而放大到宇宙学的尺度上, 并且正是由于温度的涨落造成宇宙物质分布的不均匀性, 最终得以形成诸如星系团等的一类大尺度结构。

其中对银河系相对微波背景运动速度的确立来自于观测到的微波背景辐射温度在偶极上的微小各向异性。具体来说, 星系作为膨胀宇宙中的一粒“沙子”, 除了跟着宇宙膨胀以外, 必定还有其自身的运动, 通常称之为本动。银河系当然也不例外, 因此我们是在运动的参考系中观测背景辐射的, 这样就会有 Doppler 效应。不妨设想银河系在向右运动, 设出发时频率为 ν_1 的光子在没有本动的星系上观测到的频率是 ν_0 (因为宇宙膨胀带来的宇宙红移, ν_0 当然小于 ν_1), 则由于银河系在向右运动, 右边来的光子的宇宙红移会被银河系的本动抵消一些, 从而它被接收到时的频率将比 ν_0 略大, 相反, 左边来的光子的红移将被本动加剧一些, 从而测到的频率将比 ν_0 略小, 这就等效于改变了不同方向上宇宙微波背景的温度, 使之产生一个微小的各向异性, 可以证明, 这种由于银河系本动所产生的各向异性是一种偶极各向异性。

除了偶极各向异性以外, 微波背景的温度还有微小的多极各向异性, 这种多极各向异性常常用所谓的角功率谱来刻画 (关于角功率谱的简短介绍, 见参考文献)。正如前面所说, 它起源于宇宙在形成初期极小尺度上的量子涨落, 它随着宇宙的暴胀而放大到宇宙学的尺度上, 最终导致星系等结构的形成。

最后讨论一点引申出来的物理观念。一百多年前，人们曾经认为有一种弥漫在全空间的介质，称作以太，它可以作为静止或运动的绝对基准。天体相对于以太的运动就称之为绝对运动，迈克尔逊和莫雷曾试图测出地球的绝对速度，然而始终得到零结果。从那以后人们就抛弃了以太的概念，并认为绝对运动是没有意义的，这也就是相对性原理的基础。然而，现在宇宙学肯定了弥漫全宇宙的背景辐射的存在，那么以背景辐射为基准的“绝对”运动就有了意义，比如 COBE 就测出了银河系相对于背景辐射的“绝对”速度。

实际上，在宇宙任何地点的任何一个像银河系这样的局域参考系中都能测到宇宙微波背景。如果发现从某个局域参考系中看来没有温度的偶极各向异性，那就可以认为这个局域参考系是“绝对静止”的，或者说本动速度为零。而上面所谈的银河系的“绝对速度”（本动速度）就是相对于这样的静止局域参考系而言的。当然，值得注意的是，由于宇宙本身在膨胀，所以宇宙中两个不同地点的静止局域系之间的距离是在膨胀增加的，或者说，空间不同点上的局域静止系之间是相对运动的。这就好比吹气球，气球表面任意一点都是局域静止的，但是随着气球的膨胀，两点之间的距离却在增加，我们的宇宙空间就好比气球的表面。这也说明，这种局域静止系的存在其实并不意味着牛顿绝对时空观的复活。

宇宙学原理告诉我们，不同空间点的局域静止系是完全平权的，也就是均匀的，同时，整个空间相对于任何一个局域静止系都是各向同性的。

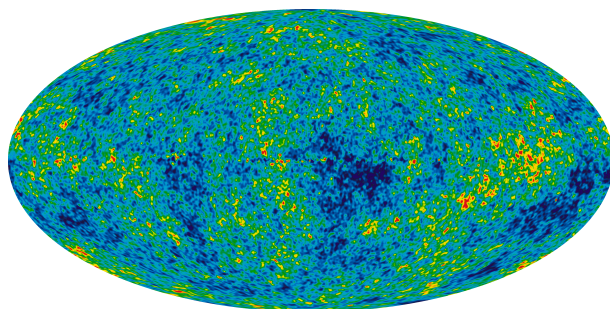


Figure 16.1: 宇宙微波背景辐射。

他们用坚硬的黑
透明的亮
在太空建起一座金字塔
空中飘来百亿年前的那幅名画
塔心那颗小小芥子
是生在画师之前的宇宙吧？

16.2 宇宙膨胀动力学

16.2.1 FLRW 度规

在宇宙学中，需要定义一些基本观察者，也就是相对于宇宙微波背景局域静止的观察者，也称作共动观察者。这些基本观察者的世界线互不相交，因此时空中每一个非奇异的点（奇异点比如有大爆炸开始的点）都有唯一一条世界线穿过，因此所有这些世界线的线束就构成了整个时空。利用这些共动观察者，我们可以定义 $t = \text{常数}$ 的空间超曲面，这只需要它与基本观察者的协变速度相正交即可。接下来我们可以取基本观察者的固有时为整个宇宙同步的时间坐标，记作 t ，此外我们还可以给每一条世界线指定三个空间坐标 $x^i, i = 1, 2, 3$ （沿着世界线是常值），因此每一个共动观察者都有一个固定的 x^i ，称作**共动坐标**，它参数化了 $t = \text{常数}$ 的空间超曲面。由于这些空间超曲面与共动观察者世界线相垂直，因此宇宙的时空度规可以写成

$$ds^2 = -dt^2 + g_{ij}dx^i dx^j, \quad (16.5)$$

式中 g_{ij} 是 t, x^i 的函数，它为空间超面上的诱导度规。

根据宇宙学原理，宇宙的空间几何（不包括时间）是均匀各向同性的，而根据第十一章关于最大对称空间的知识又可以知道，均匀各向同性的空间必定是所谓的最大对称空间，因此宇宙的空间几何是一个最大对称空间。进一步，根据第十二章的知识可知，3 维最大对称空间只有平坦的欧几里德空间 \mathbb{R}^3 、闭合的三维球面 S^3 、以及开放的三维双曲空间 H^3 这三种。

对于 S^3 ，我们可以把它嵌入在 4 维欧几里德空间 \mathbb{R}^4 中作为一张超曲面，满足方程

$$(z^1)^2 + (z^2)^2 + (z^3)^2 + (z^4)^2 = L^2, \quad (16.6)$$

S^3 上的度规就是将 $(dz^1)^2 + (dz^2)^2 + (dz^3)^2 + (dz^4)^2$ 限制在这张超曲面上。下面对前三个分量引入球坐标，即令 $(z^1)^2 + (z^2)^2 + (z^3)^2 = L^2 r^2$ ，从而

$$(dz^1)^2 + (dz^2)^2 + (dz^3)^2 = L^2(dr^2 + r^2 d\Omega^2), \quad (16.7)$$

其中 $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$ 为两维球面上的立体角。另外，根据 S^3 的方程 (16.6)，有 $(z^4)^2 = L^2(1 - r^2)$ ，从而 $z^4 dz^4 = -L^2 r dr$ ，进而即有

$$(dz^4)^2 = \frac{(-L^2 r dr)^2}{(z^4)^2} = \frac{(-L^2 r dr)^2}{L^2(1 - r^2)} = L^2 \frac{r^2}{1 - r^2} dr^2. \quad (16.8)$$

将 (16.7) 式与 (16.8) 式加起来，即可以将 S^3 上的度规 $ds^2 = [(dz^1)^2 + (dz^2)^2 + (dz^3)^2 + (dz^4)^2]_{|S^3}$ 写成

$$ds_{S^3}^2 = L^2 \left(\frac{1}{1 - r^2} dr^2 + r^2 d\Omega^2 \right). \quad (16.9)$$

对于 H^3 ，根据第十二章的知识，我们可以把它嵌入 1+3 维闵可夫斯基时空 $\mathbb{R}^{1,3}$ ，作为其中一张双曲面，满足如下方程

$$-(z^0)^2 + (z^1)^2 + (z^2)^2 + (z^3)^2 = -L^2, \quad (16.10)$$

H^3 上的度规就是将 $\mathbb{R}^{1,3}$ 的度规 $-(dz^0)^2 + (dz^1)^2 + (dz^2)^2 + (dz^3)^2$ 限制在这张双曲面上而得到。类似的, 对后三个分量引入球坐标, 即令 $(z^1)^2 + (z^2)^2 + (z^3)^2 = L^2 r^2$, 因此当然也有 (16.7) 式。另一方面, 根据 (16.10) 式, 显然有 $(z^0)^2 = L^2(1+r^2)$, 从而 $z^0 dz^0 = L^2 r dr$, 类似的, 进而即有

$$-(dz^0)^2 = -L^2 \frac{r^2}{1+r^2} dr^2. \quad (16.11)$$

将 (16.11) 式与 (16.7) 式加起来, 即可以将 H^3 上的度规写成

$$ds_{H^3}^2 = L^2 \left(\frac{1}{1+r^2} dr^2 + r^2 d\Omega^2 \right). \quad (16.12)$$

对于平坦的 \mathbb{R}^3 , 通过引入球坐标, 其度规显然可以写成

$$ds_{\mathbb{R}^3}^2 = L^2 (dr^2 + r^2 d\Omega^2). \quad (16.13)$$

因此我们可以将这三种情况综合性地写作

$$ds_{\text{Space}}^2 = h_{ij} dx^i dx^j = L^2 \left(\frac{1}{1-kr^2} dr^2 + r^2 d\Omega^2 \right), \quad (16.14)$$

式中 h_{ij} 只是空间坐标 x^i 的函数, 其中对于正曲率的 S^3 , $k=1$, 对于平坦的 \mathbb{R}^3 , $k=0$, 对于负曲率的 H^3 , $k=-1$ 。综合三种情况, 空间的截面曲率 K 可以写成

$$K = \frac{k}{L^2}. \quad (16.15)$$

对于 $k=1$ 情形。此时, 当 $r \rightarrow 1$ 时, 度规 (16.14) 看起来是奇异的。我们可以定义

$$r = \sin \chi, \quad 0 \leq \chi \leq \pi. \quad (16.16)$$

进而将度规 (16.14) 重写成

$$ds_{S^3}^2 = L^2 (d\chi^2 + \sin^2 \chi d\Omega^2). \quad (16.17)$$

对于 $k=0$ 情形。可以直接令 $r = \chi$, 进而将度规 (16.14) 重写成

$$ds_{\mathbb{R}^3}^2 = L^2 (d\chi^2 + \chi^2 d\Omega^2). \quad (16.18)$$

对于 $k=-1$ 情形。我们令

$$r = \sinh \chi, \quad (16.19)$$

则可以将度规 (16.14) 重写成

$$ds_{H^3}^2 = L^2 (d\chi^2 + \sinh^2 \chi d\Omega^2). \quad (16.20)$$

常常也将这三种情形统一写作

$$ds_{\text{Space}}^2 = L^2 (d\chi^2 + S_k^2(\chi) d\Omega^2), \quad (16.21)$$

式中

$$S_k(\chi) = \begin{cases} \sin \chi, & k = 1, \\ \chi, & k = 0, \\ \sinh \chi, & k = -1. \end{cases} \quad (16.22)$$

当然，由于宇宙是在膨胀的，其空间大小在整体性地随时间膨胀，所以 (16.14) 在作为宇宙的空间几何时，要乘上一个随时间演化的缩放因子 $a^2(t)$ ，其中 $a(t)$ 就是所谓的尺度因子，它在宇宙的今天 (即 t_0 时刻) 取 1，即 $a(t_0) = 1$ 。从而 (16.5) 中的空间度规 g_{ij} 必定可以写成

$$g_{ij}(t, x) = a^2(t)h_{ij}(x). \quad (16.23)$$

根据 (16.5) 式，从而可以将整个宇宙时空的度规取成

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t)L^2\left(\frac{1}{1-kr^2}dr^2 + r^2d\Omega^2\right) \\ &= -dt^2 + a^2(t)L^2[d\chi^2 + S_k^2(\chi)d\Omega^2] \end{aligned} \quad (16.24)$$

这就是著名的弗里德曼-勒梅特-罗伯逊-沃尔克度规，简称 FLRW 度规。其中， t 时刻宇宙空间的尺寸即是 $a(t)L$ ，由于 $a(t_0) = 1$ ，所以 L 即是宇宙在今天的尺寸。

哈勃定律

由于宇宙空间的均匀各向同性，所以我们可以取任意一个共动观察者为中心，令其 $\chi = 0$ ，则另一共动观察者到他的共动距离就是 χ ，很显然， t 时刻这两者之间的固有距离 $d_P(t)$ 则是

$$d_P(t) = a(t)L\chi. \quad (16.25)$$

很明显，随着宇宙的膨胀，这两个共动观察者之间的固有距离在增加，换言之，一个观察者在相对于另一个观察者退行，不难得到，退行速度 $u(t) = \dot{d}_P(t)$ 为

$$u(t) = \dot{d}_P(t) = \dot{a}L\chi = \frac{\dot{a}}{a}d_P(t). \quad (16.26)$$

习惯上常常定义

$$H(t) \equiv \frac{\dot{a}}{a}, \quad (16.27)$$

称之为哈勃参数，它具有时间倒数的量纲。则退行速度满足

$$u(t) = H(t)d_P(t). \quad (16.28)$$

也即是说，固有距离越远则相对退行的速度就越大，这就是哈勃定律。哈勃正是通过观察到远处的星系在远离我们 (也就是相对于我们退行)，从而发现宇宙在膨胀的。哈勃参数

$H(t)$ 依赖于宇宙时间 t , 当 $t = t_0$ 时, 也就是宇宙今天, 相应的 $H(t_0)$ 常常记作 H_0 , 称之为哈勃常数。

从哈勃定律 (16.28) 可知, 只要 d_P 足够大, 则退行速度 $u(t)$ 完全有可能超过光速 1。那这是不是违反了相对论的基本原理呢? 并非如此, 因为相对论原理说的其实是, 任何粒子的世界线都应该是类时或者类光曲线, 不能类空。这句话换种说法其实应该是, 任何粒子相对于当地观察者的速率都应该小于等于光速 1。注意, 这里的关键是, 这个速度是相对于当地观察者而言的。而遥远星系的退行速度并不是相对星系当地的观察者的, 而是相对于远方的我们而言的, 那这当然是可以超过 1 的, 并不违反相对论原理。

关于哈勃常数 H_0 的测量值, 目前不同的测量方法会得到不一致的测量结果 (主要是通过晚期宇宙的观测 (如通过对 Ia 型超新星的观测) 得出来的结果, 与宇宙微波背景 (CMB) 数据对宇宙学标准模型的全局拟合得出来的结果不一致), 为什么不一致今天人们还没搞清楚, 有时候人们也称之为哈勃常数危机。典型的测量值是在 $67 \sim 74(\text{km/s})/\text{Mpc}$ 之间。Mpc 即是百万秒差距, 这是天文测量中常用的距离单位, 1Mpc 大约为 330 万光年。如果取 $H_0 = 73(\text{km/s})/\text{Mpc}$, 那这就表示相距 330 万光年的两个星系, 在以大约每秒 73 公里的速度彼此远离。这一速度是地球围绕太阳公转速度的两倍多。

16.2.2 联络和曲率

FLRW 度规中只有尺度因子 $a(t)$ 随时间的演化需要求解, $a(t)$ 是由爱因斯坦场方程决定的, 为了应用场方程, 我们又需要先计算出与 FLRW 度规对应的联络系数以及里奇曲率。下面就是具体计算过程。

为了计算出联络系数, 我们对如下世界线作用量 (λ 为世界线参数) 进行变分

$$S = \frac{1}{2} \int d\lambda \left[- \left(\frac{dt}{d\lambda} \right)^2 + a^2(t) h_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} \right]. \quad (16.29)$$

比方说对 $t(\lambda)$ 的变分会给出

$$\delta S = \int d\lambda \left[\frac{d^2 t}{d\lambda^2} + a\dot{a} h_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} \right] \delta t + \dots \quad (16.30)$$

式中 $\dot{a} = \frac{da}{dt}$, 由这个变分结果即有方程

$$\frac{d^2 t}{d\lambda^2} + a\dot{a} h_{ij} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} = 0, \quad (16.31)$$

进而即可以读出如下非零的联络系数

$$\Gamma_{ij}^0 = a\dot{a} h_{ij}. \quad (16.32)$$

类似的, 对 x^k 的变分将给出如下方程

$$\frac{d^2 x^k}{d\lambda^2} + 2 \frac{\dot{a}}{a} \frac{dt}{d\lambda} \frac{dx^k}{d\lambda} + \frac{1}{2} h^{kl} \{ \partial_i h_{lj} + \partial_j h_{li} - \partial_l h_{ij} \} \frac{dx^i}{d\lambda} \frac{dx^j}{d\lambda} = 0. \quad (16.33)$$

式中矩阵 h^{ij} 为矩阵 h_{ij} 的逆。进而又可以读出如下非零联络系数

$$\Gamma_{0j}^i = \frac{\dot{a}}{a} \delta_j^i, \quad \Gamma_{ij}^k = \frac{1}{2} h^{kl} \{ \partial_i h_{lj} + \partial_j h_{li} - \partial_l h_{ij} \}. \quad (16.34)$$

其余联络系数均为零。很显然，其中 Γ_{ij}^k 就是按照空间度规 $h_{ij}(x)$ 算出来的标准联络系数。

特别的，对于共动观察者，其本动速度 $\frac{dx^i}{d\lambda} \equiv 0$ ，由此再求导一次，当然也有 $\frac{d^2x^i}{d\lambda^2} \equiv 0$ ，从而自动满足测地线方程 (16.33)，而且这时候测地线方程 (16.31) 简化成了 $\frac{d^2t}{d\lambda^2} = 0$ ，由于 t 是共动观察者的固有时，而根据测地线方程的定义， λ 的自然选取也是共动观察者的固有时，即取 $\lambda = t$ ，所以测地线方程 (16.31) 也是自动满足的。由此可见，在 FLRW 时空中，共动观察者走的就是测地线。

有了联络系数以后，我们就可以进一步计算黎曼曲率张量，当然，由于我们的最终目的是应用爱因斯坦场方程，所以实际上只需计算里奇曲率张量。根据黎曼曲率张量的计算公式，有

$$R_{\mu\nu} = R^{\rho}{}_{\mu\rho\nu} = \partial_{\rho}\Gamma_{\nu\mu}^{\rho} - \partial_{\nu}\Gamma_{\rho\mu}^{\rho} + \Gamma_{\rho\sigma}^{\rho}\Gamma_{\nu\mu}^{\sigma} - \Gamma_{\nu\sigma}^{\rho}\Gamma_{\rho\mu}^{\sigma}. \quad (16.35)$$

代入上面得出的联络系数，立即有

$$R_{0i} = 0. \quad (16.36)$$

以及

$$R_{00} = -\partial_t\Gamma_{0i}^i - \Gamma_{0j}^i\Gamma_{0i}^j = -3\frac{\ddot{a}}{a}. \quad (16.37)$$

以及

$$R_{ij} = \partial_t\Gamma_{ji}^0 + \Gamma_{k0}^k\Gamma_{ji}^0 - \Gamma_{jk}^0\Gamma_{0i}^k - \Gamma_{j0}^k\Gamma_{ki}^0 + \tilde{R}_{ij}, \quad (16.38)$$

式中 \tilde{R}_{ij} 为按照空间度规 $h_{ij}(x)$ 算出来的里奇张量，它由下式给出

$$\tilde{R}_{ij} = \partial_k\Gamma_{ji}^k - \partial_j\Gamma_{ki}^k + \Gamma_{lk}^l\Gamma_{ji}^k - \Gamma_{jk}^l\Gamma_{li}^k. \quad (16.39)$$

由于空间度规 h_{ij} 描述的是一个最大对称空间，其截面曲率为 $K = k/L^2$ ，根据第十一章的知识可知， $h_{ij}(x)$ 给出的黎曼曲率张量为

$$\tilde{R}_{lkij} = K(h_{li}h_{kj} - h_{lj}h_{ki}) \Rightarrow \tilde{R}_{ij} = 2Kh_{ij}. \quad (16.40)$$

将这个结果代入 (16.38) 式，同时代入 Γ_{ij}^0 和 Γ_{0j}^i 的结果，即可以得到

$$R_{ij} = \left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{K}{a^2} \right] g_{ij}, \quad (16.41)$$

式中 $g_{ij} = a^2(t)h_{ij}$ 。

根据以上结果，不难得到里奇标量为

$$R = 6\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{K}{a^2} \right]. \quad (16.42)$$

进而即能得到爱因斯坦张量 $G_{\mu\nu}$ 的非零分量如下

$$G_{00} = 3\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{K}{a^2} \right], \quad G_{ij} = -\left[2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{K}{a^2} \right] g_{ij}. \quad (16.43)$$

16.2.3 弗里德曼方程

在大尺度上宇宙的物质内容可以看成是理想流体，其能量动量张量为第十四章中引入的

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu}. \quad (16.44)$$

由于星系的本动速度相对于宇宙膨胀速度来说都很小，所以我们可以假定，在共动观察者看来，宇宙物质的这理想流体是局域静止的，也即是说，在 FLRW 坐标中，这流体的协变速度 $u^\mu = (1, 0, 0, 0)$ 。从而 $T^{00} = \rho$, $T^{0i} = 0$, $T^{ij} = g^{ij}p$, 换言之，

$$T^\mu_\nu = \text{diag}(-\rho, p, p, p). \quad (16.45)$$

宇宙空间的均匀性意味着 ρ 和 p 均只依赖于时间 t 。

能动量张量的守恒方程 $D_\mu T^{\mu\nu} = 0$ 实际上反映了流体在时空中的演化，其中 $D_\mu T^\mu_i = 0$ 会给出物质组分满足的测地线方程，也就是共动观察者的测地线方程，正如前面说过的，这是自动满足的。而能量守恒方程 $D_\mu T^\mu_0 = 0$ 将给出

$$\begin{aligned} 0 &= \partial_\mu T^\mu_0 + \Gamma^\mu_{\mu\nu} T^\nu_0 - \Gamma^\rho_{\mu 0} T^\mu_\rho \\ &= -\dot{\rho} + \Gamma^i_{i0} T^0_0 - \Gamma^i_{j0} T^j_i \\ &= -\dot{\rho} - 3\frac{\dot{a}}{a}(\rho + p). \end{aligned} \quad (16.46)$$

这个能量守恒方程有一个直观解释，为此我们把它改写成

$$d(\rho a^3) = -pd(a^3), \quad (16.47)$$

考虑到一个空间区域的体积正比于 $a^3(t)$ ，所以这个方程的意思就是，某空间区域膨胀向外做的功等于区域内物质总能量的减少。

另一方面，根据爱因斯坦场方程，我们有

$$G_{00} = 8\pi G T_{00}, \Rightarrow 3\left[\left(\frac{\dot{a}}{a}\right)^2 + \frac{K}{a^2}\right] = 8\pi G \rho. \quad (16.48)$$

重新整理一下，即可以写成

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{K}{a^2}. \quad (16.49)$$

这就是著名的弗里德曼方程，它告诉我们宇宙是如何膨胀的。

爱因斯坦场方程的空间分量 $G_{ij} = 8\pi G T_{ij}$ 将给出

$$\begin{aligned} -2\left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{K}{a^2}\right] &= 8\pi G p \\ \Rightarrow \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}(\rho + 3p). \end{aligned} \quad (16.50)$$

但是这不是一个独立的方程，因为你可以通过将弗里德曼方程对时间求导，然后代入能量守恒方程 (16.46) 来得到它，虽然人们有时候也称这个方程为第二弗里德曼方程。

总之，要描述宇宙的演化，我们需要两个相互独立的基本方程，分别是能量守恒方程 (16.46)，以及弗里德曼方程 (16.49)。

16.2.4 宇宙学红移

前面说过，哈勃观测到远处的星系在远离我们（退行），从而知道了宇宙在膨胀。但是，远处星系的退行速度是如何观测的呢？当然，是通过发现远处星系发给我们的光信号有红移，如果这星系离我们的距离不是过于遥远，从而退行速度不大的话，那么这红移可以解释为多普勒红移。但是，一般来说，宇宙学红移和多普勒红移并不是一回事。

为了讲清楚远方星系光信号的宇宙学红移，不妨假设我们的观察者位于共动坐标的原点 $\chi = 0$ ，远方的星系位于共动坐标 χ_A ，假设光信号沿着径向从 χ_A 传播到原点处的观察者。由于光信号走的是类光测地线，满足 $ds^2 = 0$ ，进而根据 FLRW 度规，我们有

$$-dt^2 + a^2(t)L^2d\chi^2 = 0 \Leftrightarrow \frac{dt}{a(t)} = -Ld\chi, \quad (16.51)$$

式中注意到了沿着光路径 χ 在减少。假设光信号是 t_1 时刻发出的， t_2 时刻被接收到，则将上式积分，即有

$$\int_{t_1}^{t_2} \frac{dt}{a(t)} = L \int_0^{\chi_A} d\chi = L\chi_A. \quad (16.52)$$

经过一个周期以后，信号发出的时刻变成了 $t_1 + \delta t_1$ ，接收时刻变成了 $t_2 + \delta t_2$ ，其中 δt_1 就是光发出时的时间周期， δt_2 则是光被接收时的时间周期。很显然，依然有

$$\int_{t_1 + \delta t_1}^{t_2 + \delta t_2} \frac{dt}{a(t)} = L\chi_A. \quad (16.53)$$

将 (16.53) 式减去 (16.52) 式，并考虑到时间周期 δt_1 和 δt_2 均很小，从而即有

$$\frac{\delta t_2}{a(t_2)} - \frac{\delta t_1}{a(t_1)} = 0. \quad (16.54)$$

假设光发出时的波长为 λ_1 ，则很显然 $\lambda_1 = \delta t_1$ ，同样，光被接收到时的波长 $\lambda_2 = \delta t_2$ ，则根据上面的式子我们有

$$\frac{\lambda_2}{\lambda_1} = \frac{\delta t_2}{\delta t_1} = \frac{a(t_2)}{a(t_1)}, \quad (16.55)$$

显然，波长变了，也就是发生了红移，这就是宇宙学红移。

在实际应用中，接收光信号的当然就是我们，我们是在今天（即 t_0 时刻）接收到光信号的，而光信号发出的时刻是过去的某个 t 时刻，发出时的波长可以记为 λ ，接收时的波长记为 λ_0 ，则可以通过下式定义光信号的宇宙学红移 z

$$1 + z \equiv \frac{\lambda_0}{\lambda} = \frac{a(t_0)}{a(t)} = \frac{1}{a(t)}, \quad (16.56)$$

式中最后一个等于号利用了 $a(t_0) = 1$ 。这个式子告诉我们，只要我们观测到远方星系光信号的红移量 $z > 0$ ，那就说明过去的 $a(t) < 1$ ，从而说明宇宙是在膨胀！特别的，宇宙开始于大爆炸奇点，即 $a(0) = 0$ ，这对应于 $z = \infty$ 的红移。而宇宙今天对应 $a(t_0) = 1$ ，这相应于 $z = 0$ 。不仅如此，星系形成以后的宇宙相应于 $z < 10$ ，而光子脱耦形成宇宙微波背景辐射的时刻相应于 $z \simeq 1000$ 。

有了宇宙学红移的概念以后，就可以讨论哈勃如何通过天文观测发现哈勃定律的了。实际上，哈勃是对远方星系进行观测，而且这个远方星系不是太远（相比于宇宙的大小），因此光子从它发出的时间 t 可以写成 $t = t_0 - \delta t$ ， t_0 就是我们接收到光子的时间，也就是宇宙的今天，由于被观测的星系离我们不是太远，所以它发出光子的时刻也不会比 t_0 早很多，因此 $\delta t \ll t_0$ 。由此即可以对尺度因子进行展开

$$\begin{aligned} a(t) &= a(t_0 - \delta t) \\ &= a(t_0) - \delta t \dot{a}(t_0) + \frac{1}{2}(\delta t)^2 \ddot{a}(t_0) + \dots \\ &= a(t_0) \left[1 - \delta t H(t_0) - \frac{1}{2}(\delta t)^2 q(t_0) H^2(t_0) + \dots \right], \end{aligned} \quad (16.57)$$

当然式中的 $a(t_0)$ 其实就是 1，而 $q(t)$ 是所谓的减速参数，其定义是

$$q(t) = -\frac{\ddot{a}(t)a(t)}{\dot{a}^2(t)}. \quad (16.58)$$

也即是说，如果 $q(t) > 0$ ，那就说明宇宙在减速膨胀，反之，如果 $q(t) < 0$ ，那宇宙就在加速膨胀。习惯上常常记 $a_0 = a(t_0) = 1$ ， $H(t_0) = H_0$ ， $q(t_0) = q_0$ ，其中 q_0 就是宇宙当前的减速参数。

利用 H_0 ， q_0 ，就可以把远方星系的红移因子表示为

$$\begin{aligned} z = \frac{a(t_0)}{a(t)} - 1 &= \left[1 - \delta t H(t_0) - \frac{1}{2}(\delta t)^2 q(t_0) H^2(t_0) + \dots \right]^{-1} - 1 \\ &= \delta t H_0 + (\delta t)^2 \left(1 + \frac{1}{2} q_0 \right) H_0^2 + \dots \end{aligned} \quad (16.59)$$

由于 δt 为一个小量，所以红移 z 也为一个小量， $z \ll 1$ ，进而即有

$$\delta t = \frac{1}{H_0} \left[z - \left(1 + \frac{1}{2} q_0 \right) z^2 + \dots \right], \quad (16.60)$$

注意，这个式子只依赖于当前的参数 H_0 和 q_0 ，与宇宙的演化历史无关。

另一方面，根据 (16.52) 式，可以得到这个星系离我们的共动距离 $d_P = a_0 L \chi_A$ 为

$$\begin{aligned} d_P = a_0 L \chi_A &= a_0 \int_t^{t_0} \frac{dt}{a(t)} = \int_t^{t_0} dt \left[1 - (t_0 - t) H_0 + \dots \right]^{-1} \\ &= \left[\delta t + \frac{1}{2}(\delta t)^2 H_0 + \dots \right]. \end{aligned} \quad (16.61)$$

代入 δt 的表达式 (16.60)，即有

$$d_P = \frac{1}{H_0} \left[z - \frac{1}{2} (1 + q_0) z^2 + \dots \right]. \quad (16.62)$$

特别的，对于不太远的星系， $z \ll 1$ ，所以作为一阶近似，即有

$$z = H_0 d_P. \quad (16.63)$$

对于这种不太远的星系，我们可以将其红移看作是多普勒红移，从而可得其退行速度 $u = z$ ，进而 (16.63) 式给出的就是哈勃定律。

但是，如果我们保留到距离红移关系 (16.62) 的二阶近似，我们就会发现它和宇宙减速参数 q_0 有关，给定远方星系的红移量 z ，如果宇宙在减速膨胀 $q_0 > 0$ ，则相对于 $q_0 = 0$ 的情形，远方星系的距离就会偏小，相反，如果宇宙在加速膨胀， $q_0 < 0$ ，则远方星系的距离就会偏大。1998 年，Perlmutter 团队和 Riess 与 Schimidt 团队观测了红移大约在 0.5 左右的 Ia 型超新星（一种亮度非常稳定，可以用作“标准烛光”的天体），发现这些超新星比预期要暗，意味着它们离我们的距离偏大，进而推断出宇宙在加速膨胀。具体来说，就是通过拟合这些超星系的距离红移关系（严格来说，拟合的是光度距离和红移的关系，但与上面的共动距离其实密切相关），进而推断出 $q_0 < 0$ 。

更一般地，由于哈勃参数 $H(t)$ 是时间的函数，而通过尺度因子与时间的关系 $a(t)$ 又可以把时间看成是红移的函数，因此哈勃参数完全可以看作是红移的函数，记作 $H(z)$ 。进而有

$$\begin{aligned} dz &= d(1+z) = d\left(\frac{a_0}{a}\right) = -\frac{a_0}{a} \frac{\dot{a}}{a} dt \\ &= -(1+z)H(z)dt. \end{aligned} \quad (16.64)$$

因此，

$$t_0 - t = \int_t^{t_0} dt = \int_0^z \frac{dz'}{(1+z')H(z')}. \quad (16.65)$$

特别的，如果取 $t = 0$ ，相应 $z = \infty$ ，则有宇宙年龄的计算公式

$$t_0 = \int_0^\infty \frac{dz}{(1+z)H(z)}. \quad (16.66)$$

类似的，也有精确的距离-红移关系

$$d_p = a_0 \int_t^{t_0} \frac{dt}{a(t)} = \int_0^z \frac{dz'}{H(z')}. \quad (16.67)$$

很显然，无论是为了计算出宇宙年龄也好，还是为了得到精确的距离-红移关系也好，我们都需要先知道 $H(z)$ 的表达式，为此就需要求助于弗里德曼方程，下一小节我们再来讨论这个问题。

16.2.5 宇宙学参数及其确定

物质、辐射以及宇宙学常数

宇宙的不同物质组分有不同的物态方程，通常可以刻画为

$$p = w\rho. \quad (16.68)$$

其中 w 是一个与时间无关的常数，称之为物态参数。如果要求物质内容满足主能量条件，那么 $|w| \leq 1$ 。一般来说，这些物质组分可以分成如下三类：**1.** $w = 0$ ，即 $p = 0$ ，这是一些非相对论性的物质，即所谓的尘埃物质，后文也简称为**物质**（和前文的物质概念有所不同

同)。比如所有星系构成的理想流体就可以看成是这样的物质，一个星系就是这尘埃中的一个“颗粒”。**2.** $w = 1/3$, 根据第十四章的知识, 这种物质组分是极端相对论性的, 后文简称为**辐射**。比如充满宇宙空间的微波背景辐射。**3.** $w = -1$, 即 $p = -\rho$, 比方说宇宙学常数 (根据第十四章的知识)。

将物态方程 (16.68) 代入能量守恒方程 (16.46), 即可以得到 $\dot{\rho}/\rho = -3(1+w)\dot{a}/a$, 积分即可以得到

$$\rho \propto a^{-3(1+w)}. \quad (16.69)$$

很显然, 对于物质

$$\rho_M \propto a^{-3}, \quad (16.70)$$

由于空间区域的体积 $\propto a^3$, 所以很显然, 对于物质, 随着宇宙的膨胀, 一个空间区域内的总能量保持为常数。但是, 对于辐射

$$\rho_R \propto a^{-4}, \quad (16.71)$$

很显然, 随着宇宙的膨胀 (即 a 的增加), 辐射的能量密度衰减得更快, 以至于一个空间区域内的辐射总能量不可能保持为常数。这是因为随着宇宙的膨胀, 光子的波长也膨胀了, 从而光子的频率就降低了, 也就是所谓的红移了。而对于宇宙学常数 Λ , 由于 $w = -1$, 所以很显然,

$$\rho_\Lambda \propto a^0, \quad (16.72)$$

当然, 这是我们早就知道的, 宇宙学常数的能量密度是一个常数, 不随宇宙膨胀而衰减。

从以上分析可知, 随着宇宙的膨胀, 辐射的能量密度衰减得最快, 物质次之, 而宇宙学常数 (真空能) 的能量密度则是不变的。当今的天文学观测发现, 宇宙中的可见物质约占 4%, 而暗能量 (可以用宇宙学常数来代表) 则约占 68%。由此可以想见, 在极早期宇宙中物质和辐射的能量应该占据绝大部分的比例, 而暗能量的比重则微乎其微, 但是随着宇宙的膨胀, 最终却是暗能量取得了主导地位。进一步推论, 如果宇宙继续膨胀下去, 暗能量的比重将越来越大, 其它物质成分将可以被忽略, 从而宇宙将进入一个渐进德西特时空。但是, 由于辐射随着宇宙膨胀衰减得最快, 所以倒推回去, 最早的时候, 宇宙必定是由极端相对论性的辐射主导的, 然后过渡到物质主导, 最后可能再过渡到暗能量主导。物质和辐射相等的时刻大约对应于红移 $z \simeq 10^4$ 。

在辐射主导的时期, $\rho \sim a^{-4}$, 代入弗里德曼方程 (16.49), 由于此时 a 很小, 而空间曲率能 $-\frac{k}{a^2 L^2} \sim a^{-2}$, 因此相比于辐射能可以忽略, 进而可将弗里德曼方程 (16.49) 近似为

$$\left(\frac{\dot{a}}{a}\right)^2 \sim a^{-4} \Rightarrow \dot{a} \sim a^{-1}, \quad (16.73)$$

积分即可以得到

$$a(t) \sim t^{1/2}. \quad (16.74)$$

特别的, $t = 0$ 时, $a(0) = 0$, 相应的空间曲率 ($\sim a^{-2}$) 即为无穷大, 物质的能量密度也为无穷大。这说明 $t = 0$ 时, 时空是奇性的, 这就是所谓的大爆炸奇点。

在物质主导的时期, $\rho \sim a^{-3}$, 代入弗里德曼方程 (16.49) 并忽略空间曲率能, 即有

$$\left(\frac{\dot{a}}{a}\right)^2 \sim a^{-3} \Rightarrow \dot{a} \sim a^{-1/2}, \quad (16.75)$$

积分即可以得到

$$a(t) \sim t^{2/3}. \quad (16.76)$$

不难看出, 无论是辐射主导的时期还是物质主导的时期, 虽然宇宙都在膨胀, 但是均有

$$\ddot{a} < 0, \quad (16.77)$$

因此宇宙是减速膨胀的。

但是在宇宙学常数主导的时期, $\rho = \frac{\Lambda}{8\pi G}$, 代入弗里德曼方程 (16.49) 并忽略空间曲率能, 即有

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\Lambda}{3}, \quad (16.78)$$

积分即得

$$a(t) \sim e^{\sqrt{\frac{\Lambda}{3}}t}. \quad (16.79)$$

很显然这时宇宙不仅加速膨胀, 而且是指数膨胀。实际上, 正是因为观测到当前宇宙在加速膨胀, 人们才推断出暗能量存在并进一步推断出其在当前约占 68% 的, 至于这个值如何可以推断出来我们后文再讨论。

利用 $a^{-1} = 1 + z$, 我们也可以将上述各组分量能量密度对 a 的关系式重新表示成

$$\rho_M = \rho_{M,0}(1+z)^3, \quad \rho_R = \rho_{R,0}(1+z)^4, \quad \rho_\Lambda = \rho_{\Lambda,0}, \quad (16.80)$$

式中 $\rho_{M,0}, \rho_{R,0}, \rho_{\Lambda,0}$ 分别表示当今的物质能量密度、当今的辐射能量密度、以及当今的宇宙学常数能量密度。从而宇宙内含物的总能量密度 $\rho = \rho_R + \rho_M + \rho_\Lambda$ 可以表示成

$$\rho = \rho_{R,0}(1+z)^4 + \rho_{M,0}(1+z)^3 + \rho_{\Lambda,0}. \quad (16.81)$$

宇宙学参数

下面我们把弗里德曼方程 (16.49) 写成如下形式

$$H^2 = \frac{8\pi G}{3}\rho - K(1+z)^2. \quad (16.82)$$

习惯上常常定义临界密度 $\rho_{c,0}$ 如下

$$\rho_{c,0} \equiv \frac{3H_0^2}{8\pi G} \approx 9.2 \times 10^{-27} \text{kg} \cdot \text{m}^{-3}, \quad (16.83)$$

数值上就是大约每立方米 5.5 个质子。进而即可以把弗里德曼方程进一步改写成

$$H^2 = H_0^2 \left[\frac{\rho}{\rho_{c,0}} - \frac{K}{H_0^2} (1+z)^2 \right] \quad (16.84)$$

将前面的 (16.81) 代入这个式子, 即有

$$H^2 = H_0^2 [\Omega_R(1+z)^4 + \Omega_M(1+z)^3 + \Omega_\Lambda + \Omega_k(1+z)^2]. \quad (16.85)$$

式中

$$\Omega_R \equiv \frac{\rho_{R,0}}{\rho_{c,0}}, \quad \Omega_M \equiv \frac{\rho_{M,0}}{\rho_{c,0}}, \quad \Omega_\Lambda \equiv \frac{\rho_{\Lambda,0}}{\rho_{c,0}}, \quad \Omega_k \equiv -\frac{K}{H_0^2}. \quad (16.86)$$

(16.85) 式正是我们一直想找的 $H(z)$ 的表达式。特别的, 在 (16.85) 中取 $z=0$ (今天), 则有

$$\Omega_R + \Omega_M + \Omega_\Lambda + \Omega_k = 1. \quad (16.87)$$

所以, 参数 $\Omega_R, \Omega_M, \Omega_\Lambda, \Omega_k$ 可以解释成今天宇宙各组分能量密度所占的百分比, 称作宇宙的密度参数。特别的, Ω_k 是空间曲率所占的百分比。不过值得注意的是, Ω_k 可以取负数, 当空间曲率为正时, Ω_k 就是负数。

很显然, Ω_k 是正是负完全取决于 $\Omega_R + \Omega_M + \Omega_\Lambda$ 是大于 1 还是小于 1, 也就是取决于 $\rho_0 = \rho_{R,0} + \rho_{M,0} + \rho_{\Lambda,0}$ 是大于 $\rho_{c,0}$ 还是小于 $\rho_{c,0}$ 。如果 $\rho_0 > \rho_{c,0}$, 则 $\Omega_k < 0$, 空间曲率为正, 宇宙为闭宇宙。如果 $\rho_0 = \rho_{c,0}$, 则 $\Omega_k = 0$, 空间曲率为零, 宇宙为平坦宇宙。如果 $\rho_0 < \rho_{c,0}$, 则 $\Omega_k > 0$, 空间曲率为负, 宇宙为开宇宙。对宇宙微波背景角功率谱第一声学峰位置 (见参考文献) 的拟合告诉我们 Ω_k 非常接近于 0, 即 $\Omega_k \approx 0$, 宇宙非常接近于平坦的。

所以, 不同的宇宙学模型基本可以通过如下四个参数来标志

$$H_0, \quad \Omega_R, \quad \Omega_M, \quad \Omega_\Lambda. \quad (16.88)$$

观测宇宙学的一个重要目标就是确定这四个参数的值, 一般来说, 这是一件非常复杂的事情, 要结合多种不同的观测结果。 H_0 的观测结果前面已经说过了, 剩下三个密度参数的观测结果大约为

$$\Omega_R \approx 5 \times 10^{-5}, \quad \Omega_M \approx 0.3, \quad \Omega_\Lambda \approx 0.7, \quad (16.89)$$

这也就是前面所说的暗能量约占 68% 的来源, 但观测如何可以定出这个结果呢? 考虑到辐射的百分比很低, 可以忽略, 那观测是如何定出 $\Omega_M \approx 0.3, \Omega_\Lambda \approx 0.7$ 的呢? 这个问题我们稍后再讨论。

不仅如此, 当前的减速参数 q_0 也可以用这些密度参数表达出来。为此我们先把第二弗里德曼方程 (16.50) 写成如下形式

$$\frac{a\ddot{a}}{a^2} = -\frac{4\pi G}{3H^2}(\rho + 3p) = -\frac{4\pi G}{3H^2}(2\rho_R + \rho_M - 2\rho_\Lambda), \quad (16.90)$$

式中最后一个等于号我们代入了各组分的物态方程。进而根据减速参数的定义，即有

$$q_0 = \frac{1}{2}\Omega_M + \Omega_R - \Omega_\Lambda. \quad (16.91)$$

很显然，如果宇宙中没有暗能量（宇宙学常数），即 $\Omega_\Lambda = 0$ ，则必定有 $q_0 > 0$ ，从而宇宙必定减速膨胀。因此观测到宇宙加速膨胀，就意味着宇宙学常数必定大于零，或者说必定有大于零的暗能量存在。实际上，由 $q_0 < 0$ ，就必定有 $\Omega_\Lambda > \frac{1}{2}\Omega_M$ 。

另一方面，利用 (16.85) 式给出的 $H(z)$ 表达式，代入宇宙年龄的计算公式 (16.66)，即有

$$t_0 = \frac{1}{H_0} \int_0^\infty \frac{dz}{(1+z)E(z)}, \quad (16.92)$$

式中

$$E(z) \equiv \sqrt{\Omega_R(1+z)^4 + \Omega_M(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\Lambda}. \quad (16.93)$$

代入上面给出的各宇宙学参数值（其中 $\Omega_k = 0$ ）并进行数值积分，就可以得到宇宙年龄 $t_0 \approx 138$ 亿年，常常也记作 $t_0 \approx 13.8 \text{Gyr}$ ，Gyr 即是 Gigayear 的缩写，1Gyr = 十亿年。粗略来说，也有 $t_0 \approx 1/H_0$ 。

同样的，假如把 $H(z)$ 的表达式代入距离-红移关系的计算公式 (16.67)，即有

$$d_P(z) = \frac{1}{H_0} \int_0^z \frac{dz'}{E(z')}. \quad (16.94)$$

特别的，如果所考察的红移不太大，比如 $z < 6$ （通常观测的 Ia 型超新星都处于这个红移范围），那注意到 Ω_R 和 Ω_k 都很接近于零，进而即可以在上述积分中忽略 $E(z)$ 中的 Ω_R 和 Ω_k 相关项，进而得到

$$d_P(z) = \frac{1}{H_0} \int_0^z \frac{dz'}{\sqrt{\Omega_M(1+z')^3 + \Omega_\Lambda}}, \quad z < 6. \quad (16.95)$$

以上结果告诉了我们一个如何通过拟合 Ia 型超新星的距离-红移关系，进而确定 $\Omega_\Lambda \approx 0.7$ 的办法。为此还需要注意到，由于 Ω_R 和 Ω_k 相比可以忽略，所以 (16.87) 式可以近似为

$$\Omega_M + \Omega_\Lambda = 1. \quad (16.96)$$

代入上面的 (16.95) 式，即可以得到

$$d_P(z) = \frac{1}{H_0} \int_0^z \frac{dz'}{\sqrt{(1-\Omega_\Lambda)(1+z')^3 + \Omega_\Lambda}}, \quad z < 6. \quad (16.97)$$

也即是说，Ia 型超新星的距离红移关系取决于 Ω_Λ 的值，因此只需要对它们的距离-红移关系进行拟合，就能够定出 Ω_Λ 的值，如图 (16.2)。这就是确定 Ω_Λ 观测值的重要方法之一。

宇宙命运

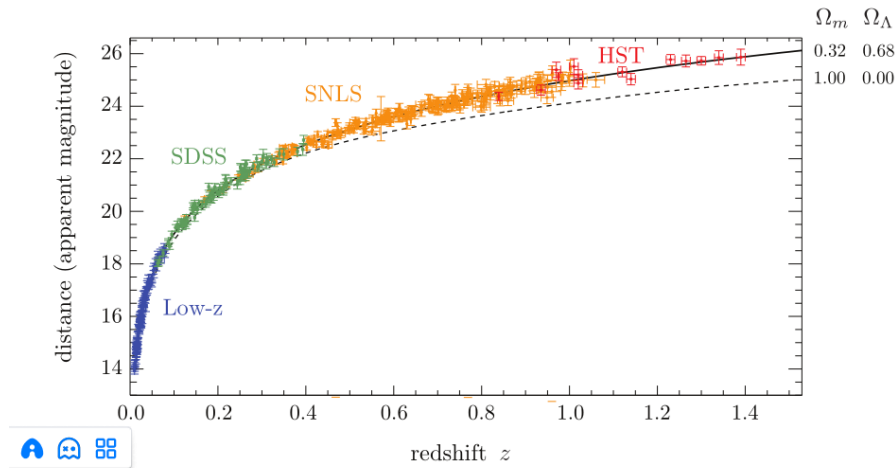


Figure 16.2: Ia 型超新星的距离红移关系 (当然严格来说这里的距离是光度距离, 但它其实完全可以由共动距离 d_p 决定) 与暗能量的发现。如果我们假设宇宙是平坦的, 那么超新星显然比在仅由物质构成的宇宙 $\Omega_M = 1$ 中预测的要暗淡 (或更远)。(SDSS = Sloan Digital Sky Survey; SNLS = SuperNova Legacy Survey; HST = Hubble Space Telescope.)

当然, 如果是想预言宇宙的未来, 那红移这个概念就不好用了, 这时候我们可以把方程 (16.85) 恢复成

$$H^2 = H_0^2 [\Omega_R a^{-4} + \Omega_M a^{-3} + \Omega_\Lambda + \Omega_k a^{-2}]. \quad (16.98)$$

利用哈勃参数 H 的定义, 显然也有

$$\frac{1}{2} \dot{a}^2 - \frac{1}{2} H_0^2 [\Omega_R a^{-2} + \Omega_M a^{-1} + \Omega_\Lambda a^2 + \Omega_k] = 0. \quad (16.99)$$

可以把这个方程看作是一个沿着 a 轴运动的一维粒子的机械能守恒方程, 这个粒子的质量为 1, 势能为 $V(a) = -\frac{1}{2} H_0^2 [\Omega_R a^{-2} + \Omega_M a^{-1} + \Omega_\Lambda a^2 + \Omega_k]$, 总机械能为零。

通过分析这个一维粒子在势能曲线 $V(a)$ 中的运动, 我们就能搞明白宇宙未来的命运 (当前这个粒子位于 $a_0 = 1$ 处, 速度 $\dot{a} > 0$), 是继续膨胀下去, 还是被 $V(a)$ 的势垒部分挡住, 然后调头开始收缩 (在物理上即是由于物质之间的万有引力 (为吸引力) 占据了主导地位, 使得宇宙在某个时刻开始收缩)。由于这种分析不过就是一道略微复杂些的力学习题, 所以我们留给读者自行完成。

16.2.6 宇宙学视界

宇宙学视界 (Cosmological Horizon) 是现代宇宙学中的一个重要概念, 它描述了我们在宇宙中能够观测到的最远距离。由于宇宙的膨胀和光速的限制, 我们无法观测到无限远的宇宙, 因此宇宙学视界定义了可观测宇宙的边界。视界的存在是因为宇宙的膨胀速度可能超过光速, 因此某些区域的光可能永远无法到达我们。

为了讲清楚视界的概念，让我们考虑一个位于 $\chi = 0$ 处的共动观察者，光信号在 t_E 时刻从远处的 χ_A 处发送过来，在 t_R 时刻被这个观察者接收到，则根据前面的 (16.52) 式，必定有

$$\int_{t_E}^{t_R} \frac{dt}{a(t)} = L\chi_A. \quad (16.100)$$

当然，这个式子也可以解释成， $\chi = 0$ 处的这个共动观察者在 t_E 时刻向远处的 χ_A 位置发送光信号， t_R 时刻这个光信号到达 χ_A 。

粒子视界

所谓的粒子视界，指的就是 $\chi = 0$ 位置的这个共动观察者在 t 时刻能够看到的最远距离，很显然这个最远距离处的光一定是在宇宙诞生之初的 $t = 0$ 时刻发出的，因此根据 (16.100) 式，这个最远距离的 χ_A 值为

$$\int_0^t \frac{dt'}{a(t')} = L\chi_A. \quad (16.101)$$

由于这是共动观察者 t 时刻观测到的，所以对应的固有距离就应该是 $d_{ph}(t) = a(t)L\chi_A$ ，根据上式这也就是

$$d_{ph}(t) = a(t) \int_0^t \frac{dt'}{a(t')}. \quad (16.102)$$

$d_{ph}(t)$ 就是 $\chi = 0$ 处的这个共动观察者在 t 时刻的粒子视界！

当然，由于 $t \rightarrow 0$ 时， $a(t) \rightarrow 0$ ，所以上面这个积分在下限处的贡献可能发散，如果是这种情况，那粒子视界就是无穷大，否则粒子视界就是有限的。比方说，按照前面的标准宇宙学模型，最早期的宇宙是辐射主导的，因此是作减速膨胀，那时 $a(t) \propto t^{1/2}$ ，代入 (16.102) 式不难发现，积分下限处的贡献是有限的。进一步，根据前面的 (16.67) 式，可得当前的粒子视界的大小为

$$d_{ph}(t_0) = a_0 \int_0^{t_0} \frac{dt}{a(t)} = \int_0^\infty \frac{dz}{H(z)} = \frac{1}{H_0} \int_0^\infty \frac{dz}{E(z)} \quad (16.103)$$

代入 $E(z)$ 的表达式，原则上就能算出这个结果。

另外，将 $d_{ph}(t)$ 对时间求导，不难导出如下结果

$$\dot{d}_{ph}(t) = H(t)d_{ph}(t) + 1, \quad (16.104)$$

也即是说，随着宇宙的膨胀 ($H(t) > 0$)，粒子视界的大小也是在膨胀的，而且膨胀的速度满足某种哈勃定律。

粒子视界的意义不仅是说明可观测范围。注意光速是一切信号速度的上限，当两个点的空间距离超过视界大小，它们之间就无法用信号建立联系，因此，这两个点上的物理情况是不可能存在因果关系的。从这个意义上讲， t 时刻粒子视界的大小也反映了此时可能有因果联系的区域的大小。

粒子视界的一个直接后果就是宇宙学模型中的视界问题。简而言之，我们观测到即使对于分得非常开的区域，它们的宇宙微波背景辐射温度也是高度一致的。然而在前面的标准宇宙学模型中，这样两个分得很开的点在光子脱耦形成微波背景的时候它们之间的距离可以超过当时的粒子视界大小，因此那时候这两个点之间就不可能有因果联系。而这两点的热平衡也不可能是此后在宇宙继续膨胀到今天的过程中形成的，因为此后光子已经脱耦变成没有相互作用的自由光子了。那为什么这样的两个点有相同的温度呢？

事件视界

虽然粒子视界会随着宇宙的膨胀而膨胀，但这并不意味着 $\chi = 0$ 处的这个共动观察者在未来可以看到 t 时刻的任何事件，实际上，即使这个观测者一直等到无穷远未来， t 时刻的宇宙也总存在一些他观测不到的区域。实际上，根据 (16.100) 式，这个共动观察者在无穷远将来能够观察到的最远坐标 χ_A 为

$$\int_t^\infty \frac{dt'}{a(t')} = L\chi_A. \quad (16.105)$$

(当然，这个 χ_A 坐标也可以解释成， t 时刻 $\chi = 0$ 处的观察者向远处发送光信号，这光信号在无穷远未来能够到达的最大 χ 值。)

由于 χ_A 是宇宙在 t 时刻的坐标，所以相应固有距离 $d_{eh}(t)$ 当然就是 $d_{eh}(t) = a(t)L\chi_A$ ，根据上一个式子，这也就是

$$d_{eh}(t) = a(t) \int_t^\infty \frac{dt'}{a(t')}. \quad (16.106)$$

很显然 $d_{eh}(t)$ 就是共动观察者在无穷远将来能看到的宇宙在 t 时刻的最大半径，称作这个共动观察者的事件视界。

不难导出，事件视界满足

$$\dot{d}_{eh}(t) = H(t)d_{eh}(t) - 1. \quad (16.107)$$

这也就是说，如果某个时刻的事件视界 $d_{eh}(t) < 1/H$ ，则随着时间的推移，它是在缩小的。

哈勃视界

除了以上两种常用宇宙学视界以外，还有一种视界也很常用，即所谓的哈勃视界 d_H ，其定义就是退行速度等于光速的那个距离。

具体来说，根据哈勃定律 (16.28)，远处星系相对于 $\chi = 0$ 处共动观察者的退行速度为 $u(t) = H(t)d_P(t)$ ， $d_P(t)$ 为远处星系离共动观察者的固有距离，所谓的哈勃视界 d_H ，就是使得 $u(t) = 1$ 的那个距离 $d_P(t)$ ，换言之，

$$d_H = \frac{1}{H}. \quad (16.108)$$

当然由于我们取了光速 $c = 1$ 的单位制，如果恢复成常用单位制的话应该是 $d_H = c/H$ 。常常也把哈勃视界称作哈勃半径。

16.3 暗物质与暗能量

暗物质

我们来考察一下当前宇宙物质的成分，即 $\rho_{M,0}$ 的成分。暂且假定它主要由星系组成。由此就可以通过天文观测数据对 $\rho_{M,0}$ 进行一个独立的估计，如果由此得出来的 Ω_M 的确约为前文确定的 0.3，那就说明这个假设基本是成立的。

设当今宇宙星系的数密度为 n ，每个星系的平均质量为 \bar{M} ，则 $\rho_{M,0} = n\bar{M}$ 。设宇宙空间中单位体积的光度（也就是光度密度）为 \mathcal{L} ，目前它已经有比较可靠的观测值了。再设每个星系的平均光度为 \bar{L} ，则 $\mathcal{L} = n\bar{L}$ ，进而即有

$$\rho_{M,0} = \mathcal{L} \frac{\bar{M}}{\bar{L}}, \quad (16.109)$$

式中 \bar{M}/\bar{L} 叫做星系的平均质光比。很显然，为了估计 $\rho_{M,0}$ ，我们只需估计 \bar{M}/\bar{L} 。实际的测量当然是对具体星系进行的，得到的只是该星系的质光比 M/L ，很显然，只有当这个星系具有相当的代表性时其质光比才可以作为 \bar{M}/\bar{L} 的估计值。好在虽然不同星系的质量差别很大，可以差好几个数量级，但是星系质光比的差别却要小很多。因此通常星系质光比的值都可以作为 \bar{M}/\bar{L} 的估计值。

要测星系的质光比，关键是测出星系的总质量。下面介绍测量漩涡星系（如银河系）质量的动力学方法。在漩涡星系中，恒星除无规运动外主要就是以星系的引力为向心力绕星系中心旋转。为简化讨论，可以把漩涡星系看成一个圆盘。则由牛顿引力理论可知，距离星系中心 r 处的恒星的旋转速度 $v(r)$ 由下式决定

$$v^2(r) = G \frac{M(r)}{r}, \quad (16.110)$$

其中 $M(r)$ 是星系在半径 r 之内的质量。函数 $v(r)$ 的曲线就称之为星系的旋转曲线。天文学家把许多星系的旋转曲线都测出来了。

假设 R_0 代表星系光度消失处的半径，则 $M(R_0)$ 就是星系内发光物质的总质量。如果星系在 $r > R_0$ 之外没有质量，则由 (16.110) 式可知，在 $r > R_0$ 处 $v(r)$ 曲线应该按照 $r^{-1/2}$ 的规律下降，如图 (16.3) 中的曲线 A(虚线)。然而实测的星系旋转曲线并非如此，而是先从星系中心陡然升起，然后近似水平地延伸，直到超出 R_0 以外很远以至于无法测量为止，如图 (16.3) 中的曲线 B。这就表明，漩涡星系在发光部分之外有一个半径比 R_0 大得多的暗晕，由暗物质组成，其质量约为发光部分的 3–10 倍。除漩涡星系之外还有其他星系，例如椭圆星系，有证据表明椭圆星系也有数量可观的暗物质。

实际上，如果仅仅只考虑星系中的发光物质，并由此估算出 $\rho_{M,0}$ ，那么由此得出的 Ω_M 将仅仅为

$$\Omega_M(\text{发光物质}) < 1\% \quad (\text{约为} 0.5\%), \quad (16.111)$$

很显然这远远小于 0.3(即 30%)，这也说明了应该有大量暗物质存在。总之，可以肯定，宇宙中的暗物质在质量上大大超过了发光物质。

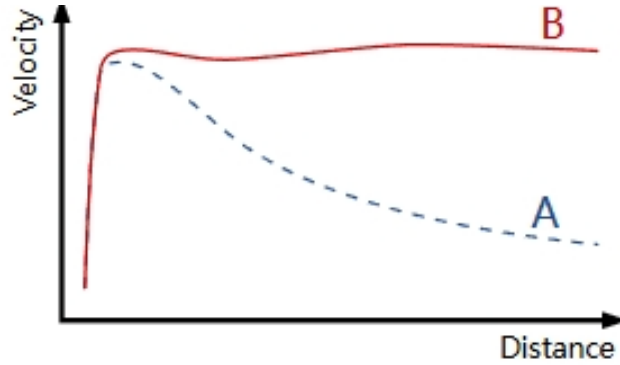


Figure 16.3: 星系旋转曲线示意图。其中 B 为实测的曲线，A 为我们预期的曲线。

暗物质的存在本来不值得惊讶，因为很容易举出暗物质的例子：行星、暗淡的小恒星、白矮星、中子星、黑洞、瓦解了的星系以及星际和星系际稀薄气体.... 这些例子中的暗物质都由重子（主要是质子和中子）组成。然而原初核合成理论对当今宇宙重子物质的百分比是有限制的。具体来说如下。

以 n_{b0} 代表当今重子（主要是核子）数密度， m_N 代表每个核子的质量，则当今重子物质的质量密度 $\rho_{b,0} = n_{b0}m_N$ ，它对应的 $\Omega_b = \rho_{b,0}/\rho_{c,0}$ 为

$$\Omega_b = \frac{8\pi G}{3H_0^2} n_{b0} m_N = \frac{8\pi G}{3H_0^2} m_N n_{\gamma 0} \eta, \quad (16.112)$$

式中 $n_{\gamma 0}$ 为当今的光子数密度，其值可由对宇宙微波背景的观测推算出来， η 是前面原初核合成小节中引入的参数， $\eta = n_b/n_\gamma = n_{b0}/n_{\gamma 0}$ 。前面原初核合成小节中给出了 η 的范围。

由于宇宙微波背景是一个黑体辐射谱，所以根据斯忒藩-玻尔兹曼定律，其能量密度正比于温度的四次方，从而光子数密度正比于温度的三次方，根据观测到的宇宙微波背景温度为 2.728K 即可以推算出

$$n_{\gamma 0} \approx 4.1 \times 10^8 \text{ 个}/m^3. \quad (16.113)$$

结合前面给出的 η 的取值范围，就可以得出

$$2.3\% < \Omega_b < 5.1\%. \quad (16.114)$$

显然 Ω_b 的值大于发光物质值 (0.5%)，由此可知，重子物质中只有一小部分是发光物质，大部分重子物质为暗物质。

但是， $\Omega_b < 5.1\% < 30\%$ 也告诉我们，宇宙中的暗物质大部分是非重子暗物质！当然，非重子暗物质远不如重子暗物质好想象，对非重子暗物质候选者的探究已经成为当今物理学的热门课题。实际上，我们通常所谓的暗物质问题更多指的就是探究这些非重子暗物质。

非重子暗物质大约应为非常稳定的非重子粒子。中微子如果有质量的话，将是很明显的候选者。从 1997 年开始人们倾向于相信某些中微子有大约 0.1eV 量级的质量，这些中微子对 Ω_M 的贡献约为 0.3% ~ 15%。当然，它们是所谓的热暗物质，热指的是在它们退

耦时运动速度依然很大。除了热暗物质以外，还有所谓的冷暗物质，冷指的是粒子以低速运动，它们曾活跃于温度很高的极早期宇宙，并随温度下降而退耦，退耦时速度已经不高，并作为背景遗迹留存至今，所以是冷暗物质。但是，虽然一些超越粒子物理标准模型的理论预言有这样的有质量非重子粒子作为候选者而存在，但目前人们还没有观测到它们。

今天的宇宙在大尺度上是均匀各向同性的，但在小一点的尺度上却呈现出有层次的结构，也就是存在恒星、星系、星系团以及超星系团。正如前面介绍热大爆炸基本概念那一小节(第16.1.1小节)中所讲的，今天的复杂结构起源于极早期宇宙中非常微小的密度涨落，然后通过引力系统的不稳定性而被放大形成。暗物质的存在对于具体解释结构形成也是必须的，因为如果认为宇宙中的物质主要是重子物质，那结构形成理论就会遇到严重困难。

根据非重子暗物质是热暗物质还是冷暗物质的区分，我们有两种结构形成模型。在热暗物质模型中，结构的形成是“自上而下”的，即先形成超星系团，然后再逐级破裂为星系团和星系。相反，在冷暗物质模型中，结构的形成是“自下而上”的，即先形成星系，然后逐级形成星系团和超团。这两类模型中，冷暗物质模型已经取得了很大的成功。今天人们广泛地认为，宇宙物质以非重子暗物质为主，而非重子暗物质中冷暗物质又是必不可少的。

不过，由于观测上一直还没有发现暗物质粒子，所以目前也有一些非主流的理论，其中人们通过修改万有引力规律来避免对暗物质的需要。

暗能量

正如前文讲过的，宇宙加速膨胀的发现，使得人们推断暗能量的存在，对暗能量的最简单也是最广为接受的解释就是宇宙学常数。通常认为宇宙学常数就是量子场的真空能，不过，正如前面的章节中提到过的，根据目前量子场论算出来的真空能比观测宇宙学确定的宇宙学常数大太多了，完全不在一个量级，这到底是怎么回事目前人们还不清楚，这就是著名的宇宙学常数问题。当然，在本书作者看来，最大的可能性是，当前的量子场论只在局域小尺度上才通过了实验检验，这样的局域量子场论很可能不能适用于宇宙学的大尺度，因此我们通过局域场论来计算整个宇宙学尺度上的真空能是不对的，因为万有引力的存在修改了量子场论在大尺度上的性质。但，关键问题在于，正确的做法是什么？这就要留给勇敢的读者们自己去探索了。

不过，暗能量也可以不是宇宙学常数，比如也有人提出暗能量密度是某种动态标量场的能量密度，等等。总之，关于暗能量的理论很多，其中宇宙学常数是最简单最广为认可的。但是，根据暗能量光谱仪(Dark Energy Spectroscopic Instrument, DESI)最近的观测数据，也的确有迹象表明暗能量不是一个宇宙学常数，而是会随时间演化的东西。

如果认为暗能量是宇宙学常数，暗物质是冷暗物质(cold dark matter, CDM)，由此得到的宇宙学模型就是所谓的 Λ CDM模型，它不仅与观测结果吻合得比较好，而且也是最广为接受的宇宙学模型，通常也称之为标准宇宙学模型。

16.4 宇宙暴胀

16.4.1 暴胀的基本概念

热大爆炸宇宙学虽然很成功，但它本身也有一些严重的问题。其中**视界问题**我们前面介绍粒子视界时已经讲过了。简单来说，即是按照标准宇宙学模型，光子脱耦时的粒子视界不够大，以至于无法解释超出视界范围的两个点为什么有相同的温度（正如前面讲到过的，宇宙微波背景的温度涨落幅度只有大约百万分之五）。宇宙为什么会如此均匀和各向同性呢？（以至于在宇宙早期即使天空中两个因果无关的点也几乎完全均匀。）

在 1970 年代之前，一些物理学家说，这是由于热大爆炸开始时的初始条件非常特殊，而初始条件要用量子引力来研究，因此这个问题超出了当前理论物理的能力范围。

热大爆炸无法解释的第二个问题是，**平坦性问题**。即宇宙空间为何如此平坦，为何 Ω_k 如此接近于零，通常来说，这同样意味着需要对初始条件进行极端精细的微调。

第三个问题是所谓的**磁单极子问题**。具体来说即是，在粒子物理大统一理论的框架下，早期宇宙在温度极高时会发生对称性破缺，导致磁单极子的产生。理论预测磁单极子的数密度为 $n_{monopole} \propto T_G^3$ ，其中 T_G 是大统一相变的温度（约为 10^{16} GeV）。根据这一预测，现今宇宙中应该有大量磁单极子。然而，尽管人们进行了大量实验和观测，至今未发现磁单极子的存在。观测上限表明，磁单极子的数密度应该远低于刚才说到的理论预言。那么，这是为什么呢？磁单极子为何如此稀少呢？

为了解决热大爆炸的这些问题，阿兰·古斯于 1980 年提出，在早期辐射主导的宇宙减速膨胀阶段之前，宇宙还经历了一个很短暂的指数级快速膨胀阶段，称作暴胀阶段，它将极早期宇宙的一个小区域迅速拉伸到整个可观测宇宙的范围。暴胀发生在大约 10^{-36} s 到 10^{-32} s 之间，这个阶段宇宙大约膨胀了 e^{50} 到 e^{60} 倍，这个暴胀阶段结束以后宇宙才开始按照原来的热大爆炸宇宙学演化。

由于粒子视界 $d_{ph}(t) = a(t) \int_0^t \frac{dt'}{a(t')}$ ，所以如果 $a(t)$ 指数膨胀了 e^{50} 到 e^{60} 倍，那么不难算出， $d_{ph}(t)$ 也会指数膨胀这么多倍。这就能够解决热大爆炸模型中光子脱耦时的粒子视界不够大的问题，因为原来按照热大爆炸模型算出来的没有因果联系的区域现在就可以被包含在同一个粒子视界中了，那这就解决了热大爆炸的视界问题。

同样的，暴胀阶段的存在也能够解决平坦性问题，为了看清楚这一点，假设我们定义 t 时刻的曲率密度参数 $\Omega_{k,t}$ 为 $\Omega_{k,t} \equiv -\frac{K}{a^2(t)H^2}$ ，则不难看出，一个指数膨胀的 $a(t)$ 会把 $\Omega_{k,t}$ 指数压低，从而相当于精细地把宇宙调节到了非常接近于平坦。直观上也就是急剧的膨胀把空间曲率给拉平了。

宇宙暴胀对磁单极子为何如此稀少的解释就更加简单了，因为暴胀导致的宇宙体积的急剧增大一下子把磁单极子给稀释了！

这里有一个问题，暴胀阶段宇宙急剧膨胀这么多倍，那温度应该就急剧降低到接近于零了，那暴胀结束以后宇宙怎么又按照原来的热大爆炸演化了呢，毕竟热大爆炸是需要高温的，这高温从哪里来呢？别着急，这个问题我们后文会给出解释。

先让我们来说说暴胀理论的一个巨大成就。前面讲到过，宇宙的结构形成需要密度涨落的存在，问题是原来的热大爆炸宇宙学很难解释原初的密度涨落从哪里来。暴胀理论能给出自然的解释！当然，你可能想到原初的密度涨落来自于量子涨落，但问题是量子涨落

完全是微观的，它和宇宙学不在一个尺度上。但是，暴胀理论恰恰能自然地连接这两个尺度，因为暴胀恰好可以把原来的微观尺度拉大到宇宙学尺度！而且具体进行的计算也表明这个想法是完全能够成立的，与观测可以吻合。当然，这样的计算对于我们的极简宇宙学而言过于专门了，所以我们这里略过。

16.4.2 暴胀场

一个重要的问题是，什么驱动了宇宙暴胀。它不可能是我们今天观测到的暗能量，虽然暗能量的确能让宇宙加速膨胀，但是，暗能量在早期宇宙中占比是可以忽略的。但是暗能量启发我们，只要我们在那个时期的宇宙中引入一种 $w \approx -1$ 的常数能量密度，并让它占主导地位，那它就能和暗能量一样让宇宙加速膨胀甚至指数膨胀。

在粒子物理中，产生常数能量密度的最简单办法就是引入一个不随空间坐标变化的标量场，并让它有一个比较大的势能密度，那这个势能密度就会产生一个常数能量密度。基于这些想法，人们引入了暴胀场，这是一个标量场，它驱动了宇宙暴胀。

作为标量场，暴胀场的作用量为

$$S_I = \int \sqrt{-g} d^4x \left[-\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - \mathcal{U}(\phi) \right]. \quad (16.115)$$

根据第 14 章的相关知识可知，作为弯曲时空中的标量场，暴胀场对应于一种理想流体，其能量密度 ρ_I 和压强 p_I 分别为

$$\rho_I = -\frac{1}{2} (\partial\phi)^2 + \mathcal{U}(\phi), \quad p_I = -\frac{1}{2} (\partial\phi)^2 - \mathcal{U}(\phi). \quad (16.116)$$

为了将注意力集中在时间演化上，可以假设暴胀场仅仅是时间 t 的函数，与空间坐标无关。代入 FLRW 度规，进而即有

$$\rho_I = \frac{1}{2} \dot{\phi}^2 + \mathcal{U}(\phi), \quad p_I = \frac{1}{2} \dot{\phi}^2 - \mathcal{U}(\phi). \quad (16.117)$$

不难看出，暴胀场的物态参数 $w_I = \frac{p_I}{\rho_I}$ 是在 -1 到 $+1$ 之间的，并且如果暴胀场的动能远远小于势能，即

$$\dot{\phi}^2 \ll \mathcal{U}(\phi). \quad (16.118)$$

那么就有 $w_I \approx -1$ 。从而这时候暴胀场就可能像宇宙学常数那样驱动宇宙指数膨胀。

进一步，把 FLRW 度规代入暴胀场作用量 S_I 的表达式，进而即有

$$S_I = \int \sqrt{h} d^3x dt a^3(t) \left[\frac{1}{2} \dot{\phi}^2 - \mathcal{U}(\phi) \right]. \quad (16.119)$$

式中 h 是空间度规 h_{ij} 的行列式。对这个作用量变分，就能得到暴胀场的运动方程

$$\ddot{\phi} + 3H\dot{\phi} + \mathcal{U}'(\phi) = 0, \quad (16.120)$$

式中 \mathcal{U}' 表示函数 \mathcal{U} 对自变量 ϕ 的导数。很显然，这个方程可以看成是描述一个沿着 ϕ 轴运动的粒子，它受到势能曲线 $\mathcal{U}(\phi)$ 的作用，方程中的 $3H\dot{\phi}$ 项相当于一种阻尼项，也即是说，宇宙的膨胀会阻碍暴胀场随时间的变化。

把 ρ_I 的表达式代入弗里德曼方程, 并注意到暴胀场是主导的 (从而可以忽略其它一些项), 即有

$$H^2 = \frac{8\pi G}{3} \left[\frac{1}{2} \dot{\phi}^2 + \mathcal{U}(\phi) \right]. \quad (16.121)$$

如果 (16.118) 式得到满足, 那就有

$$H^2 \approx \frac{1}{3M_{pl}^2} \mathcal{U}(\phi). \quad (16.122)$$

式中 $M_{pl} = 1/\sqrt{8\pi G} = 2.4353 \times 10^{18} \text{GeV}$, 称作约化的普朗克质量。很显然, 只要 ϕ 场随时间变化得足够缓慢, 从而使得 $\mathcal{U}(\phi) > 0$ 近似为常数, 那宇宙近似就是指数膨胀的。为此我们进一步要求

$$\ddot{\phi} \ll \mathcal{U}'(\phi). \quad (16.123)$$

从而 (16.120) 就可以近似成

$$3H\dot{\phi} \simeq -\mathcal{U}'(\phi). \quad (16.124)$$

这时候 $\mathcal{U}(\phi)$ 可以看作是一个有效的宇宙学常数, 只不过相比于今天观测到的有效宇宙学常数, 它要大好多个数量级, 以至于可以在暴胀阶段占据了主导地位。

(16.118) 式和 (16.123) 式合起来就是所谓的慢滚条件。如果慢滚条件满足, 我们就可以利用 (16.122) 式和 (16.124) 式, 进而有 $\dot{\phi}^2 \sim (\mathcal{U}'/H)^2 \sim M_{pl}^2 \mathcal{U}'^2/\mathcal{U}$, 进而即可以把条件 $\dot{\phi}^2 \ll \mathcal{U}$ 重写成 $M_{pl}^2(\mathcal{U}'/\mathcal{U})^2 \ll 1$ 。进一步, 既然 $\dot{\phi} \sim -M_{pl}\mathcal{U}'/\mathcal{U}^{1/2}$, 则有 $\ddot{\phi} \sim M_{pl}^2(\mathcal{U}'\mathcal{U}''/\mathcal{U}) - M_{pl}^2(\mathcal{U}'^3/(2\mathcal{U}^2))$, 又有 $M_{pl}^2(\mathcal{U}'/\mathcal{U})^2 \ll 1$, 从而条件 $\ddot{\phi} \ll \mathcal{U}'$ 相当于 $M_{pl}^2\mathcal{U}''/\mathcal{U} \ll 1$ 。根据这些, 习惯上常常定义

$$\varepsilon \equiv \frac{1}{2} M_{pl}^2 \left(\frac{\mathcal{U}'}{\mathcal{U}} \right)^2, \quad \eta \equiv M_{pl}^2 \frac{\mathcal{U}''}{\mathcal{U}}. \quad (16.125)$$

从而由慢滚条件即有 $\varepsilon \ll 1, |\eta| \ll 1$, 这就是对暴胀场势能函数 $\mathcal{U}(\phi)$ 的一个限制。

前面说了, 暴胀阶段的有效宇宙学常数很大, 那它后来是怎么变成像今天观测到的那么小的呢? 另一个相关的问题是, 暴胀阶段是如何结束的呢?

这两个问题都可以通过选取合适的势能函数 $\mathcal{U}(\phi)$ 来进行解答。比如说, 可以设想暴胀场的势能函数像图 (16.4) 中所示的那样, 图中左边的平台区满足慢滚条件要求的 $\varepsilon \ll 1, |\eta| \ll 1$, 所以可以设想开始时暴胀场在这个平台区往下慢滚, 从而驱使宇宙指数膨胀。但是, 一旦暴胀场进入图中的势阱区域, 慢滚条件就不满足了, 暴胀也就结束了。尤其是暴胀场可以掉入阱底的极小值位置, 这个位置对应的势能密度可能比平台区小好多个数量级, 从而暴胀场掉到阱底的话, 有效的宇宙学常数就会马上变小好多个数量级。

不仅如此, 暴胀场一旦掉入势阱, 它原来的势能就会转化成动能, 从而使得暴胀场在势阱底部来回振荡, 这种振荡会耦合到辐射场以及其它物质粒子场上, 从而使得振荡的动能转换成辐射场的激发, 以及正反粒子对的产生, 并且多余的暴胀场振荡能量还可以把辐射和物质粒子加热到很高的温度, 从而开始热大爆炸。这个过程就是所谓的宇宙重加热过程, 它解决了前面提到过的暴胀阶段和热大爆炸衔接的问题。

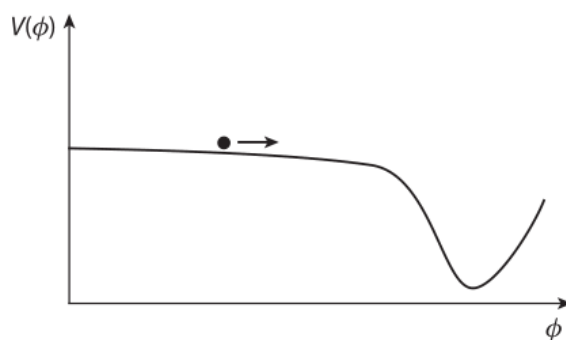


Figure 16.4: 暴胀场势能函数，图中的 $V(\phi)$ 即是正文中的 $\mathcal{U}(\phi)$ 。

16.5 参考文献

关于什么是宇宙微波背景角功率谱，以及其第一声学峰如何确定宇宙非常接近于平坦，最简短而清楚的介绍可以参见 A. Zee, *Einstein Gravity in a Nutshell*, 章节 VIII.3，或者也可以直接问人工智能，比如 deepseek。

本章参考了俞允强《物理宇宙学讲义》以及梁灿彬《微分几何与广义相对论》的相关章节。尤其是俞允强老师的《物理宇宙学讲义》给了本章很大启发。

17. 引力波

17 世纪末，牛顿坐在花园里，目睹了一颗苹果的坠落。这一瞬间的灵感，让他意识到地球与苹果之间存在着一种无形的力——引力。牛顿的万有引力定律成功解释了行星的运动、潮汐的涨落，甚至预言了海王星的存在。然而，牛顿的引力是一种“超距作用”，它瞬间传递，无需时间。这种瞬时性的假设，与 19 世纪电磁学的发展产生了深刻的矛盾。

19 世纪中叶，麦克斯韦建立了电磁场理论，预言了电磁波的存在，并证明光就是一种电磁波。电磁波以有限的速度传播，这一发现动摇了牛顿引力的根基。如果电磁相互作用需要时间传递，那么引力相互作用是否也是如此？这个问题困扰了物理学家数十年，直到爱因斯坦的出现。

1905 年，爱因斯坦提出了狭义相对论，彻底否定了“超距作用”的可能性。任何相互作用的传递速度都不能超过光速，引力也不例外。1915 年，爱因斯坦进一步提出了广义相对论，将引力解释为时空的弯曲。物质告诉时空如何弯曲，时空告诉物质如何运动。在这一全新的理论框架下，引力不再是力，而是时空几何的表现。

广义相对论的方程中隐藏着一个惊人的预言：当物质加速运动时，它会扰动周围的时空，产生一种以光速传播的波动——这就是引力波。引力波是时空的涟漪，是宇宙的琴弦，它携带着源头的信息，穿越浩瀚的宇宙，抵达地球。然而，引力波的信号极其微弱，爱因斯坦本人也曾怀疑人类是否能够探测到它。起初科学家们只能依靠间接证据来推测引力波的存在，比如双星系统中轨道能量损失的现象，这符合引力波辐射带走能量的预期。

一个世纪之后的 2015 年 9 月 14 日，激光干涉引力波天文台（LIGO）首次直接探测到了引力波信号。这一信号来自 13 亿光年外的两个黑洞的并合，标志着人类正式进入引

力波天文学的时代。引力波的探测不仅验证了爱因斯坦的预言，更为我们打开了一扇观测宇宙的新窗口。不同于传统的电磁波观测手段，引力波不受尘埃、气体或其他物质的影响，能够穿透几乎所有的障碍直达地球。这意味着我们可以窥探那些传统望远镜无法触及的极端天体事件，如黑洞碰撞、超新星爆发甚至可能是宇宙大爆炸本身的余晖。通过引力波，我们可以“聆听”黑洞的碰撞、中子星的并合，甚至追溯宇宙的起源。

17.1 线性化的场方程

引力波的理论研究当然需要求解爱因斯坦场方程，但是如果引力场很强，那除了少数特殊情况之外，就没有一般性的解析求解方法。出路之一是，寻求数值解，或者说发展数值模拟，在实际应用中，这甚至是最为重要的出路。但是，作为一本讲述理论原理而不是侧重于应用的书，转去讲数值模拟肯定是不合适的。好在，还有另一条出路，就是去求解弱引力场近似下的爱因斯坦场方程。这条出路有两个优点：第一，它可以一般性地进行求解。第二，它足以揭示引力波的存在，揭示其产生和传播的机理。

所谓的弱引力场近似，指的就是，时空几乎是平坦的，只是在这个平坦的背景上存在一些微扰，所谓时空的涟漪。具体来说，在弱场近似下，度规场 $g_{\mu\nu}$ 可以分解成如下形式

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (17.1)$$

式中 $\eta_{\mu\nu} = \text{diag}(-1, +1, +1, +1)$ 为闵可夫斯基度规，而 $h_{\mu\nu}$ 为这个平坦时空背景上的微扰小量，满足 $|h_{\mu\nu}| \ll 1$ 。

这时候，我们把 $h_{\mu\nu}$ 看成是由度规 $\eta_{\mu\nu}$ 所描述的平坦时空背景上的场（称作线性化度规场），也就是说， $h_{\mu\nu}$ 是满足洛伦兹变换的平坦时空张量，而不是弯曲时空张量，因此我们用闵可夫斯基度规来对它进行指标的升降，比方说

$$h^{\mu\nu} \equiv \eta^{\mu\rho} \eta^{\nu\sigma} h_{\rho\sigma}. \quad (17.2)$$

同样，我们也用闵可夫斯基度规来进行指标的收缩，比方说 $h \equiv \eta^{\mu\nu} h_{\mu\nu}$ 。类似的处理也将同样应用于弱引力场近似下的其它量！

根据第四章的相关知识，在如下无穷小微分同胚变换之下

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x), \quad (17.3)$$

度规场将变换为

$$g_{\mu\nu}(x) \rightarrow g'_{\mu\nu}(x) = g_{\mu\nu}(x) - [\varepsilon^\rho \partial_\rho g_{\mu\nu} + g_{\mu\rho} \partial_\nu \varepsilon^\rho + g_{\rho\nu} \partial_\mu \varepsilon^\rho]. \quad (17.4)$$

应用弱场近似 $g'_{\mu\nu} = \eta_{\mu\nu} + h'_{\mu\nu}$ ，并保留到一阶小量（注意 ε^μ 和 $h_{\mu\nu}$ 都是小量），即可以得到

$$h_{\mu\nu}(x) \rightarrow h'_{\mu\nu}(x) = h_{\mu\nu}(x) - (\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu). \quad (17.5)$$

换言之，在无穷小微分同胚之下，线性化度规场在变换前后的改变量为 $\delta h_{\mu\nu} = -(\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu)$ 。这完全类似于电动力学中电磁矢量势 A_μ 在无穷小规范变换之下的改变量 $\delta A_\mu =$

$-\partial_\mu \varepsilon$, 所以, 在线性化度规场近似下, 无穷小微分同胚变换完全可以看成是一种类似于电动力学中的规范变换, 而线性化度规场 $h_{\mu\nu}$ 就是这一规范变换之下的规范场¹。

线性化场方程

在上面的弱引力场近似下, 爱因斯坦场方程可以化简 (保留到 $h_{\mu\nu}$ 的一阶), 这个化简以后的场方程称作线性化爱因斯坦场方程, 因为其中我们只保留到小量 $h_{\mu\nu}$ 的一阶, 也就是线性阶。

首先, $g^{\mu\nu}$ 的一阶近似为

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}. \quad (17.6)$$

相应的克里斯托夫联络系数的线性化近似为

$$\Gamma_{\mu\nu}^\sigma = \frac{1}{2} \eta^{\sigma\rho} (\partial_\mu h_{\rho\nu} + \partial_\nu h_{\rho\mu} - \partial_\rho h_{\mu\nu}). \quad (17.7)$$

注意, $\Gamma_{\mu\nu}^\sigma$ 本身为一阶小量。

黎曼曲率张量为

$$R_{\sigma\mu\nu}^\rho = \partial_\mu \Gamma_{\nu\sigma}^\rho - \partial_\nu \Gamma_{\mu\sigma}^\rho + \Gamma_{\mu\lambda}^\rho \Gamma_{\nu\sigma}^\lambda - \Gamma_{\nu\lambda}^\rho \Gamma_{\mu\sigma}^\lambda. \quad (17.8)$$

在线性化引力场的近似下, 联络系数为一阶小量, 从而式中的 $\Gamma\Gamma$ 项为二阶小量, 因此可以忽略。从而, 黎曼曲率张量的线性化近似为

$$\begin{aligned} R_{\sigma\mu\nu}^\rho &= \partial_\mu \Gamma_{\nu\sigma}^\rho - \partial_\nu \Gamma_{\mu\sigma}^\rho \\ &= \frac{1}{2} \eta^{\rho\lambda} (\partial_\mu \partial_\sigma h_{\nu\lambda} - \partial_\mu \partial_\lambda h_{\nu\sigma} - \partial_\nu \partial_\sigma h_{\mu\lambda} + \partial_\nu \partial_\lambda h_{\mu\sigma}). \end{aligned} \quad (17.9)$$

注意, 这同样是一阶小量。进而里奇张量为

$$R_{\mu\nu} = \frac{1}{2} (\partial^\rho \partial_\mu h_{\nu\rho} + \partial^\rho \partial_\nu h_{\mu\rho} - \partial^2 h_{\mu\nu} - \partial_\mu \partial_\nu h), \quad (17.10)$$

式中 $\partial^2 \equiv \partial^\mu \partial_\mu$, $h \equiv h^\mu_\mu$ 。进一步可以得到里奇标量的线性化近似为

$$R = \partial^\mu \partial^\nu h_{\mu\nu} - \partial^2 h. \quad (17.11)$$

由里奇张量和里奇标量的线性化近似, 即可以得到爱因斯坦张量的线性化近似, 为

$$G_{\mu\nu} = \frac{1}{2} [\partial^\rho \partial_\mu h_{\nu\rho} + \partial^\rho \partial_\nu h_{\mu\rho} - \partial^2 h_{\mu\nu} - \partial_\mu \partial_\nu h - (\partial^\rho \partial^\sigma h_{\rho\sigma} - \partial^2 h) \eta_{\mu\nu}]. \quad (17.12)$$

爱因斯坦张量满足比安奇恒等式 $D^\mu G_{\mu\nu}$, 在线性化近似下, 这个恒等式将变成

$$\partial^\mu G_{\mu\nu} = 0. \quad (17.13)$$

¹值得注意的是, 如果把规范对称性定义成理论中的某种冗余, 那么广义相对论中一般的微分同胚对称性也可以看成是一种广义的规范对称性, 因为在量子引力中, 真正的物理可观测量要求是微分同胚不变的, 也就是说, 广义相对论的微分同胚对称性也是理论描述的某种冗余。

不难验证, 上面给出的线性化爱因斯坦张量自动满足这个恒等式。

有了爱因斯坦张量的线性化近似, 我们就可以写出爱因斯坦场方程的线性化近似了, 为

$$\partial^\rho \partial_\mu h_{\nu\rho} + \partial^\rho \partial_\nu h_{\mu\rho} - \partial^2 h_{\mu\nu} - \partial_\mu \partial_\nu h - (\partial^\rho \partial^\sigma h_{\rho\sigma} - \partial^2 h) \eta_{\mu\nu} = 16\pi G T_{\mu\nu}.$$

当然, 为了自洽, 方程右边的 $T_{\mu\nu}$ 也得近似为小量。

Fierz-Pauli 作用量

我们知道, 爱因斯坦场方程可以从爱因斯坦-希尔伯特作用量中得到, 那么, 在弱引力场近似下, 这个作用量近似为什么呢? 这个问题其实有点复杂, 复杂之处在于, 为了得到场方程的一阶近似, 我们需要将作用量近似到二阶小量, 特别的, 我们需要将里奇标量近似到二阶小量, 如果直接算的话, 这会有点复杂。不过, 我们可以根据上面给出的爱因斯坦张量的线性近似, 以及作用量原理 $\delta S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x [-G^{\mu\nu} \delta h_{\mu\nu}]$ 凑出答案 (其中 $\kappa = 8\pi G$)。经过合适的分部积分并丢掉边界项以后, 可以得到结果为

$$S_{\text{FP}} = \frac{1}{2\kappa} \int d^4x \left[-\frac{1}{4} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + \frac{1}{2} \partial_\rho h_{\mu\nu} \partial^\nu h^{\rho\mu} + \frac{1}{4} \partial_\mu h \partial^\mu h - \frac{1}{2} \partial_\nu h^{\mu\nu} \partial_\mu h \right].$$

这个结果最早是 Fierz 和 Pauli 给出来的, 所以也称作 Fierz-Pauli 作用量。

另一种快速得到 Fierz-Pauli 作用量的办法是, 首先, 要求它是 $h_{\mu\nu}$ 的二阶小量, 同时每一项都包含两个偏导运算。在这两个要求下, 作用量的一般形式为

$$S_{\text{FP}} = \int d^4x \left[c_1 \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + c_2 \partial_\rho h_{\mu\nu} \partial^\nu h^{\rho\mu} + c_3 \partial_\mu h \partial^\mu h + c_4 \partial_\nu h^{\mu\nu} \partial_\mu h \right],$$

式中 c_1, c_2, c_3, c_4 为待定的常数系数。然后, 最重要的是, 我们要求这个作用量在规范变换 (17.5) 之下保持不变 (经过合适的分部积分并丢弃边界项)。计算出规范变换 (17.5) 之下上述作用量的改变量, 经过合适的分部积分之后, 结果是

$$\begin{aligned} \delta S_{\text{FP}} = & - \int d^4x \left[(4c_1 + 2c_2) \partial^2 \partial_\mu h^{\mu\nu} \varepsilon_\nu \right. \\ & \left. + (2c_2 + 2c_4) \partial_\mu \partial_\nu \partial^\rho h^{\mu\nu} \varepsilon_\rho + (4c_3 + 2c_4) \partial^2 \partial^\mu h \varepsilon_\mu \right]. \end{aligned} \quad (17.14)$$

要求这个改变量等于零, 由此就能确定

$$4c_1 + 2c_2 = 0, \quad 2c_2 + 2c_4 = 0, \quad 4c_3 + 2c_4 = 0.$$

这组方程的通解为

$$c_1 = -\frac{1}{4}a, \quad c_2 = \frac{1}{2}a, \quad c_3 = \frac{1}{4}a, \quad c_4 = -\frac{1}{2}a, \quad (17.15)$$

式中 a 为一个未定的常数因子。也就是说, 上面这些要求可以把 Fierz-Pauli 作用量确定到只差一个整体的常数因子 a 。

以上推导也给出另一种对引力场方程进行线性近似的办法, 就是通过规范不变性的要求先确定出 Fierz-Pauli 作用量 (差一个整体的常数因子 a), 之后再将这个作用量进行牛顿

引力近似, 与牛顿引力的作用量进行比较, 得到 $a = \frac{1}{16\pi G}$ 。然后再对最终的 Fierz-Pauli 作用量进行变分, 得到线性化近似以后的爱因斯坦场方程。这种推导办法的好处是, 不需要先知道完整的爱因斯坦场方程, 甚至都不需要联络和曲率张量等几何概念, 更不要说对这些几何量进行线性近似了。事实上, 它只需要假设在弱场近似下, 引力场由 $h_{\mu\nu}$ 描述, 然后假设理论在规范变换 (17.5) 之下保持不变。这是一种完全场论化的处理办法, 它不需要那些黎曼几何的概念。

规范固定

实际上, 前面得出的线性化场方程还可以简化。为此我们定义

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}h\eta_{\mu\nu}. \quad (17.16)$$

从而易有 $\bar{h} \equiv \bar{h}^\mu{}_\mu = -h$, 进而也有

$$h_{\mu\nu} \equiv \bar{h}_{\mu\nu} - \frac{1}{2}\bar{h}\eta_{\mu\nu}. \quad (17.17)$$

所以我们只要先解出 $\bar{h}_{\mu\nu}$, 那由这个式子就能得到原来的 $h_{\mu\nu}$. 将 (17.17) 式代入前面得到的线性化场方程, 不难有

$$\partial^\rho \partial_\mu \bar{h}_{\nu\rho} + \partial^\rho \partial_\nu \bar{h}_{\mu\rho} - \partial^2 \bar{h}_{\mu\nu} - \eta_{\mu\nu} \partial^\rho \partial^\sigma \bar{h}_{\rho\sigma} = 16\pi G T_{\mu\nu}. \quad (17.18)$$

另外一方面, 理论有 $h_{\mu\nu} \rightarrow h_{\mu\nu} - (\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu)$ 的规范对称性, 用变量 $\bar{h}_{\mu\nu}$ 来说, 这个规范对称性即是

$$\bar{h}_{\mu\nu} \rightarrow \bar{h}_{\mu\nu} - (\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu - \eta_{\mu\nu} \partial^\rho \varepsilon_\rho). \quad (17.19)$$

假设原来有 $\partial^\mu \bar{h}_{\mu\nu} = f_\nu$, 其中 f_ν 是某个任意的矢量场, 那么在上述规范变换之下, 它将变成 $\partial^\mu \bar{h}_{\mu\nu} - \partial^2 \varepsilon_\nu = f_\nu$. 因此只要适当地选取 ε_ν 为方程 $-\partial^2 \varepsilon_\nu = f_\nu$ 的解, 我们就可以让最终的 $\bar{h}_{\mu\nu}$ 额外满足如下方程

$$\partial^\mu \bar{h}_{\mu\nu} = 0. \quad (17.20)$$

注意, 额外要求 $\bar{h}_{\mu\nu}$ 满足这个方程以后, 我们就不再能进行前面的规范变换了! 所以, 这个方程是一种规范固定条件, 常常称作 de Donder 规范条件, 或者简称为 de Donder 规范。当然, 规范固定条件不是唯一的, de Donder 规范只是比较好用也比较常用而已, 就如同电动力学中的洛伦兹规范一样。

很显然, 在 de Donder 规范 (17.20) 下, 线性化爱因斯坦场方程 (17.18) 可以进一步简化为

$$-\partial^2 \bar{h}_{\mu\nu} = 16\pi G T_{\mu\nu}. \quad (17.21)$$

当然, 这时候为了和方程 (17.20) 自洽, 得有

$$\partial^\mu T_{\mu\nu} = 0. \quad (17.22)$$

实际上，这就是协变能动量守恒方程 $D^\mu T_{\mu\nu} = 0$ 的弱引力场近似结果。

其实，de Donder 规范在任意非线性引力场（而不仅仅是线性弱引力场）情况下是一个对应物的，就是如下规范条件

$$g^{\mu\nu}\Gamma_{\mu\nu}^\rho = 0. \quad (17.23)$$

注意，这不是一个张量方程，因为 $\Gamma_{\mu\nu}^\rho$ 不是一个张量。事实上，这个方程只在特定的局部坐标下才成立²，它就是用来挑出这些方便的局部坐标的。读者不难自己验证，这个非线性引力场的规范条件在线性化近似下给出的就是 de Donder 规范条件。

上述非线性规范条件 (17.23) 之所以有用，是因为它有一些良好的性质。比方说如果我们考虑弯曲时空中的 $D^\mu D_\mu$ 算符，则由于 $D^\mu D_\mu = g^{\mu\nu}(\partial_\mu \partial_\nu - \Gamma_{\nu\mu}^\rho \partial_\rho)$ ，很显然，如果有规范条件 (17.23)，那么即有 $D^\mu D_\mu = g^{\mu\nu} \partial_\mu \partial_\nu$ 。类似的简化也发生在计算协变散度 $D^\mu \omega_\mu$ 时，如果有规范条件 (17.23)，则 $D^\mu \omega_\mu = g^{\mu\nu}(\partial_\mu \omega_\nu - \Gamma_{\mu\nu}^\rho \omega_\rho) = \partial^\nu \omega_\nu$ 。

17.2 真空中的引力波

2015 年 9 月 14 日，人类首次直接捕捉到了来自宇宙深处的时空涟漪——引力波。这一天，位于美国路易斯安那州和华盛顿州的两台 LIGO（激光干涉引力波天文台）探测器，记录下了一段持续约 0.2 秒的独特信号。这个编号为 GW150914 的事件，不仅验证了爱因斯坦百年前的预言，更开启了人类探索宇宙的全新窗口。

GW150914 源自 13 亿光年外的一场宇宙“婚礼”：两颗质量分别为 35 倍和 30 倍太阳质量的黑洞，在相互缠绕数十亿年后以近乎光速碰撞合并，最终形成一个 62 倍太阳质量的旋转黑洞。这一过程中，相当于 3 倍太阳质量的能量（一个巨大的能量）以引力波形式辐射而出，跨越浩瀚星海抵达地球。LIGO 通过测量两条 4 公里长激光干涉臂的微小长度差（约 10^{-18} 米），捕捉到了这场宇宙事件留下的唯一证据。

2016 年 2 月 11 日，LIGO 团队公布这一发现时，科学界为之沸腾。引力波的探测不仅填补了广义相对论实验验证的最后一块拼图，更标志着“多信使天文学”时代的来临。人类从此不仅能“看”电磁波，还能“听”引力波，以前所未有的方式感知黑洞、中子星等暗弱天体的奥秘。截至 2023 年，LIGO 与欧洲 Virgo 探测器已累计探测到百余次引力波事件，包括双中子星合并引发的“光与引力波交响曲”。

这项发现被《科学》杂志誉为“世纪突破”，其核心团队荣获 2017 年诺贝尔物理学奖。正如 LIGO 创始人之一基普·索恩所言：“我们第一次听见了宇宙的呼吸。”从爱因斯坦的纸笔推演到实验室中的激光闪烁，人类用百年坚守证明：那些藏匿于时空褶皱中的宇宙私语，终将被倾听。

本节我们来研究在弱场近似下，引力波如何在真空中传播。

17.2.1 求解波动方程

所谓在真空中，指的就是没有物质源。因此这时线性化场方程 (17.21) 将变成，

$$\partial^2 \bar{h}_{\mu\nu} = 0, \quad \partial^\mu \bar{h}_{\mu\nu} = 0. \quad (17.24)$$

²与之相反，张量方程只要在一个坐标系下成立，则就在任意坐标系下都成立。

当然, 式中第二个方程就是 de Donder 规范条件。

由于以上方程为标准的波动方程, 它当然有如下标准的平面波解,

$$\bar{h}_{\mu\nu} = \text{Re}(H_{\mu\nu}e^{ik_\rho x^\rho}), \quad (17.25)$$

式中 $H_{\mu\nu}$ 是一个复的对称极化张量, k^μ 是一个实的四维波矢量。记住, 求解的最后一步是取实部, 不过, 我们也常常省去对这一步的显式书写, 而默认人们最后必然会这么做。

将上述平面波解的假设代入方程 (17.24), 立即有

$$k_\rho k^\rho = 0, \quad k^\mu H_{\mu\nu} = 0. \quad (17.26)$$

前一个方程告诉我们, 真空中引力波的波矢量是一个类光矢量, 因此**引力波以光速传播**。后一个方程告诉我们, 引力波的极化张量与其传播方向正交。

不过, de Donder 规范条件其实并没有固定所有的规范自由度。为了看出这一点, 假设我们在规范变换 (17.19) 中取

$$\partial^2 \varepsilon_\mu = 0. \quad (17.27)$$

则不难看出, 只要变换之前的 $h_{\mu\nu}$ 满足 de Donder 规范条件, 则变换之后依然满足 de Donder 规范条件。

很显然, (17.27) 式存在如下平面波解

$$\varepsilon_\mu = \text{Re}(\lambda_\mu e^{ik_\rho x^\rho}), \quad k_\rho k^\rho = 0. \quad (17.28)$$

代入规范变换 (17.19), 即有, 满足 de Donder 规范条件的平面引力波还可以进行如下变换

$$H_{\mu\nu} \rightarrow H_{\mu\nu} - i(k_\mu \lambda_\nu + k_\nu \lambda_\mu - \eta_{\mu\nu} k^\rho \lambda_\rho). \quad (17.29)$$

因此, 我们可以进一步选择合适的 λ_μ 使得

$$H_{0\mu} = 0, \quad H^\mu{}_\mu = 0. \quad (17.30)$$

这个条件, 联合前面 (17.26) 的第二式, 就称之为**横向无迹规范**, 即 Transverse Traceless gauge 或者 TT 规范。所谓无迹, 指的就是矩阵 H 是无迹的, 或者说, 这时候 $\bar{h} = 0 = -h$ 。因此, 在横向无迹规范中, 我们有 $h_{\mu\nu} = \bar{h}_{\mu\nu}$ 。

下面我们来数一下引力波物理自由度的数量。首先, 极化张量 $H_{\mu\nu}$ 有 $\frac{4(4+1)}{2} = 10$ 个独立分量, de Donder 规范条件 ((17.26) 的第二式) 对它加上了 4 个约束条件, 而且, 在满足这些约束条件的同时, 还可以进行 4 个剩余的规范变换 (由前面的 λ_μ 刻画), 所以, 真正物理的独立极化分量只有 $10 - 4 - 4 = 2$ 个。

取类光矢量 $k^\mu = (\omega, k^i)$ (类光条件为 $\omega^2 - k^i k_i = 0$), 则根据 (17.30) 式以及 (17.26) 的第二式可知, 在横向无迹规范中, 必有

$$H_{00} = 0, \quad H_{0i} = 0, \quad k^i H_{ij} = 0, \quad H^i{}_i = 0. \quad (17.31)$$

由于 3-矢量 k^i 的方向就是波的传播方向，所以这就告诉我们，平面引力波是横波，因为其极化张量与波的传播方向相垂直。

特别的，如果取 k^i 沿着 z 轴的方向，则满足上述条件的 H 必定有如下形式

$$H_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & H_+ & H_\times & 0 \\ 0 & H_\times & -H_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (17.32)$$

式中， H_+, H_\times 就是极化张量 $H_{\mu\nu}$ 的两个独立物理分量。稍微我们会进一步解释这两种极化模式的物理含义。

17.2.2 随波起伏

假设你是一个粒子，只受引力作用，因此你走的是弯曲时空中的测地线，那么，根据等效原理，当引力波从你身上经过时，你将什么都感受不到，你觉得你是完全自由的。不过，你并不是一个粒子，更好的模型是把你看成一簇粒子，每一个都在引力场中作测地线运动，当引力波经过你时，虽然你身上每一个点的引力都没有绝对意义，但是，潮汐力有绝对意义，作为一簇粒子，你可以感受到潮汐力，它可以由第七章所讲述的测地线偏离方程来刻画。

根据第七章，测地线偏离方程为

$$\frac{D^2 \eta^\mu}{D\tau^2} = R^\mu{}_{\rho\sigma\nu} u^\rho u^\sigma \eta^\nu, \quad (17.33)$$

式中 η^μ 为偏离矢量。 u^μ 为这些粒子的协变速度，即 $\partial_\tau = u^\mu \partial_\mu$ (τ 为粒子的固有时)， η^μ 与 u^μ 正交。类似于第七章中的一些做法，我们取沿着其中一个粒子测地线自由下落的局部惯性系 (从而 $t = \tau$)，则在这个局部惯性系中，潮汐加速度可以简化为

$$\frac{D^2 \eta^\mu}{D\tau^2} = \frac{d^2 \eta^\mu}{dt^2}. \quad (17.34)$$

在此局部惯性系中， $u^\mu = (1, 0, 0, 0)$ ，因此与之正交的偏离矢量 η^μ 必然满足 $\eta^0 = 0$ ，根据第七章的相关推导，这时候可以将测地线偏离方程简化为

$$\frac{d^2 \eta^i}{dt^2} = R_{i00j} \eta^j. \quad (17.35)$$

前面推导过，线性化的黎曼曲率张量为

$$R^\rho{}_{\sigma\mu\nu} = \frac{1}{2} \eta^{\rho\lambda} (\partial_\mu \partial_\sigma h_{\nu\lambda} - \partial_\mu \partial_\lambda h_{\nu\sigma} - \partial_\nu \partial_\sigma h_{\mu\lambda} + \partial_\nu \partial_\lambda h_{\mu\sigma}). \quad (17.36)$$

注意到在 TT 规范中 $h_{\mu\nu} = \bar{h}_{\mu\nu}$ ，且 $h_{0\mu} = 0$ ，从而即有

$$R_{i00j} = \frac{1}{2} \partial_t^2 h_{ij}. \quad (17.37)$$

代入 $h_{ij} = \bar{h}_{ij} = \text{Re}(H_{ij}e^{-i\omega t + ik_j x^j})$, 并将注意力集中在基准测地线附近 (在上述局部惯性系中即取 $x^j = 0$), 即有

$$R_{i00j} = -\omega^2 \frac{1}{2} \text{Re}(H_{ij}e^{-i\omega t}). \quad (17.38)$$

为了进一步简化讨论, 我们假设引力波沿着局部惯性系的 z 轴传播, 则在将 (17.38) 式代入测地线偏离方程 (17.35) 以后, 即有:

1. H_+ 极化: 对于 $H_\times = 0$ 的 H_+ 极化模式, 有

$$\frac{d^2\eta^1}{dt^2} = -\frac{\omega^2}{2} \text{Re}(H_+e^{-i\omega t})\eta^1, \quad \frac{d^2\eta^2}{dt^2} = \frac{\omega^2}{2} \text{Re}(H_+e^{-i\omega t})\eta^2. \quad (17.39)$$

设 $H_+ = |H_+|e^{i\phi_+}$, 并注意到 $|H_+|$ 为一个微扰小量, 从而可以用微扰法求解出上述方程的近似解, 为

$$\begin{aligned} \eta^1(t) &= \eta^1(0) \left(1 + \frac{1}{2}|H_+| \cos(\omega t - \phi_+) + \dots\right) \\ \eta^2(t) &= \eta^2(0) \left(1 - \frac{1}{2}|H_+| \cos(\omega t - \phi_+) + \dots\right). \end{aligned} \quad (17.40)$$

在所考察的这一簇测地线中取定一条基准测地线, 假设它位于 $x^i = 0$ 的坐标原点, 取偏离矢量 η^i 为它到临近测地线的距离。由于引力波沿着 z 轴传播, 所以我们只需要考察 $x-y$ 平面里的情况。假设在原点周围取一圈测地线, 并让它在初始时围绕原点形成一个圆周。则上述解 (17.40) 告诉我们, 在引力波的影响下, 如果这圆周上的测地线沿着 x 轴 (η^1 方向) 向外移动, 则它们必定同时沿着 y 轴 (η^2) 向内移动, 反过来也一样。所以, 随着时间的推移, 这一圈粒子必定从圆周演化成椭圆, 然后再演化回来, 如图 (17.1) 所示。

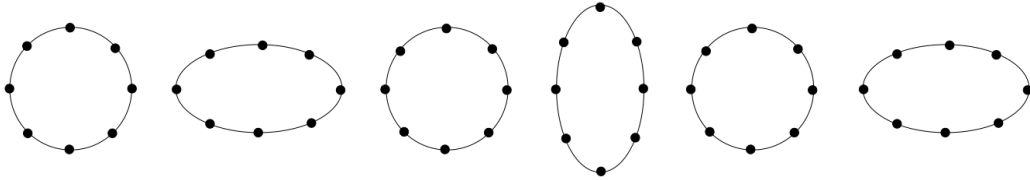


Figure 17.1: H_+ 极化的引力波影响下 $x-y$ 平面上的测地偏离演化情况。

2. H_\times 极化: 类似的, 对于 $H_+ = 0$ 的 H_\times 极化模式, 有

$$\frac{d^2\eta^1}{dt^2} = -\frac{\omega^2}{2} \text{Re}(H_\times e^{-i\omega t})\eta^2, \quad \frac{d^2\eta^2}{dt^2} = -\frac{\omega^2}{2} \text{Re}(H_\times e^{-i\omega t})\eta^1. \quad (17.41)$$

设 $H_\times = |H_\times|e^{i\phi_\times}$, 并注意到 $|H_\times|$ 为一个微扰小量, 从而可以用微扰法求解出上述方程的近似解, 为

$$\begin{aligned} \eta^1(t) &= \eta^1(0) + \eta^2(0) \frac{1}{2}|H_\times| \cos(\omega t - \phi_\times) + \dots \\ \eta^2(t) &= \eta^2(0) + \eta^1(0) \frac{1}{2}|H_\times| \cos(\omega t - \phi_\times) + \dots \end{aligned} \quad (17.42)$$

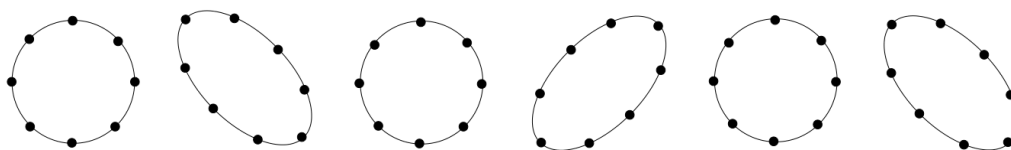


Figure 17.2: H_x 极化的引力波影响下 $x-y$ 平面上的测地偏离演化情况。

换言之， $\eta^1 \pm \eta^2$ 的演化情况才与 H_+ 极化的情形完全类似。这就相当于在 $x-y$ 平面内转过了 45° 。从而， H_x 极化的引力波影响下 $x-y$ 平面上的测地偏离演化情况如图 (17.2) 所示。

当然，我们也可以考察以上两种极化模式的线性叠加。值得注意的是，引力波影响下的偏离矢量在角度 π 的旋转之下是不变的！这与电磁波极化矢量不同，后者只在 2π 的旋转下才不变。这一事实反映的是，引力子的自旋是 2，而光子的自旋是 1。

引力波探测器

引力波探测器是干涉仪。它们通过让光线在两条互相垂直的臂之间来回反射进行探测，臂两端的镜子充当测试质量块。当引力波垂直于探测器平面传播时，会导致其中一条臂缩短而另一条臂伸长。若将探测器两臂分别沿 x 轴和 y 轴对齐，则可通过公式 (17.40) 读出臂长的最大变化量，

$$L' = L\left(1 \pm \frac{1}{2}H_+\right) \Rightarrow \frac{\delta L}{L} = H_+. \quad (17.43)$$

为了估算这一变化的量级，我们需要了解典型天体物理源可能产生的振幅 H_+ 的大小。我们将在后面的小节中详细讨论这一点。结果显示其数值极其微小：典型源的 H_+ 约为 10^{-21} 。LIGO 探测器的每条臂长约 $L \approx 3$ 公里，这意味着需要检测的臂长变化量仅为 $\delta L \approx 10^{-18}$ 米。这个数字看似小得离谱——它比质子的半径还小，并且比干涉仪所用光的波长小约 10^{12} 倍。尽管如此，探测器的灵敏度足以胜任这一任务。2015 年，LIGO 天文台首次成功探测到引力波，该合作组织的三位成员因此获得 2017 年诺贝尔奖。此后，LIGO 和 VIRGO 探测器已观测到大量涉及黑洞与中子星合并的事件。

17.2.3 精确解

我们已经找到了线性爱因斯坦方程的波动解。对于一个沿着 z 轴正方向传播的平面引力波，这种解的度规可以写成

$$ds^2 = -dt^2 + dz^2 + (\delta_{ab} + h_{ab}(t-z))dx^a dx^b, \quad (17.44)$$

式中 $a, b = 1, 2$ 指示 $x-y$ 平面内的坐标分量。由于所考察的是线性爱因斯坦方程，所以任意 $h_{ab}(t-z)$ 都是解，前文考察的只是这种解的傅里叶分解。

如果我们引入如下光锥坐标

$$u = t - z, \quad v = t + z. \quad (17.45)$$

则也可以把上述线性爱因斯坦方程的解写成

$$ds^2 = -dudv + dx^a dx^a + h_{ab}(u) dx^a dx^b. \quad (17.46)$$

问题是，对于完整的非线性爱因斯坦场方程，我们能否找到类似的波动解呢？

回答是，可以！具体来说，下述 Brinkmann 度规就是一个这样的精确解，描述沿着 z 轴正方向传播的平面引力波，

$$ds^2 = -dudv + dx^a dx^a + H_{ab}(u) x^a x^b du^2. \quad (17.47)$$

人们可以通过直接计算证明，只要 $H_{ab}(u)$ 无迹，则相应的 Brinkmann 度规就是里奇平坦的，从而是真空爱因斯坦场方程的精确解，

$$R_{\mu\nu} = 0 \Leftrightarrow H^a_a(u) = 0. \quad (17.48)$$

换言之，这类精确解也有两个独立的极化模式，为

$$H_{ab}(u) = \begin{bmatrix} H_{11}(u) & H_{12}(u) \\ H_{12}(u) & -H_{11}(u) \end{bmatrix}. \quad (17.49)$$

一般来说，人们很难找到一个非线性偏微分方程的一簇依赖于任意函数（比如这里的 $H_{11}(u)$ 和 $H_{12}(u)$ ）的精确解，但是，Brinkmann 度规例外。

17.3 引力波的产生

前面的小节中讨论了引力波在真空中的传播，然而引力波是如何产生的呢？我们知道加速运动的电荷会辐射出电磁波，类似的，加速运动的质点会辐射出引力波。

想象一个静止的球放在橡皮膜上，橡皮膜会凹陷，但不会产生波动。如果球在橡皮膜上匀速运动，凹陷的形状会移动，但仍然不会产生波动。但，如果球在橡皮膜上加速运动，凹陷的形状就会不断变化，这种变化会以波动的形式向外传播。类似的，物质的存在会引起时空的弯曲（凹陷），当物质加速运动时，时空的弯曲会发生变化，这种变化以引力波的形式向外传播。

更具体地说，由于线性化爱因斯坦方程 (17.21) 和电磁矢量势满足的方程类似，都是波动方程，方程的右边就是源，电流矢量是电磁波的源，而能动张量就是引力波的源，所以，引力波和电磁波有很大的类似之处。但是，由于电磁矢量势是一个矢量，相应的源也是矢量，而引力波的 $\bar{h}_{\mu\nu}$ 是二阶张量，相应的源也是二阶张量，所以引力波的产生和电磁波的产生也有不同之处，这反映在，偶极辐射就能产生电磁波，但是引力波并不来源于偶极辐射，要产生引力波，至少需要四极辐射，稍后我们会清楚地看到这一点。

17.3.1 求解线性化爱因斯坦场方程

我们的出发点是线性化爱因斯坦方程 (17.21), 为方便起见重写如下

$$-\partial^2 \bar{h}_{\mu\nu} = 16\pi G T_{\mu\nu}. \quad (17.50)$$

我们的基本目标就是求解这个方程。

以上方程可以用标准的格林函数法来求解, 具体求解过程与我的《经典场论新讲》第七章对电磁辐射的求解完全类似, 这里只概述一下结果。首先, 我们定义格林函数 $G(x)$ (其中 $x = (t, \mathbf{x})$ 代表时空坐标 x^μ), 它满足如下格林方程

$$-\partial^2 G(x) = \delta^4(x). \quad (17.51)$$

利用这个格林函数, 即可以把 (17.50) 式的解写成

$$\bar{h}_{\mu\nu}(x) = 16\pi G \int d^4x' G(x-x') T_{\mu\nu}(x'). \quad (17.52)$$

不难验证, 这个解的确满足 de Donder 规范条件, 验证如下

$$\begin{aligned} \partial^\mu \bar{h}_{\mu\nu}(x) &= 16\pi G \int d^4x' \partial^\mu G(x-x') T_{\mu\nu}(x') \\ &= -16\pi G \int d^4x' (\partial'^\mu G(x-x')) T_{\mu\nu}(x') \\ &= 16\pi G \int d^4x' G(x-x') \partial'^\mu T_{\mu\nu}(x') = 0, \end{aligned} \quad (17.53)$$

其中第三行我们进行了分部积分, 最后的等号是利用能动量守恒方程。值得注意的是, 虽然这个解满足 de Donder 规范条件, 但并不意味着它也满足横向无迹规范, 通常来说, 它不满足。

由于因果性的考虑, 我们应该取 $G(x)$ 为所谓的延迟格林函数, 记作 $G_{\text{ret}}(x)$, 根据《经典场论新讲》第七章的推导,

$$G_{\text{ret}}(x) = \frac{1}{2\pi} \delta(x^2) \theta(t), \quad (17.54)$$

式中 $x^2 = x_\mu x^\mu$, $\theta(t)$ 为阶跃函数。这个结果在正洛伦兹变换 $SO^+(1,3)$ 下明显保持不变。

取时空坐标 $x = (t, \mathbf{x})$, 其中三维矢量 \mathbf{x} 为空间坐标, 并取 $r = |\mathbf{x}|$, 从而 $\delta(x^2) = \delta(t^2 - r^2) = \frac{1}{2r} [\delta(t-r) - \delta(t+r)]$. 又注意到, 当 $t > 0$ 时, $\delta(t+r)$ 自动等于零 (因为 r 总大于零), 而当 $t < 0$ 时 $\delta(t-r)$ 自动等于零, 从而即可以将上面的 (17.54) 式写成

$$G_{\text{ret}}(x) = \frac{1}{2\pi} \delta(x^2) \theta(t) = \frac{1}{4\pi r} \delta(t-r). \quad (17.55)$$

将以上结果代入 (17.52) 式, 即有

$$\begin{aligned} \bar{h}_{\mu\nu}(x) &= 16\pi G \int d^4x' G_{\text{ret}}(x-x') T_{\mu\nu}(x') \\ &= 4G \int d^4x' \frac{\delta(t-t' - |\mathbf{x} - \mathbf{x}'|)}{|\mathbf{x} - \mathbf{x}'|} T_{\mu\nu}(x') \\ &= 4G \int d^3\mathbf{x}' \frac{1}{|\mathbf{x} - \mathbf{x}'|} T_{\mu\nu}(\mathbf{x}', t_{\text{ret}}). \end{aligned} \quad (17.56)$$

式中 $t_{\text{ret}} \equiv t - |\mathbf{x} - \mathbf{x}'|$, 也称作延迟时刻。这个结果的含义是, 空间 \mathbf{x} 点 t 时刻的引力场由 \mathbf{x}' 点的源在更早的延迟时刻 t_{ret} 决定, 这是因为引力波的传播是需要时间的, 它以有限大小的光速 (取为 1) 传播。当然, 这个结果与我们在电动力学里熟悉的推迟势公式基本是一样的。

如果物质源只分布在一个有限的空间区域内, 设这个区域的尺寸为 d , 并且假设我们感兴趣的是源在距离很远的地方产生的引力场, 即假设 $|\mathbf{x} - \mathbf{x}'| \gg d$, 则就可以作如下近似 (注意 $r = |\mathbf{x}|$)

$$|\mathbf{x} - \mathbf{x}'| = r - \frac{\mathbf{x} \cdot \mathbf{x}'}{r} + \dots \Rightarrow \frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{r} + \frac{\mathbf{x} \cdot \mathbf{x}'}{r^3} + \dots \quad (17.57)$$

注意延迟时刻 t_{ret} 中包含了一个 $|\mathbf{x} - \mathbf{x}'|$, 从而

$$T_{\mu\nu}(\mathbf{x}', t_{\text{ret}}) = T_{\mu\nu}(\mathbf{x}', t - r + \mathbf{x} \cdot \mathbf{x}'/r + \dots). \quad (17.58)$$

为了进一步近似展开上述结果, 我们需要假设在引力波从源点 \mathbf{x}' 传播到场点 \mathbf{x} 的时间之内, 作为源的能动张量改变很小, 也即是说, 我们要假设物质源的运动是非相对论的。有了这个假设以后, 就可以进一步将上述能动张量近似展开成

$$T_{\mu\nu}(\mathbf{x}', t_{\text{ret}}) = T_{\mu\nu}(\mathbf{x}', t - r) + \dot{T}_{\mu\nu}(\mathbf{x}', t - r) \frac{\mathbf{x} \cdot \mathbf{x}'}{r} + \dots \quad (17.59)$$

物质源作非相对论运动的假设意味着 $\dot{T}_{\mu\nu}$ 与相应 $T_{\mu\nu}$ 的比是一个小量, 简称 $\dot{T}_{\mu\nu}$ 是一个小量。

有了 (17.57) 式和 (17.59) 式这两个近似以后, 即可以对结果 (17.56) 进行近似展开了。在小量展开的领头阶 (小量包含 d/r 以及 $\dot{T}_{\mu\nu}$), 我们有

$$\bar{h}_{\mu\nu}(\mathbf{x}, t) = \frac{4G}{r} \int d^3\mathbf{x}' T_{\mu\nu}(\mathbf{x}', t - r) + \dots \quad (17.60)$$

特别的,

$$\bar{h}_{00}(\mathbf{x}, t) = \frac{4G}{r} E + \dots, \quad \text{其中 } E = \int d^3\mathbf{x}' T_{00}(\mathbf{x}', t - r) \quad (17.61)$$

根据能动量守恒, E 为源的守恒能量, 所以这意味着 $\bar{h}_{00}(\mathbf{x}, t)$ 在小量展开的领头阶是不依赖于时间的, 实际上这个结果就等价于牛顿势 $\Phi = -GE/r$ 。类似的, 也有

$$\bar{h}_{0i}(\mathbf{x}, t) = -\frac{4G}{r} P_i + \dots, \quad \text{其中 } P_i = - \int d^3\mathbf{x}' T_{0i}(\mathbf{x}', t - r) \quad (17.62)$$

当然, P_i 就是物质源的守恒动量, 因此 $\bar{h}_{0i}(\mathbf{x}, t)$ 在领头阶也不随时间变化。通常会选取一个合适的惯性系, 以使得 $P_i = 0$, 从而让 $\bar{h}_{0i}(\mathbf{x}, t) = 0$ 。

正如我们已经看到的, 在小量展开的领头阶, $\bar{h}_{00}(\mathbf{x}, t), \bar{h}_{0i}(\mathbf{x}, t)$ 均是静态的, 不随时间演化, 因此对于引力波的考察并没有多少意思。有意思的是 $\bar{h}_{ij}(\mathbf{x}, t)$, 根据 (17.60) 式, 它应该为

$$\bar{h}_{ij}(\mathbf{x}, t) = \frac{4G}{r} \int d^3\mathbf{x}' T_{ij}(\mathbf{x}', t - r) + \dots \quad (17.63)$$

注意现在等式右边的积分并不是一个守恒量，因此一般来说是依赖于时间的。

定理：

$$\int d^3\mathbf{x}T_{ij}(\mathbf{x},t) = \frac{1}{2}\ddot{I}_{ij}(t), \quad (17.64)$$

式中 $I_{ij}(t)$ 是能量的四极矩

$$I_{ij}(t) = \int d^3\mathbf{x}T^{00}(\mathbf{x},t)x_ix_j. \quad (17.65)$$

证明：注意到

$$T^{ij} = \partial_k(T^{ik}x^j) - (\partial_k T^{ik})x^j = \partial_k(T^{ik}x^j) + (\partial_0 T^{0i})x^j, \quad (17.66)$$

式中第二个等于是利用了能动量守恒方程 $\partial_\mu T^{\mu\nu} = 0$ 。对于这个结果中的 T^{0i} 项，我们再次利用同样的技巧。具体来说对于对称化的指标 (ij) ，我们有

$$\begin{aligned} T^{0(i}x^j)} &= \frac{1}{2}\partial_k(T^{0k}x^i x^j) - \frac{1}{2}(\partial_k T^{0k})x^i x^j \\ &= \frac{1}{2}\partial_k(T^{0k}x^i x^j) + \frac{1}{2}(\partial_0 T^{00})x^i x^j. \end{aligned} \quad (17.67)$$

将 (17.66) 式对三维空间积分，忽略全微分的边界项，然后再代入 (17.67) 式，即有

$$\int d^3\mathbf{x}T^{ij}(\mathbf{x},t) = \frac{1}{2}\partial_0^2 \int d^3\mathbf{x}T^{00}(\mathbf{x},t)x^i x^j, \quad (17.68)$$

从而定理得证。

利用上述定理的结论，就可以把结果 (17.63) 重写成，

$$\bar{h}_{ij}(\mathbf{x},t) = \frac{2G}{r}\ddot{I}_{ij}(t-r) + \dots \quad (17.69)$$

这个结果告诉我们，只需要晃动能量四极矩不为零的物质源，就可以在远处产生引力波。特别的，这意味着，在领头阶，引力辐射是一种四极辐射，这与电磁辐射不同，后者在领头阶是偶极辐射。

也即是说，不存在偶极的引力辐射。这可以从下面的说明中看得更清楚一些。为此我们定义能量的偶极矩

$$E\mathbf{X} = \int d^3\mathbf{x}T^{00}(\mathbf{x},t)\mathbf{x}. \quad (17.70)$$

注意到，由于能量守恒 $\dot{E} = 0$ ，所以

$$E\dot{X}^i = \int d^3\mathbf{x}(\partial_0 T^{00})x^i = - \int d^3\mathbf{x}(\partial_j T^{j0})x^i = \int d^3\mathbf{x}T^{i0} = P^i, \quad (17.71)$$

其中倒数第二个等于是进行了分部积分。但是，动量守恒意味着 $\dot{\mathbf{P}} = 0$ ，从而

$$E\ddot{\mathbf{X}} = \dot{\mathbf{P}} = 0. \quad (17.72)$$

这就告诉我们偶极辐射为零。

现在, 我们可以利用结果 (17.69) 进一步推导出 \bar{h}_{00} 和 \bar{h}_{0i} 的次领头阶贡献。为此我们注意到根据 de Donder 规范条件 $\partial^\mu \bar{h}_{\mu\nu} = 0$, 有

$$\partial_0 \bar{h}_{0i} = \partial_j \bar{h}_{ji}, \quad \partial_0 \bar{h}_{00} = \partial_i \bar{h}_{i0}. \quad (17.73)$$

这两个方程中的第一个给出

$$\partial_0 \bar{h}_{0i} = \partial_j \left(\frac{2G}{r} \ddot{I}_{ij}(t-r) \right) = -\frac{2G\hat{x}_j}{r^2} \ddot{I}_{ij}(t-r) - \frac{2G\hat{x}_j}{r} \ddot{I}_{ij}(t-r), \quad (17.74)$$

式中我们利用了 $\partial_j r = x_j/r \equiv \hat{x}_j$ 。

(17.74) 式的两项中, 哪一项的贡献更大呢? 由于我们考虑的是远场, 看起来似乎是 $1/r$ 项的贡献远比 $1/r^2$ 项的贡献为大。但是, 这里有一个细节需要小心, 即 $1/r$ 项比 $1/r^2$ 项多一阶时间导数, 对于晃动频率为 ω 的源, 多一阶时间导数相当于多一个 $\omega \sim 1/\lambda$ 因子, 这里 λ 为引力辐射的波长。因此这也即是说, 仅在 $(1/\lambda)(1/r) \gg 1/r^2$ 时, $1/r$ 项的贡献才远超过 $1/r^2$ 项的贡献。换言之, 仅在如下条件满足时

$$r \gg \lambda, \quad (17.75)$$

$1/r$ 项的贡献才远超过 $1/r^2$ 项的贡献。这个条件就是所谓的远场区条件, 或者说辐射区条件。

因此, 由 (17.74) 式可知, 在辐射区, 我们有

$$\bar{h}_{0i} = -\frac{2G\hat{x}_j}{r} \ddot{I}_{ij}(t-r) + \dots, \quad (17.76)$$

这里我们已经取了 $P^i = 0$ 的惯性参考系。

重复上述技巧, 即可以进一步决定 \bar{h}_{00} , 结果是, 在辐射区, 它可以近似由下式给出

$$\bar{h}_{00} = \frac{4GE}{r} + \frac{2G\hat{x}_i\hat{x}_j}{r} \ddot{I}_{ij}(t-r) + \dots \quad (17.77)$$

当然, 这些结果也可以通过直接将 (17.56) 展开到次领头阶来得到, 只不过用 de Donder 规范条件来推导这些结果更加快速。

17.3.2 示例: 双星系统

作为示例, 我们考察一个双星系统。假设两颗星的质量均为 M , 两者间距为 R , 在 $x-y$ 平面内旋转。用牛顿引力很容易算出旋转频率为

$$\omega^2 = \frac{2GM}{R^3}. \quad (17.78)$$

假设把两颗星均当成质点处理, 每一颗星的质量密度当然都是 δ 函数的乘积, 因此, 在非相对论近似下, 双星系统的能量密度为

$$T^{00} = M\delta(z) \left[\delta\left(x - \frac{R}{2} \cos \omega t\right) \delta\left(y - \frac{R}{2} \sin \omega t\right) + \delta\left(x + \frac{R}{2} \cos \omega t\right) \delta\left(y + \frac{R}{2} \sin \omega t\right) \right].$$

因此不难算出，能量四极矩 $I_{ij}(t)$ 为

$$\begin{aligned} I_{ij}(t) &= \frac{MR^2}{2} \begin{bmatrix} \cos^2 \omega t & \cos \omega t \sin \omega t & 0 \\ \cos \omega t \sin \omega t & \sin^2 \omega t & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \frac{MR^2}{4} \begin{bmatrix} 1 + \cos 2\omega t & \sin 2\omega t & 0 \\ \sin 2\omega t & 1 - \cos 2\omega t & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned} \quad (17.79)$$

代入 (17.69) 式，即可以算出相应的引力波扰动

$$\bar{h}_{ij} = -\frac{2GMR^2\omega^2}{r} \begin{bmatrix} \cos 2\omega t_r & \sin 2\omega t_r & 0 \\ \sin 2\omega t_r & -\cos 2\omega t_r & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (17.80)$$

式中 $t_r \equiv t - r$.

可以看出，双星系统辐射出来的引力波基本上沿着径向往外传播。但是，如果我们将注意力集中在 z 轴方向，则上面的结果告诉我们，和自由空间的平面引力波一样，这时候相应的引力波为横波，其极化模式为 H_+ 模式和 H_\times 模式的某种叠加，实际上，这是一种圆极化模式。

为了估算辐射出来的引力波到底有多强，我们代入 (17.78) 式给出的 ω^2 ，即有

$$|h_{ij}| \sim \frac{G^2 M^2}{Rr}. \quad (17.81)$$

很显然，双星的质量越大，离得越近，则辐射出来的引力波就越强。由于双星不可能是真正的质点，它们的尺寸最小都是黑洞大小，因此两者最近的距离当然就是相距一个施瓦西半径 $R \sim R_g = 2GM$ ，这时候

$$|h_{ij}| \sim \frac{GM}{r}. \quad (17.82)$$

一个质量约为几个太阳那么大的黑洞半径约为 $R_g \sim 10\text{km}$ ，假如两个这样的黑洞形成双星系统，假如这系统距离我们数百万光年，即 $r \sim 10^{18}\text{km}$ ，则我们将有 $|h_{ij}| \sim 10^{-17}$ 。而假如这系统距离我们为数十亿光年，则有 $|h_{ij}| \sim 10^{-20}$ 。当然它们都非常非常小，但是，正如前面的相应小节中提到过的，当前的引力波探测器已经有足够的灵敏度，足以探测到这么微小的引力波扰动。

17.4 诺特技巧推导引力波能动张量

考察一个正在发出引力辐射的物质源，我们想问，它的辐射功率为多少？为此，我们需要取一个很大的两维闭合曲面包围这个物质源（这个两维曲面在源区的外面，且远离源区），然后计算单位时间经过这个闭合曲面流出去的引力能为多少。但是，为了进行这种计算，我们需要知道线性化引力场的能流密度公式，或者更一般地，需要知道线性化引力场的能动张量 $t_{\mu\nu}$ 。本节将表明，如何可以用诺特技巧导出它。

根据前文, 我们把 $h_{\mu\nu}$ 看作平坦时空中的一个二阶张量场 (因此, 现在 $h_{\mu\nu}$ 不必是小量), 它具有如下 Fierz-Pauli 作用量

$$S_{\text{FP}} = \frac{1}{16\pi G} \int d^4x \left[-\frac{1}{4} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + \frac{1}{2} \partial_\rho h_{\mu\nu} \partial^\nu h^{\rho\mu} + \frac{1}{4} \partial_\mu h \partial^\mu h - \frac{1}{2} \partial_\nu h^{\mu\nu} \partial_\mu h \right].$$

很明显, 这个作用量具有时空平移不变性, 因此原则上当然可以从它出发通过标准的诺特定理得到一个正则能动张量。问题是, 这样得到的正则能动张量往往是不对称的。

好在本书第二章 (2.103) 式和 (2.104) 式的下面我们描述了一种通过诺特技巧得到对称能动张量的办法。为此和第二章中的处理一样, 我们取如下无穷小局域洛伦兹变换

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x), \quad \varepsilon^\mu(x) = \varepsilon^\mu_\nu(x) x^\nu. \quad (17.83)$$

式中 $\varepsilon_{\mu\nu}(x)$ 是无穷小的反对称张量。

作为二阶张量场, $h_{\mu\nu}(x)$ 必定满足如下洛伦兹变换规则

$$h'_{\mu\nu}(x') \Lambda^\mu_\rho(x) \Lambda^\nu_\sigma(x) = h_{\rho\sigma}(x) \Rightarrow h'_{\mu\nu}(x') (\delta^\mu_\rho + \varepsilon^\mu_\rho(x)) (\delta^\nu_\sigma + \varepsilon^\nu_\sigma(x)) = h_{\rho\sigma}(x).$$

注意到 $\varepsilon^\mu_\nu(x)$ 为无穷小量, 不难得到

$$h'_{\mu\nu}(x') = h_{\mu\nu}(x) - (h_{\rho\nu} \varepsilon^\rho_\mu + h_{\mu\rho} \varepsilon^\rho_\nu). \quad (17.84)$$

换言之, 在无穷小局域洛伦兹变换 (172) 之下, $h_{\mu\nu}$ 在变换前后的改变量为 (注意, $h_{\mu\nu}$ 不必是小量)

$$\bar{\delta} h_{\mu\nu} \equiv h'_{\mu\nu}(x') - h_{\mu\nu}(x) = -(h_{\rho\nu} \varepsilon^\rho_\mu(x) + h_{\mu\rho} \varepsilon^\rho_\nu(x)). \quad (17.85)$$

式中我们用加横杠的 $\bar{\delta}$ 以区别于前文出现过的规范变换 $\delta h_{\mu\nu} = -(\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu)$ 。

我们把 Fierz-Pauli 作用量看作基本场变量 $h_{\mu\nu}$ 的泛函, 而诸如 $h^{\mu\nu}$ 这样的场变量则看成是用 $\eta^{\mu\nu}$ 上升指标的结果, 其余上指标也作类似理解。这样一来, Fierz-Pauli 作用量就具有如下形式

$$S_{\text{FP}}[h_{\mu\nu}(x)] = \int_{\mathcal{D}} d^4x \mathcal{L}(h_{\mu\nu}(x), \partial_\rho h_{\mu\nu}(x)), \quad (17.86)$$

式中为了明晰起见, 假定只考察一个特定的时空区域 \mathcal{D} , 也即是说, 作用量的时空积分是在 \mathcal{D} 上积的。

记无穷小局域时空变换 (172) 把时空局域 \mathcal{D} 变到了 \mathcal{D}' , 即

$$\mathcal{D} \rightarrow \mathcal{D}', \quad (17.87)$$

当然, \mathcal{D}' 与 \mathcal{D} 的差别是无穷小的。则, 无穷小局域时空变换 (172) 之后的 Fierz-Pauli 作用量为

$$\begin{aligned} S_{\text{FP}}[h'_{\mu\nu}(x)] &= \int_{\mathcal{D}'} d^4x \mathcal{L}(h'_{\mu\nu}(x), \partial_\rho h'_{\mu\nu}(x)) \\ &= \int_{\mathcal{D}'} d^4x' \mathcal{L}(h'_{\mu\nu}(x'), \partial'_\rho h'_{\mu\nu}(x')), \end{aligned} \quad (17.88)$$

注意，现在这个作用量积分是在变换之后的区域 \mathcal{D}' 上积的。

可以对上面的 (175) 式继续进行如下推导

$$\begin{aligned} S_{\text{FP}}[h'_{\mu\nu}(x)] &= \int_{\mathcal{D}'} d^4x' \mathcal{L}(h'_{\mu\nu}(x'), \partial'_\rho h'_{\mu\nu}(x')), \\ &= \int_{\mathcal{D}} d^4x \det\left(\frac{\partial x'}{\partial x}\right) \mathcal{L}(h + \bar{\delta}h, \frac{\partial x^\sigma}{\partial x'^\rho} \partial_\sigma(h + \bar{\delta}h)), \end{aligned} \quad (17.89)$$

式中 h 是 $h_{\mu\nu}$ 的简记符号 (下文约定相同, 不要和 h^μ_μ 搞混了)。注意到 $\det(\frac{\partial x'}{\partial x}) = 1 + \partial_\mu \epsilon^\mu$, 以及 $\frac{\partial x^\sigma}{\partial x'^\rho} = \delta^\sigma_\rho - \partial_\rho \epsilon^\sigma$, 代入 (176) 式, 即有

$$\begin{aligned} S_{\text{FP}}[h'_{\mu\nu}(x)] &= \int_{\mathcal{D}} d^4x (1 + \partial_\alpha \epsilon^\alpha) \mathcal{L}(h + \bar{\delta}h, \partial_\rho(h + \bar{\delta}h) - \partial_\rho \epsilon^\sigma \partial_\sigma(h + \bar{\delta}h)) \\ &= \int_{\mathcal{D}} d^4x (1 + \partial_\alpha \epsilon^\alpha) \mathcal{L}(h + \bar{\delta}h, \partial_\rho(h + \bar{\delta}h) - \partial_\rho \epsilon^\sigma \partial_\sigma h) \end{aligned} \quad (17.90)$$

式中我们注意到 $\partial_\rho \epsilon^\sigma \partial_\sigma(\bar{\delta}h)$ 为二阶无穷小量, 从而可以略去。

进一步将 (177) 式展开到一阶小量, 即有

$$\begin{aligned} S_{\text{FP}}[h'_{\mu\nu}(x)] &= \int_{\mathcal{D}} d^4x \mathcal{L}(h, \partial_\rho h) + \int_{\mathcal{D}} d^4x \partial_\mu \epsilon^\mu \mathcal{L}(h, \partial_\rho h) \\ &\quad - \int_{\mathcal{D}} d^4x \frac{\partial \mathcal{L}}{\partial(\partial_\rho h)} \partial_\sigma h \partial_\rho \epsilon^\sigma + \int_{\mathcal{D}} d^4x \bar{\delta} \mathcal{L}, \end{aligned} \quad (17.91)$$

式中

$$\bar{\delta} \mathcal{L} \equiv \mathcal{L}(h + \bar{\delta}h, \partial_\rho(h + \bar{\delta}h)) - \mathcal{L}(h, \partial_\rho h). \quad (17.92)$$

注意到 $\int_{\mathcal{D}} d^4x \mathcal{L}(h, \partial_\rho h) = S_{\text{FP}}[h_{\mu\nu}(x)]$, 从而立即有变换前后作用量的改变量为

$$\begin{aligned} \delta S_{\text{FP}} &\equiv S_{\text{FP}}[h'_{\mu\nu}(x)] - S_{\text{FP}}[h_{\mu\nu}(x)] \\ &= \int_{\mathcal{D}} d^4x \partial_\mu \epsilon^\mu \mathcal{L}(h, \partial_\rho h) - \int_{\mathcal{D}} d^4x \frac{\partial \mathcal{L}}{\partial(\partial_\rho h)} \partial_\sigma h \partial_\rho \epsilon^\sigma + \int_{\mathcal{D}} d^4x \bar{\delta} \mathcal{L}, \\ &= - \int_{\mathcal{D}} d^4x \left[\frac{\partial \mathcal{L}}{\partial(\partial^\rho h_{\mu\nu})} \partial_\sigma h_{\mu\nu} - \delta^\rho_\sigma \mathcal{L} \right] \partial_\rho \epsilon^\sigma + \int_{\mathcal{D}} d^4x \bar{\delta} \mathcal{L}. \end{aligned} \quad (17.93)$$

很显然, 最后的这两项中, 前一项就是标准的正则能动张量的贡献, 不妨记

$$\theta_{\rho\sigma} \equiv - \left[\frac{\partial \mathcal{L}}{\partial(\partial^\rho h_{\mu\nu})} \partial_\sigma h_{\mu\nu} - \eta_{\rho\sigma} \mathcal{L} \right]. \quad (17.94)$$

则可以将上述结果简写成

$$\delta S_{\text{FP}} = \int_{\mathcal{D}} d^4x \theta_{\mu\nu} \partial^\mu \epsilon^\nu + \int_{\mathcal{D}} d^4x \bar{\delta} \mathcal{L}. \quad (17.95)$$

后面的 $\bar{\delta} \mathcal{L}$ 项是对正则能动张量 $\theta_{\mu\nu}$ 的修正, 正如本书第二章中根据局域洛伦兹变换证明过的, 这一修正将正好使得最终的能动张量为对称张量。

读者可能觉得, 根据最小作用量原理, $\bar{\delta} \mathcal{L}$ 的修正项不应该正好等于零吗? 然而情况并非如此。的确, 和最小作用量原理的推导一样, 经过分部积分, 我们有

$$\int_{\mathcal{D}} d^4x \bar{\delta} \mathcal{L} = \int_{\mathcal{D}} d^4x \left[\frac{\partial \mathcal{L}}{\partial h} - \partial_\rho \frac{\partial \mathcal{L}}{\partial(\partial_\rho h)} \right] \bar{\delta} h + \int_{\mathcal{D}} d^4x \partial_\rho \left[\frac{\partial \mathcal{L}}{\partial(\partial_\rho h)} \bar{\delta} h \right] \quad (17.96)$$

等式右边的第一项根据 h 场的运动方程 (欧拉-拉格朗日方程) 自动等于零, 第二项是一个全微分项, 它在区域边界 $\partial\mathcal{D}$ 上有贡献, 而且和最小作用量原理的情形不同, 现在这个边界贡献非零。这是由于 $\bar{\delta}h$ 中包含的是 ϵ^μ_ν 项, 而我们现在并没有要求它在区域边界 $\partial\mathcal{D}$ 上取零, 相反, 我们将要求 ϵ^μ_ν 的偏导在边界 $\partial\mathcal{D}$ 上取零 (即在区域边界上, 所考察的无穷小局域洛伦兹变换趋于一个整体的无穷小洛伦兹变换)。所以, 与最小作用量原理的情形不同, 这里这个全微分项的贡献不是零!

代入前面的 (173) 式, 即可以得到

$$\begin{aligned} \int_{\mathcal{D}} d^4x \bar{\delta}\mathcal{L} &= \int_{\mathcal{D}} d^4x \partial_\rho \left[\frac{\partial\mathcal{L}}{\partial(\partial_\rho h_{\mu\beta})} \bar{\delta}h_{\mu\beta} \right] = -2 \int_{\mathcal{D}} d^4x \partial_\rho \left[\frac{\partial\mathcal{L}}{\partial(\partial_\rho h_{\mu\beta})} h_{\nu\beta} \epsilon^\nu_\mu \right] \\ &= -2 \int_{\mathcal{D}} d^4x \partial_\rho \left[\frac{\partial\mathcal{L}}{\partial(\partial_\rho h_{\mu\beta})} h_{\nu\beta} \right] \epsilon^\nu_\mu - 2 \int_{\mathcal{D}} d^4x \left[\frac{\partial\mathcal{L}}{\partial(\partial_\rho h_{\mu\beta})} h_{\nu\beta} \right] \partial_\rho \epsilon^\nu_\mu \\ &= 2 \int_{\mathcal{D}} d^4x \partial_\rho \left[\frac{\partial\mathcal{L}}{\partial(\partial_\rho h_{\mu\beta})} h^\nu_\beta \right] \epsilon_{\mu\nu} + 2 \int_{\mathcal{D}} d^4x \left[\frac{\partial\mathcal{L}}{\partial(\partial_\rho h_{\mu\beta})} h^\nu_\beta \right] \partial_\rho \epsilon_{\mu\nu} \end{aligned} \quad (17.97)$$

式中我们适当地升降了指标, 并且利用了 $\epsilon_{\mu\nu}$ 的反对称性来去掉一些负号。不妨记,

$$X^{\rho\mu\nu} \equiv \frac{\partial\mathcal{L}}{\partial(\partial_\rho h_{\mu\beta})} h^\nu_\beta. \quad (17.98)$$

将上面这个结果 (184) 和 (182) 式的结果结合, 即有

$$\delta S_{\text{FP}} = \int_{\mathcal{D}} d^4x \left[\theta^{\mu\nu} \partial_\mu \epsilon_\nu + 2\partial_\rho X^{\rho\mu\nu} \epsilon_{\mu\nu} + 2X^{\rho\mu\nu} \partial_\rho \epsilon_{\mu\nu} \right]. \quad (17.99)$$

将 $X^{\rho\mu\nu}$ 的 $\mu\nu$ 指标反对称化, 并记

$$Y^{\rho\mu\nu} \equiv X^{\rho\mu\nu} - X^{\rho\nu\mu}. \quad (17.100)$$

则不难看出 (17.99) 又可以写成

$$\delta S_{\text{FP}} = \int_{\mathcal{D}} d^4x \left[\theta^{\mu\nu} \partial_\mu \epsilon_\nu + \partial_\rho Y^{\rho\mu\nu} \epsilon_{\mu\nu} + Y^{\rho\mu\nu} \partial_\rho \epsilon_{\mu\nu} \right]. \quad (17.101)$$

根据第二章的知识, 为了读出对称的能动张量 $t_{\mu\nu}$, 关键是要把上式凑成如下形式

$$\delta S_{\text{FP}} = \int_{\mathcal{D}} d^4x t^{\mu\nu} \partial_\mu \epsilon_\nu. \quad (17.102)$$

结果是, 的确能 \square 出来, 我先写出答案, 然后我们再验证它的确能倒过来给出原来的 (186) 式 (这比直接凑答案容易些)。答案是,

$$\delta S_{\text{FP}} = \int_{\mathcal{D}} d^4x \left[\theta^{\mu\nu} - \partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) \right] \partial_\mu \epsilon_\nu. \quad (17.103)$$

为了验证, 我们需要用到 $\epsilon_\nu = \epsilon_{\nu\sigma} x^\sigma$, 进而 $\partial_\mu \epsilon_\nu = (\partial_\mu \epsilon_{\nu\sigma}) x^\sigma - \epsilon_{\mu\nu}$, 并注意到 $Y^{\rho\mu\nu}$ 这样的表达式关于后两个指标是反对称的。

为了验证 (188) 式, 显然暂时可以不管 $\theta^{\mu\nu}$

$$\begin{aligned} &\int_{\mathcal{D}} d^4x \left[-\partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) \right] \partial_\mu \epsilon_\nu \\ &= \int_{\mathcal{D}} d^4x \left[-\partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) \right] ((\partial_\mu \epsilon_{\nu\sigma}) x^\sigma - \epsilon_{\mu\nu}) \\ &= \int_{\mathcal{D}} d^4x \left[\partial_\rho Y^{\rho\mu\nu} \epsilon_{\mu\nu} - \partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) (\partial_\mu \epsilon_{\nu\sigma}) x^\sigma \right]. \end{aligned} \quad (17.104)$$

这一步我们用到 $Y^{\mu\rho\nu} + Y^{\nu\rho\mu}$ 关于 $\mu\nu$ 是对称的, 而 $\varepsilon_{\mu\nu}$ 是反对称的。很显然, (186) 式中的 $\varepsilon_{\mu\nu}$ 项已经得到验证了, 下面我们把注意力放在 $\partial_\rho \varepsilon_{\mu\nu}$ 项上。由上面的推导

$$\begin{aligned} & \int_{\mathcal{D}} d^4x \left[-\partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) (\partial_\mu \varepsilon_{\nu\sigma}) x^\sigma \right] \\ &= \int_{\mathcal{D}} d^4x \left[(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) [(\partial_\rho \partial_\mu \varepsilon_{\nu\sigma}) x^\sigma + \partial_\mu \varepsilon_{\nu\rho}] \right]. \end{aligned} \quad (17.105)$$

这里是进行了分部积分 (分部积分的边界项之所以可以忽略, 是因为我们假设了 $\varepsilon_{\nu\sigma}$ 的偏导在 $\partial\mathcal{D}$ 上取零)。注意到 $(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu})$ 关于 $\mu\rho$ 反对称, 而 $(\partial_\rho \partial_\mu \varepsilon_{\nu\sigma}) x^\sigma$ 关于 $\mu\rho$ 对称, 立即有

$$\begin{aligned} & \int_{\mathcal{D}} d^4x \left[(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) [(\partial_\rho \partial_\mu \varepsilon_{\nu\sigma}) x^\sigma + \partial_\mu \varepsilon_{\nu\rho}] \right] \\ &= \int_{\mathcal{D}} d^4x \left[(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) \partial_\mu \varepsilon_{\nu\rho} \right] \\ &= \int_{\mathcal{D}} d^4x \left[(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} + Y^{\nu\mu\rho}) \partial_\mu \varepsilon_{\nu\rho} \right] \\ &= \int_{\mathcal{D}} d^4x \left[(-Y^{\mu\rho\nu}) \partial_\mu \varepsilon_{\nu\rho} \right] = \int_{\mathcal{D}} d^4x \left[Y^{\rho\mu\nu} \partial_\rho \varepsilon_{\mu\nu} \right]. \end{aligned} \quad (17.106)$$

正是 (186) 式中的 $\partial_\rho \varepsilon_{\mu\nu}$ 项, 验证完成! 上述过程中我们利用了 $Y^{\rho\mu\nu} + Y^{\nu\mu\rho}$ 关于指标 $\nu\rho$ 对称, 而 $\partial_\mu \varepsilon_{\nu\rho}$ 关于 $\nu\rho$ 反对称。

从最终的 (188) 式中, 可以读出

$$t^{\mu\nu} = \theta^{\mu\nu} - \partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}). \quad (17.107)$$

这个结果与 Weinberg 量子场论第一卷 7.4 节中给出的一般公式是一致的 (虽然温伯格并没有讨论张量场), 当然推导方法并不相同, 我们这里的推导主要是强调诺特技巧。有些时候人们也称这种对称的能动量张量为 Belinfante(贝林凡特) 张量。不过, 和正则能动张量一样, 一般来说, 这个能动张量并不具有 $\delta h_{\mu\nu} = -(\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu)$ 规范变换之下的规范不变性。

17.5 辐射功率

本节我们着重考察引力源的辐射功率, 正如前面所说, 这需要知道线性化引力场能动张量 $t_{\mu\nu}$ 的具体形式, 而且, 我们想设法让 $t_{\mu\nu}$ 规范不变, 因为这样它才是一个物理量。

17.5.1 引力波能动张量的具体形式

但是真正将 Fierz-Pauli 作用量的表达式代入上面的 (192) 式进行推导计算还是太复杂了! 为了简化问题, 我们可以进一步取横向无迹规范 (TT 规范), 即取

$$h_{00} = h_{0i} = 0, \quad \partial_i h_{ij} = 0, \quad h_{ii} = 0, \quad (17.108)$$

当然, 这只是前面 (17.31) 式的另一种写法。因此, 在这个规范中 $h = h^\mu{}_\mu = 0$, $\partial^\mu h_{\mu\nu} = 0$, 又注意到 Fierz-Pauli 作用量为

$$S_{\text{FP}} = \frac{1}{16\pi G} \int d^4x \left[-\frac{1}{4} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + \frac{1}{2} \partial_\rho h_{\mu\nu} \partial^\nu h^{\rho\mu} + \frac{1}{4} \partial_\mu h \partial^\mu h - \frac{1}{2} \partial_\nu h^{\mu\nu} \partial_\mu h \right],$$

进而通过合适的分部积分即可知, TT 规范下的 Fierz-Pauli 作用量可以写成

$$S_{\text{FP}} = \frac{1}{16\pi G} \int d^4x \left[-\frac{1}{4} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} \right]. \quad (17.109)$$

当然, 这时候 $h_{\mu\nu}$ 场的运动方程就简化为

$$\partial^2 h_{\mu\nu} = 0. \quad (17.110)$$

从 (17.109) 中可以读出相应的拉格朗日量 \mathcal{L} , 为

$$\mathcal{L} = \frac{1}{16\pi G} \left[-\frac{1}{4} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} \right]. \quad (17.111)$$

为了具体计算 (192) 式给出的能动张量, 我们先算一下 $X^{\rho\mu\nu}$, 代入前面的定义, 即可得

$$X^{\rho\mu\nu} = -\frac{1}{32\pi G} (\partial^\rho h^{\mu\beta}) h^\nu_\beta. \quad (17.112)$$

进而即有

$$Y^{\rho\mu\nu} = -\frac{1}{32\pi G} [(\partial^\rho h^{\mu\beta}) h^\nu_\beta - (\partial^\rho h^{\nu\beta}) h^\mu_\beta]. \quad (17.113)$$

利用运动方程, 不难得 $\partial_\rho Y^{\rho\mu\nu} = 0$, 但是, $\partial_\rho (Y^{\mu\rho\nu} + Y^{\nu\rho\mu}) \neq 0$, 具体表达式当然可以直接算出来, 不过, 我们这里就不写这个表达式了, 因为正如后文将表明的, 它对我们本节想得到的最终结果没有贡献。另一方面, 代入前面的一般公式 (181), 即可以得出 $\theta_{\mu\nu}$, 为 (经过了一些指标升降)

$$\theta_{\mu\nu} = \frac{1}{16\pi G} \left[\frac{1}{2} \partial_\mu h_{\alpha\beta} \partial_\nu h^{\alpha\beta} - \frac{1}{4} \eta_{\mu\nu} \partial_\rho h_{\alpha\beta} \partial^\rho h^{\alpha\beta} \right]. \quad (17.114)$$

由于所考察作用量的特殊性, 这已经是一个对称张量了。综上, 可知

$$t^{\mu\nu} = \theta^{\mu\nu} + \partial_\rho (Y^{\mu\rho\nu} + Y^{\nu\rho\mu}). \quad (17.115)$$

的确是一个对称张量。这个能动张量是在 TT 规范下给出的, 它当然不是规范不变的! 然而我们希望让它规范不变, 如此才能作为一个物理量。

为了得到规范不变的能动张量, 一种不太严格但是简单的处理办法是这样的。假设所考察的引力波的特征波长为 λ , 下面我们考察一个以时空点 x 为中心的典型尺寸为 a 的时空区域 V , $a \gg \lambda$ 。然后在 V 上对 $t_{\mu\nu}$ 进行某种粗粒化平均, 即将 $t_{\mu\nu}$ 在 V 上的多次振荡平均掉, 具体来说, 我们定义

$$\langle t_{\mu\nu}(x) \rangle \equiv \int_V d^4y W(x-y) t_{\mu\nu}(y), \quad (17.116)$$

其中权函数 W 在 V 上光滑变化, 满足 $\int_V d^4y W(x-y) = 1$, 以及在区域的边界 ∂V 上 $W = 0$ 。这种粗粒化意味着, 对于一个全微分 ∂X , 其平均将有 $\langle \partial X \rangle \sim 1/a$, 由于相对来说 a 很大 ($a \gg \lambda$), 因此 $1/a$ 可以忽略, 所以在相差 $1/a$ 量级的意义上, 我们有

$$\langle \partial X \rangle = 0. \quad (17.117)$$

由此进一步, 我们有如下分部积分关系 (在相差 $1/a$ 量级的意义上, 以后不再说明)

$$\langle X \partial Y \rangle = -\langle (\partial X) Y \rangle. \quad (17.118)$$

利用上面的关系, 立即有

$$\langle \partial_\rho (Y^{\mu\rho\nu} + Y^{\nu\rho\mu}) \rangle = 0, \quad (17.119)$$

因为它平均的是一个全微分。类似的, 也有

$$\langle \partial_\rho h_{\alpha\beta} \partial^\rho h^{\alpha\beta} \rangle = -\langle (\partial^2 h_{\alpha\beta}) h^{\alpha\beta} \rangle = 0, \quad (17.120)$$

其中最后的等于零是利用了 $h_{\mu\nu}$ 的运动方程。综合这些结果, 以及前面给出的能动张量表达式 (17.115), 立即有

$$\langle t_{\mu\nu} \rangle = \frac{1}{32\pi G} \langle \partial_\mu h_{\rho\sigma} \partial_\nu h^{\rho\sigma} \rangle. \quad (17.121)$$

不难验证, 平均以后的能动张量依然满足守恒方程, 验证如下

$$\partial^\mu \langle t_{\mu\nu} \rangle = \frac{1}{32\pi G} \left[\langle \partial^2 h_{\rho\sigma} \partial_\nu h^{\rho\sigma} \rangle + \frac{1}{2} \langle \partial_\nu (\partial_\mu h_{\rho\sigma} \partial^\mu h^{\rho\sigma}) \rangle \right] = 0, \quad (17.122)$$

等式右边第一项等于零是由于运动方程, 而第二项等于零是因为平均的是一个全微分。

不仅如此, 平均以后的能动张量 (17.121) 还是规范不变的! 为此, 我们考察规范变换 $\delta h_{\mu\nu} = -(\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu)$, 在此变换之下

$$\begin{aligned} \delta \langle t_{\mu\nu} \rangle &= \frac{1}{32\pi G} \left[\langle \partial_\mu \delta h_{\rho\sigma} \partial_\nu h^{\rho\sigma} \rangle + \langle \partial_\mu h_{\rho\sigma} \partial_\nu \delta h^{\rho\sigma} \rangle \right] \\ &= -\frac{1}{32\pi G} \left[\langle \partial_\mu (\partial_\rho \varepsilon_\sigma + \partial_\sigma \varepsilon_\rho) \partial_\nu h^{\rho\sigma} \rangle + \langle \partial_\mu h_{\rho\sigma} \partial_\nu (\partial^\rho \varepsilon^\sigma + \partial^\sigma \varepsilon^\rho) \rangle \right]. \end{aligned}$$

通过分部积分并利用 de Donder 规范条件 $\partial^\rho h_{\rho\sigma} = 0$, 可以看出最后这些项都等于零, 因此最后的结果等于零, 即

$$\delta \langle t_{\mu\nu} \rangle = 0. \quad (17.123)$$

所以, 平均以后的能动张量 (17.121) 才是我们最终要找的物理量。

17.5.2 再谈 TT 规范

对于无源区的一个波矢量为 \mathbf{k} 的平面引力波解 $\bar{h}_{\mu\nu}$, 假设它已经满足 de Donder 规范, 但是不满足 TT 规范, 如何才能得到一个相应的 TT 规范下的解呢?

为此我们注意到波的传播方向为 $\mathbf{n} = \mathbf{k}/|\mathbf{k}|$, \mathbf{n} 为单位矢量, TT 规范当然要满足 $n^i h_{ij} = 0$ 。所以我们引入

$$P_{ij}(\mathbf{n}) \equiv \delta_{ij} - n_i n_j, \quad (17.124)$$

显然, 这是一个对称横向张量, 满足 $n^i P_{ij} = 0$, 且很明显其迹为 2, $P_{ii} = 2$ 。并且, 不难看出 P_{ij} 是一个投影算符, 即满足

$$P_{ik} P_{kj} = P_{ij}. \quad (17.125)$$

进一步，我们构建如下张量

$$\Lambda_{ij,kl}(\mathbf{n}) = P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl}, \quad (17.126)$$

显然，它关于指标对 (ij) 和 (kl) 是对称的。不难验证，它也是投影算符，满足

$$\Lambda_{ij,kl}\Lambda_{kl,mn} = \Lambda_{ij,mn}. \quad (17.127)$$

并且，它的所有指标均是横向的，即有 $n^i\Lambda_{ij,kl} = 0, n^j\Lambda_{ij,kl} = 0$ 等等。且它的 (ij) 和 (kl) 指标对均是无迹的，即

$$\Lambda_{ii,kl} = \Lambda_{ij,kk} = 0. \quad (17.128)$$

得到 TT 规范下的解的办法很简单，就是利用 $\bar{h}_{\mu\nu}$ 的空间分量 \bar{h}_{ij} ,

$$h_{ij}^{\text{TT}} = \Lambda_{ij,kl}\bar{h}_{kl}, \quad (17.129)$$

式中，为了以示区分，我们用上标 TT 来明确标示 TT 规范下的解。更一般的，对于任何对称张量 S_{ij} ，均可以定义其横向无迹部分为

$$S_{ij}^{\text{TT}} = \Lambda_{ij,kl}S_{kl}. \quad (17.130)$$

现在，我们可以将上一小节得到的 $\langle t_{\mu\nu} \rangle$ 公式 (17.121) 重记为 (注意 $h_{0\mu}^{\text{TT}} = 0$)

$$\langle t_{\mu\nu} \rangle = \frac{1}{32\pi G} \langle \partial_\mu h_{ij}^{\text{TT}} \partial_\nu h_{ij}^{\text{TT}} \rangle. \quad (17.131)$$

特别的，能量密度 $\langle t^{00} \rangle$ 为

$$\langle t^{00} \rangle = \frac{1}{32\pi G} \langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \rangle. \quad (17.132)$$

能流密度 $\langle t^{0k} \rangle$ 为 (负号来源于 0 指标的升降)

$$\langle t^{0k} \rangle = -\frac{1}{32\pi G} \langle \dot{h}_{ij}^{\text{TT}} \partial_k h_{ij}^{\text{TT}} \rangle. \quad (17.133)$$

有时候需要将涉及单位矢量 \mathbf{n} 的一些量在单位球面上平均，这里收集一些结果

$$\int \frac{d\Omega}{4\pi} n_i n_j = \frac{1}{3} \delta_{ij}, \quad (17.134)$$

式中 $d\Omega$ 为球面上的立体角。类似的，也有

$$\int \frac{d\Omega}{4\pi} n_i n_j n_k n_l = \frac{1}{15} (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \quad (17.135)$$

而且类似的恒等式还可以推广到任意多个 n 乘积的平均，结果是：对于奇数个 n 的乘积，由于被积函数在空间反演下为奇宇称，所以在球面上积分的结果等于零。而对于偶数个 n 的乘积，注意到 $n_{i_1} n_{i_2} \dots n_{i_{2l}}$ 是全对称的，所以其平均的最终结果也只能依赖于克隆内克符

号的全对称乘积, 需要确定的是这个结果前面的整体常数因子, 这可以通过把所有指标成对地收缩掉来确定, 最终结果是

$$\int \frac{d\Omega}{4\pi} n_{i_1} n_{i_2} \cdots n_{i_{2l}} = \frac{1}{(2l+1)!!} (\delta_{i_1 i_2} \delta_{i_3 i_4} \cdots \delta_{i_{2l-1} i_{2l}} + \cdots). \quad (17.136)$$

根据定义式 (17.126), $\Lambda_{ij,kl}(\mathbf{n})$ 也可以展开成如下形式

$$\begin{aligned} \Lambda_{ij,kl}(\mathbf{n}) = & \delta_{ik} \delta_{jl} - \frac{1}{2} \delta_{ij} \delta_{kl} - n_j n_l \delta_{ik} - n_i n_k \delta_{jl} \\ & + \frac{1}{2} n_k n_l \delta_{ij} + \frac{1}{2} n_i n_j \delta_{kl} + \frac{1}{2} n_i n_j n_k n_l. \end{aligned} \quad (17.137)$$

进一步运用上一段的结果, 即可以得到

$$\int \frac{d\Omega}{4\pi} \Lambda_{ij,kl}(\mathbf{n}) = \frac{1}{30} (11 \delta_{ik} \delta_{jl} - 4 \delta_{ij} \delta_{kl} + \delta_{il} \delta_{jk}). \quad (17.138)$$

17.5.3 辐射功率

有了以上这些准备, 现在即可以计算一个有限尺寸引力源辐射引力波的功率了。为此, 我们需要以这个源为中心, 在离源很远的地方取一个大的球面 S^2 包围这个源 (假设这个球面的半径为 $r = |\mathbf{x}|$), 然后计算单位时间从这个球面流出去的能流, 换言之, 辐射功率 \mathcal{P} 为

$$\mathcal{P} = \int_{S^2} dA n_i \langle t^{0i} \rangle = \int_{S^2} dA \langle t^{0r} \rangle, \quad (17.139)$$

式中 $\mathbf{n} = \mathbf{x}/|\mathbf{x}|$ 为径向的单位矢量, $dA = r^2 d\Omega$ 为球面的面积元。

根据上一小节, 我们知道

$$\langle t^{0r} \rangle = -\frac{1}{32\pi G} \langle \dot{h}_{ij}^{\text{TT}} \frac{\partial}{\partial r} h_{ij}^{\text{TT}} \rangle. \quad (17.140)$$

另一方面, 根据引力波的产生那一节中的公式 (17.69) 可知, h_{ij}^{TT} 必定具有如下形式

$$h_{ij}^{\text{TT}}(t, r) = \frac{1}{r} f_{ij}(t-r). \quad (17.141)$$

因此, 立即有

$$\frac{\partial}{\partial r} h_{ij}^{\text{TT}} = -\frac{1}{r^2} f_{ij}(t-r) + \frac{1}{r} \frac{\partial}{\partial r} f_{ij}(t-r). \quad (17.142)$$

而很显然 $\frac{\partial}{\partial r} f_{ij}(t-r) = -\frac{\partial}{\partial t} f_{ij}(t-r)$, 所以

$$\frac{\partial}{\partial r} h_{ij}^{\text{TT}} = -\dot{h}_{ij}^{\text{TT}} + O(1/r^2). \quad (17.143)$$

由于我们考察的球面离源非常远, 所以相比来说, 后面的 $O(1/r^2)$ 项可以忽略。进一步结合 (17.140) 式, 即有

$$\langle t^{0r} \rangle = \frac{1}{32\pi G} \langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \rangle = \langle t^{00} \rangle. \quad (17.144)$$

所以，引力源的辐射功率为

$$\mathcal{P} = \frac{r^2}{32\pi G} \int_{S^2} d\Omega \langle \dot{h}_{ij}^{\text{TT}} \dot{h}_{ij}^{\text{TT}} \rangle. \quad (17.145)$$

为了得到 h_{ij}^{TT} 的表达式，下面我们要引用引力波的产生那一节中的公式 (17.69) 了，不妨将之重写在下面

$$\bar{h}_{ij}(\mathbf{x}, t) = \frac{2G}{r} \ddot{I}_{ij}(t-r). \quad (17.146)$$

张量 I_{ij} 可以分解成无迹部分和迹部分的和，

$$I_{ij} = \left(I_{ij} - \frac{1}{3} \delta_{ij} I_{kk} \right) + \frac{1}{3} \delta_{ij} I_{kk}. \quad (17.147)$$

式中 $I_{ij} - \frac{1}{3} \delta_{ij} I_{kk}$ 就是无迹部分，常常记作 Q_{ij} ，根据 I_{ij} 的定义，显然有

$$Q_{ij}(t) = I_{ij} - \frac{1}{3} \delta_{ij} I_{kk} = \int d^3 \mathbf{x} T^{00}(t, \mathbf{x}) \left(x_i x_j - \frac{1}{3} \delta_{ij} r^2 \right). \quad (17.148)$$

为了得到引力辐射的 h_{ij}^{TT} ，我们只需要沿着引力波传播的径向 (\mathbf{n} 方向) 对 (17.146) 作一个横向无迹投影，注意横向无迹投影的作用之下， I_{ij} 的迹部分可以丢弃，因此有

$$h_{ij}^{\text{TT}}(\mathbf{x}, t) = \frac{2G}{r} \Lambda_{ij,kl}(\mathbf{n}) \ddot{Q}_{kl}(t-r). \quad (17.149)$$

代入辐射功率的 (17.145) 式，并利用投影算符的性质，不难有

$$\mathcal{P} = \frac{G}{8\pi} \int_{S^2} d\Omega \Lambda_{ij,kl}(\mathbf{n}) \langle \ddot{Q}_{ij} \ddot{Q}_{kl} \rangle. \quad (17.150)$$

注意到 \dot{I}_{ij} 仅仅是 $t-r$ 的函数，因此 \ddot{Q}_{ij} 也仅仅是 $t-r$ 的函数，与方向 \mathbf{n} 无关。所以，上述球面积分其实仅仅只需对 $\Lambda_{ij,kl}(\mathbf{n})$ 进行，利用前面的 (17.138) 式，并注意到 Q_{ij} 是无迹的，不难得到

$$\mathcal{P} = \frac{G}{5} \langle \ddot{Q}_{ij} \ddot{Q}_{ij} \rangle. \quad (17.151)$$

这就是四极近似下引力源辐射功率的最终表达式。

17.5.4 估算辐射功率的量级

下面我们来对辐射功率进行一个快速的估算。还是假设一个双星系统，两星的质量均为 M ，间距为 R ，则根据牛顿引力，它们相互旋转的频率 ω 为

$$\omega^2 R \sim \frac{GM}{R^2}. \quad (17.152)$$

能量四极矩 $Q \sim MR^2$ ，因此 $\ddot{Q} \sim \omega^3 MR^2$ ，由此即可得此双星系统的辐射功率约为

$$\mathcal{P} \sim G \ddot{Q}^2 \sim \frac{G^4 M^5}{R^5}. \quad (17.153)$$

为了估算出量级, 我们需要恢复光速 c , 注意到施瓦西半径为 $R_g = 2GM/c^2$, 牛顿引力常数的量纲为 $[G] = M^{-1}L^3T^{-2}$, 从而即知, c^5/G 为功率的量纲 (即能量除以时间的量纲)。不妨定义如下普朗克亮度

$$L_{\text{Plank}} \equiv \frac{c^5}{G} \approx 3.6 \times 10^{52} \text{J/s}. \quad (17.154)$$

进而即可以把上面的 (17.153) 式恢复为

$$\mathcal{P} \sim \left(\frac{R_g}{R}\right)^5 L_{\text{Plank}}. \quad (17.155)$$

普朗克亮度到底是一个多大的功率呢? 这么说吧, 太阳发光的功率为 $L_{\odot} \approx 10^{-26} L_{\text{Plank}}$, 一个星系大约有 10^{11} 颗恒星, 其总亮度为 $L_{\text{galaxy}} \approx 10^{-15} L_{\text{Plank}}$, 可观测宇宙中大约有 10^{10} 个星系, 因此宇宙中所有恒星的总亮度大约为 $10^{-5} L_{\text{Plank}}$ 。

但是, 当两个黑洞相互旋转着彼此靠近, 当它们之间的距离小到差不多等于施瓦西半径时, 上面的公式 (17.155) 告诉我们, 它们辐射引力波的功率可以达到一个 L_{Plank} ! 也就是说, 这时候这两个黑洞辐射引力波的功率超过整个可观测宇宙所有恒星发光的总功率。

那么, 行星绕日运动也是一个双星系统, 也会辐射引力波, 这个辐射功率大约为多少呢? 当然, 由于这时候我们考察的两个星星质量不等, $M_1 \gg M_2$, 所以这时候公式 (17.153) 应该修正为

$$\mathcal{P} \sim \frac{G^4 M_1^3 M_2^2}{R^5}. \quad (17.156)$$

以木星绕日为例, 木星的质量约为 $10^{-3} M_{\odot}$, 距离太阳约为 10^9km , 太阳的施瓦西半径约为 3km , 可以根据上面公式估算出木星绕日辐射引力波的功率, 约为

$$\mathcal{P} \sim 10^{-50} L_{\text{Plank}} \approx 10^{-24} L_{\odot}. \quad (17.157)$$

显然, 这是一个小到可以忽略的功率。之所以和前面黑洞结合过程辐射功率差这么多, 是因为引力辐射的功率反比于距离 5 次方, 因此它随着距离的增加下降得非常快。

17.6 参考文献

关于引力波, 本章当然只是一个导引, 更多的专门内容可以参见 Michele Maggiore 所著的两卷本:

Gravitational Waves, Volume 1 Theory and Experiments

Gravitational Waves, Volume 2 Astrophysics and Cosmology

非常好的两本书, 强烈推荐给想进一步了解或者深入研究相关内容的读者。

18. 哈密顿表述与 ADM 能量

本章主要是想讨论如何将哈密顿力学的框架应用于广义相对论的研究。这不仅在时空动力学演化的数值模拟中很重要，而且还可以给出引力场哈密顿量的表达式，以及引力场能量的一个合理定义，也就是所谓的 ADM 能量 (Arnowitt-Deser-Misner 能量)。但是为了系统地处理这些问题，需要在数学和物理两方面做一些准备，数学上的准备就是要系统地研究如何将时空流形分解为不同时刻的一系列空间切片，也就是所谓的 3+1 分解或者 ADM 分解，而物理上的准备就是要引入引力场作用量的边界项，也就是所谓的 Gibbons-Hawking-York 边界项。本章将从数学准备开始，讲清楚所有这些问题。

18.1 时空中的超曲面

18.1.1 超曲面与 ADM 分解

考虑 n 维时空中的 $n-1$ 维超曲面，记作 Σ ，通常也称作余 1 维超曲面 (由于 $n-(n-1)=1$)。这样的超曲面常常可以由一个光滑函数 $S(x)$ 按照如下方程刻画

$$S(x) = \sigma, \quad (18.1)$$

式中 σ 为常数。不同的常数 σ 将给出不同的余 1 维超曲面，它们形成一个超曲面簇，通常称这样的簇为时空流形的页层化。局域地，我们总可以用 $n-1$ 维坐标 $y^a, a=1, 2, \dots, n-1$ 参数化 Σ ，进而可以将时空的局域坐标选取为 (σ, y^a) 。

如果 Σ 本身是某个时空区域 \mathcal{D} 的边界，记作 $\Sigma = \partial\mathcal{D}$ ，则 Σ 必定是一个闭合超曲面，也即是说， Σ 自己将没有边界。这当然是因为边界的边界总等于零，即

$$\partial\partial\mathcal{D} = 0. \quad (18.2)$$

$S(x)$ 的梯度将给出 Σ 的余法矢量, 具体来说, Σ 的余法矢量 n_μ 正比于 $\partial_\mu S(x)$, 即

$$n_\mu = f(x)\partial_\mu S(x). \quad (18.3)$$

式中 $f(x)$ 为任意标量函数。当然, 将 n_μ 的指标升上去就是法矢量 n^μ 。如果 n^μ 不是类光矢量, 则我们总可以通过适当地选取 $f(x)$, 使得 n^μ 成为单位法矢量, 即满足

$$g_{\mu\nu}n^\mu n^\nu = \varepsilon, \quad (18.4)$$

式中 $\varepsilon = +1, -1$ 分别对应 n^μ 为类空和类时矢量的情形。如果 n^μ 处处类时, 则称相应的超曲面 Σ 为类空超曲面, 反之, 如果法矢量 n^μ 处处类空, 则称相应 Σ 为类时超曲面。

利用方程 (18.3) 不难证明, 超曲面的法矢量满足所谓的 Frobenius 定理

$$n_{[\mu}D_\nu n_{\rho]} = 0. \quad (18.5)$$

证明很简单, 只注意到联络系数关于两个下指标是对称的, 从而在全反对称操作下贡献为零即可。换言之, 我们有 $n_{[\mu}D_\nu n_{\rho]} = n_{[\mu}\partial_\nu n_{\rho]}$, 进而利用方程 (18.3) 很快就能得出结果等于零。反过来, 如果一个矢量场 n_μ 满足如上 Frobenius 定理, 那么它必定是某个余 1 维超曲面的法矢量场, 具体证明请读者们参阅微分流形相关书籍。

当用 y^a 参数化了 Σ 以后, $\frac{\partial}{\partial y^a} = \frac{\partial x^\mu}{\partial y^a}\partial_\mu$ 当然就是 Σ 的切矢量场, 常常将其指标形式记为 e^μ_a , 也即是说

$$e^\mu_a \equiv \frac{\partial x^\mu}{\partial y^a}. \quad (18.6)$$

作为 Σ 的切矢量场, e^μ_a 当然和法矢量 n^μ 正交, 即有

$$n_\mu e^\mu_a = 0. \quad (18.7)$$

定义

$$h^\mu_\nu \equiv \delta^\mu_\nu - \varepsilon n^\mu n_\nu, \quad (18.8)$$

不难验证,

$$h^\mu_\nu h^\nu_\rho = h^\mu_\rho, \quad (18.9)$$

从而 h^μ_ν 是所谓的投影算符 (不要和引力波那一章中的符号混淆了, 两者之间没有关系)。同样不难验证, 在这个投影算符的作用下

$$h^\mu_\nu e^\nu_a = e^\mu_a, \quad h^\mu_\nu n^\nu = 0, \quad h^\mu_\nu n_\mu = 0. \quad (18.10)$$

也即是说, 这个投影算符的作用是向着超曲面 Σ 的切向进行投影。

记 h^μ_ν 降指标以后的结果为 $h_{\mu\nu} = g_{\mu\nu} - \varepsilon n_\mu n_\nu$, 根据投影恒等式以及指标升降关系不难验证, $h_{\mu\nu}$ 正好就是原来的度规张量 $g_{\mu\nu}$ 向着超曲面 Σ 的切向投影的结果,

$$h_{\mu\nu} \equiv g_{\mu\nu} - \varepsilon n_\mu n_\nu = h^\rho_\mu h^\sigma_\nu g_{\rho\sigma}. \quad (18.11)$$

因此, 与 $h_{\mu\nu}$ 相关的, 可以引入超面上的诱导度规 h_{ab} , 定义如下

$$h_{ab} \equiv e_a^\mu e_b^\nu h_{\mu\nu} = e_a^\mu e_b^\nu g_{\mu\nu}, \quad (18.12)$$

其中第二个等于号是利用了 $n_\mu e_a^\mu = 0$ 。

h_{ab} 的逆矩阵记为 h^{ab} , 它们可以对超曲面的指标 a, b, \dots 进行升降或者缩并。值得注意的是, 虽然 h^{ab} 为 h_{ab} 的逆矩阵, 但是, $h^{\mu\nu}$ 并不是 $h_{\mu\nu}$ 的逆矩阵, 而是它用逆度规张量 $g^{\mu\nu}$ 升指标的结果。事实上, $h_{\mu\nu}$ 是一个秩为 $n-1$ 的 $n \times n$ 矩阵, 它没有逆矩阵。

不过, 我们有如下完备性关系式

$$g^{\mu\nu} = \varepsilon n^\mu n^\nu + h^{ab} e_a^\mu e_b^\nu. \quad (18.13)$$

为了验证这个式子, 我们只需验证等式右边能给出和左边一样的各种内积关系, 比如, 等式右边显然可以给出 n_μ 与 n_ν 的内积为 ε , 也能给出 $e_{\mu a}$ 与 $e_{\nu b}$ 的内积为 h_{ab} , n_μ 与 $e_{\nu a}$ 的内积为零。根据这个结果, 我们有

$$h^{ab} e_a^\mu e_b^\nu = g^{\mu\nu} - \varepsilon n^\mu n^\nu = h^{\mu\nu}. \quad (18.14)$$

更一般地, 任何与超曲面 Σ 相切的张量场 $A^{\mu\nu\dots}$ 均可以作如下分解

$$A^{\mu\nu\dots} = A^{ab\dots} e_a^\mu e_b^\nu \dots, \quad (18.15)$$

很显然, 这个式子的确意味着 $n_\mu A^{\mu\nu\dots} = n_\nu A^{\mu\nu\dots} = 0$, 确证了 $A^{\mu\nu\dots}$ 是与 Σ 相切的张量场。反过来也有

$$A_{ab\dots} = A_{\mu\nu\dots} e_a^\mu e_b^\nu \dots \quad (18.16)$$

因此, 给定时空流形中任意一个张量场, 我们可以用投影算符 h^μ_ν 作用于其每一个指标, 进而得到一个与 Σ 相切的张量场, 然后再利用上式, 即可以得到一个相应的定义在 Σ 上的诱导张量场。值得注意的是, 诱导张量场 $A_{ab\dots}$ 仅仅相对于超面上的坐标变换 $y^a \rightarrow y'^a$ 来说才是张量, 对于时空中的坐标变换 $x^\mu \rightarrow x'^\mu$ 来说, 它是标量。

当在时空流形上取局域坐标 (σ, y^a) 以后, 则除了超面上的切矢量 e_a^μ 以外, 还可以进一步考察时空流形上的切矢量 s^μ , 其定义为

$$s^\mu \equiv \frac{\partial x^\mu}{\partial \sigma}. \quad (18.17)$$

一般来说, s^μ 应该是 n^μ 与 e_a^μ 的线性组合, 写作

$$s^\mu = N n^\mu + N^a e_a^\mu. \quad (18.18)$$

进而即可以将时空流形上的坐标微分写成

$$dx^\mu = s^\mu d\sigma + e_a^\mu dy^a = N n^\mu d\sigma + (N^a d\sigma + dy^a) e_a^\mu. \quad (18.19)$$

因此, 原来时空流形上的线元即可以写成

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = \varepsilon N^2 d\sigma^2 + h_{ab} (dy^a + N^a d\sigma)(dy^b + N^b d\sigma). \quad (18.20)$$

这称之为时空流形度规的 ADM 分解。

特别的，当 Σ 为类空超曲面，从而 $\varepsilon = -1$ 时，变量 σ 就起到时间坐标的作用，这时候 N 就称为时移函数 (lapse function), N^a 就称为位移矢量 (shift vector).

不难看出，如果用矩阵形式，则上述度规的 ADM 分解可以写成

$$(g_{\mu\nu}) = \begin{bmatrix} \varepsilon N^2 + N^a N_a & N_a \\ N_a & h_{ab} \end{bmatrix}. \quad (18.21)$$

所以有

$$\sqrt{|g|} = N\sqrt{|h|}, \quad \text{其中 } h \equiv \det(h_{ab}). \quad (18.22)$$

18.1.2 外曲率

超曲面 Σ 在时空流形中的弯曲情况经常由其外曲率刻画。外曲率张量是一个对称张量，其定义为

$$K_{ab} \equiv \frac{1}{2} e_a^\mu e_b^\nu \mathcal{L}_n g_{\mu\nu}, \quad (18.23)$$

式中 \mathcal{L}_n 表示沿着向量场 n^μ 的李导数。根据第四章中的相关知识， $\frac{1}{2} \mathcal{L}_n g_{\mu\nu} = \frac{1}{2} (D_\mu n_\nu + D_\nu n_\mu) = D_{(\mu} n_{\nu)}$ ，所以上述定义又可以重写为

$$K_{ab} \equiv \frac{1}{2} e_a^\mu e_b^\nu \mathcal{L}_n g_{\mu\nu} = e_a^\mu e_b^\nu D_{(\mu} n_{\nu)}. \quad (18.24)$$

进一步， K_{ab} 的迹为

$$K \equiv h^{ab} K_{ab}. \quad (18.25)$$

利用 (18.13) 式，即有

$$K = g^{\mu\nu} D_{(\mu} n_{\nu)} - \varepsilon n^\mu n^\nu D_{(\mu} n_{\nu)} \quad (18.26)$$

注意到 $n^\mu n_\mu = \varepsilon$ ，从而 $n^\mu D_\nu n_\mu = 0$ ，进而又可以把这个结果写成

$$K = D_\mu n^\mu. \quad (18.27)$$

注意到 $[\frac{\partial}{\partial y^a}, \frac{\partial}{\partial y^b}] = 0$ ，利用 $[X, Y] = 0 \Rightarrow D_X Y - D_Y X = 0$ ，从而 $D_{\partial_{y^a}} e_b^\mu - D_{\partial_{y^b}} e_a^\mu = 0$ ，也就是

$$e_a^\nu D_\nu e_b^\mu = e_b^\nu D_\nu e_a^\mu. \quad (18.28)$$

再利用 $n_\mu e_a^\mu = 0$ ，从而 $(D_\nu n_\mu) e_a^\mu = -n_\mu D_\nu e_a^\mu$ ，不难有

$$\begin{aligned} e_a^\mu e_b^\nu D_\mu n_\nu &= -e_a^\mu (D_\mu e_b^\nu) n_\nu \\ &= -e_b^\mu (D_\mu e_a^\nu) n_\nu = e_b^\mu e_a^\nu D_\mu n_\nu. \end{aligned} \quad (18.29)$$

也即是说， $e_a^\mu e_b^\nu D_\mu n_\nu$ 关于 a, b 指标已经是对称的了，因此，

$$K_{ab} = e_a^\mu e_b^\nu D_{(\mu} n_{\nu)} = e_a^\mu e_b^\nu D_\mu n_\nu. \quad (18.30)$$

外曲率描写的是超曲面在流形中的相对弯曲状况，它不是曲面本身内在的曲率，相同的超曲面 (具有相同的诱导度规) 放在不同的流形中，其外曲率往往不相同。

18.1.3 超曲面上的协变导数

我们想研究任意与超曲面 Σ 相切的张量场 $A^{\mu\nu\dots}$ 是如何求协变导数的, 具体来说, 我们想把 $A^{\mu\nu\dots}$ 在时空流形中的协变导数与 $A^{ab\dots}$ 在超曲面 Σ 中的协变导数 (联络系数由诱导度规 h_{ab} 决定) 联系起来。为了简单起见, 我们考察一个与 Σ 相切的矢量场 A^μ , 它满足

$$n_\mu A^\mu = 0, \quad A^\mu = A^a e^\mu_a, \quad A_a = A_\mu e^\mu_a. \quad (18.31)$$

我们定义 A_a 在超曲面上的内禀协变导数 $\bar{D}_a A_b$ 为时空流形上的协变导数 $D_\mu A_\nu$ 向着超曲面切向的投影, 也即

$$\bar{D}_a A_b \equiv e^\mu_a e^\nu_b (D_\mu A_\nu). \quad (18.32)$$

我们将证明, \bar{D}_a 的联络系数正是与诱导度规 h_{ab} 相容的克里斯托夫联络。

为此, 我们把 (18.32) 式右边改写成

$$\begin{aligned} e^\mu_a e^\nu_b (D_\mu A_\nu) &= e^\mu_a D_\mu (e^\nu_b A_\nu) - (e^\mu_a D_\mu e^\nu_b) A_\nu \\ &= e^\mu_a D_\mu A_b - (e^\mu_a D_\mu e^\nu_b) A^\nu \\ &= \frac{\partial x^\mu}{\partial y^a} \frac{\partial A_b}{\partial x^\mu} - (e^\mu_a D_\mu e^\nu_b) e^{\nu c} A_c \\ &= \frac{\partial}{\partial y^a} A_b - \bar{\Gamma}^c_{ab} A_c \equiv \partial_a A_b - \bar{\Gamma}^c_{ab} A_c. \end{aligned} \quad (18.33)$$

式中我们定义了

$$\bar{\Gamma}^c_{ab} \equiv (e^\mu_a D_\mu e^\nu_b) e^{\nu c} = (e^\mu_a D_\mu e^\nu_b) e^{\nu c}. \quad (18.34)$$

利用 (18.28) 式, 不难看出 $\bar{\Gamma}^c_{ab}$ 关于两个下指标 a, b 对称。

另一方面, 根据定义

$$\begin{aligned} \bar{D}_a h_{bc} &\equiv e^\mu_a e^\nu_b e^\rho_c (D_\mu h_{\nu\rho}) = e^\mu_a e^\nu_b e^\rho_c [D_\mu g_{\nu\rho} - \varepsilon D_\mu (n_\nu n_\rho)] \\ &= -\varepsilon e^\mu_a e^\nu_b e^\rho_c [(D_\mu n_\nu) n_\rho + n_\nu (D_\mu n_\rho)] = 0. \end{aligned} \quad (18.35)$$

式中最后一个等号是利用了 $n_\mu e^\mu_a = 0$ 。所以, 协变导数 \bar{D}_a 与诱导度规 h_{bc} 相容, 而刚才又证明了它的联络系数 $\bar{\Gamma}^c_{ab}$ 是对称的。进而根据第四章中的相关讨论可知, 必定有

$$\bar{\Gamma}^c_{ab} = \frac{1}{2} h^{ce} (\partial_a h_{be} + \partial_b h_{ae} - \partial_e h_{ab}), \quad (18.36)$$

即 $\bar{\Gamma}^c_{ab}$ 为与诱导度规 h_{ab} 相容的克里斯托夫联络。

Gauss-Weingarten 方程

任给与超曲面 Σ 相切的向量场 A^μ , 我们已经定义了 $\bar{D}_a A_b \equiv e^\mu_a e^\nu_b (D_\mu A_\nu)$ 为向量场 $e^\mu_a D_\mu A_\nu$ 的切向投影。现在, 我们想问的是, $e^\mu_a D_\mu A_\nu$ 的法向分量是多少?

为此, 我们将 $e^\mu_a D_\mu A^\nu$ 写作 $e^\mu_a D_\mu A^\rho g^\nu_\rho$ (当然, g^ν_ρ 实际上就是 δ^ν_ρ), 代入 (18.13) 式, 即有

$$\begin{aligned} e^\mu_a D_\mu A^\nu &= (e^\mu_a D_\mu A^\rho) g^\nu_\rho = (e^\mu_a D_\mu A^\rho) (\varepsilon n^\nu n_\rho + h^{bc} e^{\nu}_b e_{\rho c}) \\ &= \varepsilon (e^\mu_a D_\mu A^\rho n_\rho) n^\nu + h^{bc} e^{\nu}_b (e^\mu_a e^{\rho}_c D_\mu A_\rho). \end{aligned} \quad (18.37)$$

注意到 $n_\mu A^\mu = 0$, 从而 $D_\mu A^\rho n_\rho = -A^\rho D_\mu n_\rho$, 代入上式即有

$$\begin{aligned} e^\mu_a D_\mu A^\nu &= -\varepsilon (e^\mu_a A^\rho D_\mu n_\rho) n^\nu + h^{bc} e^{\nu}_b \bar{D}_a A_c \\ &= e^{\nu}_b \bar{D}_a A^b - \varepsilon (e^\mu_a e^{\rho}_b D_\mu n_\rho) A^b n^\nu \\ &= e^{\nu}_b \bar{D}_a A^b - \varepsilon K_{ab} A^b n^\nu. \end{aligned} \quad (18.38)$$

最后一行是代入了关于 K_{ab} 的 (18.30) 式, 这一行的第一项就是切向分量, 第二项则是法向分量。很显然, 法向分量正比于外曲率, 当且仅当外曲率等于零时法向分量才等于零。

特别的, 取 $A^\nu = e^{\nu}_b$, 从而 $A^a = \delta^a_b$, 代入 (18.38) 式, 即有

$$e^\mu_a D_\mu e^{\nu}_b = e^{\nu}_c \bar{\Gamma}^c_{ab} - \varepsilon K_{ab} n^\nu. \quad (18.39)$$

这就是所谓的 Gauss-Weingarten 方程。

18.1.4 Gauss-Codazzi 方程

根据诱导度规 h_{ab} , 当然可以算出超曲面 Σ 的内禀曲率张量 \bar{R}^c_{dab} , 和所有黎曼曲率张量一样, 它由下式定义

$$[\bar{D}_a, \bar{D}_b] A^c = \bar{R}^c_{dab} A^d, \quad (18.40)$$

根据协变导数 \bar{D}_a 的定义, 即可以得出

$$\bar{R}^c_{dab} = \partial_a \bar{\Gamma}^c_{bd} - \partial_b \bar{\Gamma}^c_{ad} + \bar{\Gamma}^c_{ae} \bar{\Gamma}^e_{bd} - \bar{\Gamma}^c_{be} \bar{\Gamma}^e_{ad}. \quad (18.41)$$

另一方面, 时空流形本身也有一个黎曼曲率张量 $R^\rho_{\sigma\mu\nu}$, 它是由时空上的度规张量 $g_{\mu\nu}$ 决定的。我们想问, 这两种黎曼曲率张量之间的关系是什么?

为此, 我们从 Gauss-Weingarten 方程开始, 由它可得

$$e^\mu_a D_\mu (e^{\nu}_b D_\nu e^{\rho}_c) = e^\mu_a D_\mu (e^{\rho}_d \bar{\Gamma}^d_{bc} - \varepsilon K_{bc} n^\rho). \quad (18.42)$$

对于这个方程的左手边 (LHS), 我们有

$$\begin{aligned} \text{LHS} &= e^\mu_a D_\mu (e^{\nu}_b D_\nu e^{\rho}_c) = e^\mu_a e^{\nu}_b (D_\mu D_\nu e^{\rho}_c) + (e^\mu_a D_\mu e^{\nu}_b) (D_\nu e^{\rho}_c) \\ &= e^\mu_a e^{\nu}_b (D_\mu D_\nu e^{\rho}_c) + (e^{\nu}_d \bar{\Gamma}^d_{ab} - \varepsilon K_{ab} n^\nu) (D_\nu e^{\rho}_c) \\ &= e^\mu_a e^{\nu}_b (D_\mu D_\nu e^{\rho}_c) + \bar{\Gamma}^d_{ab} e^{\nu}_d D_\nu e^{\rho}_c - \varepsilon K_{ab} (n^\nu D_\nu e^{\rho}_c) \\ &= e^\mu_a e^{\nu}_b (D_\mu D_\nu e^{\rho}_c) + \bar{\Gamma}^d_{ab} (e^{\rho}_e \bar{\Gamma}^e_{dc} - \varepsilon K_{dc} n^\rho) - \varepsilon K_{ab} (n^\nu D_\nu e^{\rho}_c). \end{aligned}$$

将这个结果减去 ab 置换以后的结果, 即得

$$e^\mu_a e^\nu_b R^\rho_{\sigma\mu\nu} e^\sigma_c. \quad (18.43)$$

而对于方程 (18.42) 的右手边 (RHS), 我们有

$$\begin{aligned} \text{RHS} &= e^\mu_a D_\mu (e^\rho_d \bar{\Gamma}^d_{bc} - \varepsilon K_{bc} n^\rho) \\ &= (e^\mu_a D_\mu e^\rho_d) \bar{\Gamma}^d_{bc} + e^\rho_d (\partial_a \bar{\Gamma}^d_{bc}) - \varepsilon (\partial_a K_{bc}) n^\rho - \varepsilon K_{bc} (e^\mu_a D_\mu n^\rho) \\ &= (e^\rho_e \bar{\Gamma}^e_{ad} - \varepsilon K_{ad} n^\rho) \bar{\Gamma}^d_{bc} + e^\rho_e (\partial_a \bar{\Gamma}^e_{bc}) - \varepsilon (\partial_a K_{bc}) n^\rho - \varepsilon K_{bc} (e^\mu_a D_\mu n^\rho). \end{aligned}$$

同样, 将这个结果减去 ab 置换以后的类似结果, 即得

$$e^\rho_e \bar{R}^e_{cab} - \varepsilon (\bar{D}_a K_{bc} - \bar{D}_b K_{ac}) n^\rho + \varepsilon K_{ac} (e^\mu_b D_\mu n^\rho) - \varepsilon K_{bc} (e^\mu_a D_\mu n^\rho). \quad (18.44)$$

令等式左边的 (18.43) 与等式右边的 (18.44) 相等, 即得

$$\begin{aligned} &e^\mu_a e^\nu_b R^\rho_{\sigma\mu\nu} e^\sigma_c \\ &= e^\rho_e \bar{R}^e_{cab} - \varepsilon (\bar{D}_a K_{bc} - \bar{D}_b K_{ac}) n^\rho + \varepsilon K_{ac} (e^\mu_b D_\mu n^\rho) - \varepsilon K_{bc} (e^\mu_a D_\mu n^\rho) \end{aligned}$$

将这个结果向切向投影, 即得 (用到关于 K_{ab} 的 (18.30) 式)

$$R_{\mu\nu\rho\sigma} e^\mu_a e^\nu_b e^\rho_c e^\sigma_d = \bar{R}_{abcd} + \varepsilon (K_{ad} K_{bc} - K_{ac} K_{bd}). \quad (18.45)$$

而如果沿着法向 n_ν 投影, 则得 (用了一些黎曼曲率张量的代数性质)

$$e^\mu_a e^\nu_b e^\sigma_c R_{\mu\nu\rho\sigma} n^\rho = \bar{D}_a K_{bc} - \bar{D}_b K_{ac}. \quad (18.46)$$

推导过程中我们消去了等式两边整体的负号。

方程 (18.30) 和方程 (18.46) 就是所谓的 Gauss-Codazzi 方程。它们揭示了时空流形中的黎曼曲率张量的一些分量可以用超曲面的内禀曲率与外曲率共同来表达。剩下的分量是 $R_{\mu\rho\nu\sigma} e^\mu_a n^\rho e^\nu_b n^\sigma$, 它们不能用仅仅与 h_{ab} 、 K_{ab} 相关的量来表达。

缩并形式

为了将 Gauss-Codazzi 方程揭示的曲率分解应用于爱因斯坦场方程, 我们需要将黎曼曲率张量进行缩并。

首先, 里奇张量 $R_{\rho\sigma} = g^{\mu\nu} R_{\mu\rho\nu\sigma}$, 代入 (18.13) 式, 即有

$$\begin{aligned} R_{\rho\sigma} &= (\varepsilon n^\mu n^\nu + h^{ab} e^\mu_a e^\nu_b) R_{\mu\rho\nu\sigma} \\ &= \varepsilon R_{\mu\rho\nu\sigma} n^\mu n^\nu + h^{ab} R_{\mu\rho\nu\sigma} e^\mu_a e^\nu_b. \end{aligned} \quad (18.47)$$

注意到 $R_{\mu\rho\nu\sigma} n^\mu n^\rho n^\nu n^\sigma = 0$ (因为 $R_{\mu\rho\nu\sigma}$ 关于指标 μ, σ 反对称), 所以立即有

$$R_{\rho\sigma} n^\rho n^\sigma = h^{ab} R_{\mu\rho\nu\sigma} e^\mu_a n^\rho e^\nu_b n^\sigma. \quad (18.48)$$

同样注意到 $R_{\mu\rho\nu\sigma}n^\mu n^\nu n^\sigma = 0$, 所以有

$$R_{\rho\sigma}e^\rho{}_c n^\sigma = h^{ab}R_{\mu\rho\nu\sigma}e^\mu{}_a e^\rho{}_c e^\nu{}_b n^\sigma. \quad (18.49)$$

进一步, 里奇张量 $R = g^{\rho\sigma}R_{\rho\sigma}$, 利用上面的 (18.47), 即有

$$\begin{aligned} R &= (\varepsilon n^\rho n^\sigma + h^{cd}e^\rho{}_c e^\sigma{}_d)(\varepsilon R_{\mu\rho\nu\sigma}n^\mu n^\nu + h^{ab}R_{\mu\rho\nu\sigma}e^\mu{}_a e^\nu{}_b) \\ &= 2\varepsilon h^{ab}R_{\mu\rho\nu\sigma}e^\mu{}_a n^\rho e^\nu{}_b n^\sigma + h^{ab}h^{cd}R_{\mu\rho\nu\sigma}e^\mu{}_a e^\rho{}_c e^\nu{}_b e^\sigma{}_d \\ &= 2\varepsilon R_{\rho\sigma}n^\rho n^\sigma + h^{ab}h^{cd}R_{\mu\rho\nu\sigma}e^\mu{}_a e^\rho{}_c e^\nu{}_b e^\sigma{}_d. \end{aligned} \quad (18.50)$$

上式最后一行我们代入了 (18.48) 式。

注意到爱因斯坦张量 $G_{\rho\sigma} = R_{\rho\sigma} - \frac{1}{2}g_{\rho\sigma}R$, 进而即可以得到

$$\begin{aligned} G_{\rho\sigma}n^\rho n^\sigma &= R_{\rho\sigma}n^\rho n^\sigma - \frac{1}{2}\varepsilon R \\ &= -\frac{1}{2}\varepsilon[\bar{R} + \varepsilon(K_{ab}K^{ab} - K^2)]. \end{aligned} \quad (18.51)$$

式中我们利用了上面的 (18.48) 式和 (18.50) 式, 并代入了 Gauss-Codazzi 方程。式中的 \bar{R} 就是由诱导度规 h_{ab} 算出来的曲率标量。类似的, 利用 (18.49) 式, 并代入 Gauss-Codazzi 方程, 还可以得到

$$G_{\rho\sigma}e^\rho{}_a n^\sigma = R_{\rho\sigma}e^\rho{}_a n^\sigma = \bar{D}_b K^b{}_a - \bar{D}_a K^b{}_b. \quad (18.52)$$

方程 (18.51) 和方程 (18.52) 的重要性在于, 它们可以用来把爱因斯坦场方程的部分分量用超曲面 Σ 上的相关变量来重新表达。不过, 值得注意的是, 剩下的分量 $G_{\rho\sigma}e^\rho{}_a e^\sigma{}_b$ 不能用仅仅与 h_{ab} 、 K_{ab} 相关的量来表达。

如果 Σ 为类空超曲面, 从而 n^μ 类时, 则通常会定义

$$\rho \equiv T_{\mu\nu}n^\mu n^\nu, \quad j_a \equiv T_{\rho\sigma}e^\rho{}_a n^\sigma. \quad (18.53)$$

则根据爱因斯坦场方程中的 $G_{\rho\sigma}n^\rho n^\sigma = \kappa T_{\rho\sigma}n^\rho n^\sigma$ ($\kappa = 8\pi G$), 以及 $G_{\rho\sigma}e^\rho{}_a n^\sigma = \kappa T_{\rho\sigma}e^\rho{}_a n^\sigma$, 立即有

$$\begin{aligned} \bar{R} + K^2 - K_{ab}K^{ab} &= (2\kappa)\rho \\ \bar{D}_b K^b{}_a - \bar{D}_a K^b{}_b &= \kappa j_a. \end{aligned} \quad (18.54)$$

式中我们利用了 (18.51) 式和 (18.52) 式。

里奇标量

下面对里奇标量的分解形式进行进一步推导, 我们的出发点是上面的 (18.50) 式, 重写如下

$$R = 2\varepsilon R_{\rho\sigma}n^\rho n^\sigma + h^{ab}h^{cd}R_{\mu\rho\nu\sigma}e^\mu{}_a e^\rho{}_c e^\nu{}_b e^\sigma{}_d. \quad (18.55)$$

其中, 利用 Gauss-Codazzi 方程, 立即有

$$h^{ab}h^{cd}R_{\mu\rho\nu\sigma}e^{\mu}_ae^{\rho}_ce^{\nu}_be^{\sigma}_d = \bar{R} + \varepsilon(K_{ab}K^{ab} - K^2). \quad (18.56)$$

利用黎曼曲率张量的定义 $[D_\nu, D_\sigma]n_\rho = -n^\mu R_{\mu\rho\nu\sigma}$, 立即有

$$\begin{aligned} R_{\mu\nu}n^\mu n^\nu &= -n^\nu [D_\nu, D_\rho]n^\rho = -n^\nu (D_\nu D_\rho n^\rho) + n^\nu (D_\rho D_\nu n^\rho) \\ &= -D_\nu (n^\nu D_\rho n^\rho) + (D_\nu n^\nu)(D_\rho n^\rho) + D_\rho (n^\nu D_\nu n^\rho) - (D_\rho n^\nu)(D_\nu n^\rho). \end{aligned}$$

注意到 $K = D_\mu n^\mu$, 所以第二行的第二项就是 K^2 , 下面我们关注第二行的最后一项。

$$\begin{aligned} (D_\rho n^\nu)(D_\nu n^\rho) &= g^{\mu\nu}g^{\rho\sigma}(D_\rho n_\mu)(D_\nu n_\sigma) \\ &= (\varepsilon n^\mu n^\nu + h^{\mu\nu})(\varepsilon n^\rho n^\sigma + h^{\rho\sigma})(D_\rho n_\mu)(D_\nu n_\sigma) \\ &= (\varepsilon n^\mu n^\nu + h^{\mu\nu})h^{\rho\sigma}(D_\rho n_\mu)(D_\nu n_\sigma) = h^{\mu\nu}h^{\rho\sigma}(D_\rho n_\mu)(D_\nu n_\sigma) \\ &= h^{bc}e^{\mu}_be^{\nu}_ce^{\rho}_de^{\sigma}_d(D_\rho n_\mu)(D_\nu n_\sigma) \\ &= h^{ad}h^{bc}(e^{\rho}_ae^{\mu}_bD_\rho n_\mu)(e^{\nu}_ce^{\sigma}_dD_\nu n_\sigma) \\ &= h^{ad}h^{bc}K_{ab}K_{cd} = K^{ab}K_{ab}. \end{aligned}$$

式中第三行是利用了类似 $n^\sigma D_\nu n_\sigma = 0$ (因为 $D_\nu(n^\sigma n_\sigma) = 0$) 这样的结果。因此, 最终我们有

$$R_{\mu\nu}n^\mu n^\nu = -D_\nu (n^\nu D_\rho n^\rho) + D_\rho (n^\nu D_\nu n^\rho) + K^2 - K^{ab}K_{ab}. \quad (18.57)$$

综合上面两段的结果, 即有

$$\begin{aligned} R &= \bar{R} + \varepsilon(K^2 - K^{ab}K_{ab}) + 2\varepsilon[-D_\nu (n^\nu D_\rho n^\rho) + D_\rho (n^\nu D_\nu n^\rho)] \\ &= \bar{R} + \varepsilon(K^2 - K^{ab}K_{ab}) + 2\varepsilon D_\rho (n^\nu D_\nu n^\rho - n^\rho D_\nu n^\nu). \end{aligned} \quad (18.58)$$

这个结果在后文讨论广义相对论的哈密顿表述时会很有用。

18.2 引力作用量的边界项

在本书第五章中, 我们通过对爱因斯坦-希尔伯特作用量进行变分, 得出爱因斯坦场方程。但是, 在那里我们也提到过, 关于变分的边界项实际上还遗留有一个微妙的问题有待处理。现在我们来处理这个问题。

首先让我们简单回顾一下对爱因斯坦-希尔伯特作用量的变分计算。注意到 $S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x \sqrt{-g} R_{\mu\nu} g^{\mu\nu}$ ($\kappa = 8\pi G$), 从而即有

$$\delta S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x [\sqrt{-g} R_{\mu\nu} \delta g^{\mu\nu} + \sqrt{-g} (\delta R_{\mu\nu}) g^{\mu\nu} + \delta(\sqrt{-g}) R]. \quad (18.59)$$

利用 $\delta(\sqrt{-g}) = \frac{1}{2}\sqrt{-g}g^{\mu\nu}\delta g_{\mu\nu}$ 以及 $\delta g^{\mu\nu} = -g^{\mu\rho}g^{\sigma\nu}\delta g_{\rho\sigma}$, 即可以得到

$$\delta S_{\text{EH}} = \frac{1}{2\kappa} \int d^4x \sqrt{-g} [- (R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R) \delta g_{\mu\nu} + (\delta R_{\mu\nu}) g^{\mu\nu}]. \quad (18.60)$$

所以最后要算的就是 $\delta R_{\mu\nu}$, 这已经在第五章中算过了, 这里不再重复, 结果是

$$\delta R_{\mu\nu} = D_\rho(\delta\Gamma_{\nu\mu}^\rho) - D_\nu(\delta\Gamma_{\rho\mu}^\rho). \quad (18.61)$$

利用以上结果, 容易看出 (18.60) 式中的 $\delta R_{\mu\nu}$ 项可以化成

$$\begin{aligned} \frac{1}{2\kappa} \int d^4x \sqrt{-g} (\delta R_{\mu\nu}) g^{\mu\nu} &= \frac{1}{2\kappa} \int d^4x \sqrt{-g} [D_\rho(\delta\Gamma_{\nu\mu}^\rho) - D_\nu(\delta\Gamma_{\rho\mu}^\rho)] g^{\mu\nu} \\ &= \frac{1}{2\kappa} \int d^4x \sqrt{-g} [D_\rho(g^{\mu\nu} \delta\Gamma_{\nu\mu}^\rho) - D_\nu(g^{\mu\nu} \delta\Gamma_{\rho\mu}^\rho)] \\ &= \frac{1}{2\kappa} \int d^4x \sqrt{-g} D_\mu (g^{\rho\sigma} \delta\Gamma_{\rho\sigma}^\mu - g^{\mu\nu} \delta\Gamma_{\rho\nu}^\rho). \end{aligned} \quad (18.62)$$

习惯上, 会令

$$v^\mu \equiv g^{\rho\sigma} \delta\Gamma_{\rho\sigma}^\mu - g^{\mu\nu} \delta\Gamma_{\rho\nu}^\rho. \quad (18.63)$$

从而可以将 (18.60) 式写成

$$(2\kappa)\delta S_{\text{EH}} = \int_{\mathcal{D}} d^4x \sqrt{-g} [-G^{\mu\nu} \delta g_{\mu\nu}] + \int_{\mathcal{D}} d^4x \sqrt{-g} (D_\mu v^\mu), \quad (18.64)$$

式中我们已经假定原来的爱因斯坦-希尔伯特作用量是定义在某个时空区域 \mathcal{D} 上的。

记时空区域 \mathcal{D} 的边界为 $\partial\mathcal{D}$, 它当然是一个闭合的时空超曲面 (为了简单起见, 假定 $\partial\mathcal{D}$ 不存在类光的部分), 记 $\partial\mathcal{D}$ 上的坐标为 y^a , 其上的诱导度规为 h_{ab} , 外曲率为 K_{ab} , 法向量为 n^μ , 满足 $n^\mu n_\mu = \varepsilon$, 其中 ε 在 $\partial\mathcal{D}$ 的类空部分取 -1 , 而在类时部分取 $+1$ 。

进而利用弯曲时空中的高斯定理 (见第四章中的讲述), 即可以把 (18.64) 式进一步写成

$$(2\kappa)\delta S_{\text{EH}} = \int_{\mathcal{D}} d^4x \sqrt{-g} [-G^{\mu\nu} \delta g_{\mu\nu}] + \int_{\partial\mathcal{D}} d^3y \sqrt{|h|} \varepsilon (n_\mu v^\mu), \quad (18.65)$$

很容易看到第二项其实是一个表面积分项, 其中 $\varepsilon = -1$ 的情形是把第四章中关于高斯定理的讲述推广到 n^μ 类时的情形。由于在利用最小作用量原理时我们总是假定场位形在区域边界的变分等于零, 即假定

$$\delta g_{\mu\nu}|_{\partial\mathcal{D}} = 0. \quad (18.66)$$

(注意, 这意味着边界上诱导度规 h_{ab} 的变分也等于零。) 所以你可能认为这个表面积分项会自动等于零。然而情况并没有这么简单, 因为正如我们马上就可以看清楚的, v^μ 并不完全由 $\delta g_{\mu\nu}$ 决定, 它还含有 $\delta g_{\mu\nu}$ 的偏导项, 虽然 $\delta g_{\mu\nu}$ 在 $\partial\mathcal{D}$ 上为零, 但是, $\delta g_{\mu\nu}$ 的偏导在 $\partial\mathcal{D}$ 上并不一定等于零。所以, (18.65) 式中的表面积分项其实并不等于零! 这就是第五章中我们提到过的那个非常微妙的细节问题。

为了看清楚这个细节问题, 让我们回到 v^μ 的定义 (18.63) 式。为此我们需要计算 $\delta\Gamma_{\rho\sigma}^\mu$, 注意到我们关心的是边界 $\partial\mathcal{D}$ 上的情况, 这时候有 $\delta g_{\rho\sigma} = 0 = \delta g^{\rho\sigma}$, 所以

$$\delta\Gamma_{\rho\sigma}^\mu|_{\partial\mathcal{D}} = \frac{1}{2} g^{\mu\nu} (\partial_\rho \delta g_{\sigma\nu} + \partial_\sigma \delta g_{\rho\nu} - \partial_\nu \delta g_{\rho\sigma}). \quad (18.67)$$

代入 v^μ 的定义式 (18.63), 可以得到

$$v_\mu = g^{\rho\sigma}(\partial_\sigma \delta g_{\mu\rho} - \partial_\mu \delta g_{\sigma\rho}). \quad (18.68)$$

进而即有

$$\begin{aligned} (n^\mu v_\mu)|_{\partial\mathcal{D}} &= n^\mu(\varepsilon n^\rho n^\sigma + h^{\rho\sigma})(\partial_\sigma \delta g_{\mu\rho} - \partial_\mu \delta g_{\sigma\rho}) \\ &= n^\mu h^{\rho\sigma}(\partial_\sigma \delta g_{\mu\rho} - \partial_\mu \delta g_{\sigma\rho}), \end{aligned} \quad (18.69)$$

式中我们注意到了 $(\partial_\sigma \delta g_{\mu\rho} - \partial_\mu \delta g_{\sigma\rho})$ 关于 σ, μ 反对称。又由于 $h^{\rho\sigma}$ 是与超曲面 $\partial\mathcal{D}$ 相切的张量, 而 $\delta g_{\rho\sigma}$ 在超曲面 $\partial\mathcal{D}$ 上恒等于零, 因此, $\delta g_{\rho\sigma}$ 沿着切向的导数也等于零, 从而 $h^{\rho\sigma} \partial_\sigma \delta g_{\mu\rho} = 0$, 因此最终我们有

$$(n^\mu v_\mu)|_{\partial\mathcal{D}} = -h^{\rho\sigma} n^\mu \partial_\mu \delta g_{\rho\sigma}, \quad (18.70)$$

注意这里涉及的是 $\delta g_{\rho\sigma}$ 沿着法向的偏导, 而不是切向的, 因此不等于零。

综合以上所有结果, 即有

$$(2\kappa)\delta S_{\text{EH}} = \int_{\mathcal{D}} d^4x \sqrt{-g} [-G^{\mu\nu} \delta g_{\mu\nu}] - \int_{\partial\mathcal{D}} d^3y \sqrt{|h|} \varepsilon h^{\rho\sigma} n^\mu \partial_\mu \delta g_{\rho\sigma}. \quad (18.71)$$

这里涉及一个非零的边界积分项, 而在应用最小作用量原理时, 这样的项是我们希望能消除的。之所以出现这样的非零边界项, 是因为里奇标量中 (从而 S_{EH} 中) 包含了度规场的二阶偏导, 而不仅仅是一阶偏导, 这是与通常场论不同的地方。

为了消除这样的变分边界项, 我们在引力场的作用量中额外添加一个边界作用量 S_B (称作 Gibbons-Hawking-York 边界项), 其定义是

$$S_B = \frac{1}{\kappa} \int_{\partial\mathcal{D}} d^3y \sqrt{|h|} \varepsilon K. \quad (18.72)$$

下面我们来证明, 这个边界作用量的变分正好能抵消掉 (18.71) 中多出来的边界项。

注意到

$$\begin{aligned} K &= D_\mu n^\mu = g^{\mu\nu} D_\mu n_\nu = (\varepsilon n^\mu n^\nu + h^{\mu\nu}) D_\mu n_\nu \\ &= h^{\mu\nu} D_\mu n_\nu = h^{\mu\nu} (\partial_\mu n_\nu - \Gamma_{\mu\nu}^\rho n_\rho). \end{aligned} \quad (18.73)$$

由于 $h^{\mu\nu}$ 是边界上的度规, 其变分等于零, 从而

$$\begin{aligned} \delta K &= -h^{\mu\nu} \delta \Gamma_{\mu\nu}^\rho n_\rho \\ &= -\frac{1}{2} h^{\mu\nu} n^\rho (\partial_\mu \delta g_{\nu\rho} + \partial_\nu \delta g_{\mu\rho} - \partial_\rho \delta g_{\mu\nu}) \\ &= \frac{1}{2} h^{\mu\nu} n^\rho \partial_\rho \delta g_{\mu\nu}. \end{aligned} \quad (18.74)$$

最后一行同样是利用了度规场变分沿着切向的偏导等于零。根据这个结果, 即有

$$(2\kappa)\delta S_B = 2 \int_{\partial\mathcal{D}} d^3y \sqrt{|h|} \varepsilon \delta K = \int_{\partial\mathcal{D}} d^3y \sqrt{|h|} \varepsilon h^{\rho\sigma} n^\mu \partial_\mu \delta g_{\rho\sigma}. \quad (18.75)$$

很显然, 这正好抵消了 (18.71) 中多出来的边界项。

因此, 只要我们取引力场的作用量为 $S_G = S_{\text{EH}} + S_B$, 式中

$$S_{\text{EH}} = \frac{1}{2\kappa} \int_{\mathcal{D}} d^4x \sqrt{-g} R, \quad S_B = \frac{1}{\kappa} \int_{\partial\mathcal{D}} d^3y \sqrt{|h|} \varepsilon K. \quad (18.76)$$

则 S_G 的变分将正好给出

$$\delta S_G = \frac{1}{2\kappa} \int_{\mathcal{D}} d^4x \sqrt{-g} [-G^{\mu\nu} \delta g_{\mu\nu}]. \quad (18.77)$$

进而就补上了第五章中推导爱因斯坦场方程时的细节性漏洞。

但是, 这里还有一个细节问题, 那就是即使对于平坦的闵可夫斯基时空 (因此里奇标量 $R = 0$), 随着时空区域 \mathcal{D} 趋于无穷大, 也即 $\partial\mathcal{D}$ 趋于无穷远, 计算发现上面定义的 S_B 也会发散, 从而使得整个时空的 S_G 发散! 为了克服这个问题, 通常会将 S_B 修正为

$$S_B \equiv \frac{1}{\kappa} \int_{\partial\mathcal{D}} d^3y \sqrt{|h|} \varepsilon (K - K_0), \quad (18.78)$$

式中 K_0 为同一个超曲面 $\partial\mathcal{D}$ (有相同的诱导度规) 嵌入参考的平坦时空流形时的外曲率 (K 当然为 $\partial\mathcal{D}$ 在待考察的这个时空中的外曲率)。很显然, 这样修正以后, 对于平坦时空即有 $S_B = 0$, 从而也有 $S_G = 0$ 。更一般地, 对于任何渐进平坦时空, 修正以后引力作用量 S_G 都将保持有限, 这就解决了发散问题。注意, 由于 K_0 不依赖于待考察时空的度规, 因此它的变分等于零, 从而这种修正不会改变前面关于变分边界项的所有讨论。

18.3 哈密顿表述与 ADM 能量

现在开始讨论本章的正题, 广义相对论的哈密顿表述, 也即是说, 在哈密顿力学的框架下讨论广义相对论。这需要将四维时空流形进行 ADM 分解, 也称作 3+1 分解, 即将时空流形切片成三维类空超曲面簇, 只不过对应前面 (18.1.1) 小节中的 σ 现在取作时间参数 t 。利用哈密顿表述, 我们可以给引力场整体的能量下一个合适的定义, 称作 ADM 能量。需要说明的是, ADM 能量仅适用于渐近平坦的时空, 因此需要宇宙学常数 $\Lambda = 0$ 。

为了建立哈密顿表述, 我们需要把时空切片成余 1 维的类空超曲面簇, 其中每一个超曲面对应一个时间参数 t , 给定 t 的类空超曲面记作 Σ_t , 它由方程 $t(x) = t$ 定义, 等式左边的 $t(x)$ 相应于前面 (18.1.1) 小节中的 $S(x)$ 。 Σ_t 的法向量记作 n^μ ($n_\mu \propto \partial_\mu t(x)$), 当然 n^μ 是类时的, 满足

$$n^\mu n_\mu = -1. \quad (18.79)$$

同样也可以用坐标 y^a 参数化 Σ_t 本身, 如此一来, (t, y^a) 就构成了时空的一组局域坐标。

根据 (18.1.1) 小节中讨论的 ADM 分解, 可以将时空流形的度规分解成

$$ds^2 = -N^2 dt^2 + h_{ab} (dy^a + N^a dt) (dy^b + N^b dt), \quad (18.80)$$

其中 h_{ab} 为 Σ_t 上的诱导度规, 其余符号的含义也都和前文相同, 比如 Σ_t 的外曲率还是记作 K_{ab} 。

以 t 为参数的等 y^a 曲线与 Σ_t 相交, 但通常不是正交。具体来说, 这些曲线的切矢量 $t^\mu \equiv \left(\frac{\partial x^\mu}{\partial t}\right)_{y^a}$ (相应于 (18.1.1) 小节中的 s^μ) 可以展开为

$$t^\mu = Nn^\mu + N^a e^\mu_a. \quad (18.81)$$

这个方程解释了时移函数 N 以及位移矢量 N^a 的含义。从这个方程我们可以得到

$$n_\mu t^\mu = -N. \quad (18.82)$$

另一方面, 又有

$$(\partial_\mu t) t^\mu = \frac{\partial t}{\partial x^\mu} \frac{\partial x^\mu}{\partial t} = \frac{\partial t}{\partial t} = 1. \quad (18.83)$$

而且, 我们也知道 $n_\mu \propto \partial_\mu t$, 进而比较上面两个式子, 即知

$$n_\mu = -N \partial_\mu t. \quad (18.84)$$

如果我们取局部坐标 $x^\mu = (t, y^a)$, 则很显然 $t^\mu = (1, 0, 0, 0)$, $e^\mu_a = \delta^\mu_a$, 由此立即有

$$\mathcal{L}_t e^\mu_a = 0. \quad (18.85)$$

式中李导数 \mathcal{L}_t 的下标 t 表示向量场 t^μ , 由于这是一个张量方程, 所以它当然在任意局部坐标中均成立!

18.3.1 构造类时边界

在 Σ_t 上面取一个余 1 维闭合超曲面 S_t , 注意, 这里余 1 维是指在 Σ_t 中余 1 维, 同样, 闭合也是指在 Σ_t 中闭合。记 S_t 在 Σ_t 中的法矢量为 r^a , 在时空中, 这个法矢量当然就是 $r^\mu = r^a e^\mu_a$, 显然 $r^\mu n_\mu = 0$ 。随着沿法向 n^μ 的推移, 不同时刻 Σ_t 上的 S_t 会扫出一个在时空中余 1 维的类时超曲面 \mathcal{B} , 如图 (18.1) 所示。很显然, r^μ 也是 \mathcal{B} 在时空中的法矢量, 因此当然是一个类空矢量, 满足 $r^\mu r_\mu = 1$ 。

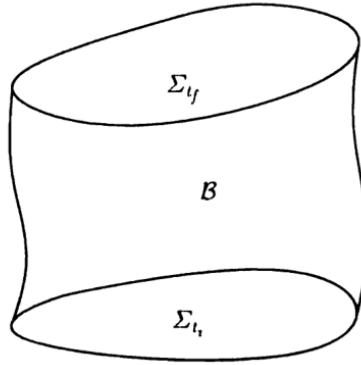


Figure 18.1: 以 $\Sigma_{t_f} \cup (-\Sigma_{t_i}) \cup \mathcal{B}$ 为边界的时空区域 \mathcal{D} 。

超曲面 S_t 可以看成是 Σ_t 与 \mathcal{B} 的交, 在时空中, 它当然是余 2 维的。以坐标 θ^A 参数化 S_t , 引入 S_t 上的切矢量场

$$e_A^\mu \equiv \frac{\partial x^\mu}{\partial \theta^A}. \quad (18.86)$$

这些切矢量场当然同时也是 Σ_t 上的切矢量 (因此有 $n_\mu e_A^\mu = 0$), 因此可以用 e_a^μ 展开, 设 $e_A^\mu = e_a^\mu e^a_A$, 不难看出 $e^a_A = \frac{\partial y^a}{\partial \theta^A}$ 。

S_t 上的诱导度规 σ_{AB} 显然满足

$$\sigma_{AB} = g_{\mu\nu} e_A^\mu e_B^\nu = h_{ab} e^a_A e^b_B. \quad (18.87)$$

注意到 r^a 是 S_t 在 Σ_t 中的法向量, 所以 S_t 在 Σ_t 中的完备性关系为 $h^{ab} = r^a r^b + \sigma^{AB} e^a_A e^b_B$, 又由于 Σ_t 在时空中的完备性关系为 $g^{\mu\nu} = -n^\mu n^\nu + h^{ab} e^a_\mu e^b_\nu$, 进而可得 S_t 最终在时空中的完备性关系为

$$g^{\mu\nu} = -n^\mu n^\nu + r^\mu r^\nu + \sigma^{AB} e^a_A e^b_B, \quad (18.88)$$

这当然是因为 S_t 在时空中是余 2 维超曲面。 S_t 在 Σ_t 中的外曲率 (相对于 Σ_t 的弯曲程度) 及其迹分别为

$$k_{AB} = e^a_A e^b_B D_\mu r_\nu = e^a_A e^b_B \bar{D}_a r_b, \quad k = \sigma^{AB} k_{AB}. \quad (18.89)$$

式中我们代入了 $e_A^\mu = e^a_A e^a_\mu$ 。

从上述讨论来看, 参数化不同时刻 S_t 的坐标 θ^A 似乎可以是相互独立的。但是, 前面定义 \mathcal{B} 时说了, 是沿着 Σ_t 的法向随时间推移 S_t 。因此, 可以设想以 t 为参数的曲线是正交地穿过不同时刻的 S_t 的 (也即是说在 S_t 上, 位移矢量 $N^a = 0$), 自然的定义是, 取同一条 t 曲线上所有点的 θ^A 坐标相同。也即是说, 我们有 (即在方程 (18.81) 中取 $N^a = 0$)

$$\left(\frac{\partial x^\mu}{\partial t} \right)_{\theta^A} = N n^\mu. \quad (18.90)$$

假设用任意坐标 z^i 参数化 \mathcal{B} , 并在 \mathcal{B} 上引入切向量场 e_i^μ , 定义为

$$e_i^\mu \equiv \frac{\partial x^\mu}{\partial z^i}. \quad (18.91)$$

满足 $r_\mu e_i^\mu = 0$ 。因此 \mathcal{B} 上的诱导度规 γ_{ij} 满足

$$\gamma_{ij} = g_{\mu\nu} e_i^\mu e_j^\nu. \quad (18.92)$$

类似于前面的 (18.13), 当然也有如下完备性关系

$$g^{\mu\nu} = r^\mu r^\nu + \gamma^{ij} e_i^\mu e_j^\nu. \quad (18.93)$$

类似的, \mathcal{B} 的外曲率为

$$\mathcal{K}_{ij} = e_i^\mu e_j^\nu D_\mu r_\nu. \quad (18.94)$$

其迹为 $\mathcal{K} = \gamma^{ij} \mathcal{K}_{ij}$ 。

虽然原则上 z^i 可以任意选取，但是上面关于如何定义不同时刻 S_t 上坐标的讨论告诉我们，选取 $z^i = (t, \theta^A)$ 是特别方便的。在这种选取中， \mathcal{B} 上的一个小位移 dx^μ 将由下式给出

$$dx^\mu = \left(\frac{\partial x^\mu}{\partial t} \right)_{\theta^A} dt + \left(\frac{\partial x^\mu}{\partial \theta^A} \right)_t d\theta^A = Nn^\mu dt + e^\mu_A d\theta^A. \quad (18.95)$$

这里我们用了 (18.86) 式和 (18.90) 式。因此 \mathcal{B} 上的线元将为

$$\begin{aligned} ds_{\mathcal{B}}^2 &= g_{\mu\nu} dx^\mu dx^\nu|_{\mathcal{B}} \\ &= g_{\mu\nu} (Nn^\mu dt + e^\mu_A d\theta^A) (Nn^\nu dt + e^\nu_B d\theta^B) \\ &= (g_{\mu\nu} n^\mu n^\nu) N^2 dt^2 + (g_{\mu\nu} e^\mu_A e^\nu_B) d\theta^A d\theta^B, \end{aligned} \quad (18.96)$$

式中用到 $n_\mu e^\mu_A = 0$ 。由这个结果立即有

$$\gamma_{ij} dz^i dz^j = -N^2 dt^2 + \sigma_{AB} d\theta^A d\theta^B. \quad (18.97)$$

这意味着 $\sqrt{-\gamma} = N\sqrt{\sigma}$ ，其中 $\gamma = \det(\gamma_{ij})$ ， $\sigma = \det(\sigma_{AB})$ 。

18.3.2 引力作用量的 ADM 分解

为了寻找广义相对论的哈密顿表述，我们需要把引力场的作用量进行 ADM 分解。为此我们先将引力场作用量重写在下面

$$S_G = \frac{1}{2\kappa} \int_{\mathcal{D}} d^4x \sqrt{-g} R + \frac{1}{\kappa} \int_{\partial\mathcal{D}} d^3y \sqrt{|h|} \epsilon K. \quad (18.98)$$

式中我们暂时忽略了减除项 K_0 。

我们让 \mathcal{D} 为类空超曲面 Σ_{t_i} 和 Σ_{t_f} 以及类时超曲面 \mathcal{B} 围起来的时空区域，如上面的图 (18.1) 所示，换言之

$$\partial\mathcal{D} = \Sigma_{t_f} \cup (-\Sigma_{t_i}) \cup \mathcal{B}, \quad (18.99)$$

式中 $-\Sigma_{t_i}$ 前面的负号是用来提醒我们作为 \mathcal{D} 的边界 $-\Sigma_{t_i}$ 的法向是向着过去方向的，与 Σ_{t_i} 的法向相反，这当然是因为 $\partial\mathcal{D}$ 的法向按照定义就是从区域内部指向外部的。进而利用上一小节的符号，我们可以把区域 \mathcal{D} 上引力场的作用量更清楚地写成

$$\begin{aligned} S_G &= \frac{1}{2\kappa} \int_{\mathcal{D}} d^4x \sqrt{-g} R - \frac{1}{\kappa} \int_{\Sigma_{t_f}} d^3y \sqrt{h} K \\ &\quad + \frac{1}{\kappa} \int_{\Sigma_{t_i}} d^3y \sqrt{h} K + \frac{1}{\kappa} \int_{\mathcal{B}} d^3z \sqrt{-\gamma} \mathcal{K}, \end{aligned} \quad (18.100)$$

式中我们已经考虑 $-\Sigma_{t_i}$ 前面那个额外的负号了，即用了 $\int_{-\Sigma_{t_i}} d^3y \cdots = -\int_{\Sigma_{t_i}} d^3y \cdots$ 。

当然，时空区域 \mathcal{D} 被切片成了类空超曲面 Σ_t 簇，根据前面的公式 (18.58)，在 Σ_t 上

$$R = \bar{R} + K^{ab} K_{ab} - K^2 - 2D_\rho (n^\nu D_\nu n^\rho - n^\rho D_\nu n^\nu). \quad (18.101)$$

进而即有

$$\begin{aligned} \frac{1}{2\kappa} \int_{\mathcal{D}} d^4x \sqrt{-g} R &= \frac{1}{2\kappa} \int_{t_i}^{t_f} dt \int_{\Sigma_t} d^3y N \sqrt{h} (\bar{R} + K^{ab} K_{ab} - K^2) \\ &\quad - \frac{1}{\kappa} \int_{\partial\mathcal{D}} d^3y \sqrt{|h|} \varepsilon m_\rho (n^\nu D_\nu n^\rho - n^\rho D_\nu n^\nu). \end{aligned} \quad (18.102)$$

式中我们应用了高斯定理, m_ρ 为 $\partial\mathcal{D}$ 的法向量, 注意它不一定为 n_ρ 。同时我们还用到 $d^4x \sqrt{|g|} = d^3y N \sqrt{h}$ 。

同样, 上式中的边界积分项要分解成分别在 $-\Sigma_{t_i}$, Σ_{t_f} , 以及 \mathcal{B} 上的积分。其中在 $-\Sigma_{t_i}$ 上的积分可以转换成 (考虑到了这时候 $m_\rho = -n_\rho$, 以及 $\varepsilon = -1$)

$$\begin{aligned} &\frac{1}{\kappa} \int_{\Sigma_{t_i}} d^3y \sqrt{h} (-) n_\rho (n^\nu D_\nu n^\rho - n^\rho D_\nu n^\nu) \\ &= -\frac{1}{\kappa} \int_{\Sigma_{t_i}} d^3y \sqrt{h} D_\nu n^\nu = -\frac{1}{\kappa} \int_{\Sigma_{t_i}} d^3y \sqrt{h} K. \end{aligned} \quad (18.103)$$

式中我们用到了 $n_\rho n^\rho = -1$, 以及 $n_\rho D_\nu n^\rho = 0$ 。很显然, 结果正好与 (18.100) 中边界作用量的相应项抵消。类似的, Σ_{t_f} 上的所有边界积分项也会刚好抵消。因此只剩下 \mathcal{B} 上的贡献 (\mathcal{B} 的 $\varepsilon = 1$, 法向量为 r^μ)

$$\begin{aligned} &-\frac{1}{\kappa} \int_{\mathcal{B}} d^3z \sqrt{-\gamma} r_\rho (n^\nu D_\nu n^\rho - n^\rho D_\nu n^\nu) \\ &= -\frac{1}{\kappa} \int_{\mathcal{B}} d^3z \sqrt{-\gamma} r_\rho n^\nu D_\nu n^\rho = \frac{1}{\kappa} \int_{\mathcal{B}} d^3z \sqrt{-\gamma} n^\nu n^\rho (D_\nu r_\rho). \end{aligned} \quad (18.104)$$

式中我们利用了 $r_\rho n^\rho = 0$ 并进行了分部积分。

把所有这些结果放在一起, 即有

$$\begin{aligned} S_G &= \frac{1}{2\kappa} \int_{t_i}^{t_f} dt \int_{\Sigma_t} d^3y N \sqrt{h} (\bar{R} + K^{ab} K_{ab} - K^2) \\ &\quad + \frac{1}{\kappa} \int_{\mathcal{B}} d^3z \sqrt{-\gamma} (\mathcal{K} + n^\mu n^\nu D_\mu r_\nu). \end{aligned} \quad (18.105)$$

下面我们改写一下 \mathcal{K} ,

$$\begin{aligned} \mathcal{K} &= \gamma^{ij} \mathcal{K}_{ij} = \gamma^{ij} (e^\mu_i e^\nu_j D_\mu r_\nu) \\ &= (\gamma^{ij} e^\mu_i e^\nu_j) D_\mu r_\nu = (g^{\mu\nu} - r^\mu r^\nu) D_\mu r_\nu, \end{aligned} \quad (18.106)$$

其中我们利用了完备性关系 (18.93)。进而即有

$$\begin{aligned} \mathcal{K} + n^\mu n^\nu D_\mu r_\nu &= (g^{\mu\nu} - r^\mu r^\nu + n^\mu n^\nu) D_\mu r_\nu \\ &= (\sigma^{AB} e^\mu_A e^\nu_B) D_\mu r_\nu = \sigma^{AB} (e^\mu_A e^\nu_B D_\mu r_\nu) \\ &= \sigma^{AB} k_{AB} = k. \end{aligned} \quad (18.107)$$

其中我们利用了完备性关系 (18.88)。

有了上面的结果, 再利用 $d^3z \sqrt{-\gamma} = dt d^2\theta N \sqrt{\sigma}$, 进而即有

$$\begin{aligned} S_G &= \int_{t_i}^{t_f} dt \left[\frac{1}{2\kappa} \int_{\Sigma_t} d^3y N \sqrt{h} (\bar{R} + K^{ab} K_{ab} - K^2) \right. \\ &\quad \left. + \frac{1}{\kappa} \int_{S_t} d^2\theta N \sqrt{\sigma} (k - k_0) \right]. \end{aligned} \quad (18.108)$$

式中通过包含进一个额外的 $-k_0$, 我们已经恢复了被忽略的减除项。 k_0 是 S_t 嵌入在平坦空间中的外曲率。有了减除项 k_0 以后, 就可以使得平坦时空的作用量为零, 进一步, 对于渐进平坦时空, 它可以保证作用量在 $S_t \rightarrow \infty$ 时保持有限。

注意, 在上面的结果中, 边界积分项对拉格朗日量是有贡献的, 但是对于拉格朗日体密度当然没有贡献。事实上, 我们可以从 (18.108) 中读出引力场的拉格朗日体密度 \mathcal{L}_G (不要和李导数混淆了), 为

$$\mathcal{L}_G = \frac{1}{2\kappa} N \sqrt{h} (\bar{R} + K^{ab} K_{ab} - K^2). \quad (18.109)$$

而拉格朗日量 L_G 为

$$L_G = \int_{\Sigma_t} d^3y \mathcal{L}_G + \frac{1}{\kappa} \int_{S_t} d^2\theta N \sqrt{\sigma} (k - k_0). \quad (18.110)$$

当然, 一般来说时空中还会有物质场, 物质场的作用量也要进行 3+1 分解, 但是由于给定物质场作用量以后, 这种分解没有原则上的困难, 所以我们不再进行专门的讨论。同样, 后文也都会仅仅将注意力集中在引力场上, 忽略物质场。

18.3.3 引力场的哈密顿量

为了得到引力场的哈密顿表述, 我们定义诱导度规 h_{ab} 的时间导数为

$$\dot{h}_{ab} \equiv \mathcal{L}_t h_{ab}, \quad (18.111)$$

根据诱导度规的定义, 有

$$\dot{h}_{ab} = \mathcal{L}_t (g_{\mu\nu} e^\mu_a e^\nu_b) = (\mathcal{L}_t g_{\mu\nu}) e^\mu_a e^\nu_b, \quad (18.112)$$

式中我们利用了 (18.85) 式。直接计算度规场的李导数

$$\begin{aligned} \mathcal{L}_t g_{\mu\nu} &= D_\mu t_\nu + D_\nu t_\mu = D_\mu (N n_\nu + N_\nu) + D_\nu (N n_\mu + N_\mu) \\ &= (\partial_\mu N) n_\nu + (\partial_\nu N) n_\mu + N (D_\mu n_\nu + D_\nu n_\mu) + D_\mu N_\nu + D_\nu N_\mu. \end{aligned}$$

其中 $N^\mu \equiv N^a e^\mu_a$ 。进而即有

$$\dot{h}_{ab} = 2N K_{ab} + \bar{D}_a N_b + \bar{D}_b N_a \Rightarrow K_{ab} = \frac{1}{2N} (\dot{h}_{ab} - \bar{D}_a N_b - \bar{D}_b N_a). \quad (18.113)$$

上述分析表明, 引力场作用量 (18.108) 中的时间导数全部都包含在外曲率中, 进而利用拉格朗日体密度 \mathcal{L}_G 的定义 (18.109), 即可以得到场变量 h_{ab} 的共轭动量 π^{ab} 为

$$\pi^{ab} \equiv \frac{\partial \mathcal{L}_G}{\partial \dot{h}_{ab}} = \frac{1}{2\kappa} \sqrt{h} (K^{ab} - K h^{ab}). \quad (18.114)$$

所以根据勒让德变换, 哈密顿体密度 \mathcal{H}_G 为

$$\begin{aligned}
\mathcal{H}_G &\equiv \pi^{ab} \dot{h}_{ab} - \mathcal{L}_G \\
&= \frac{1}{2\kappa} \sqrt{h} (K^{ab} - Kh^{ab}) (2NK_{ab} + \bar{D}_a N_b + \bar{D}_b N_a) \\
&\quad - \frac{1}{2\kappa} N \sqrt{h} (\bar{R} + K^{ab} K_{ab} - K^2) \\
&= \frac{1}{2\kappa} N \sqrt{h} (K^{ab} K_{ab} - K^2 - \bar{R}) + \frac{1}{\kappa} \sqrt{h} (K^{ab} - Kh^{ab}) (\bar{D}_a N_b) \\
&= \frac{1}{2\kappa} N \sqrt{h} (K^{ab} K_{ab} - K^2 - \bar{R}) + \frac{1}{\kappa} \sqrt{h} \bar{D}_a [(K^{ab} - Kh^{ab}) N_b] \\
&\quad - \frac{1}{\kappa} \sqrt{h} N_b \bar{D}_a (K^{ab} - Kh^{ab}). \tag{18.115}
\end{aligned}$$

但是, 哈密顿量 H_G 其实是拉格朗日量的勒让德变换, 定义为 $H_G \equiv \int d^3 y \pi^{ab} \dot{h}_{ab} - L_G$, 因此根据 (18.110) 式, 其实还要加上边界积分项的贡献, 最终即有

$$\begin{aligned}
H_G &= \int_{\Sigma_t} d^3 y \mathcal{H}_G - \frac{1}{\kappa} \int_{S_t} d^2 \theta N \sqrt{\sigma} (k - k_0) \\
&= \int_{\Sigma_t} d^3 y (N \mathcal{H} + N_b \mathcal{H}^b) - \frac{1}{\kappa} \int_{S_t} d^2 \theta N \sqrt{\sigma} (k - k_0) \\
&\quad + \frac{1}{\kappa} \int_{S_t} d^3 y \sqrt{h} \bar{D}_a [(K^{ab} - Kh^{ab}) N_b], \tag{18.116}
\end{aligned}$$

式中

$$\mathcal{H} \equiv \frac{1}{2\kappa} \sqrt{h} (K^{ab} K_{ab} - K^2 - \bar{R}) \tag{18.117}$$

$$\begin{aligned}
\mathcal{H}^b &\equiv -\frac{1}{\kappa} \sqrt{h} \bar{D}_a (K^{ab} - Kh^{ab}) \Rightarrow \\
\mathcal{H}_a &= -\frac{1}{\kappa} \sqrt{h} (\bar{D}_b K^b_a - \bar{D}_a K^b_b). \tag{18.118}
\end{aligned}$$

另一方面, 利用高斯定理, 有

$$\frac{1}{\kappa} \int_{\Sigma} d^3 y \sqrt{h} \bar{D}_a [(K^{ab} - Kh^{ab}) N_b] = \frac{1}{\kappa} \int_{S_t} d^2 \theta \sqrt{\sigma} r_a [(K^{ab} - Kh^{ab}) N_b].$$

代入上面的 (18.116) 式, 即有

$$\begin{aligned}
H_G &= \int_{\Sigma_t} d^3 y (N \mathcal{H} + N_b \mathcal{H}^b) + \\
&\quad - \frac{1}{\kappa} \int_{S_t} d^2 \theta \sqrt{\sigma} [N(k - k_0) - r_a N_b (K^{ab} - Kh^{ab})]. \tag{18.119}
\end{aligned}$$

常常也用共轲动量密度 π^{ab} 来把 \mathcal{H} 重新表达成

$$\mathcal{H} = \frac{2\kappa}{\sqrt{h}} (\pi^{ab} \pi_{ab} - \frac{1}{2} \pi^2) - \frac{\sqrt{h}}{2\kappa} \bar{R}. \tag{18.120}$$

式中 $\pi = h_{ab} \pi^{ab}$. 引入如下所谓的 de Witt 度规

$$G_{abcd} \equiv \frac{1}{2} (h_{ac} h_{bd} + h_{ad} h_{bc} - h_{ab} h_{cd}). \tag{18.121}$$

进而即可以进一步将 \mathcal{H} 重写成

$$\mathcal{H} = \frac{2\kappa}{\sqrt{h}} G_{abcd} \pi^{ab} \pi^{cd} - \frac{\sqrt{h}}{2\kappa} \bar{R}. \quad (18.122)$$

另外, 利用共轭动量密度 π^{ab} 也可以把 \mathcal{H}^b 重新表达成

$$\mathcal{H}^b = -2\sqrt{h} \bar{D}_a \left(\frac{1}{\sqrt{h}} \pi^{ab} \right). \quad (18.123)$$

注意, 在拉格朗日体密度 \mathcal{L}_G 中, 并未出现 N 和 N^a 的时间导数, 这意味着 N 和 N_a 是拉格朗日乘子, 它们的共轭动量密度为零, 因此根据哈密顿正则方程, 即有如下约束方程

$$0 = \frac{\delta H_G}{\delta N} = \mathcal{H}, \quad 0 = \frac{\delta H_G}{\delta N^a} = \mathcal{H}_a. \quad (18.124)$$

利用上面关于 \mathcal{H} 和 \mathcal{H}_a 的定义式 (18.117)、(18.118), 即有

$$\bar{R} + K^2 - K^{ab} K_{ab} = 0, \quad \bar{D}_b K^b_a - \bar{D}_a K^b_b = 0. \quad (18.125)$$

这与前面的结果 (18.54) 在忽略物质场时一致。如果进一步考虑物质场, 那么就要在 \mathcal{H} 和 \mathcal{H}_a 中添加上物质场的贡献, 最终得到的约束方程依然是 (18.54) 式。其中 $\mathcal{H} = 0$ 称作哈密顿约束, 而 $\mathcal{H}_a = 0$ 称作动量约束, 两种约束加起来一共四个分量约束方程, 正好对应四个独立的微分同胚变换, 实际上, 微分同胚不变性就是这四个分量约束方程的起源。

进一步当然还可以根据 h_{ab} 和 π^{ab} 的哈密顿正则方程 (如下) 推导它们所满足的运动方程

$$\dot{h}_{ab} = \frac{\delta H_G}{\delta \pi^{ab}}, \quad \dot{\pi}^{ab} = -\frac{\delta H_G}{\delta h_{ab}}, \quad (18.126)$$

只不过具体推导 H_G 对 h_{ab} 和 π^{ab} 的变分有点繁琐, 这里从略。特别的, 给定 h_{ab} 和 π^{ab} 在某个初始类空超曲面 Σ_{t_0} 的初始值, 我们就可以通过求解它们的哈密顿正则方程, 得到它们在此后任意时刻的 Σ_t 上的值, 而且这种求解还可以通过数值方法进行, 这在时空动力学演化的数值模拟中很有用。

很显然, 对于满足哈密顿约束和动量约束的物理场位形, 相应的哈密顿量 H_G 的体密度贡献其实等于零, 只有边界积分有贡献, 有时候也把这个边界贡献记作 H_{ADM}

$$H_{\text{ADM}} = -\frac{1}{\kappa} \int_{S_t} d^2\theta \sqrt{\sigma} \left[N(k - k_0) - r_a N_b (K^{ab} - Kh^{ab}) \right]. \quad (18.127)$$

对于物理场位形, $H_G = H_{\text{ADM}}$.

18.3.4 ADM 能量

本小节讨论如何定义渐近平坦时空的总能量, 尤其是我们要把它和 H_{ADM} 联系起来。

前面说过, 对于物理的引力场位形, 其哈密顿量由 H_{ADM} 给出。但是, H_{ADM} 的定义依赖于类空超曲面 Σ_t 在很远的渐近区域的行为, 同时也依赖于时移函数 N 以及位移矢量 N^a 的渐近行为。虽然 N 和 N^a 总可以任意选, 但时空渐近平坦的事实会给出 Σ_t 渐近行为

的合适要求。具体来说，我们要求 Σ_t 渐近地与闵可夫斯基时空的某个等时面重合：假如 $(\tilde{t}, \tilde{x}, \tilde{y}, \tilde{z})$ 是无穷远处的一个洛伦兹参考系，则 Σ_t 必须渐近地与 $\tilde{t} = \text{constant}$ 的超曲面重合。因此， Σ_t 上任意选取的坐标 y^a 会渐近地与闵可夫斯基空间坐标联系起来，即有渐近关系 $y^a \rightarrow y^a(\tilde{x}, \tilde{y}, \tilde{z})$ 。类似的，也有渐近关系 $x^\mu \rightarrow x^\mu(\tilde{t}, \tilde{x}, \tilde{y}, \tilde{z})$ 。

注意到 \tilde{t} 为渐近区域的静止观察者的固有时，此观察者的协变速度为 $u^\mu = \frac{\partial x^\mu}{\partial \tilde{t}}$ ，由于 u^μ 是归一的，且与 $\tilde{t} = \text{constant}$ 的超曲面正交，因此它必定就是渐近区域的 n^μ ，换言之，我们有渐近关系 $n^\mu \rightarrow \frac{\partial x^\mu}{\partial \tilde{t}}$ 。又注意到 $t^\mu = Nn^\mu + N^a e^\mu_a$ ，所以必然也有如下渐近关系

$$t^\mu \rightarrow N \left(\frac{\partial x^\mu}{\partial \tilde{t}} \right)_{y^a} + N^a \left(\frac{\partial x^\mu}{\partial y^a} \right)_{\tilde{t}} \quad (18.128)$$

也即是说，按照我们对 Σ_t 渐近行为的要求，选取渐近区域的 N 、 N^a 与选取相应区域的 t^μ 是一回事。特别的， H_{ADM} 可以看成是依赖于渐近区域的 t^μ 。

为了定义渐近平坦时空的总能量，我们取 S_t 为空间无穷远处的 2 维球面，记作 S_t^∞ ，同时，我们在无穷远处选取 $N = 1$ 以及 $N^a = 0$ 。我们把总能量 E_{ADM} 定义成如此计算出来的 H_{ADM} ，即

$$E_{\text{ADM}} \equiv -\frac{1}{\kappa} \int_{S_t^\infty} d^2\theta \sqrt{\sigma} (k - k_0). \quad (18.129)$$

之所以说 E_{ADM} 是总能量，原因如下：首先 $N = 1, N^a = 0$ 相应于 $t^\mu \rightarrow \frac{\partial x^\mu}{\partial \tilde{t}}$ ，也即是说， t^μ 渐近于无穷远处的一个时间平移生成元。 E_{ADM} 正是这个时间平移生成元对应的引力哈密顿量，从而就是时空的总能量。

类似的， H_{ADM} 也可以提供时空总角动量与渐近旋转之间的联系。只要我们选取 $t^\mu \rightarrow \phi^\mu \equiv \frac{\partial x^\mu}{\partial \phi}$ ，这里 ϕ 是渐近区域相对于笛卡尔坐标 $(\tilde{x}, \tilde{y}, \tilde{z})$ 的一个旋转角度。当然，这种选取相应于 $N = 0, N^a \equiv \phi^a = \frac{\partial y^a}{\partial \phi}$ 。也即是说，时空的总角动量 J_{ADM} 可以定义成

$$J_{\text{ADM}} \equiv -\frac{1}{\kappa} \int_{S_t^\infty} d^2\theta \sqrt{\sigma} \left[r^a \phi^b (K_{ab} - K h_{ab}) \right], \quad (18.130)$$

相对于原来 H_{ADM} 的公式，我们在式中多加了一个额外的负号，这是为了使得计算出来的角动量的正负与通常约定吻合。值得注意的是，总角动量的定义是相对于某个转轴而言的， ϕ 就是相对于这个转轴的转角。

从上面的定义可以看出，渐近平坦时空的总能量和总角动量只依赖于引力场在空间无穷远处的渐近行为，这是一个相当奇妙的性质，因为一般来说，能量总量这样的东西应该是某个能量体密度的体积积分，应该是依赖于整个空间体的。不过，在电动力学中我们也见过类似的现象，即空间的电荷总量只依赖于空间无穷远处的电通量。

在渐近平坦时空中，ADM 能量（也包括 ADM 角动量）是时间无关的守恒量。这是因为：首先，时空的渐近对称性（如时间平移不变性）导致能量守恒。其次，为了保证时空渐近平坦，引力辐射在无穷远处的能量流需满足特定衰减条件，从而对 ADM 能量无贡献。

值得说明的是，这里给出的 ADM 能量的公式和通常所见的 ADM 公式并不相同，推导方式当然也不同。但是，可以证明，两个不同的公式结果其实是一致的，证明可以参见 S. W. Hawking, Gary T. Horowitz, The Gravitational Hamiltonian, Action, Entropy and Surface Terms, arXiv: gr-qc/9501014.

正质量定理 (Schoen-Yau-Witten 定理)

既然 ADM 能量 E_{ADM} 为渐近平坦时空的总能量, 一个重要的问题是, E_{ADM} 是否有下界? 甚至是否有 $E_{\text{ADM}} \geq 0$? 这个问题之所以重要, 是因为如果总能量没有下界, 那么任何时空都将不稳定, 会无限地往能量更低的时空衰变。

好在 Schoen 和 Yau 给出了这个问题的肯定回答 (Yau 就是丘成桐)。Schoen-Yau 证明, 对于物理上合理的物质场 (满足主能量条件), 总有

$$E_{\text{ADM}} \geq 0, \quad (18.131)$$

且 $E_{\text{ADM}} = 0$ 当且仅当时空全局为平坦的闵可夫斯基时空。这表明, 引力系统的总能量非负, 且“无引力”的真空即为能量零点, 所以通常称这个定理为正质量定理 (质量即是能量)。后来 Witten 又给出了这个定理的更优雅而且大大简化的证明, 所以也称之为 Schoen-Yau-Witten 定理。

轴对称稳态时空的能量和角动量

下面我们考察一个具体的渐近时空, 并计算其 ADM 能量和角动量。我们考察的时空就是轴对称稳态时空。

正如第十三章 (讲述克尔黑洞的章节) 中提到过的, 轴对称的稳态时空在空间无穷远 ($r \rightarrow +\infty$) 处度规的渐近形式为

$$\begin{aligned} ds^2 \rightarrow & -\left(1 - \frac{2GM}{r}\right)dt^2 + \left(1 + \frac{2GM}{r}\right)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \\ & - \frac{4GJ}{r} \sin^2\theta dt d\phi. \end{aligned} \quad (18.132)$$

通过重新定义 r , 具体来说就是将 r 替换成 $r + GM$, 我们也可以将上面的渐近度规改写成 (注意 $r \rightarrow +\infty$)

$$\begin{aligned} ds^2 \rightarrow & -\left(1 - \frac{2GM}{r}\right)dt^2 + \left(1 + \frac{2GM}{r}\right)(dr^2 + r^2 d\Omega^2) \\ & - \frac{4GJ}{r} \sin^2\theta dt d\phi. \end{aligned} \quad (18.133)$$

式中 $d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2$ 。在第十三章中我们提到过, 上面的 M 是整个时空的总能量, J 是总角动量, 现在我们就通过具体计算来确定这一点。

首先, 选取 Σ_t 为 t 取常数值 of 的超曲面, 不难看出上面的这个渐近度规 (18.133) 符合我们对 Σ_t 的要求。显然 $n_\mu = -(1 - \frac{GM}{r})\partial_\mu t$ (注意到 r 很大, 并且我们的计算仅仅保留到 GM/r 的一阶, 下同)。很显然, Σ_t 上的诱导度规为

$$h_{ab} dy^a dy^b = \left(1 + \frac{2GM}{r}\right)(dr^2 + r^2 d\Omega^2). \quad (18.134)$$

取空间边界 S_t 为 $r = R$ 处的两维球面, 最终我们要取 $R \rightarrow \infty$ 的极限。不难看出, S_t 的法向量 r_a 为 $r_a = (1 + \frac{GM}{r})\partial_a r$, 并且 S_t 上的诱导度规为

$$\sigma_{AB} d\theta^A d\theta^B = \left(1 + \frac{2GM}{R}\right)R^2 d\Omega^2. \quad (18.135)$$

为了计算 ADM 能量 M , 我们需要计算 $k = \bar{D}_a r^a$, 计算的结果是 $k = 2(1 - 2GM/R)/R$. 另外还需要从 k 中减除掉 k_0 , k_0 是同一个 S_t 嵌入平坦空间的外曲率。记 S_t 在平坦空间的诱导度规为

$$\sigma_{AB}^0 d\theta^A d\theta^B = R'^2 d\Omega^2, \quad (18.136)$$

很显然, 为了保证 $\sigma_{AB}^0 = \sigma_{AB}$, 需要取 $R' = R(1 + GM/R)$. 进而可算得 $k_0 = 2/R' = 2(1 - GM/R)/R$. 所以 $k - k_0 = -2GM/R^2$. 另一方面 $\sqrt{\sigma} d^2\theta = R^2(1 + 2GM/R) \sin\theta d\theta d\phi$. 将这些结果代入公式 (18.140), 即得

$$E_{\text{ADM}} = M. \quad (18.137)$$

正如我们所期望的。

为了计算 ADM 角动量 J_{ADM} , 我们注意到 $K_{ab} r^a \phi^b = K_b^a r_a \phi^b = K_b^a (1 + GM/r) \frac{\partial r}{\partial y^a} \frac{\partial y^b}{\partial \phi} = (1 + GM/r) K_{\phi}^r = (1 + GM/r) g^{rr} K_{r\phi} = (1 - GM/r) K_{r\phi}$, 利用定义 $K_{ab} = e_a^\mu e_b^\nu D_\mu n_\nu$ 可以算得 $K_{r\phi} = (1 - GM/r) \Gamma_{r\phi}^t$. 利用

$$g^{tt} = -(1 + 2GM/r), \quad g^{t\phi} = -2GJ/r^3. \quad (18.138)$$

进而可以算得 $\Gamma_{r\phi}^t = -3GJ \sin^2\theta / r^2$, 进而即有 $K_{r\phi} = -3GJ \sin^2\theta / R^2$. 另一方面, 公式 (18.130) 中的第二项 ($K_{ab} r^a \phi^b$ 项) 其实等于零, 因为 r^a 和 ϕ^a 是正交的。将这些结果代入公式 (18.130), 即得 $J_{\text{ADM}} = (3J/4) \int_0^\pi \sin^3\theta d\theta = J$, 即

$$J_{\text{ADM}} = J, \quad (18.139)$$

正是我们期待的结果。

引力能量的非定域性

ADM 能量定义的是整个时空的总能量, 而且它只取决于度规的渐近行为, 也就是说, 这样定义出来的能量是非定域的, 即不是由一个能量密度进行体积积分给出的。除了 ADM 能量之外, 关于引力场全局能量的定义还有 Bondi 能量等。

那么, 可不可能定义引力场的能量密度呢? 前面引力波的章节中好像我们给出过引力场的能量密度 (能动张量), 但是, 那里是把时空背景取作平坦时空, 把引力场看作时空背景上的微扰场, 而不是采用更一般的引力场对应于时空弯曲的观点, 尤其是, 那里我们并没有用到等效原理。那么, 一般来说, 对于非微扰的引力场, 可不可以定义它的局域能量密度呢? 或者更广泛地说, 可不可以定义引力场本身的能动张量呢?

类比于电磁场的话, 我们知道, 电磁场的能动张量取决于场强的平方, 因此人们可能同样猜想, 引力场的能动张量也取决于场强的平方, 但问题是, 引力场的场强由克里斯托夫联络来刻画, 而根据等效原理, 我们总可以取局域惯性系, 在局域惯性系中, 克里斯托夫联络为零, 也就是引力场强为零, 如此一来就会导致引力场的能动张量恒等于零! 但是, 引力场当然是有能量的, 所以这样想的话就会自相矛盾。

问题可能还是出在引力能量的非定域性，也即是说，引力场有能量，但是这能量并不由局域的能量密度来刻画。换言之，可能不存在引力场能动张量的一个合理定义！当然，人们也作过一些尝试，想用某种赝能动张量来定义引力场的能量密度，比如朗道和栗弗席兹的《场论》中就提出过一个这样的赝张量。但总体来说，这些尝试都不太成功。

另一方面，人们也尝试过定义引力场的准局域能量，即通过一个包围有限空间区域 Σ 的两维表面 $\partial\Sigma$ 来定义能量，这避免了 ADM 能量定义中需要取无限远处边界条件的问题。这种准局域能量的典型例子是所谓的布朗-约克能量，其定义和 ADM 能量很像，为

$$E_{\text{BY}}(\Sigma) = -\frac{1}{\kappa} \int_{\partial\Sigma} d^2\theta \sqrt{\sigma} (k - k_0). \quad (18.140)$$

另外，在 AdS/CFT 对应的框架中，引力场的能量可以全息地编码在边界共形场论的能动张量中，但这一方法仅仅限制于渐近 AdS 时空，而且它定义的也是一种全局能量，因为整个时空的能量还是由远处的渐近边界给出，只是这时候由于是渐近 AdS 而不是渐近平坦，所以有一个类时的共形边界而已。

诺特技巧与对称的能动张量

任何一个定义在四维闵可夫斯基空间的场论系统都有一个守恒的能动张量。根据著名的诺特定理，任何一个连续的对称性都必然有一个相应的守恒流四矢量，能量-动量张量就是与时空平移对称性相对应的守恒流，由于时空平移本身是一个四维矢量，所以它对应的守恒流就是一个二阶张量。

但是，这样定义出来的能动张量(称作正则能动张量)一般来说并不是一个对称张量。有些人可能会认为，简单地用诺特定理无法得出一个对称的能动张量。本文就是要表明，情况并非如此，我们将表明，诺特技巧的一个巧妙延伸可以让我们很自然地得出任何场论系统的对称能动张量。

不仅如此，我们还可以得出这个对称能动张量与正则能动张量之间的显式关系，实际上，这个关系与 Weinberg 量子场论第一卷 7.4 节中给出的公式完全一致，当然推导方法并不相同，我们这里的推导主要是强调诺特技巧。有些时候人们也称这种对称的能动张量为 Belinfante(贝林凡特)张量。

.1 诺特技巧与能动张量

.1.1 时空平移与正则能动张量

为了简单起见，我们首先讲述标量场系统的能量-动量张量。

假设有一个标量场 $\phi(x)$ ，其拉格朗日密度可以写成 $\mathcal{L}(\phi, \partial_\mu \phi)$ ，特别的，这个拉格朗日密度不显含时空坐标 x 。现在，假定将整个场论系统进行一个时空平移，使得 x 点的场平移到 x' 点，

$$x^\mu \rightarrow x'^\mu = x^\mu + a^\mu, \quad (141)$$

其中 a^μ 为某常数四矢量。记平移之后的场为 $\phi'(x)$ ，很显然 ϕ' 在 x' 点的场值来自于平移之前 ϕ 在 x 点的场值，即

$$\phi'(x') = \phi(x). \quad (142)$$

很容易验证作用量 $S[\phi] = \int_{\mathcal{D}} d^4x \mathcal{L}(\phi, \partial_\mu \phi)$ 在此平移之下保持不变，具体验证过程如下

$$\begin{aligned} S[\phi'] &= \int_{\mathcal{D}'} d^4x \mathcal{L}(\phi'(x), \partial_\mu \phi'(x)) \\ &= \int_{\mathcal{D}'} d^4x' \mathcal{L}(\phi'(x'), \partial'_\mu \phi'(x')) = \int_{\mathcal{D}'} d^4x' \mathcal{L}(\phi(x), \partial'_\mu \phi(x)) \\ &= \int_{\mathcal{D}} d^4x \mathcal{L}(\phi(x), \partial_\mu \phi(x)) = S[\phi] \end{aligned} \quad (143)$$

式中第二个等号只是将 x 变量重写成了 x' , 倒数第二个等号是用了 $d^4x' = d^4x$ 以及 $\partial'_\mu = \partial_\mu$ 。上述推导过程中, 为了明晰起见, 假定了只考察一个特定的时空区域 \mathcal{D} , 也即是说, 作用量的时空积分是在 \mathcal{D} 上积的。并记平移操作把时空局域 \mathcal{D} 变到了 \mathcal{D}' , 即

$$\mathcal{D} \rightarrow \mathcal{D}'. \quad (144)$$

为了通过诺特定理引入相应的守恒流, 下面考察无穷小时空平移, 即将 a^μ 取成无穷小量 ε^μ 。进一步, 我们使用一个**关键技巧** (诺特技巧), 即将 ε^μ 变成依赖于时空坐标 x 的无穷小量 $\varepsilon^\mu(x)$, 即考察如下无穷小时空平移

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x). \quad (145)$$

当然我们依然有

$$\phi'(x') = \phi(x). \quad (146)$$

但是, 在这种依赖于时空点 x 的局域平移之下, 作用量当然无法保持不变, 因为这种局域平移根本不是系统的对称性。

很显然, 在一阶近似上有

$$\frac{\partial x'^\mu}{\partial x^\nu} = \delta_\nu^\mu + \partial_\nu \varepsilon^\mu, \quad \frac{\partial x^\nu}{\partial x'^\mu} = \delta_\mu^\nu - \partial_\mu \varepsilon^\nu. \quad (147)$$

另外, 记坐标变换的雅可比行列式为 $\det(\frac{\partial x'}{\partial x})$, 利用矩阵恒等式 $\det(A) = e^{\text{Tr}(\ln A)}$, 易得在一阶近似上有

$$\det\left(\frac{\partial x'}{\partial x}\right) = 1 + \partial_\mu \varepsilon^\mu(x). \quad (148)$$

下面我们来计算在局域平移 (145) 之前和之后, 系统作用量的改变量。平移以后的作用量 $S[\phi'(x)]$ 为

$$\begin{aligned} S[\phi'(x)] &= \int d^4x \mathcal{L}(\phi'(x), \partial_\nu \phi'(x)) \\ &= \int d^4x' \mathcal{L}(\phi'(x'), \partial'_\mu \phi'(x')) = \int d^4x' \mathcal{L}(\phi(x), \partial'_\mu \phi(x)) \\ &= \int d^4x \det\left(\frac{\partial x'}{\partial x}\right) \mathcal{L}(\phi(x), \frac{\partial x^\nu}{\partial x'^\mu} \partial_\nu \phi(x)) \\ &= \int d^4x (1 + \partial_\mu \varepsilon^\mu(x)) \mathcal{L}(\phi(x), \partial_\mu \phi(x) - \partial_\mu \varepsilon^\nu \partial_\nu \phi(x)) \\ &= \int d^4x \left[-\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi(x) + \delta_\nu^\mu \mathcal{L}(\phi(x), \partial_\rho \phi(x)) \right] \partial_\mu \varepsilon^\nu \\ &\quad + \int d^4x \mathcal{L}(\phi(x), \partial_\nu \phi(x)). \end{aligned} \quad (149)$$

由上面的推导易知, 变换前后作用量的改变量为

$$\begin{aligned} \delta S &= S[\phi'(x)] - S[\phi(x)] \\ &= - \int d^4x \left[\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial_\nu \phi(x) - \delta_\nu^\mu \mathcal{L} \right] \partial_\mu \varepsilon^\nu \\ &= \int d^4x T^\mu{}_\nu \partial_\mu \varepsilon^\nu = \int d^4x T_{\mu\nu} \partial^\mu \varepsilon^\nu. \end{aligned} \quad (150)$$

式中

$$T^{\mu\nu} = - \left[\frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \partial^\nu \phi(x) - \eta^{\mu\nu} \mathcal{L} \right]. \quad (151)$$

其中 $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$ 是闵可夫斯基时空的度规。

注意 (150) 式, 当 ϵ^ν 是一个不依赖于时空点的整体无穷小平移时, 即有 $\delta S = 0$, 这当然是因为这种整体时空平移是系统的对称性。正因为如此, 不仅对于标量场系统可以找到这样一个 $T^{\mu\nu}$, 实际上对闵可夫斯基时空的任何物质系统都能找到一个类似的 $T^{\mu\nu}$, 因为任何物质系统的作用量在局域平移 (145) 前后的改变量都必定能写成如下形式

$$\delta S = \int d^4x T_{\mu\nu} \partial^\mu \epsilon^\nu. \quad (152)$$

而这又是因为, 当将 ϵ^ν 取成常矢量 (即考察无穷小整体时空平移时) 时, 由于时空平移对称性, 必有 $\delta S = 0$, 所以在局域化平移情形, δS 只能依赖于 ϵ^ν 的偏导 $\partial^\mu \epsilon^\nu$, 从而 δS 必定能写成 (152) 式的形式!

下面进一步假定 $\epsilon^\nu(x)$ 在时空区域的边界 $\partial \mathcal{D}$ 上趋于零。另外, 上面的 (152) 式对于任意场位形都成立, 下面我们考察真实的满足场运动微分方程的场位形, 那这时候由于这些场位形满足最小作用量原理, 当然就有 $\delta S = 0$ 对于任意的在区域边界 $\partial \mathcal{D}$ 上趋于零的场变分都成立, 那当然也对局域时空平移 (145) 引起的场改变成立。即是说, 对于真实场位形, 必有

$$0 = \delta S = \int_{\mathcal{D}} d^4x T_{\mu\nu} \partial^\mu \epsilon^\nu. \quad (153)$$

将上式分部积分, 即有

$$0 = \int_{\mathcal{D}} d^4x T_{\mu\nu} \partial^\mu \epsilon^\nu = - \int_{\mathcal{D}} d^4x \partial^\mu T_{\mu\nu} \epsilon^\nu(x). \quad (154)$$

由于 $\epsilon^\nu(x)$ 为任意无穷小函数, 从而即有

$$\partial^\mu T_{\mu\nu} = 0. \quad (155)$$

即是说, $T_{\mu\nu}$ 就是与时空平移对称性相应的守恒流! 特别的, 如果取 $\epsilon^\mu = \epsilon \delta^{\mu 0}$, 即假设考察的是时间平移, 那相应的守恒流 $T^{\mu 0}$ 当然就是能量流四矢量, 换言之, T^{00} 必定为能量密度, $T^{0i} (i=1, 2, 3)$ 为能量流密度。而如果取 $\epsilon^\mu = \epsilon \delta^{\mu j}$ 为空间平移, 那相应的守恒流 $T^{\mu j}$ 当然就是动量流四矢量, 即是说 T^{0j} 必为动量密度, 而 T^{ij} 为动量流密度。将所有这些合起来, 我们就称 $T_{\mu\nu}$ 为能量动量张量! 上面的讨论也说明, 闵可夫斯基时空的任何场论系统都存在这么一个守恒的能量动量张量。

比如, 假设我们考虑 $\mathcal{L} = -\frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \mathcal{U}(\phi)$ 的场论模型, 则容易算得,

$$\begin{aligned} T^{\mu\nu} &= \partial^\mu \phi \partial^\nu \phi + \eta^{\mu\nu} \mathcal{L} \\ &= \partial^\mu \phi \partial^\nu \phi - \frac{1}{2} \eta^{\mu\nu} \partial_\rho \phi \partial^\rho \phi - \eta^{\mu\nu} \mathcal{U}(\phi). \end{aligned} \quad (156)$$

很明显, 这个 $T^{\mu\nu}$ 的两个指标是对称的。实际上, 可以证明, 对于任何洛伦兹不变的标量场论, 其 $T^{\mu\nu}$ 都必定是对称张量。但是当我们的考察范围超出标量场论时, 其按照上面类似的办法求出来的 $T^{\mu\nu}$ 就不一定为对称张量了, 比方说对于矢量场, 它的 $T^{\mu\nu}$ 就不是对称张量。不过, $T^{\mu\nu}$ 的流守恒方程 $\partial_\mu T^{\mu\nu} = 0$ 告诉我们, $T^{\mu\nu}$ 的定义不是唯一的, 实际上人们很容易看出, 对于任何 $\Lambda^{\rho\mu\nu}$, 只要 $\Lambda^{\rho\mu\nu} = -\Lambda^{\mu\rho\nu}$, 则 $T^{\mu\nu} + \partial_\rho \Lambda^{\rho\mu\nu}$ 同样满足流守恒方程, 因此可以定义为新的能动量张量。从而只要我们合适地选取 $\Lambda^{\rho\mu\nu}$, 我们总可以让重新定义以后的能动量张量为一个对称张量。实际上, 马上我们就会给出一个找到这样一个对称能量动量张量的巧妙办法。

.1.2 洛伦兹对称性与对称能动张量

我们考察的经典场论都是具有洛伦兹不变性的场论, 即拉格朗日密度为洛伦兹标量的理论, 具体来说即是拉氏密度在如下坐标变换下保持不变的理论,

$$x^\mu \rightarrow x'^\mu = \Lambda^\mu_\nu x^\nu, \quad (157)$$

Λ^μ_ν 就是所谓的洛伦兹变换。那么, 洛伦兹变换对应的守恒流是什么呢?

为此我们需要考察无穷小洛伦兹变换, 即取

$$\Lambda^\mu_\nu = \delta^\mu_\nu + \varepsilon^\mu_\nu, \quad (158)$$

式中 ε^μ_ν 为无穷小量。很显然, 无穷小洛伦兹变换由于可以和恒等变换连续过渡, 从而必定是正洛伦兹变换, 即满足¹

$$\det(\Lambda) = 1 \Rightarrow \varepsilon^\mu_\mu = 0. \quad (159)$$

进一步, 利用洛伦兹变换的定义 $\eta_{\mu\nu} \Lambda^\mu_\alpha \Lambda^\nu_\beta = \eta_{\alpha\beta}$, 易得

$$\eta_{\mu\nu} (\delta^\mu_\alpha + \varepsilon^\mu_\alpha) (\delta^\nu_\beta + \varepsilon^\nu_\beta) = \eta_{\alpha\beta} \Rightarrow \varepsilon_{\alpha\beta} + \varepsilon_{\beta\alpha} = 0. \quad (160)$$

即 $\varepsilon^{\mu\nu}$ 是一个二阶反对称张量。

注意到无穷小洛伦兹变换为 $x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu_\nu x^\nu$ 。为了利用诺特定理, 我们使用诺特技巧, 即将无穷小参数 ε^μ_ν 变成依赖于时空坐标的 $\varepsilon^\mu_\nu(x)$, 进而考察如下无穷小局域时空坐标变换,

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x), \quad \text{这里 } \varepsilon^\mu(x) = \varepsilon^\mu_\nu(x) x^\nu. \quad (161)$$

则完全类似于前面对 (152) 式的论证, 可知在此时空变换之下作用量的改变量必定可以写成

$$\begin{aligned} \delta S &= \int d^4x T^\mu_\nu \partial_\mu \varepsilon^\nu(x) \\ &= \int d^4x T^\mu_\nu (\varepsilon^\nu_\mu + x^\rho \partial_\mu \varepsilon^\nu_\rho). \end{aligned} \quad (162)$$

¹利用矩阵恒等式 $\det(A) = e^{\text{Tr}(\ln A)}$.

(注意, 对于比方说矢量场, 它的各指标分量也要按照无穷小局域洛伦兹变换 $\Lambda^\mu_\nu(x) = \delta^\mu_\nu + \varepsilon^\mu_\nu(x)$ 的规则变, 这是不同于前面时空平移情形的。在时空平移情形, 矢量场的分量指标并不变换。因此上式中的 $T_{\mu\nu}$ 一般不同于前面按照时空平移直接求出来的 $T_{\mu\nu}$ 。) 注意到 $\varepsilon_{\mu\nu}$ 关于指标反对称, 从而有

$$\begin{aligned} \delta S &= \frac{1}{2} \int d^4x \left[(T^{\mu\nu} - T^{\nu\mu}) \varepsilon_{\nu\mu} + (x^\rho T^{\mu\nu} - x^\nu T^{\mu\rho}) \partial_\mu \varepsilon_{\nu\rho} \right] \\ &= -\frac{1}{2} \int d^4x \left[(T^{\mu\nu} - T^{\nu\mu}) \varepsilon_{\mu\nu} \right] - \frac{1}{2} \int d^4x \left[(x^\mu T^{\rho\nu} - x^\nu T^{\rho\mu}) \partial_\rho \varepsilon_{\mu\nu} \right]. \end{aligned} \quad (163)$$

如果我们将 $\varepsilon_{\mu\nu}$ 取回常数, 则 $\partial_\rho \varepsilon_{\mu\nu} = 0$, 从而上式最后一行只剩下前面那项, 但是洛伦兹不变性告诉我们, 当 $\varepsilon_{\mu\nu}$ 为常数时作用量应该不变, 即这时候必有 $\delta S = 0$, 由此可知

$$T^{\mu\nu} - T^{\nu\mu} = 0, \quad (164)$$

即这样求出来的能动量张量必定是对称张量! 这就是我们给出的如何寻找对称能量动量张量的办法。

在 (163) 式中代入 $T^{\mu\nu} = T^{\nu\mu}$, 即有

$$\delta S = -\frac{1}{2} \int d^4x \left[(x^\mu T^{\rho\nu} - x^\nu T^{\rho\mu}) \partial_\rho \varepsilon_{\mu\nu} \right]. \quad (165)$$

完全类似于前面导出能动量张量守恒的讨论, 这个结果意味着 $M^{\rho\mu\nu} = x^\mu T^{\rho\nu} - x^\nu T^{\rho\mu}$ 为守恒流, 满足守恒方程

$$\partial_\rho M^{\rho\mu\nu} = 0. \quad (166)$$

综合可知, 在无穷小局域洛伦兹变换之下

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x), \quad \varepsilon^\mu(x) = \varepsilon^\mu_\nu(x) x^\nu, \quad (167)$$

任何场论系统作用量的改变量必定可以写成

$$\delta S = \int d^4x T^{\mu\nu} \partial_\mu \varepsilon_\nu(x) = \frac{1}{2} \int d^4x T^{\mu\nu} (\partial_\mu \varepsilon_\nu + \partial_\nu \varepsilon_\mu). \quad (168)$$

其中我们利用了 $T^{\mu\nu}$ 是一个对称张量。

.2 诺特技巧推导对称能动张量

本节就是要将上一节讨论的技巧具体用于推导一个洛伦兹不变的任意场论系统的对称能动张量。

.2.1 任意场

为了分析洛伦兹不变的任意场系统, 考虑一个无穷小洛伦兹变换

$$x^\mu \rightarrow x'^\mu = \Lambda^\mu_\nu x^\nu, \quad \Lambda^\mu_\nu = \delta^\mu_\nu + \varepsilon^\mu_\nu. \quad (169)$$

洛伦兹不变的任意场都要构成这一洛伦兹变换下的群表示, 对于一个 N 维表示, 不妨一般地将这场记作 N 维列矢量 $\psi^l(x)$, 它在上述洛伦兹变换之下变换为 $\psi'^l(x)$, 假设它满足

$$\psi'^l(x') = [\exp(\frac{1}{2}\varepsilon^{\mu\nu}\Sigma_{\mu\nu})]^l_m \psi^m(x). \quad (170)$$

式中 $\Sigma_{\mu\nu}$ 为 $N \times N$ 的矩阵 (同时关于时空指标 $\mu\nu$ 反对称), 不依赖于时空坐标, 实际上它们就是洛伦兹群表示的生成元。换言之, 在上述无穷小洛伦兹变换之下

$$\bar{\delta}\psi^l(x) \equiv \psi'^l(x') - \psi^l(x) = \frac{1}{2}\varepsilon^{\mu\nu}[\Sigma_{\mu\nu}]^l_m \psi^m(x). \quad (171)$$

用场变量 $\psi^l(x)$ 我们可以构造出一个具有洛伦兹不变性的任意场系统, 它的拉格朗日密度为 $\mathcal{L}(\psi, \partial_\mu\psi)$, 式中 ψ 是 $\psi^l(x)$ 的简略记法。下面我们要做的, 就是用诺特技巧推导这样一个场论系统的对称能动张量。

.2.2 诺特技巧与对称能动张量

前文我们描述了一种通过诺特技巧得到对称能动张量的办法。为此取如下无穷小局域洛伦兹变换

$$x^\mu \rightarrow x'^\mu = x^\mu + \varepsilon^\mu(x), \quad \varepsilon^\mu(x) = \varepsilon^\mu_\nu(x)x^\nu. \quad (172)$$

式中 $\varepsilon_{\mu\nu}(x)$ 是无穷小的反对称张量。

根据上一小节可知, 在此无穷小局域洛伦兹变换 (172) 之下, 场 $\psi^l(x)$ 在变换前后的改变量为

$$\bar{\delta}\psi^l(x) \equiv \psi'^l(x') - \psi^l(x) = \frac{1}{2}\varepsilon^{\mu\nu}(x)[\Sigma_{\mu\nu}]^l_m \psi^m(x). \quad (173)$$

场的作用量泛函为

$$S[\psi(x)] = \int_{\mathcal{D}} d^4x \mathcal{L}(\psi(x), \partial_\rho\psi(x)), \quad (174)$$

式中为了明晰起见, 假定只考察一个特定的时空区域 \mathcal{D} , 也即是说, 作用量的时空积分是在 \mathcal{D} 上积的。式中 $\psi(x)$ 是 $\psi^l(x)$ 的简记符号 (下文约定相同)。

记无穷小局域时空变换 (172) 把时空局域 \mathcal{D} 变到了 \mathcal{D}' , 即 $\mathcal{D} \rightarrow \mathcal{D}'$, 当然, \mathcal{D}' 与 \mathcal{D} 的差别是无穷小的。则, 无穷小局域时空变换 (172) 之后的作用量为

$$\begin{aligned} S[\psi'(x)] &= \int_{\mathcal{D}'} d^4x \mathcal{L}(\psi'(x), \partial_\rho\psi'(x)) \\ &= \int_{\mathcal{D}'} d^4x' \mathcal{L}(\psi'(x'), \partial'_\rho\psi'(x')), \end{aligned} \quad (175)$$

注意, 现在这个作用量积分是在变换之后的区域 \mathcal{D}' 上积的。

可以对上面的 (175) 式继续进行如下推导

$$\begin{aligned} S[\psi'(x)] &= \int_{\mathcal{D}'} d^4x' \mathcal{L}(\psi'(x'), \partial'_\rho\psi'(x')), \\ &= \int_{\mathcal{D}} d^4x \det\left(\frac{\partial x'}{\partial x}\right) \mathcal{L}(\psi + \bar{\delta}\psi, \frac{\partial x^\sigma}{\partial x'^\rho} \partial_\sigma(\psi + \bar{\delta}\psi)), \end{aligned} \quad (176)$$

注意到 $\det(\frac{\partial x'}{\partial x}) = 1 + \partial_\mu \varepsilon^\mu$, 以及 $\frac{\partial x^\sigma}{\partial x'^\rho} = \delta_\rho^\sigma - \partial_\rho \varepsilon^\sigma$, 代入 (176) 式, 即有

$$\begin{aligned} S[\psi'(x)] &= \int_{\mathcal{D}} d^4x (1 + \partial_\alpha \varepsilon^\alpha) \mathcal{L}(\psi + \bar{\delta}\psi, \partial_\rho(\psi + \bar{\delta}\psi) - \partial_\rho \varepsilon^\sigma \partial_\sigma(\psi + \bar{\delta}\psi)) \\ &= \int_{\mathcal{D}} d^4x (1 + \partial_\alpha \varepsilon^\alpha) \mathcal{L}(\psi + \bar{\delta}\psi, \partial_\rho(\psi + \bar{\delta}\psi) - \partial_\rho \varepsilon^\sigma \partial_\sigma \psi) \end{aligned} \quad (177)$$

式中我们注意到 $\partial_\rho \varepsilon^\sigma \partial_\sigma(\bar{\delta}\psi)$ 为二阶无穷小量, 从而可以略去。

进一步将 (177) 式展开到一阶小量, 即有

$$\begin{aligned} S[\psi'(x)] &= \int_{\mathcal{D}} d^4x \mathcal{L}(\psi, \partial_\rho \psi) + \int_{\mathcal{D}} d^4x \partial_\mu \varepsilon^\mu \mathcal{L}(\psi, \partial_\rho \psi) \\ &\quad - \int_{\mathcal{D}} d^4x \frac{\partial \mathcal{L}}{\partial(\partial_\rho \psi)} \partial_\sigma \psi \partial_\rho \varepsilon^\sigma + \int_{\mathcal{D}} d^4x \bar{\delta}\mathcal{L}, \end{aligned} \quad (178)$$

式中

$$\bar{\delta}\mathcal{L} \equiv \mathcal{L}(\psi + \bar{\delta}\psi, \partial_\rho(\psi + \bar{\delta}\psi)) - \mathcal{L}(\psi, \partial_\rho \psi). \quad (179)$$

注意到 $\int_{\mathcal{D}} d^4x \mathcal{L}(\psi, \partial_\rho \psi) = S[\psi(x)]$, 从而立即有变换前后作用量的改变量为

$$\begin{aligned} \delta S &\equiv S[\psi'(x)] - S[\psi(x)] \\ &= \int_{\mathcal{D}} d^4x \partial_\mu \varepsilon^\mu \mathcal{L}(\psi, \partial_\rho \psi) - \int_{\mathcal{D}} d^4x \frac{\partial \mathcal{L}}{\partial(\partial_\rho \psi)} \partial_\sigma \psi \partial_\rho \varepsilon^\sigma + \int_{\mathcal{D}} d^4x \bar{\delta}\mathcal{L}, \\ &= - \int_{\mathcal{D}} d^4x \left[\frac{\partial \mathcal{L}}{\partial(\partial_\rho \psi')} \partial_\sigma \psi' - \delta_\sigma^\rho \mathcal{L} \right] \partial_\rho \varepsilon^\sigma + \int_{\mathcal{D}} d^4x \bar{\delta}\mathcal{L}. \end{aligned} \quad (180)$$

很显然, 最后的这两项中, 前一项就是标准的正则能动张量的贡献, 不妨记

$$\Theta_{\rho\sigma} \equiv - \left[\frac{\partial \mathcal{L}}{\partial(\partial^\rho \psi')} \partial_\sigma \psi' - \eta_{\rho\sigma} \mathcal{L} \right]. \quad (181)$$

则可以将上述结果简写成

$$\delta S = \int_{\mathcal{D}} d^4x \Theta_{\mu\nu} \partial^\mu \varepsilon^\nu + \int_{\mathcal{D}} d^4x \bar{\delta}\mathcal{L}. \quad (182)$$

后面的 $\bar{\delta}\mathcal{L}$ 项是对正则能动张量 $\Theta_{\mu\nu}$ 的修正, 正如上一节中根据局域洛伦兹变换证明过的, 这一修正将正好使得最终的能动张量为对称张量。

读者可能觉得, 根据最小作用量原理, $\bar{\delta}\mathcal{L}$ 的修正项不应该正好等于零吗? 然而情况并非如此。的确, 和最小作用量原理的推导一样, 经过分部积分, 我们有

$$\int_{\mathcal{D}} d^4x \bar{\delta}\mathcal{L} = \int_{\mathcal{D}} d^4x \left[\frac{\partial \mathcal{L}}{\partial \psi} - \partial_\rho \frac{\partial \mathcal{L}}{\partial(\partial_\rho \psi)} \right] \bar{\delta}\psi + \int_{\mathcal{D}} d^4x \partial_\rho \left[\frac{\partial \mathcal{L}}{\partial(\partial_\rho \psi)} \bar{\delta}\psi \right] \quad (183)$$

等式右边的第一项根据 ψ 场的运动方程 (欧拉-拉格朗日方程) 自动等于零, 第二项是一个全微分项, 它在区域边界 $\partial\mathcal{D}$ 上有贡献, 而且和最小作用量原理的情形不同, 现在这个边界贡献非零。这是由于 $\bar{\delta}\psi$ 中包含的是 ε^μ_ν 项, 而我们现在并没有要求它在区域边界 $\partial\mathcal{D}$ 上取零, 相反, 我们将要求 ε^μ_ν 的偏导在边界 $\partial\mathcal{D}$ 上取零 (即在区域边界上, 所考察的无穷小局域洛伦兹变换趋于一个整体的无穷小洛伦兹变换)。所以, 与最小作用量原理的情形不同, 这里这个全微分项的贡献不是零!

代入前面的 (173) 式, 即可以得到

$$\begin{aligned}\int_{\mathcal{D}} d^4x \bar{\delta} \mathcal{L} &= \int_{\mathcal{D}} d^4x \partial_\rho \left[\frac{\partial \mathcal{L}}{\partial (\partial_\rho \psi^l)} \bar{\delta} \psi^l \right] = \int_{\mathcal{D}} d^4x \partial_\rho \left[\frac{\partial \mathcal{L}}{\partial (\partial_\rho \psi^l)} \frac{1}{2} \varepsilon_{\mu\nu} [\Sigma^{\mu\nu}]^l{}_m \psi^m \right] \\ &= \int_{\mathcal{D}} d^4x \partial_\rho \left[\frac{\partial \mathcal{L}}{\partial (\partial_\rho \psi)} \frac{1}{2} \Sigma^{\mu\nu} \psi \right] \varepsilon_{\mu\nu} + \int_{\mathcal{D}} d^4x \left[\frac{\partial \mathcal{L}}{\partial (\partial_\rho \psi)} \frac{1}{2} \Sigma^{\mu\nu} \psi \right] \partial_\rho \varepsilon_{\mu\nu}\end{aligned}\quad (184)$$

不妨记,

$$Y^{\rho\mu\nu} \equiv \frac{\partial \mathcal{L}}{\partial (\partial_\rho \psi)} \frac{1}{2} \Sigma^{\mu\nu} \psi \equiv \frac{\partial \mathcal{L}}{\partial (\partial_\rho \psi^l)} \frac{1}{2} [\Sigma^{\mu\nu}]^l{}_m \psi^m. \quad (185)$$

将上面这个结果 (184) 和 (182) 式的结果结合, 即有

$$\delta S = \int_{\mathcal{D}} d^4x \left[\Theta^{\mu\nu} \partial_\mu \varepsilon_\nu + \partial_\rho Y^{\rho\mu\nu} \varepsilon_{\mu\nu} + Y^{\rho\mu\nu} \partial_\rho \varepsilon_{\mu\nu} \right]. \quad (186)$$

根据上一节的知识, 为了读出对称的能动张量 $T_{\mu\nu}$, 关键是要把上式凑成如下形式

$$\delta S = \int_{\mathcal{D}} d^4x T^{\mu\nu} \partial_\mu \varepsilon_\nu. \quad (187)$$

结果是, 的确能 \square 出来, 我先写出答案, 然后我们再验证它的确能倒过来给出原来的 (186) 式 (这比直接凑答案容易些)。答案是,

$$\delta S = \int_{\mathcal{D}} d^4x \left[\Theta^{\mu\nu} - \partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) \right] \partial_\mu \varepsilon_\nu. \quad (188)$$

为了验证, 我们需要用到 $\varepsilon_\nu = \varepsilon_{\nu\sigma} x^\sigma$, 进而 $\partial_\mu \varepsilon_\nu = (\partial_\mu \varepsilon_{\nu\sigma}) x^\sigma - \varepsilon_{\mu\nu}$, 并注意到 $Y^{\rho\mu\nu}$ 这样的表达式关于后两个指标是反对称的。

为了验证 (188) 式, 显然暂时可以不管 $\Theta^{\mu\nu}$

$$\begin{aligned}& \int_{\mathcal{D}} d^4x \left[-\partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) \right] \partial_\mu \varepsilon_\nu \\ &= \int_{\mathcal{D}} d^4x \left[-\partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) \right] ((\partial_\mu \varepsilon_{\nu\sigma}) x^\sigma - \varepsilon_{\mu\nu}) \\ &= \int_{\mathcal{D}} d^4x \left[\partial_\rho Y^{\rho\mu\nu} \varepsilon_{\mu\nu} - \partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) (\partial_\mu \varepsilon_{\nu\sigma}) x^\sigma \right].\end{aligned}\quad (189)$$

这一步我们用到 $Y^{\mu\rho\nu} + Y^{\nu\rho\mu}$ 关于 $\mu\nu$ 是对称的, 而 $\varepsilon_{\mu\nu}$ 是反对称的。很显然, (186) 式中的 $\varepsilon_{\mu\nu}$ 项已经得到验证了, 下面我们把注意力放在 $\partial_\rho \varepsilon_{\mu\nu}$ 项上。由上面的推导

$$\begin{aligned}& \int_{\mathcal{D}} d^4x \left[-\partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) (\partial_\mu \varepsilon_{\nu\sigma}) x^\sigma \right] \\ &= \int_{\mathcal{D}} d^4x \left[(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) [(\partial_\rho \partial_\mu \varepsilon_{\nu\sigma}) x^\sigma + \partial_\mu \varepsilon_{\nu\rho}] \right].\end{aligned}\quad (190)$$

这里是进行了分部积分 (分部积分的边界项之所以可以忽略, 是因为我们假设了 $\varepsilon_{\nu\sigma}$ 的偏导在 $\partial\mathcal{D}$ 上取零)。注意到 $(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu})$ 关于 $\mu\rho$ 反对称, 而 $(\partial_\rho \partial_\mu \varepsilon_{\nu\sigma}) x^\sigma$ 关于 $\mu\rho$ 对称, 立即有

$$\begin{aligned}& \int_{\mathcal{D}} d^4x \left[(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) [(\partial_\rho \partial_\mu \varepsilon_{\nu\sigma}) x^\sigma + \partial_\mu \varepsilon_{\nu\rho}] \right] \\ &= \int_{\mathcal{D}} d^4x \left[(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}) \partial_\mu \varepsilon_{\nu\rho} \right] \\ &= \int_{\mathcal{D}} d^4x \left[(Y^{\rho\mu\nu} - Y^{\mu\rho\nu} + Y^{\nu\mu\rho}) \partial_\mu \varepsilon_{\nu\rho} \right] \\ &= \int_{\mathcal{D}} d^4x \left[(-Y^{\mu\rho\nu}) \partial_\mu \varepsilon_{\nu\rho} \right] = \int_{\mathcal{D}} d^4x \left[Y^{\rho\mu\nu} \partial_\rho \varepsilon_{\mu\nu} \right].\end{aligned}\quad (191)$$

正是 (186) 式中的 $\partial_\rho \epsilon_{\mu\nu}$ 项, 验证完成! 上述过程中我们利用了 $Y^{\rho\mu\nu} + Y^{\nu\mu\rho}$ 关于指标 $\nu\rho$ 对称, 而 $\partial_\mu \epsilon_{\nu\rho}$ 关于 $\nu\rho$ 反对称。

从最终的 (188) 式中, 可以读出

$$T^{\mu\nu} = \Theta^{\mu\nu} - \partial_\rho (Y^{\rho\mu\nu} - Y^{\mu\rho\nu} - Y^{\nu\rho\mu}). \quad (192)$$

这个结果与 Weinberg 量子场论第一卷 7.4 节中给出的公式完全一致, 当然推导方法并不相同, 我们这里的推导主要是强调诺特技巧。有些时候人们也称这种对称的能动量张量为 Belinfante(贝林凡特) 张量。

参考文献

首先，对于大学低年级的同学如果了解广义相对论，但是读我这个讲义感觉有难度的话，请参考：L. Susskind, *The Theoretical Minimum* 系列的，*General Relativity*. 以及 James B. Hartle, *Gravity: An Introduction to Einstein's General Relativity*. 这两本书都足够简单足够入门，讲得又好。

其次，如果读完我这个讲义意犹未尽，比方说对霍金和彭罗斯的经典工作产生了兴趣，也就是有关奇性定理和黑洞面积不减定理的那些东西，那么我觉得最好的读物是：E. Witten 的经典综述文章 *Light Rays, Singularities, and All That*. 这篇文章有中文译本（经过 Witten 授权的），叫《光线、奇点，以及其它》，是由我们一些人翻译的，网上能搜到，我的主页上就有。而如果想对黑洞热力学有更多了解，也同样请参阅 E. Witten 的综述文章 *Introduction to Black Hole Thermodynamics*, arXiv: 2412.16795

另外本书参考过：Eric Poisson, *A Relativist's Toolkit—The Mathematics of Black-Hole Mechanics*. 和我讲义重叠的部分不太多，那些我讲义中没有涉及的略微高阶一点的知识 Eric Poisson 这本书都会详细讲！关于黑洞，还有一份资料也很值得推荐：P.K. Townsend, *Black Holes*, Lecture notes, arXiv: gr-qc/9707012。

关于宇宙学的全面讨论，请参考 Daniel Baumann, *Cosmology*. 很现代，全面和细致的书籍，非常适合用来入门宇宙学，当然书稍微有点厚。

关于引力波，本书的介绍当然也只是导引，更多的介绍可以参见卢昌海老师的高级科普书《时空的乐章——引力波百年漫谈》，非常好的书。更多的专业技术内容则可以参见 Michele Maggiore 所著的两卷本：*Gravitational Waves, Volume 1 Theory and Experiments* 以及 *Gravitational Waves, Volume 2 Astrophysics and Cosmology*. 非常好的两本书，强烈推荐给想进一步了解或者深入研究相关内容的读者。

如果你还想了解一点量子引力，那么推荐：Thomas Hartman, *Lectures on Quantum*

Gravity and Black Holes

本书还参考过 David Tong 的在线讲义 General Relativity, 在他的页面上能找到, 不过总体上 David Tong 的讲法和本书是两个路子。中文教材里面, 最著名的肯定是梁灿彬老师的《微分几何入门与广义相对论》三卷本, 我当然也参考过, 但是参考得不多, 和本书的写法基本也是两个路子。如果本讲义的讲述方式不合你的口味, 如果你喜欢更数学一些的介绍方式, 那么也许你可以试试梁老师的这三卷书。