

初识经典场论

陈童

2026 年 1 月 16 日

目录

第一章 为何要引入场?	3
1.1 从粒子到场	3
1.2 粒子与场所构成的系统	5
1.3 何为场?	6
1.4 历史上是如何引入场的?	8
1.5 法拉第简介	9
第二章 相对性原理	13
2.1 相对性原理	13
2.1.1 信息传播的最大速度	13
2.1.2 间隔不变性	15
2.1.3 洛伦兹变换	19
2.2 四维时空的矢量和张量	23
2.2.1 四维矢量和张量	23
2.3 相对性原理与经典场论	26
2.3.1 相对论粒子的作用量以及相互作用	27
2.3.2 场的作用量	28
第三章 电磁场	32
3.1 规范不变性	32
3.2 “发现”麦克斯韦方程	34
3.3 电磁场携带能量	39
3.4 光是一种电磁波	41
3.5 电荷量子化	42

目录	2
3.6 电磁对偶与磁单极子	43
3.6.1 狄拉克量子化条件	45
第四章 粒子源自于场	49
4.1 粒子作为场方程的一种解	49
4.1.1 场位形的拓扑等价类	50
4.1.2 Bogomolny能限和Bogomolny方程	52
4.1.3 sine-Gordon模型和孤立子	53
4.2 粒子作为场的量子	56

第一章 为何要引入场？

1.1 从粒子到场

经典力学的学习告诉我们， N 个相互作用粒子所构成的系统的力学规律可以由最小作用量原理描述，其中，系统的作用量可以写成

$$S[\mathbf{x}_1(t), \mathbf{x}_2(t), \dots, \mathbf{x}_N(t)] = \int dt \left[\sum_{i=1}^N \frac{1}{2} m_i \dot{\mathbf{x}}_i^2 - U(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \right], \quad (1.1)$$

式中 $\mathbf{x}_i(t)$ 为第 i 个粒子 t 时刻的位置矢量。式中的函数 $U(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ 称为系统的相互作用势能，它描述各粒子之间的相互作用。

但是，经典力学对相互作用粒子系统的这种描述隐藏了一个重要的假设，即粒子间的相互作用是瞬时的。反映在函数 $U(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ 中即是：这个函数取决于 N 个粒子在 t 时刻的瞬时位置，粒子的位置变了，则相互作用瞬时就会改变，无论各粒子之间距离有多远。这就是所谓的超距相互作用，指的就是两个分离的、不接触的物体，其相互作用（如引力）是瞬间发生的，不需要时间，也不需要任何中间介质传递。这与日常经验（如推车、声音传播）完全相悖，在哲学和物理上都难以接受。牛顿本人已经注意到这种超距作用的假设在哲学上不令人满意，但是，他不知道如何修正这一假设。而且在牛顿的时代，超距的万有引力势能足以解释太阳和行星之间的万有引力相互作用，能够完美描述行星运动和潮汐等现象。

但牛顿本人对引力作用的机制深感困扰。他在《自然哲学的数学原理》中写道：“引力应该是物质内在的、固有的和根本的，因此一个物体可以通过真空超距作用于另一个物体，无需任何中介……这对我来说是巨大的荒谬。”

为了消除“超距作用”，许多科学家（包括牛顿后期的思考）设想宇

宙充满一种看不见、摸不着的弹性介质——“以太”。他们认为，引力可能是通过以太的挤压或涡旋来传递的。然而，所有构建机械以太模型的尝试都失败了。这些模型要么在数学上复杂且矛盾，要么无法与天文观测（如行星运行无阻力）相容。以太成了一个尴尬的“补丁”，但超距作用的困境依然存在。

后来，爱因斯坦发现，超距作用必定是错的！因为，宇宙中的一切相互影响存在一个速度上限，不妨记这个速度上界为 c （当然，正如后面的章节中将会证明的， c 就是真空中的光速），爱因斯坦发现，宇宙中一切的相互影响不能以超过 c 的速度传播，当然更不能瞬时传播（不能瞬移！）。因此，一个粒子的位置改变不能瞬时影响另一个粒子，而是必须有一个时间延迟，这个时间延迟就是影响从前一个粒子传播到后一个粒子所需要的时间。在这个发现的基础上，爱因斯坦建立起了狭义相对论。

考虑到相对论，前面的(1.1)式就不成立了，或者说，它只在粒子的速度 x_i 远远小于速度上界 c 时才近似成立，因为这时候我们可以近似认为相互作用的传播速度是无穷大的，从而近似可以使用相互作用瞬时发生的超距作用观点。

在相对论中，相互作用不是瞬时的，因此无法通过一个相互作用势能函数来刻画，而必须引入动力学的、传递相互作用的某种中介物，称之为场。两粒子发生相互作用的过程实际上是，一个粒子激发出一个场，这个场本身也是有动力学的，当它传播到另一个粒子的位置处时，就作用在这个粒子上。同时后面这个粒子也会激发场，这个场反过来传播到前一个粒子位置时，也会作用上去。通过这种方式，后一个粒子就和前一个粒子发生了相互作用。比方说两个带电粒子通过电磁力相互作用，如果考虑到相对论的话，那就不能用库伦势来刻画两粒子之间的相互作用，而必须引入整个电磁场的动力学，那就到经典场论了。因此，相对论必然会把我们引导到经典场论。

当然，狭义相对论必定导致场论这个结论你可能听说过，但你听说过的也许是量子力学加狭义相对论一定会导致量子场论，而我们这里说的是经典场论，所以两者还稍微有些区别。区别之处在于，在量子场论的情形中（也就是量子力学加狭义相对论的情形中），一切都是场，也即是说，物质粒子和传递相互作用的规范粒子本质上都是场。而在不考虑量子力学只

考虑狭义相对论的情形中，可以允许物质为粒子（而不是场），只有传递相互作用的中介物才必须是动力学场，所以这是一种多粒子通过动力学场耦合的情形。

1.2 粒子与场所构成的系统

那么，在经典相对论理论中，这种“多粒子+场”的图像具体是如何实现的呢？关键的一点在于，场本身是要传播要演化的，换言之，**场本身是一个动力学系统**。这意味着场可以脱离激发它的粒子源而独立存在、传播和演化，就像电磁波可以脱离电荷源在真空中传播一样。场本身就是一种动力学自由度，场的演化本身就对整个系统的作用量有贡献。当我们把粒子与场耦合在一起时，粒子的运动会激发场，场的演化又会反过来影响别的粒子的运动。这种耦合关系可以通过一个统一的作用量来描述，其中既包含粒子的自由度，也包含场的自由度。

对于这种多粒子和场共同构成的系统，根据作用量的可加性，我们可以假设总作用量能写成：

$$S = S_{\text{粒子}} + S_{\text{场}} + S_{\text{相互作用}}$$

其中 $S_{\text{粒子}}$ 描述自由粒子的运动， $S_{\text{场}}$ 描述场本身的动力学，而 $S_{\text{相互作用}}$ 描述粒子与场之间的耦合。通过变分原理，我们可以同时导出粒子的运动方程和场的运动方程。这样，整个系统就成为一个自洽的动力学理论。

这种描述方式彻底摒弃了超距作用：一个粒子的位置变化会激发场的变化，这个变化以速度 c 向周围传播，当它传播到另一个粒子处时，才会对后者施加力的作用。这种“作用传播需要时间”的图像，正是相对论因果性的体现。

以两粒子通过场相互作用为例，“多粒子+场”的总图像如图(1.1)所示。图中标示粒子1和粒子2的两条曲线是两个粒子在时空中运动的路径，称之为粒子的**世界线**。图中的阴影部分表示作为中介物的场，这里我们仅考虑了粒子1在某个时间段内激发的场，并且只表示出了这个场中最终作用在粒子2上的那一部分。请注意，图中场的传播需要一定的时间，换言之，相互作用不是瞬时的。

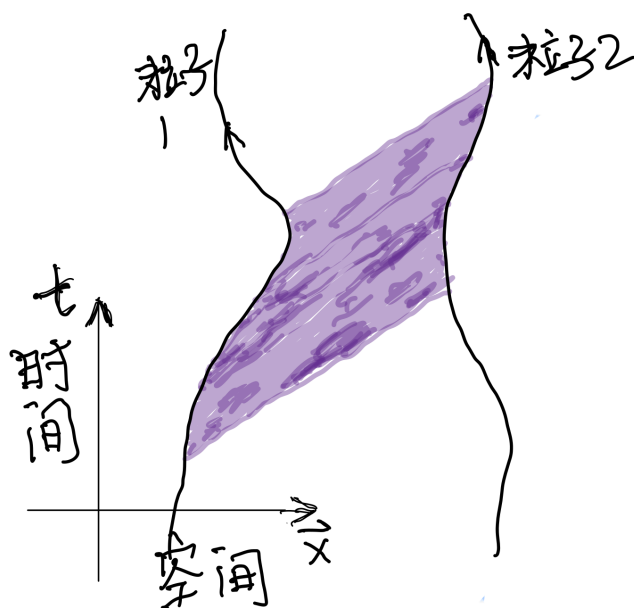


图 1.1: 两粒子通过作为中介物的场相互作用, 图中阴影部分表示场, 这里只考虑粒子1在某个时间段内激发的场。

所以, 引入场不仅是为了数学上的方便, 更是为了满足相对论的基本要求——相互作用的传播速度不能超过速度上限 c 。在这个意义上, 场是相对论性相互作用不可或缺的载体。这也为我们理解自然界中各种力(电磁力、引力等)提供了统一的框架: 它们都是通过某种特定的场来传递的。

1.3 何为场?

到现在为止我们只知道场是传递粒子间相互作用的某种中介物, 而且它本身是可以独立演化的。那么, 什么是场? 假设某个粒子某时刻激发起

一个场，那这个场是要向四周传播的，某时刻它传到哪个位置，就会作用在这个位置处的其它粒子上，但是，由于这些其它粒子的位置并不是预先给定的，所以场必定不能和粒子一样只沿着某条世界线传播，不能只传播到某些特定的位置，而是必须向四周传播。

既然场随时间演化传播，就意味着场必定是时间 t 的函数，同时场向四周的不同位置传播又意味着场必定同时以空间位置坐标 $\mathbf{x} = (x, y, z)$ 为自变量，所以什么是场？场就是一种时空中的分布，典型的，比如 $\phi(t, \mathbf{x}) = \phi(t, x, y, z)$ 就是一个场。

空间的温度分布当然也是一个场，称之为温度场。但是，这个场不是我们经典场论中关心的场，原因有二：第一，温度并不是一种传递相互作用的中介物，第二，温度是粒子热运动的平均动能，它是粒子的一种性质，不能离开粒子而独立存在。

经典场论中关心的场是某种动力学自由度，也就是说，经典场论中的 $\phi(t, \mathbf{x})$ 是与某粒子系统的运动路径 $q_a(t)$ (这里写的是粒子系统的广义坐标)一样的动力学变量。实际上，你可以把经典场论中的 $\phi(t, \mathbf{x})$ 重写成 $\phi_{\mathbf{x}}(t)$ ，把空间位置 \mathbf{x} 看成是指标，那么 $\phi_{\mathbf{x}}(t)$ 就是一个位于 \mathbf{x} 点的动力学变量。所以，一个经典场系统，就是在每个空间点上都放上一个动力学变量的系统，由于每个空间点上都有一个动力学变量，动力学变量的总数目当然是无穷，所以一个经典场系统就是一个无穷自由度的动力学系统。

粒子的作用量是广义坐标 $q_a(t)$ 的泛函，形如 $S_{\text{粒子}}[q_a(t)]$ 。同样，场的作用量也必然是其动力学变量 $\phi_{\mathbf{x}}(t)$ 的泛函，形如 $S_{\text{场}}[\phi(t, \mathbf{x})]$ 。怎么决定这些作用量的具体表达式呢？这个问题我们留在后面的章节中讨论。

从物理图像上看，我们可以将场想象为一种遍布全空间的“氛围”。粒子沉浸在这种“氛围”中，其运动（作为“源”）会扰动“氛围”，产生“涟漪”或“波动”（即场的变化）。这些扰动以有限速度 c 向外传播，当传播到其他粒子所在位置时，就会施加力于该粒子。这样，相互作用便通过场的激发与传播得以实现，并严格遵守因果律：原因（第一个粒子的运动）必须早于结果（第二个粒子感受到力）发生，其时间差至少为两者空间距离除以 c 。

1.4 历史上是如何引入场的?

在物理学史上，场的概念并非一蹴而就，它源于实验家的直观想象，最终由数学家的方程所确立。迈克尔·法拉第是场的概念真正的开创者，他的贡献来自深刻的物理直觉和对实验的洞察。

19世纪30-40年代人们在研究电磁现象时，发现库仑定律、安培定律在形式上与万有引力定律相似，都是超距作用的平方反比律。但电磁现象中复杂的力（吸引、排斥、旋转）让法拉第感到，必须有一种更直观的方式来理解。法拉第在做磁铁实验时，通过在磁铁的周围撒上铁屑，观察铁屑的分布，“看到”了磁力线的形状。他认为这些力线不是数学抽象，而是真实的物理存在。他提出：电荷或磁极并不是直接超距作用于另一个电荷或磁极，而是首先在其周围空间建立起一种紧张的状态（a state of tension），即“场”。这种场由布满空间的力线构成，力线的疏密代表力的强弱，力线的切线方向代表力的方向。

尤其是，法拉第发现通过磁力线的直观概念，以及相应的磁通量概念，他就可以将神奇的电磁感应现象概括成一条简单的定律，即著名的法拉第电磁感应定律。而如果不用磁力线和磁通的概念，只用力的超距作用概念，那人们甚至无法将各种电磁感应现象归纳成一条定律。

法拉第提出，相互作用是通过场来传递的。例如，磁铁A的力线分布到空间，当磁铁B进入这个空间时，就受到力线上张力的作用。法拉第推测，力的传递需要时间，是局域的、连续的。法拉第用“力线”和“场”的概念，第一次将物理作用的核心从粒子本身转移到了粒子之间的空间。空间不再是被动的真空，而是充满活性、具有物理属性的中介物，即场。

詹姆斯·克拉克·麦克斯韦是法拉第思想的“翻译者”和升华者。他将模糊的力线图像转化为精确的数学语言。麦克斯韦早期（1855-1864年）试图用复杂的机械以太模型（idle wheels and vortices）来类比法拉第的力线，目的是找到描述场的数学关系。最终，麦克斯韦超越了机械类比，直接从实验定律（法拉第电磁感应、安培环路定律等）出发，引入了电磁场的数学描述。尤其是，麦克斯韦引入一个革命性的概念——“位移电流”。这是为了保持电流的连续性而提出的假设，意味着变化的电场本身就能产生磁场。

1864年，麦克斯韦发表了著名的方程组。这组方程表明，变化的电场

产生磁场，变化的磁场产生电场，两者相互激发，形成一个可以在空间中自持传播的波动——电磁波。他从方程中推导出电磁波的速度，发现与已知的光速惊人地接近，于是大胆预言：光就是一种电磁波。

至此，“场”不再仅仅是一种描述力的辅助工具，而是一个独立的动力学实体。它可以脱离源（电荷、磁铁）而存在，在空间中以有限速度（光速）传播，并携带能量和动量。1887年，海因里希·赫兹通过实验产生了并检测到了麦克斯韦预言的电磁波，证实了它的速度、反射、折射、衍射等波动性质与光完全相同。赫兹实验是场的概念的终极实证。它证明电磁场是真实的物理存在，完全可以脱离产生它的仪器而独立传播。“场”作为物理学基本实体的地位从此不可动摇。

狭义相对论（1905年）彻底否定了“以太”作为绝对静止参考系的概念，但极大地强化了场的地位。电磁场本身就成了一个基本实体，其规律（麦克斯韦方程）在惯性系中协变。广义相对论（1915年）则是场概念的巅峰之作。爱因斯坦将引力解释为时空度规的弯曲，而引力场就是这个弯曲的时空几何本身。物质告诉时空如何弯曲，弯曲的时空告诉物质如何运动。这完全实现了用“场”（时空度规场）来解释引力的梦想，彻底终结了引力的超距作用解释。

1.5 法拉第简介

迈克尔·法拉第（1791年9月22日 - 1867年8月25日）是英国物理学家和化学家，被广泛誉为科学史上最具影响力的实验天才之一。尽管出身贫寒，几乎没有接受过正规教育，但他凭借非凡的动手能力、直觉和勤奋，在电磁学和电化学领域做出了奠基性的贡献。他的核心特点在于：**实验大师**：他的许多伟大发现源于精巧而深刻的实验设计。**直观思维者**：他创造了“力线”（场线）的概念来思考电磁作用，这是一种超越当时数学框架的物理图像思维。**沟通者**：他在英国皇家研究院主办的“星期五晚间讲座”和“圣诞少年讲座”极为成功，使其科学思想得以普及。他的讲座集《蜡烛的故事》成为科普经典。

法拉第的主要科学贡献（分领域）：

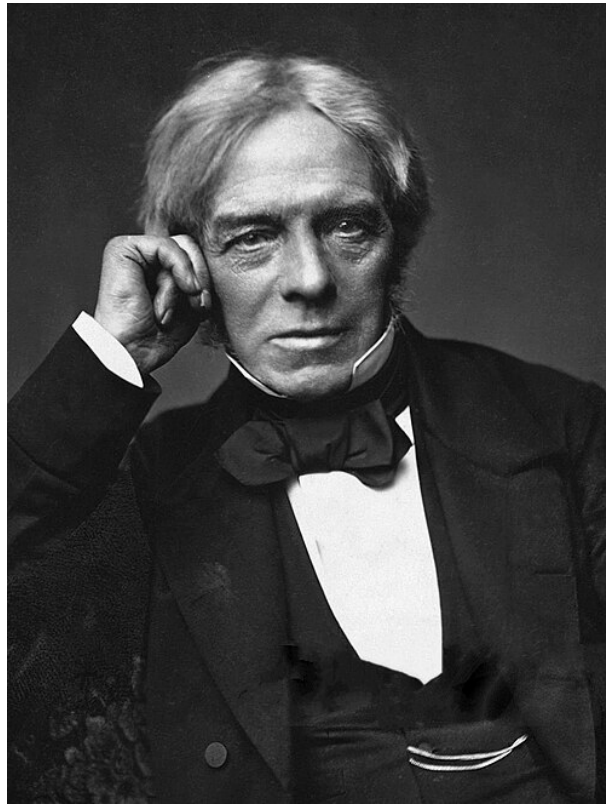


图 1.2: 法拉第

A. 电磁学领域（核心贡献）

这是法拉第对现代世界影响最深远的领域，他的工作直接催生了电气时代。

1. 电磁感应（1831年）：法拉第发现，当闭合线圈的**磁通发生变化**时（无论是磁铁移动，还是线圈移动），线圈中会产生感生电流。这就是法拉第电磁感应定律。这是人类历史上第一次将机械能（运动）高效地转化为电能。这一定律是几乎所有发电机、变压器和电动机工作原理的基础。没有这个发现，就没有现代的电力系统。

2. 场论的先驱：为了解释超距作用（如磁铁为何能隔空吸引铁屑），法拉第提出了革命性的“场”概念。他认为空间并非虚空，而是布满了一种传递力的媒介，他形象地称之为“力线”（即今天的电场线和磁场线）。

场的概念是物理学思想的巨大飞跃。它打破了牛顿“超距作用”的框架，为后来詹姆斯·克拉克·麦克斯韦建立统一的电磁场方程组提供了关键的物理图像和概念基础。法拉第的“场”是现代物理学（包括引力场、量子场）的核心概念起源。

3. 其他电磁学发现：抗磁性（1845年）：发现所有物质在强磁场中都会表现出微弱的磁性（大部分为抗磁性），拓展了对物质磁性的理解。法拉第效应（磁致旋光效应，1845年）：发现磁场可以使通过某些透明物质（如玻璃）的光的偏振面发生旋转。这是光与电磁现象之间存在直接联系的第一个实验证据，同样为麦克斯韦的电磁波理论提供了支持。

B. 电化学领域

法拉第是电化学的奠基人之一，他系统性地研究了电流与化学反应的关系。

1. 法拉第电解定律（1830年代）：第一定律：电解过程中，电极上析出（或溶解）的物质的质量与通过的电荷量成正比。第二定律：当通过相同的电荷量时，不同物质析出的质量与其化学当量（即摩尔质量与化合价之比）成正比。这些定律精确定量地连接了物理学（电量）和化学（物质质量）。它们不仅确立了电荷的“原子性”概念（暗示了电的分立性，为电子论的提出埋下伏笔），还为后来的化学分析（电解分析法）和工业电镀工艺提供了理论基础。“法拉第常数”就是以他的名字命名的基本物理常数。

2. 引入关键术语：他创造了我们今天仍在使用的大量电学术语，如电极（阳极、阴极）、离子（阳离子、阴离子）、电解质、电解等。这些术语极大地规范并推动了化学的发展。

C. 化学及其他领域

1. 发现苯（1825年）：从鱼油中分离出一种新的碳氢化合物，命名为“氢的重碳化合物”，即苯。苯是芳香烃化学的母体，对有机化学和化学工业（染料、制药、塑料）的发展至关重要。2. 气体液化的先驱：在低温高压下，成功液化了多种被认为的“永久气体”，如氯气、二氧化硫等，为低温物理学开辟了道路。3. 发明“法拉第笼”（1836年）：演示了静电屏蔽效应，即封闭的金属导体外壳可以隔绝外部的静电场。这是电磁屏蔽原理的基础，广泛应用于现代电子设备保护和高压作业安全。

法拉第的科学哲学与思想遗产

自然力的统一观：法拉第坚信自然界各种力（电、磁、光、引力等）是相互联系、可以相互转化的。他毕生都在寻找这些联系（如研究磁对光的影响），这种思想深刻影响了后来的物理学家，特别是麦克斯韦和爱因斯坦。实践与理论的结合：他是一个纯粹的实验家，但他的物理直觉和图像化思维催生了最深远的理论概念（场）。

法拉第的贡献可以概括为：他用实验开启了电气时代的大门，用思想播下了场物理学的种子。总而言之，迈克尔·法拉第是一位从装订学徒成长为科学巨人的传奇。他的贡献既是实用性的、革命性的，也是概念性的、哲学性的。他不仅告诉我们电如何产生，更从根本上改变了我们理解自然力作用方式的世界观。

第二章 相对性原理

上一章说过，狭义相对论的要求迫使人们引入动力学场来传递粒子间的相互作用。因此，为了进一步讨论经典场论，本章先系统地讨论一下狭义相对论。

在本章中，对于一个带指标的量，将使用上、下两种指标，并且两者通常是有区别的。我们引入一个所谓的**求和约定**，即当一个表达式中某个上指标和某个下指标符号相同时，就默认对这个指标的所有可能性进行求和。

2.1 相对性原理

2.1.1 信息传播的最大速度

伽利略最早论述了相对性原理，它说的是，在所有惯性系中力学规律都是一样的。力学规律当然也要包括物体之间相互作用的规律，但是在伽利略和牛顿时代建立的理论中，两个物体之间的相互作用是瞬时进行的，比如牛顿的万有引力定律就是这样，这就是所谓的超距相互作用。正如牛顿已经注意到的，这肯定是有问题的，一个物体要和另一个物体相互作用，也就是要对另一个物体施加影响，这种影响不可能即时，而必定有一个时间的延迟，但是伽利略的力学相对性原理没有考虑到这一点。

爱因斯坦将相对性原理进行了推广，爱因斯坦提出，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。这个相互作用传播速度有限的相对性原理就是爱因斯坦的相对性原理，今天谈相对性原理时都默认指的是这条原理。

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 ，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 $c = 1$ 。

同时的相对性

信息传播的最大速度是常数的一个立即推论即是，两个不同地点的事件是否同时发生是依赖于参考系的，在一个惯性系中同时发生的两件事，在另一个惯性系中往往不同时，这就是同时的相对性。同时的相对性意味着，在狭义相对论中，时间不是绝对的，不同的惯性系将有不同的时间！

为了理解同时的相对性，让我们考虑两个惯性系，分别称为 S 系和 S' 系，假设 S' 沿着 S 的 x 轴正方向匀速运动。设想在 S 系的 x 轴上两个不同点 A 和 B （假设 $x_A < x_B$ ）同时发生了两个事件，并且在事件发生的同时， A 、 B 两地分别向着 AB 的中点以最大传播速度 c 发出一个信号，很显然，在 S 系位于 AB 中点的观察者看来，这两束信号将同时到达。并且正是根据这两者的同时到达，这个观察者才可以推断 A 、 B 两点的事件是同时发生的。现在假设在 S' 系的 AB 中点也有一个观察者，同样的道理，这个观察者对 A 、 B 两地的事件是否同时发生的判断也是依据他是否同时接收到事件发生时发出的那两个信号。

那么 S' 系的这个观察者是否会同时接收到那两个信号呢？根据 c 的不变性，答案将是否定的，实际上这个观察者将先接收到 B 地发出的信号。这是因为，根据最大信息传播速度的不变性， A 、 B 两地发出的信号在 S' 的观察者看来依然以同样的速度 c 传播，但是，现在这两个信号将要到达的目的地——也就是 AB 在 S' 系中的中点——在朝着 B 地发出的信号运动（同时也在背着 A 地发出的信号运动），因此 S' 系位于中点的这个观察者无疑会先接收到 B 地发出的信号。因此，在 S' 系的这个观察者就会进而得出结论，即 A 、 B 两地的这两个事件不是同时发生的，而是 B 点的事件先发生， A 点的事件后发生，因此同时是相对的。请千万要记住，在整个分析过程中，你心中

绝对不能预设A、B两地的信号在 S' 系中也是同时发出的。

2.1.2 间隔不变性

依然考察 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。

人们通常约定 $t = ct = x^0, x = x^1, y = x^2, z = x^3$, 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, 我们就可以将间隔的计算公式重写为

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

式中我们默认了求和约定(后文也都默认求和约定), 而 $\eta_{\mu\nu}$ 为, $-\eta_{00} = \eta_{11} = \eta_{22} = \eta_{33} = 1$, 其它指标分量都等于零。 $\eta_{\mu\nu}$ 称为四维闵可夫斯基时空的度规张量, 所谓的闵可夫斯基时空指的就是狭义相对论中的平直时空。值得注意的是, $\eta_{\mu\nu}$ 关于它的两个指标是对称的, 即满足 $\eta_{\mu\nu} = \eta_{\nu\mu}$ 。

我们可以将 $\eta_{\mu\nu}$ 看成是一个 4×4 矩阵的分量形式(记这个矩阵为 η), 进而引入这个矩阵的逆矩阵 η^{-1} , 其分量形式记为 $\eta^{\mu\nu}$,

$$\eta_{\alpha\beta} \eta^{\beta\gamma} = \delta_\alpha^\gamma, \quad (2.9)$$

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

$$(\Delta s)^2 = \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu = -(\Delta t)^2 + (\Delta \mathbf{x})^2. \quad (2.10)$$

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

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

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,

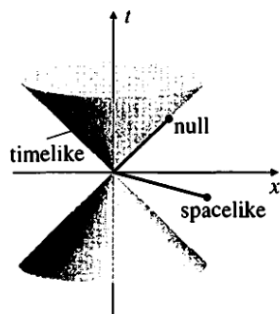


图 2.1: 时空图上原点处的光锥。

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。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

$$\eta_{\mu\nu} dx^\mu dx^\nu = -d\tau^2. \quad (2.12)$$

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。

假设粒子的速度为 \mathbf{v} ，即 $d\mathbf{x} = \mathbf{v}dt$ ，则根据(2.12)式即有

$$d\tau^2 = dt^2 - d\mathbf{x}^2 = dt^2(1 - \mathbf{v}^2) \Rightarrow dt = \frac{d\tau}{\sqrt{1 - \mathbf{v}^2}}. \quad (2.13)$$

即是说，坐标时总是比固有时长的。测固有时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。

那么假设有两只钟，一只是静止的，另一只沿着闭合路径运动一圈再回到起点与静止的钟比较，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。

2.1.3 洛伦兹变换

狭义相对论中，两个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

$$x'^{\mu} = \Lambda^{\mu}_{\nu} x^{\nu}, \quad (2.14)$$

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

$$\eta_{\mu\nu} dx'^{\mu} dx'^{\nu} = \eta_{\mu\nu} \Lambda^{\mu}_{\alpha} \Lambda^{\nu}_{\beta} dx^{\alpha} dx^{\beta} = \eta_{\alpha\beta} dx^{\alpha} dx^{\beta}, \quad (2.15)$$

从而即有

$$\eta_{\mu\nu} \Lambda^{\mu}_{\alpha} \Lambda^{\nu}_{\beta} = \eta_{\alpha\beta}. \quad (2.16)$$

更一般地，人们也把任何满足上式的时空坐标变换 Λ^{μ}_{ν} 称作洛伦兹变换，哪怕它不是源于不同惯性系之间的变换。

我们当然可以将(2.16)式写成矩阵形式，即

$$\Lambda^T \eta \Lambda = \eta. \quad (2.17)$$

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(2.17)式两边求行列式，并注意到 $\det(\eta) = -1$ ，从而即可得

$$[\det(\Lambda)]^2 = 1 \Rightarrow \det(\Lambda) = \pm 1. \quad (2.18)$$

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(2.16)式中取 $\alpha = \beta = 0$ ，即可得

$$(\Lambda_0^0)^2 = 1 + \sum_{i=1,2,3} (\Lambda_0^i)^2 \geq 1 \Rightarrow \Lambda_0^0 \geq 1 \text{ or } \Lambda_0^0 \leq -1. \quad (2.19)$$

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

$$\det(\Lambda) = 1, \quad \Lambda_0^0 \geq 1. \quad (2.20)$$

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 $SO^+(1,3)$ 。

一般来说，洛伦兹变换依赖于某些连续参数，比如两个参考系的相对速度以及坐标轴的相对角度等等。最简单的洛伦兹变换就是 $\Lambda_\nu^\mu = \delta_\nu^\mu$ 的恒等变换，它描写两个始终完全重合的参考系。但是，我们可以设想连续地改变洛伦兹变换的参数，使得这两个参考系变得不再重合，从而使得恒等变换变成非平凡的洛伦兹变换。而正洛伦兹变换的集合正是能够由恒等变换连续地变化过来的所有洛伦兹变换。这是因为，恒等变换显然满足 $\det(\Lambda) = 1$ ， $\Lambda_0^0 \geq 1$ ，而连续的变化不可能使得 $\det(\Lambda)$ 从+1突变到-1，也不可能使得 Λ_0^0 从 ≥ 1 突变到 ≤ -1 ，所以能和恒等变换连续过渡的洛伦兹变换一定是正洛伦兹变换。

另外，按照洛伦兹变换的一般定义， $t' = -t$ ， $\mathbf{x}' = \mathbf{x}$ 的时间反演变换显然也是一种洛伦兹变换，它的变换矩阵 Λ 显然满足， $\Lambda_0^0 = -1$ ， $\det(\Lambda) = -1$ 。因此，在任何正洛伦兹变换的基础上再进行一个时间反演变换，得到的就是 $\det(\Lambda) = -1$ ， $\Lambda_0^0 \leq -1$ 的洛伦兹变换。类似的，空间反演变换 $t' = t$ ， $\mathbf{x}' = -\mathbf{x}$ 也是洛伦兹变换，它的变换矩阵满足 $\Lambda_0^0 = 1$ ， $\det(\Lambda) = -1$ ，因此，在正洛伦兹变换的基础上再进行一个空间反演变换，得到的就是 $\det(\Lambda) = -1$ ， $\Lambda_0^0 \geq 1$ 的洛伦兹变换。而在正洛伦兹变换的基础上再同时进行时间和空间的反演，得到的就是 $\det(\Lambda) = 1$ ， $\Lambda_0^0 \leq -1$ 的洛伦兹变换。

正洛伦兹变换的一个例子就是三维空间旋转，即如下洛伦兹变换

$$\Lambda_0^0 = 1, \quad \Lambda_i^0 = \Lambda_0^i = 0, \quad \Lambda_j^i = R_{ij}, \quad (2.21)$$

式中 $i, j = 1, 2, 3$, 而 3×3 矩阵 R 是特殊正交矩阵, 即满足 $R^T R = 1, \det(R) = 1$ 。不过, 通常我们谈到洛伦兹变换的时候不是指这种旋转变换, 虽然它的确很重要。

一些重要的洛伦兹变换

下面我们来构造一个最为通常而重要的洛伦兹变换。为此, 让我们考虑一种特殊情况, 假如 S' 系沿着 S 系的 x 轴正方向以匀速 v 运动。假设我们这么选取 S' 的坐标轴, 以使得初始时两个坐标系的坐标轴完全重合。这样一来, 由于两坐标系的相对运动只发生在 x 方向上, 与 y, z 方向无关, 所以显然有 $y' = y, z' = z$ 。因此, 我们只需要考虑事件的 t 坐标和 x 坐标在参考系变换下如何变换, 为此不妨暂时忽略间隔公式(2.6)中的 y, z 坐标, 进而将间隔公式简化为

$$ds^2 = -dt^2 + dx^2 = dx^+ dx^-, \quad (2.22)$$

式中 $x^+ = x + t, x^- = x - t$ 。间隔不变性告诉我们, 参考系变换以后的 $x'^+ = x' + t', x'^- = x' - t'$ 必然满足 $dx'^+ dx'^- = dx^+ dx^-$ 。显然这就意味着必定有某个参数 ω 使得 $x'^+ = e^{-\omega} x^+, x'^- = e^{\omega} x^-$, 即有

$$x' = \cosh(\omega)x - \sinh(\omega)t, \quad (2.23)$$

$$t' = -\sinh(\omega)x + \cosh(\omega)t. \quad (2.24)$$

考虑 S' 系的坐标原点 $x' = 0$, 它在 S 中的运动速度为 v , 因此我们有 $\tanh(\omega) = v$, 利用双曲函数的相关公式, 容易得到 $\cosh(\omega) = 1/\sqrt{1-v^2}$, $\sinh(\omega) = v/\sqrt{1-v^2}$, 也即是说, x, t 坐标的变换公式为

$$\begin{aligned} t' &= \frac{t - vx}{\sqrt{1-v^2}}, \\ x' &= \frac{x - vt}{\sqrt{1-v^2}}, \end{aligned} \quad (2.25)$$

再加上前面的 $y' = y, z' = z$, 这四个式子就是最常用的一组洛伦兹变换。而且, 反过来, S 系也在以 $-v$ 的速度相对 S' 系运动, 所以必定有

$$\begin{aligned} t &= \frac{t' + vx'}{\sqrt{1-v^2}}, \\ x &= \frac{x' + vt'}{\sqrt{1-v^2}}. \end{aligned} \quad (2.26)$$

另外, 也容易得到 $e^{-\omega} = \cosh \omega - \sinh \omega = \left(\frac{1-v}{1+v}\right)^{\frac{1}{2}}$, 从而即有

$$x' + t' = \left(\frac{1-v}{1+v}\right)^{\frac{1}{2}}(x+t), \quad x' - t' = \left(\frac{1+v}{1-v}\right)^{\frac{1}{2}}(x-t). \quad (2.27)$$

如果先将 S 系变换到以速度 v_1 沿着 x 轴运动的 S' 系, 从而即有 $x' + t' = \left(\frac{1-v_1}{1+v_1}\right)^{\frac{1}{2}}(x+t)$, 接着再变换到以速度 v_2 相对 S' 系沿 x 运动的 S'' 系, 即有 $x'' + t'' = \left(\frac{1-v_2}{1+v_2}\right)^{\frac{1}{2}}(x' + t')$ 。则 S'' 系相对 S 系的变换将是, $x'' + t'' = \left(\frac{1-v}{1+v}\right)^{\frac{1}{2}}(x+t) = \left(\frac{1-v_2}{1+v_2}\right)^{\frac{1}{2}} \left(\frac{1-v_1}{1+v_1}\right)^{\frac{1}{2}}(x+t)$, 式中 v 为 S'' 系相对于 S 系的速度。从而即有,

$$\left(\frac{1-v}{1+v}\right)^{\frac{1}{2}} = \left(\frac{1-v_1}{1+v_1}\right)^{\frac{1}{2}} \cdot \left(\frac{1-v_2}{1+v_2}\right)^{\frac{1}{2}}. \quad (2.28)$$

很明显, 只要 $v_1 < 1, v_2 < 1$, 则上面等式右边就是两个正实数的乘积, 因此结果也必定是正实数, 从而必有 $v < 1$, 即是说, 我们不可能通过多次洛伦兹变换使得相对运动速度大于光速。另外, 从(2.28)式也容易得到

$$\frac{1-v}{1+v} = \frac{1-v_1-v_2+v_1v_2}{1+v_1+v_2+v_1v_2}. \quad (2.29)$$

也即有

$$v = \frac{v_1 + v_2}{1 + v_1 v_2}, \quad (2.30)$$

这就是相对论的速度合成公式。

下面, 我们对(2.25)式给出的洛伦兹变换进行一个推广。依然考虑两个参考系 S 与 S' 之间的变换, 同样假设初始时这两个参考系的坐标轴完全重合。不过, 现在 S' 系相对于 S 系以任意方向的速度 \mathbf{v} 运动。以 $v = |\mathbf{v}|$ 表示速度的大小, 以 $v_i, i = 1, 2, 3$ 表示速度 \mathbf{v} 的各分量。为了利用(2.25)式, 我们将坐标矢量 \mathbf{x} 正交分解成平行于速度 \mathbf{v} 的 \mathbf{x}_{\parallel} 和垂直于速度 \mathbf{v} 的 \mathbf{x}_{\perp} 两部分, 它们满足

$$\mathbf{x}_{\parallel} = \frac{\mathbf{v}(\mathbf{v} \cdot \mathbf{x})}{v^2}, \quad \mathbf{x}_{\perp} = \mathbf{x} - \mathbf{x}_{\parallel}. \quad (2.31)$$

则根据(2.25)式, 容易写出下面洛伦兹变换关系

$$\begin{aligned} t' &= \frac{t - \mathbf{v} \cdot \mathbf{x}}{\sqrt{1-v^2}}, \\ \mathbf{x}'_{\parallel} &= \frac{\mathbf{x}_{\parallel} - \mathbf{v}t}{\sqrt{1-v^2}}, \\ \mathbf{x}'_{\perp} &= \mathbf{x}_{\perp}. \end{aligned} \quad (2.32)$$

由此即可以得到,

$$\begin{aligned}\mathbf{x}' &= \mathbf{x}'_{\perp} + \mathbf{x}'_{\parallel} = (\mathbf{x} - \mathbf{x}_{\parallel}) + \frac{\mathbf{x}_{\parallel} - \mathbf{v}t}{\sqrt{1-v^2}} \\ &= \mathbf{x} + \left(\frac{1}{\sqrt{1-v^2}} - 1\right) \frac{\mathbf{v}(\mathbf{v} \cdot \mathbf{x})}{v^2} - \frac{\mathbf{v}}{\sqrt{1-v^2}} t.\end{aligned}\quad (2.33)$$

将以上结果用洛伦兹变换矩阵 Λ_{ν}^{μ} 写出来即是,

$$\begin{aligned}\Lambda_0^0 &= \frac{1}{\sqrt{1-v^2}} \\ \Lambda_i^0 &= -\frac{v_i}{\sqrt{1-v^2}} = \Lambda_0^i \\ \Lambda_j^i &= \delta_{ij} + \left(\frac{1}{\sqrt{1-v^2}} - 1\right) \frac{v_i v_j}{v^2}.\end{aligned}\quad (2.34)$$

很显然, 当速度 $\mathbf{v} \rightarrow 0$ 时, 这种洛伦兹变换就退化为恒等变换, 所以它可以与恒等变换连续过渡, 从而必定是正洛伦兹变换。实际上, 任何正洛伦兹变换都可以表示成以上洛伦兹变换和空间旋转变换 R 的乘积。

利用上面的洛伦兹变换, 我们可以计算一下两个事件在不同参考系中的时间差, 即有

$$\Delta t' = \frac{\Delta t - v|\Delta \mathbf{x}_{\parallel}|}{\sqrt{1-v^2}} = \Delta t \left(\frac{1 - v \frac{|\Delta \mathbf{x}_{\parallel}|}{\Delta t}}{\sqrt{1-v^2}} \right). \quad (2.35)$$

如果两事件类时相间, 则有 $|\Delta \mathbf{x}_{\parallel}| \leq |\Delta \mathbf{x}| < |\Delta t|$, 即 $\frac{|\Delta \mathbf{x}_{\parallel}|}{|\Delta t|} < 1$, 从而

$$1 - v \frac{|\Delta \mathbf{x}_{\parallel}|}{\Delta t} \geq 1 - v \frac{|\Delta \mathbf{x}_{\parallel}|}{|\Delta t|} \geq 1 - \frac{|\Delta \mathbf{x}_{\parallel}|}{|\Delta t|} > 0. \quad (2.36)$$

从而, 对于这种类时相间的情况, (2.35)式子左边 $\Delta t'$ 的正负符号与右边 Δt 的正负符号相同, 即是说, 这时候洛伦兹变换不会改变事件的先后顺序, 也就是会保持事件之间的因果关系! 这个结论当然是我们已经知道了的。并且, 从上面的讨论也可以看出, 如果两事件类空相间, 那它们之间的先后顺序就是相对的, 在不同的参考系看来可能有不同的结论。

2.2 四维时空的矢量和张量

2.2.1 四维矢量和张量

上一节说过, 在不同的参考系中, 时空坐标按照下面的洛伦兹变换而

变换,

$$dx'^{\mu} = \Lambda^{\mu}_{\nu} dx^{\nu}, \quad (2.37)$$

式中变换矩阵 Λ 满足

$$\Lambda^T \eta \Lambda = \eta \Leftrightarrow \Lambda^{-1} = \eta^{-1} \Lambda^T \eta. \quad (2.38)$$

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

$$A'^{\mu} = \Lambda^{\mu}_{\nu} A^{\nu}, \quad (2.39)$$

则我们称 A^{μ} 为一个四维时空的矢量, 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

$$A^2 = \eta_{\mu\nu} A^{\mu} A^{\nu} = -(A^0)^2 + \mathbf{A}^2 = -(A^0)^2 + (A^1)^2 + (A^2)^2 + (A^3)^2. \quad (2.40)$$

很明显, A^2 在洛伦兹变换下是不变的。

我们也可以定义下指标的四分量量 A_{μ} 为,

$$A_{\mu} = \eta_{\mu\nu} A^{\nu}, \quad (2.41)$$

写得更清楚一点就是

$$A_{\mu} = (A_0, A_1, A_2, A_3) = (-A^0, A^1, A^2, A^3) = (A_0, \mathbf{A}). \quad (2.42)$$

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

$$A'_{\mu} A'^{\mu} = A_{\mu} A^{\mu}. \quad (2.43)$$

注意到 A^{μ} 在洛伦兹变换下按照(2.39)式变换, 因此上式就意味着 A_{μ} 必然按照下式变换

$$A'_{\mu} = (\Lambda^{-1})^{\nu}_{\mu} A_{\nu}. \quad (2.44)$$

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

$$A^\mu = \eta^{\mu\nu} A_\nu. \quad (2.45)$$

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

$$\partial'_\mu = (\Lambda^{-1})^\nu_\mu \partial_\nu. \quad (2.46)$$

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。两个四矢量的内积是洛伦兹不变的，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

$$B'^{\mu\nu} = \Lambda^\mu_\alpha \Lambda^\nu_\beta B^{\alpha\beta}, \quad (2.47)$$

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，

$$B'_{\mu\nu} = (\Lambda^{-1})^\alpha_\mu (\Lambda^{-1})^\beta_\nu B_{\alpha\beta}. \quad (2.48)$$

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，

$$B'^\mu_\nu = \Lambda^\mu_\alpha (\Lambda^{-1})^\beta_\nu B^\alpha_\beta. \quad (2.49)$$

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, 这就叫做张量的缩并。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 $C_{\mu_1\mu_2\ldots\mu_p}$, 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, $C_{\mu_1\mu_2\ldots\mu_p}$ 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 $C_{\mu_1\mu_2\ldots\mu_p} = 0$, 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。

2.3 相对性原理与经典场论

现在我们可以将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。因此本书要考察的经典场论当然也要满足这种洛伦兹不变性, 进一步, 从第一章我们已经知道, “粒子+场”系统的规律(也就是运动方程)可以由最小作用量原理导出, 因此这就意味着“多粒子+场”系统的作用量泛函必须在洛伦兹变换下保持不变! 这就意味着这些作用量泛函必须是洛伦兹标量。

从第一章我们知道“多粒子+场”系统的作用量可以分成三部分,

$S_{\text{粒子}}$, $S_{\text{相互作用}}$, 以及 $S_{\text{场}}$ 。根据上面的讨论, 每一部分都得在洛伦兹变换之下保持不变。下面我们先来讨论 $S_{\text{粒子}}$ 。

2.3.1 相对论粒子的作用量以及相互作用

相对论不变性告诉我们, 相对论粒子的作用量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(因为 $c = 1$), 从而我们知道, 相对论粒子的作用量必定可以写成

$$S_{\text{粒子}}[x(s)] = -m \int d\tau = -m \int ds \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{ds} \frac{dx^\nu}{ds}}. \quad (2.50)$$

式中 s 为粒子运动路径的参数, 第一个式子的负号有多个理解角度, 这里我们想指出的是, 有了这个负号, 这个作用量的非相对论极限才会对, 请读者在待会儿我们讨论非相对论极限时留心。

值得指出的是, 上述作用量(2.50)显然具有重参数不变性, 即在 $s \rightarrow \tilde{s}(s)$ 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参数 s 。最常见的选择有两种, 第一种是, 取 $s = \tau$, 即取固有时本身为路径的参数, 从而作用量(2.50)可以写成

$$S_{\text{粒子}}[x(s)] = -m \int d\tau = -m \int d\tau \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}}. \quad (2.51)$$

第二种选择是取 $s = x^0 = t$, 即取通常的时间坐标为参数, 这时候注意到

$$d\tau^2 = dt^2 - d\mathbf{x}^2 = dt^2(1 - \mathbf{v}^2). \quad (2.52)$$

式中 $\mathbf{v} = \frac{d\mathbf{x}}{dt}$ 是粒子的速度。从而就可以将作用量(2.50)写成

$$S_{\text{粒子}}[x(s)] = -m \int dt \sqrt{1 - \mathbf{v}^2}. \quad (2.53)$$

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(2.53)近似成

$$S_{\text{粒子}}[x(s)] = -m \int dt + \int dt \frac{1}{2} m \mathbf{v}^2 + \dots \quad (2.54)$$

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！

为了进一步考察 $S_{\text{相互作用}}$ ，我们假设与粒子耦合的场是一个洛伦兹标量场 $\phi(x)$ (x 表示时空点)，很显然，最简单的 $S_{\text{相互作用}}$ 具有如下形式

$$S_{\text{相互作用}} = \int d\tau \phi(x(\tau)), \quad (2.55)$$

式中的积分沿着粒子的世界线进行。

对 $S_{\text{粒子}} + S_{\text{相互作用}}$ 中的粒子坐标 $x^\mu(\tau)$ 进行变分，并应用最小作用量原理，就能得到粒子在场作用下的运动方程。具体推导如下

$$\begin{aligned} \delta S_{\text{粒子}} + \delta S_{\text{相互作用}} &= \int (-m + \phi(x(\tau))) \delta(d\tau) + \int d\tau \delta\phi \\ &= \int (-m + \phi(x(\tau))) \frac{1}{2} \frac{\delta(d\tau^2)}{d\tau} + \int d\tau \partial_\mu \phi \delta x^\mu \\ &= \int (m - \phi(x(\tau))) \frac{1}{2} \delta(\eta_{\mu\nu} dx^\mu dx^\nu) / d\tau + \int d\tau \partial_\mu \phi \delta x^\mu \\ &= \int (m - \phi(x(\tau))) \eta_{\mu\nu} \frac{dx^\mu}{d\tau} d(\delta x^\nu) + \int d\tau \partial_\mu \phi \delta x^\mu \\ &= \int d\tau \frac{d}{d\tau} [(m - \phi(x(\tau))) \eta_{\mu\nu} \frac{dx^\mu}{d\tau}] \delta x^\nu + \int d\tau \partial_\nu \phi \delta x^\nu. \end{aligned} \quad (2.56)$$

最后一行我们进行了分部积分。根据最小作用量原理，可以得到如下关于粒子的运动方程

$$(m - \phi) \eta_{\mu\nu} \frac{d^2 x^\mu}{d\tau^2} = \partial_\rho \phi \eta_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\rho}{d\tau} + \partial_\nu \phi. \quad (2.57)$$

适当升降指标，也即是

$$(m - \phi) \frac{d^2 x^\mu}{d\tau^2} = \partial_\nu \phi \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} + \partial^\mu \phi. \quad (2.58)$$

这个方程告诉我们，场的存在会对粒子的运动产生影响，也就是产生力的作用。

2.3.2 场的作用量

下面我们讨论场本身的作用量 $S_{\text{场}}$ 。由于场是一个时空中的分布，而不是和粒子一样仅集中在一根世界线上，因此这里有一个重要的假设(通常称

之为局域性原理), 即作用量 $S_{\text{场}}$ 总可以写成某拉格朗日密度 \mathcal{L} 对整个时空体积的积分, 即 $S_{\text{场}} = \int d^4x \mathcal{L}$, 注意到体积元 d^4x 是洛伦兹不变的, 因此 $S_{\text{场}}$ 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!

进一步, 假设我们考虑的是一个标量场论, 场变量记为 ϕ , 则局域性原理还假设拉氏密度是 ϕ 和 $\partial_\mu \phi$ 的函数, 记为 $\mathcal{L} = \mathcal{L}(\phi, \partial_\mu \phi)$ 。

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

$$\partial_\mu \phi \partial^\mu \phi = \eta^{\mu\nu} \partial_\mu \phi \partial_\nu \phi = -(\partial_t \phi)^2 + (\nabla \phi)^2. \quad (2.59)$$

要求动能项为正, 我们总能将 $\partial_\mu \phi$ 对拉氏密度最简单的贡献写作

$$-\frac{1}{2e^2} \partial_\mu \phi \partial^\mu \phi. \quad (2.60)$$

式中 $\frac{1}{2e^2}$ 是某个正的比例常数。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

$$\begin{aligned} \mathcal{L} &= -\frac{1}{2e^2} \partial_\mu \phi \partial^\mu \phi - \mathcal{U}(\phi) \\ &= \frac{1}{2e^2} [(\partial_t \phi)^2 - (\nabla \phi)^2] - \mathcal{U}(\phi). \end{aligned} \quad (2.61)$$

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,

$$\mathcal{L} = -\frac{1}{2} g(\phi) \partial_\mu \phi \partial^\mu \phi - \mathcal{U}(\phi), \quad (2.62)$$

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 $g_0 = 1/e^2$, 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

$$\mathcal{L} = -\frac{1}{2} \partial_\mu \phi \partial^\mu \phi + f(\phi) (\partial_\mu \partial_\nu \phi) (\partial^\mu \partial^\nu \phi). \quad (2.63)$$

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。

以上面最简单的模型为例，这时候整个“多粒子+场”的系统总的作用量可以写成

$$S = - \sum_n m_n \int d\tau_n + \sum_n \int d\tau_n \phi(x_n(\tau_n)) - \int d^4x \left[\frac{1}{2e^2} \partial_\mu \phi \partial^\mu \phi + \mathcal{U}(\phi) \right].$$

式中指标 n 表示第 n 个粒子。

为了利用最小作用量原理导出场所满足的运动方程，我们不妨先改写一下作用量中的相互作用部分 $\sum_n \int d\tau_n \phi(x_n(\tau_n))$ 。为此，我们得引入狄拉克德尔塔函数 $\delta(y)$ ，我们定义四维时空中的德尔塔函数

$$\delta^4(x) \equiv \delta(x^0)\delta(x^1)\delta(x^2)\delta(x^3). \quad (2.64)$$

显然满足 $\int d^4x \delta^4(x - x') \phi(x) = \phi(x')$ 。进而有

$$\sum_n \int d\tau_n \phi(x_n(\tau_n)) = \int d^4x \phi(x) \sum_n \int d\tau_n \delta^4(x - x_n(\tau)) \equiv \int d^4x \phi(x) J(x).$$

式中 $J(x) \equiv \sum_n \int d\tau_n \delta^4(x - x_n(\tau))$ 。

进而将整个“粒子+场”系统的总作用量 S 对场 $\phi(x)$ 进行变分，即有

$$\begin{aligned} \delta S &= \int d^4x J(x) \delta \phi(x) - \int d^4x \left[\frac{1}{e^2} \partial^\mu \phi \partial_\mu \delta \phi + \delta \mathcal{U}(\phi) \right] \\ &= \int d^4x \left[\frac{1}{e^2} \partial_\mu \partial^\mu \phi - \mathcal{U}'(\phi) + J(x) \right] \delta \phi. \end{aligned} \quad (2.65)$$

第二行我们进行了分部积分，并丢掉了边界项¹，式中 \mathcal{U}' 表示 $\mathcal{U}(\phi)$ 对自变量 ϕ 的导数。进而由最小作用量原理(即对于满足运动方程的场，作用量的变分为零)，可以得到如下场方程

$$\frac{1}{e^2} \partial_\mu \partial^\mu \phi - \mathcal{U}'(\phi) + J(x) = 0. \quad (2.66)$$

¹在非相对论力学中，最小作用量原理要求广义坐标在路径的端点处固定，即在路径的端点处变分为零。这里也一样，我们要求场的取值在时空的边界处是固定的，即边界处场的变分为零，从而可以丢掉边界项。

其中的 $J(x)$ 根据定义是由粒子贡献的，所以这个方程告诉我们，粒子可以在时空中激发起场。

综合粒子的运动方程(2.58)和场的运动方程(2.66)我们可以看到，粒子的存在会在时空中激发起场，反过来，场又会对粒子施加作用力，进而影响粒子的运动。

以上只考虑了粒子耦合到一个标量场的系统。粒子当然也有可能与一个矢量场 A_μ 进行耦合，相应的“粒子+场”系统我们放在后面的章节中进行讨论。

第三章 电磁场

上一章考察了粒子与一个标量场的耦合，本章我们考察多粒子通过一个洛伦兹矢量场 $A_\mu(x)$ 进行相互作用的系统，通常称这样的系统为带电粒子系统。当然，其中自由粒子的作用量 $S_{\text{粒子}}$ 与上一章给出的完全一样。

3.1 规范不变性

依然是先考察粒子与 A_μ 耦合的作用量 $S_{\text{相互作用}}$ 。先考察一个粒子与 A_μ 的耦合。首先 A_μ 必须和一个洛伦兹矢量收缩才能形成一个洛伦兹标量，而由粒子的运动坐标构造出来的洛伦兹矢量只有 $\frac{dx^\mu}{d\tau}$ ，其中 τ 为粒子的固有时。由此可知，最简单的 $S_{\text{相互作用}}$ 可以写成

$$S_{\text{相互作用}} = \int d\tau A_\mu(x(\tau)) \frac{dx^\mu}{d\tau} = \int A_\mu(x(\tau)) dx^\mu, \quad (3.1)$$

式中的积分同样是沿着粒子的世界线进行。

以上 $S_{\text{相互作用}}$ 意味着系统有一种额外的对称性，即在如下变换下保持不变

$$A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \epsilon(x). \quad (3.2)$$

式中 $\epsilon(x)$ 是一个任意的标量函数。验证很简单，在上面的变换之下， $S_{\text{相互作用}}$ 将变换为

$$S_{\text{相互作用}} \rightarrow S_{\text{相互作用}} + \int_i^f \partial_\mu \epsilon(x) dx^\mu = S_{\text{相互作用}} + \epsilon(f) - \epsilon(i). \quad (3.3)$$

式中 i, f 分别表示粒子运动路径(世界线)的两个端点。很显然，变换下多出来的部分只影响粒子运动路径的端点，而根据经典力学中的最小作用量原

理,端点在变分时是要保持固定的。因此,这些变换下多出来的部分对于粒子的变分没有任何影响,从而也就不改变粒子的运动微分方程,因此也就是系统的对称性。

通常称(3.2)式的对称变换为规范变换,相应的对称性为规范对称性,并称相应的 A_μ 为规范场。并且,进一步,我们要求这样的规范对称性在场本身的作用量 $S_{\text{场}}$ 中也得以保持。不过,在详细讨论场本身的作用量之前,不妨让我们把上面的 $S_{\text{相互作用}}$ 推广到多粒子情形,结果是

$$S_{\text{相互作用}} = \sum_a n_a \int d\tau_a A_\mu(x_a(\tau_a)) \frac{dx_a^\mu}{d\tau_a} = \sum_a n_a \int A_\mu(x_a(\tau_a)) dx_a^\mu, \quad (3.4)$$

式中 n_a 为第 a 个粒子对耦合项的贡献的权重系数,稍后我们将证明 n_a 必定为整数。

将作用量(3.4)对规范势 A_μ 变分,即可得到

$$\begin{aligned} \delta S_{\text{相互作用}} &= \sum_a n_a \int d\tau_a \frac{dx_a^\mu}{d\tau_a} \delta A_\mu \\ &= \int d^4x \left[\sum_a n_a \int d\tau_a \frac{dx_a^\mu}{d\tau_a} \delta^4(x - x_a(\tau_a)) \right] \delta A_\mu. \end{aligned} \quad (3.5)$$

通常人们定义带电粒子系统的电流四矢量 J^μ ,为

$$\begin{aligned} J^\mu(x) &= \sum_a n_a \int d\tau_a \frac{dx_a^\mu}{d\tau_a} \delta^4(x - x_a(\tau_a)) \\ &= \sum_a n_a \int ds_a \frac{dx_a^\mu}{ds_a} \delta^4(x - x_a(s_a)). \end{aligned} \quad (3.6)$$

式中第二行,我们将世界线参数从固有时换成了任意的参数 s_a (请读者想想为什么可以换?)。进而即有

$$\delta S_{\text{相互作用}} = \int d^4x J^\mu \delta A_\mu. \quad (3.7)$$

通常称 $J^0(x) \equiv \rho(x)$ 为带电粒子的电荷密度,称 $\mathbf{J} = (J^1, J^2, J^3)$ 为带电粒子的电流密度。为了看清楚(3.6)式的物理意义,我们取世界线参数 $s_a = x_a^0$,并做出(3.6)式中的积分,则有

$$\begin{aligned} \rho &= \sum_a n_a \delta^3(\mathbf{x} - \mathbf{x}_a(t)) \\ \mathbf{J} &= \sum_a n_a \mathbf{v}_a \delta^3(\mathbf{x} - \mathbf{x}_a(t)). \end{aligned} \quad (3.8)$$

3.2 “发现” 麦克斯韦方程

为了找到规范场 A_μ 本身的作用量, 和标量场情形一下, 我们首先假设有局域性原理, 即 $S_{\text{场}}$ 可以写成 $S_{\text{场}} = \int d^4x \mathcal{L}(A_\rho, \partial_\mu A_\nu)$ 这样的形式。其次, 更重要的是要注意到这样的作用量必须保持规范不变性, 即它得在 $A_\mu \rightarrow A_\mu + \partial_\mu \epsilon$ 的规范变换下保持不变! 同时, 当然, A_μ 的作用量还得洛伦兹不变, 即得是洛伦兹标量。

为了满足规范不变性的要求, 我们注意到如下反对称张量 $F_{\mu\nu}$ 在 $A_\mu \rightarrow A_\mu + \partial_\mu \epsilon$ 的规范变换下是不变的,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \quad (3.9)$$

因此, 我们只要用 $F_{\mu\nu}$ 而不是直接用 A_μ 来构造作用量, 就能自动满足规范不变性的要求。

利用 $F_{\mu\nu}$ 构造出来的最简单的洛伦兹标量如下,

$$F_{\mu\nu} F^{\mu\nu}, \quad (3.10)$$

因此, 最简单的规范场作用量可以写成如下形式,

$$S_{\text{场}} = - \int d^4x \frac{1}{4e^2} F_{\mu\nu} F^{\mu\nu}, \quad (3.11)$$

式中 $\frac{1}{4e^2} > 0$ 为一个实的比例常数, 式中的负号是为了使得 A_i ($i=1,2,3$)场的动能项(即 $(\partial_t A_i)^2$ 项)为正。

将上面这个作用量对 A_μ 变分, 可得

$$\begin{aligned} \delta S_{\text{场}} &= - \int d^4x \frac{1}{2e^2} F^{\mu\nu} \delta F_{\mu\nu} = - \frac{1}{e^2} \int d^4x F^{\mu\nu} \partial_\mu \delta A_\nu \\ &= \frac{1}{e^2} \int d^4x \partial_\mu F^{\mu\nu} \delta A_\nu. \end{aligned} \quad (3.12)$$

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)。

习惯上, 人们常常重新定义 A_μ 为 $A_\mu \rightarrow eA_\mu$, 进而即可以将 $S_{\text{相互作用}}$ 重写为

$$S_{\text{相互作用}} = \sum_a n_a e \int d\tau_a A_\mu(x_a(\tau_a)) \frac{dx_a^\mu}{d\tau_a}, \quad (3.13)$$

同时将 $S_{\text{场}}$ 重写为

$$S_{\text{场}} = - \int d^4x \frac{1}{4} F_{\mu\nu} F^{\mu\nu}. \quad (3.14)$$

习惯上, 人们依然保持 $\delta S_{\text{相互作用}} = \int d^4x J^\mu \delta A_\mu$ 成立, 并记 $n_a e \equiv q_a$, 从而这时候

$$J^\mu(x) = \sum_a q_a \int d\tau_a \frac{dx_a^\mu}{d\tau_a} \delta^4(x - x_a(\tau_a)). \quad (3.15)$$

相应的有

$$\begin{aligned} \rho &= \sum_a q_a \delta^3(\mathbf{x} - \mathbf{x}_a(t)) \\ \mathbf{J} &= \sum_a q_a \mathbf{v}_a \delta^3(\mathbf{x} - \mathbf{x}_a(t)). \end{aligned} \quad (3.16)$$

很显然, 如果将 $q_a = n_a e$ 理解为粒子的电荷量, 那这个结果正符合我们对电荷密度以及电流密度表达式的预期。特别的, 将电荷密度 $\rho(x)$ 在一个空间区域 V 上积分, 则有

$$\int_V d^3\mathbf{x} \rho(x) = \sum_{a \in V \text{ 内}} q_a. \quad (3.17)$$

即 ρ 对体积的积分, 即是区域内所有带电粒子的总电量。

进而即有 $S_{\text{相互作用}} + S_{\text{场}}$ 对 A_μ 的变分为

$$\delta[S_{\text{相互作用}} + S_{\text{场}}] = \int d^4x [\partial_\mu F^{\mu\nu} + J^\nu] \delta A_\nu. \quad (3.18)$$

根据最小作用量原理即可得规范场 A_μ 满足的运动微分方程,

$$-\partial_\mu F^{\mu\nu} = J^\nu. \quad (3.19)$$

另外, 根据 $F_{\mu\nu}$ 的定义(3.9), 很容易验证它还必然满足如下方程

$$\partial_\mu F_{\nu\rho} + \partial_\nu F_{\rho\mu} + \partial_\rho F_{\mu\nu} = 0. \quad (3.20)$$

以上两个方程, 其实只有方程(3.19)是真正的 A_μ 场的动力学微分方程, 因为根据 $F_{\mu\nu}$ 场的定义式(3.9), 方程(3.20)其实是一个恒等式, 是自动成立的。反过来, 满足方程(3.20)的 $F_{\mu\nu}$ 场也必定可以写成(3.9)式的形式。

特别的, 注意到 $\partial_\mu \partial_\nu$ 关于 μ, ν 是对称的, 而 $F^{\mu\nu}$ 关于 μ, ν 是反对称的, 从而 $\partial_\mu \partial_\nu F^{\mu\nu} = 0$ (请想一想为什么?), 从而根据(3.19)式即有

$$\partial_\nu J^\nu = 0, \quad \Leftrightarrow \partial_t \rho + \nabla \cdot \mathbf{J} = 0. \quad (3.21)$$

这就是著名的描述电荷守恒的连续性方程。

为了看清楚连续性方程的物理含义, 我们把它对一个空间区域 V 积分, 从而即有

$$-\frac{\partial}{\partial t} \int_V \rho(x) d^3\mathbf{x} = \int_V (\nabla \cdot \mathbf{J}) d^3\mathbf{x} = \int_{\partial V} \mathbf{J} \cdot d\mathbf{S}. \quad (3.22)$$

式中 ∂V 表示区域 V 的边界, 它是一个封闭曲面, $d\mathbf{S}$ 表示这个边界表面上的面积元矢量。式中的最后一个等于号利用了数学中的高斯定理。由于 $\int_V \rho(x) d^3\mathbf{x}$ 即是区域 V 内所有带电粒子的总电量, 因此上面这个结果告诉我们的即是, 区域 V 内电量随时间的减少率(负号表示减少)等于单位时间通过区域边界 ∂V 流出去的电荷量, 也即是 $\int_{\partial V} \mathbf{J} \cdot d\mathbf{S}$ 。

换言之, 电荷不能凭空消失, 也不能凭空增加, 它只能被转移! 特别的, 如果区域 V 是全空间, ∂V 是其无穷远边界, 则由于有限时间之内带电粒子不可能跑到无穷远处去, 从而 \mathbf{J} 在无穷远边界处等于零, 从而即有: 全空间的总电量是不随时间变化的, 是一个守恒量。

令人吃惊的是, 方程(3.19)和方程(3.20) 其实就是用一种更优雅的方式写出来的关于电磁场的麦克斯韦方程组! 为了和更常见的麦克斯韦方程组形式联系起来, 对于 $i, j, k = 1, 2, 3$, 我们定义

$$F^{0i} = E_i, \quad F^{ij} = \epsilon^{ijk} B_k. \quad (3.23)$$

注意, 这里也用了求和约定, 式中 ϵ^{ijk} 关于三个指标是全反对称的(即交换任意两个指标都会多出一个负号), 且 $\epsilon^{123} = 1$ 。特别的, 根据 $F_{\mu\nu}$ 的定义(3.9), 可以得到 \mathbf{B} 与矢量势 \mathbf{A} ($(\mathbf{A})_i = A_i$)的关系式

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (3.24)$$

用 \mathbf{E} 场和 \mathbf{B} 场表达 $F_{\mu\nu}$ 以后, 进而由方程(3.19)有

$$\begin{aligned} -\partial_t F^{0j} - \partial_i F^{ij} &= J^j \\ \Rightarrow -\partial_t E_j - \epsilon^{ijk} \partial_i B_k &= J^j \\ \Rightarrow -\partial_t \mathbf{E} + \nabla \times \mathbf{B} &= \mathbf{J} \end{aligned} \quad (3.25)$$

或者写成更常见的形式,

$$\nabla \times \mathbf{B} = \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}. \quad (3.26)$$

类似的, 由方程(3.19)也有

$$-\partial_i F^{i0} = J^0 \Rightarrow \partial_i F^{0i} = \rho \Rightarrow \nabla \cdot \mathbf{E} = \rho. \quad (3.27)$$

而由方程(3.20)有

$$\partial_1 F_{23} + \partial_2 F_{31} + \partial_3 F_{12} = 0 \Rightarrow \nabla \cdot \mathbf{B} = 0. \quad (3.28)$$

同样由方程(3.20), 也有

$$\begin{aligned} \partial_0 F_{ij} + \partial_i F_{j0} + \partial_j F_{0i} &= 0 \\ \Rightarrow \partial_t F_{ij} + \partial_i F^{0j} - \partial_j F^{0i} &= 0. \end{aligned} \quad (3.29)$$

与 $\frac{1}{2}\epsilon_{ijk}$ 进行指标收缩, 即有

$$\partial_t B_k + (\nabla \times \mathbf{E})_k = 0 \Rightarrow \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \quad (3.30)$$

方程(3.26)、(3.27)、(3.28)、(3.30)就是人们熟悉的麦克斯韦方程组的四个方程! \mathbf{E} 就是电场强度, \mathbf{B} 就是磁场强度。当然, 它们和国际单位制中的麦克斯韦方程组还有些微的系数出入, 但这完全是单位制的问题, 因为我们这里采用的不是国际单位制。不妨把这四个方程概括在下面

$$\nabla \cdot \mathbf{E} = \rho, \quad \nabla \cdot \mathbf{B} = 0 \quad (3.31)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \times \mathbf{B} = \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t}. \quad (3.32)$$

将上面四个方程中的前两个在空间区域 V 上积分, 并利用数学中的高斯定理, 即有

$$\int_{\partial V} \mathbf{E} \cdot d\mathbf{S} = Q_{in}, \quad \int_{\partial V} \mathbf{B} \cdot d\mathbf{S} = 0. \quad (3.33)$$

式中 Q_{in} 表示区域 V 内的电荷总量。这两个方程, 前者叫做电场的高斯定理, 后者叫做磁场的高斯定理。根据电场的高斯定理, 一个区域内部的电

荷总量等于电场在区域边界上的总通量(简称电通量)。而根据磁场的高斯定理, 任何一个闭合曲面 ∂V 上磁场的总通量(简称磁通量)都等于零, 也就是说, 磁场在某地方通入就必定会在另一地方通出, 通入为负, 通出为正。

对于麦克斯韦方程组的第三个方程, 我们将它在一个有限曲面 S (S 不是闭合曲面)上进行面积分, 并利用数学中的斯托克斯公式, 即有

$$\oint_{\partial S} \mathbf{E} \cdot d\mathbf{l} = \int_S (\nabla \times \mathbf{E}) \cdot d\mathbf{S} = -\frac{\partial}{\partial t} \int_S \mathbf{B} \cdot d\mathbf{S}. \quad (3.34)$$

式中 ∂S 表示曲面 S 的边界, 它是一条闭合回路。这个结果告诉我们, 电场强度在一条闭合回路上的积分(也即是闭合回路上的电动势)等于回路所包围面积上磁通量变化率的负值。根据这个结果可知, 磁场的变化会在它周围感应出电场(磁生电)。实际上这就是著名的法拉第电磁感应定律, 即: 闭合回路磁通的变化率等于回路上的感应电动势。

类似的, 将麦克斯韦方程组的最后一个方程在一个有限曲面 S 上进行面积分, 并利用数学中的斯托克斯公式, 即有

$$\oint_{\partial S} \mathbf{B} \cdot d\mathbf{l} = \int_S \mathbf{J} \cdot d\mathbf{S} + \frac{\partial}{\partial t} \int_S \mathbf{E} \cdot d\mathbf{S}. \quad (3.35)$$

这个结果告诉我们, 变化的电场同样也在它周围感应出磁场(电生磁)。不过, 与法拉第电磁感应定律不同, 这个结果最早并不是由实验现象归纳而来的, 而是麦克斯韦的天才理论发现, 麦克斯韦称变化的电场 $\frac{\partial \mathbf{E}}{\partial t}$ 为“位移电流”。早期的实验精度不足以揭示这个现象, 这是因为, 在通常的条件下, 电现象比磁现象要容易得多, 磁生电比较容易观察到, 但是电生磁就太微弱而难以观测了。特别的, 如果电场 E 是不随时间变化的恒定电场, 那上面的结果就退化成

$$\oint_{\partial S} \mathbf{B} \cdot d\mathbf{l} = \int_S \mathbf{J} \cdot d\mathbf{S} = I_{in}. \quad (3.36)$$

这就是安培环路定理, 式中的 I_{in} 就是穿过回路 ∂S 的总电流强度。

就这样, 如果要求粒子通过一个洛伦兹矢量场进行相互作用, 并要求规范对称性, 我们最终就“独立发现”了著名的麦克斯韦方程! 这说明, 电磁场起源于规范对称性, 电场和磁场是一个整体, 它们共同构成二阶反对称张量场 $F_{\mu\nu}$, 因此电场和磁场在洛伦兹变换下会相互转变。

为了得到带电粒子在 A_μ 场中的运动微分方程，我们将作用量 $S_{\text{粒子}} + S_{\text{相互作用}}$ 对粒子坐标变分，其中 $S_{\text{粒子}}$ 对粒子坐标的变分与第二章中的相关讨论完全类似，这里不再重复。关键是要将 $S_{\text{相互作用}}$ ，即与规范势 A_μ 耦合的项对粒子坐标进行变分，则有

$$\begin{aligned}
 \sum_a q_a \delta \int A_\mu dx_a^\mu &= \sum_a q_a \int \delta A_\nu dx_a^\nu + \sum_a q_a \int A_\mu d(\delta x_a^\mu) \\
 &= \sum_a q_a \int (\partial_\mu A_\nu \delta x_a^\mu dx_a^\nu - dA_\mu \delta x_a^\mu) \\
 &= \sum_a q_a \int (\partial_\mu A_\nu \delta x_a^\mu dx_a^\nu - \partial_\nu A_\mu dx_a^\nu \delta x_a^\mu) \\
 &= \sum_a q_a \int d\tau_a F_{\mu\nu} \frac{dx_a^\nu}{d\tau_a} \delta x_a^\mu. \tag{3.37}
 \end{aligned}$$

其中第二个等号进行了分部积分，并注意到边界项等于零。从而即可得到，

$$\delta[S_{\text{粒子}} + S_{\text{相互作用}}] = \sum_a \int d\tau_a \left[-m_a \frac{d^2 x_a^\mu}{d\tau_a^2} + q_a F^\mu{}_\nu \frac{dx_a^\nu}{d\tau_a} \right] \delta x_{a\mu}. \tag{3.38}$$

根据最小作用量原理，即可得粒子的运动微分方程，为

$$m_a \frac{d^2 x_a^\mu}{d\tau_a^2} = q_a F^\mu{}_\nu \frac{dx_a^\nu}{d\tau_a}. \tag{3.39}$$

很显然，假如定义 $p^\mu = (p^0, \mathbf{p}) = m \frac{dx^\mu}{d\tau}$ ，则方程(3.39)也可以写作

$$\frac{dp_a^\mu}{dt} = q_a F^\mu{}_\nu \frac{dx_a^\nu}{dt}. \tag{3.40}$$

从这里读者不难得到

$$\frac{d\mathbf{p}_a}{dt} = q_a (\mathbf{E} + \mathbf{v}_a \times \mathbf{B}). \tag{3.41}$$

这正是通常所谓的洛伦兹力公式。

3.3 电磁场携带能量

电磁场并不是一种数学虚构，而是一种物理实体，虽然它不同于粒子，但它也同样携带能量。不过粒子的能量集中在一个点上，而电磁场的能量

则是分布在空间中的，因为电磁场本身就是空间分布。所以，与粒子不同，对于电磁场，我们可以讨论每单位体积内的能量，称之为能量密度，记作 $w(x)$ 。

为了说服你电磁场携带能量，我们不妨考虑 $J^\mu = 0$ 的这种特殊情形，也就是电磁场远离电荷与电流而在真空中传播的情形。这时候麦克斯韦的四个方程可以写为

$$\begin{aligned}\nabla \cdot \mathbf{E} &= 0, & \nabla \cdot \mathbf{B} &= 0 \\ \nabla \times \mathbf{B} &= \frac{\partial \mathbf{E}}{\partial t}, & -\nabla \times \mathbf{E} &= \frac{\partial \mathbf{B}}{\partial t}.\end{aligned}\quad (3.42)$$

我们可以把后两个方程分别点乘以 \mathbf{E} 和 \mathbf{B} ，并将两式加起来，则有

$$\mathbf{E} \cdot \frac{\partial \mathbf{E}}{\partial t} + \mathbf{B} \cdot \frac{\partial \mathbf{B}}{\partial t} = \mathbf{E} \cdot (\nabla \times \mathbf{B}) - \mathbf{B} \cdot (\nabla \times \mathbf{E}). \quad (3.43)$$

利用数学公式 $\nabla \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\nabla \times \mathbf{a}) - \mathbf{a} \cdot (\nabla \times \mathbf{b})$ ，即有

$$\frac{\partial}{\partial t} \left[\frac{1}{2} (\mathbf{E}^2 + \mathbf{B}^2) \right] = -\nabla \cdot [\mathbf{E} \times \mathbf{B}]. \quad (3.44)$$

引入定义

$$w \equiv \frac{1}{2} (\mathbf{E}^2 + \mathbf{B}^2), \quad \mathbf{Y} \equiv \mathbf{E} \times \mathbf{B}. \quad (3.45)$$

则刚才的结果可以简写为

$$\frac{\partial w}{\partial t} + \nabla \cdot \mathbf{Y} = 0. \quad (3.46)$$

类比于电荷守恒的相关讨论可知，这就是一个连续性方程，它反映的是某种守恒定律。这种守恒定律就是能量守恒定律， w 就称作电磁场的能量密度， \mathbf{Y} 就称作电磁场的能流密度，也称作坡印廷矢量。所以，电磁场是携带能量的，可以定义它的能量密度。

读者可能会说，但是你这里的讨论都是忽略了电荷与电流的存在的啊，也就是说，是忽略了带电粒子的啊，如果把带电粒子也包括进行情况会怎么样呢？答案是，这时候能量守恒将表现如下

$$-\frac{\partial}{\partial t} [\text{区域} V \text{内带电粒子的总能量}] + \int_V d^3\mathbf{x} w(x) = \int_{\partial V} \mathbf{Y} \cdot d\mathbf{S}. \quad (3.47)$$

这个式子的物理意义是显然的，这里不再作展开。至于这个式子的证明，作为一个有点挑战的问题，我们把它留给读者。

实际上，不仅电磁场，任何传递相互作用的场都携带能量，不仅携带能量，它们还携带动量以及角动量。不过，限于篇幅，我们就不在这个初识讲义中进一步讨论了。读者可以参阅我的《经典场论新讲》。

3.4 光是一种电磁波

将 $F_{\mu\nu}$ 场的定义式(3.9)代入动力学方程(3.19)，即可得

$$-\partial_\mu \partial^\mu (A^\nu) + \partial^\nu (\partial_\mu A^\mu) = J^\nu. \quad (3.48)$$

而且，这个方程还可以进一步简化。简化的关键点就在于注意到 A_μ 场本身并非物理上可直接观测的，原因在于 A_μ 和它规范变换以后的 $A_\mu + \partial_\mu \epsilon$ 描述的是同样的物理内容，换言之，对于同样物理上可测量的内容，描述它的 A_μ 不唯一，而是可以相差一个规范变换。因此 A_μ 本身并不是物理的，真正物理的场是规范不变的 $F_{\mu\nu}$ 。

由于 A_μ 有一个规范变换的自由度，因此对于同样的物理内容，我们总可以用一个合适的函数 $\epsilon(x)$ 来对 A_μ 场进行一个合适的规范变换，使得 $\partial_\mu A^\mu + \partial_\mu \partial^\mu \epsilon = 0$ ，即是说，总可以使得变换之后新的 A_μ 场满足下式

$$\partial_\mu A^\mu = 0. \quad (3.49)$$

这个方程就称为一种规范选择，或者说规范固定，它有一个专门名称，称为洛伦兹规范条件，额外满足这个方程的 A_μ 场就称作洛伦兹规范下的 A_μ 场。将洛伦兹规范条件代入上面的方程(3.48)，即可得

$$-\partial_\mu \partial^\mu (A^\nu) = J^\nu. \quad (3.50)$$

特别的，对于真空中的自由电磁场，即电流四矢量 J^μ 等于零时的电磁场，方程(3.50)将变成，

$$-\partial_\mu \partial^\mu (A^\nu) = 0, \Leftrightarrow (\partial_t^2 - \nabla^2) A^\nu = 0. \quad (3.51)$$

显然，对于 A^μ 场的每一个分量，这个方程描述的都是一种以 $c = 1$ 的速度传播的波动，这就是电磁波！由于在我们的单位制中 $c = 1$ 正是速度上限，也

就是光速，而由于电磁波的传播速度和它一样，因此，麦克斯韦提出光就是一种电磁波！麦克斯韦的这一设想是对历史上人们长期迷惑不解的光的本性的一次深刻洞察！这一设想后来被无数的实验证实，并在今天成为人们熟知的基本事实。同时，由于 $c = 1$ 是相对论中的最大信息传播速度，所以这也证明光——或者更一般的电磁波——是以最大速度传播的。

3.5 电荷量子化

本节我们恢复到重定义 A_μ 为 eA_μ 的写法，从而粒子与 A_μ 耦合的作用量为(3.4)的形式。本节我们就是要证明，为什么其中的权重系数 n_a 必定为整数。不妨以单粒子为例，将这个 $S_{\text{相互作用}}$ 重写如下

$$S_{\text{相互作用}} = n \int A_\mu(x(\tau)) dx^\mu. \quad (3.52)$$

因此，这里就是要证明为什么式中的 n 必须为整数。为此，我们需要一点量子力学知识。为了简单起见，这里我们假设取了合适的单位制，以使得普朗克常数 $\hbar = 1$ 。

在量子力学中，粒子要用波函数来描述，一个位于时空点 x 的粒子，其波函数取值为 $\psi(x)$ ，这是一个复函数，其模方代表粒子在 x 点出现的概率。进一步，粒子从 x_i 点运动到 x_f 点，其概率由 $\psi^*(x_f)\psi(x_i)$ 决定，称之为概率幅。

但是，考虑到粒子与电磁场有耦合，这一耦合对概率幅是有修正的，典型的修正是要乘上一个 $e^{iS_{\text{相互作用}}}$ 的相因子。因此，在电磁场中，带电粒子从 x_i 点运动到 x_f 点，其概率幅 \mathcal{A} 将由下式给出

$$\mathcal{A} = \psi^*(x_f) e^{in \int_{x_i}^{x_f} A_\mu(x(\tau)) dx^\mu} \psi(x_i). \quad (3.53)$$

相应的概率由 $|\mathcal{A}|^2$ 给出。

另外，规范势 A_μ 本身不可直接观测，因为 A_μ 和规范变换之后的 $A_\mu + \partial_\mu \epsilon$ 在物理上等价，换言之，同一个物理可以对应多个不同的 A_μ ，因此 A_μ 当然不可以直接观测。但是规范场强 $F_{\mu\nu}$ 却是可观测的，因为它规范不变，在物理上是唯一的，当然，它相应于电场强度和磁场强度。总之，人们发现，所有的物理可观测量都得规范不变！特别的，上面的 $|\mathcal{A}|^2$ 作为概率当然可

以观测，因此它必须规范不变。由于只差一个模方，所以人们往往进一步要求，概率幅 \mathcal{A} 本身必须规范不变。

容易验证，在 $A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \epsilon(x)$ 的规范变换下，我们有如下变换

$$\exp\{in \int_{x_i}^{x_f} A_\mu dx^\mu\} \rightarrow e^{in\epsilon(x_f)} \exp\{in \int_{x_i}^{x_f} A_\mu dx^\mu\} e^{-in\epsilon(x_i)}. \quad (3.54)$$

因此，为了保持(3.53)式给出的概率幅规范不变，我们必须要求波函数 $\psi(x)$ 同步地进行规范变换。不难看出，波函数的规范变换规则为

$$\psi(x) \rightarrow e^{in\epsilon(x)} \psi(x), \quad (3.55)$$

很明显，如果按照这个规则同时变换波函数，那么就能保证概率幅 \mathcal{A} 规范不变！

特别的， $e^{i\epsilon}$ 是一个相位因子，为了保证这个相位因子的单值性，我们要求 $\epsilon(x)$ 的取值以 2π 为周期！进一步，为了保证波函数 $\psi(x)$ 在规范变换之下的单值性，很显然，必须有

$$e^{in(2\pi)} = 1 \Rightarrow n \in \mathbb{Z}. \quad (3.56)$$

所以， n 必须为整数。

由于 $q = ne$ 为带电粒子的电荷，因此 n 必须为整数就说明，带电粒子的电荷量必须为某个基本电荷 e 的整数倍！这就是著名的电荷量子化。

3.6 电磁对偶与磁单极子

麦克斯韦方程组有一种深层次的对称性，称之为电磁对偶。这种对偶在无源的麦克斯韦方程中体现得最为清楚，这时候麦克斯韦方程可以写为

$$\begin{aligned} \nabla \cdot \mathbf{E} &= 0, & \nabla \times \mathbf{B} &= \frac{\partial \mathbf{E}}{\partial t} \\ \nabla \cdot \mathbf{B} &= 0, & \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}. \end{aligned} \quad (3.57)$$

显然这个方程组在如下变换下保持不变

$$\mathbf{E} \rightarrow \mathbf{B}, \quad \mathbf{B} \rightarrow -\mathbf{E}, \quad (3.58)$$

因为这个变换交换了电场和磁场，所以称作电磁对偶。

为了将这种美妙性质拓展到有源情形，与电荷和电流密度对应的，就需要引入磁荷和磁流密度。为了以示区分，不妨记电荷密度为 ρ_e ，电流密度为 \mathbf{J}_e ，记磁荷密度为 ρ_m ，磁流密度为 \mathbf{J}_m 。进而将有源的麦克斯韦方程修改为

$$\begin{aligned}\nabla \cdot \mathbf{E} &= \rho_e, & \nabla \times \mathbf{B} &= \frac{\partial \mathbf{E}}{\partial t} + \mathbf{J}_e \\ \nabla \cdot \mathbf{B} &= \rho_m, & -\nabla \times \mathbf{E} &= \frac{\partial \mathbf{B}}{\partial t} + \mathbf{J}_m.\end{aligned}\quad (3.59)$$

显然，这组方程在如下电磁对偶变换下保持不变

$$\begin{aligned}\mathbf{E} &\rightarrow \mathbf{B}, & \rho_e &\rightarrow \rho_m, & \mathbf{J}_e &\rightarrow \mathbf{J}_m \\ \mathbf{B} &\rightarrow -\mathbf{E}, & \rho_m &\rightarrow -\rho_e, & \mathbf{J}_m &\rightarrow -\mathbf{J}_e.\end{aligned}\quad (3.60)$$

而一个带电荷 q 磁荷 g 的粒子(称作双荷子)在电磁场中受到的洛伦兹力公式就应该推广成

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) + g(\mathbf{B} - \mathbf{v} \times \mathbf{E}). \quad (3.61)$$

由此可见，为了将电磁对偶不变性拓展到有源的麦克斯韦方程组，我们就需要引入磁荷。带电荷的粒子是比如电子，而带磁荷的粒子就是所谓的磁单极子。但是狄拉克证明了，同时兼容带电粒子和磁单极子的量子力学理论是可以定义的，前提是带电粒子的电荷 q 和磁单极子的磁荷 g 之间满足如下量子化条件，

$$qg = 2\pi n, \quad n \in \mathbb{Z}. \quad (3.62)$$

我们将在下一小节详细讨论狄拉克的这一理论。

关于磁单极子，尽管理论上有很多漂亮的结果，但它最大的问题就是，直到今天，实验上依然没有发现磁单极子的存在。因此，和前面的一些小节不同，本节我们讨论的主要是一种理论结构，而不是试图发展理论来描述实验上的事实。

3.6.1 狄拉克量子化条件

假设我们引入一个位于空间坐标原点的磁单极子，其磁荷为 g ，则磁场将满足如下方程

$$\nabla \cdot \mathbf{B} = g\delta^3(\mathbf{x}), \quad (3.63)$$

类比于点电荷的库伦场，可知这个方程的解为

$$\mathbf{B} = \frac{1}{4\pi} \frac{g}{r^2} \mathbf{e}_r. \quad (3.64)$$

式中 $r = |\mathbf{x}|$ ，而 $\mathbf{e}_r = \mathbf{x}/r$ 为径向指向的单位矢量。这个解在空间坐标原点是奇异的。现在，设想将这个奇异的空间坐标原点去掉，进而考察方程(3.63)在 $\mathbb{R}^3 \setminus \{0\}$ 上的解。注意， $\mathbb{R}^3 \setminus \{0\}$ 由于有一个洞，从而是拓扑非平庸的。

由于在 $\mathbb{R}^3 \setminus \{0\}$ 上，方程(3.63)变成了如下形式

$$\nabla \cdot \mathbf{B} = 0, \quad (3.65)$$

所以显而易见的想法是，在 $\mathbb{R}^3 \setminus \{0\}$ 上依然可以引入矢量势 \mathbf{A} ，使得 $\mathbf{B} = \nabla \times \mathbf{A}$ 。但这种想法其实是过于天真的，为了看清楚其问题，考察 $\mathbb{R}^3 \setminus \{0\}$ 上一个包围原点的球面 S^2 ，其包围的空间区域记为 V ，若 $\mathbf{B} = \nabla \times \mathbf{A}$ ，则有

$$\int_{S^2} \mathbf{B} \cdot d\mathbf{S} = \int_{S^2} (\nabla \times \mathbf{A}) \cdot d\mathbf{S} = \int_V d^3\mathbf{x} \nabla \cdot (\nabla \times \mathbf{A}) = 0. \quad (3.66)$$

但在另一方面，根据方程(3.63)，我们有

$$\int_{S^2} \mathbf{B} \cdot d\mathbf{S} = \int_V d^3\mathbf{x} \nabla \cdot \mathbf{B} = g \int_V d^3\mathbf{x} \delta^3(\mathbf{x}) = g \neq 0. \quad (3.67)$$

这就说明，虽然在 $\mathbb{R}^3 \setminus \{0\}$ 上有 $\nabla \cdot \mathbf{B} = 0$ ，但是由于 $\mathbb{R}^3 \setminus \{0\}$ 拓扑非平庸，我们并不能整体地有 $\mathbf{B} = \nabla \times \mathbf{A}$ 。换言之，不可能存在一个在整个 $\mathbb{R}^3 \setminus \{0\}$ 上都定义良好的矢量势 \mathbf{A} ，使得 $\mathbf{B} = \nabla \times \mathbf{A}$ 。

但是由于 $\nabla \cdot \mathbf{B} = 0$ ，根据所谓的庞加莱引理可知，在 $\mathbb{R}^3 \setminus \{0\}$ 的任何一个拓扑平庸的开子集 U 上，必定都可以找到一个 \mathbf{A}^U ，使得在 U 上有 $\mathbf{B} = \nabla \times \mathbf{A}^U$ 。比方说，对于上面的磁单极子，在球坐标中我们可以构造一个 \mathbf{A}^N ，

$$\mathbf{A}^N = A_\phi^N \mathbf{e}_\phi = \frac{g}{4\pi r} \frac{1 - \cos \theta}{\sin \theta} \mathbf{e}_\phi, \quad (3.68)$$

式中 \mathbf{e}_ϕ 为周角 ϕ 增加方向的单位矢量。根据球坐标中的旋度运算，容易验证

$$\mathbf{B} = \nabla \times \mathbf{A}^N = \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (A_\phi^N \sin \theta) \mathbf{e}_r - \frac{1}{r} \frac{\partial}{\partial r} (r A_\phi^N) \mathbf{e}_\theta, \quad (3.69)$$

代入(3.68)式，即可得 $\mathbf{B} = \frac{g}{4\pi r^2} \mathbf{e}_r$ ，正是磁单极子的磁场。这说明，(3.68)式是一个正确的构造。

进而根据前面的讨论可知， \mathbf{A}^N 必定不可能在整个 $\mathbb{R}^3 \setminus \{0\}$ 上都定义良好。事实上，(3.68)式给出的 \mathbf{A}^N 有奇异性，这个奇异性不仅在已经被扣掉了的坐标原点处，而且是沿着整个负 z 轴都是奇异的。这是因为 \mathbf{A}^N 表达式(3.68)中的 $1/\sin \theta$ 因子，这个因子在 $\theta = 0$ 和 $\theta = \pi$ 处都奇异，但是 $\theta = 0$ 处的奇异性被分子的 $1 - \cos \theta$ 经由0比0消去了，从而 \mathbf{A}^N 仅在 $\theta = \pi$ 处奇异，也就是在整个负 z 轴上奇异。也即是说， \mathbf{A}^N 在负 z 轴上没有定义，而在除此之外的区域都定义良好。限制在包围原点的球面 S^2 上来考察的话，即是， \mathbf{A}^N 在南极处没有定义，在除南极之外的区域都定义良好，特别的， \mathbf{A}^N 在整个北半球面都定义良好，事实上，这就是上标 N 所表示的含义。

现在考虑一个不同的规范势

$$\mathbf{A}^S = A_\phi^S \mathbf{e}_\phi = -\frac{g}{4\pi r} \frac{1 + \cos \theta}{\sin \theta} \mathbf{e}_\phi. \quad (3.70)$$

不难验证，它同样可以给出磁单极子的磁场(3.64)。但是现在，这个规范势在北极 $\theta = 0$ 处是奇异的，而在整个南半球都定义良好，这也就是上标 S 的含义。

很明显，在 S^2 球面赤道附近的区域， \mathbf{A}^N 和 \mathbf{A}^S 均有定义，因为它是南北半球的交叠区。在交叠区上，由于给出的场强相同，所以这两个不同的规范势当然只能相差一个规范变换。

谈到规范变换，这个里要做个注解。本节实际上是回到了 A_μ 重写成 eA_μ 的约定，从而带电粒子波函数的规范变换应该为

$$\psi(x) \rightarrow e^{iq\epsilon(x)} \psi(x), \quad (3.71)$$

同时规范场的规范变换为

$$\mathbf{A} \rightarrow \mathbf{A} + \nabla \epsilon. \quad (3.72)$$

根据 \mathbf{A}^N 的表达式(3.68) \mathbf{A}^S 的表达式(3.70)，再根据球坐标中的梯度运算，人们不难发现，在赤道附近的交叠区， \mathbf{A}^N 与 \mathbf{A}^S 间的规范变换为

$$A_\phi^N = A_\phi^S + \frac{1}{r \sin \theta} \partial_\phi \epsilon, \quad \text{其中} \quad \epsilon = \frac{g\phi}{2\pi}. \quad (3.73)$$

也即是说绕着赤道走一圈，我们有 $\epsilon(2\pi) = \epsilon(0) + g$ 。另一方面，根据规范变换的(3.71)式，绕行一圈，必有

$$\begin{aligned} q\epsilon(2\pi) &= q\epsilon(0) + 2\pi n \\ \Rightarrow \epsilon(2\pi) &= \epsilon(0) + \frac{2\pi n}{q}, \end{aligned} \quad (3.74)$$

其中 n 为整数。综上即可知，

$$qg = 2\pi n, \quad n \in \mathbb{Z}. \quad (3.75)$$

这就是所谓的狄拉克量子化条件。

狄拉克量子化条件是一个理论中任意带电粒子的电荷与任意磁单极子的磁荷间都必须满足的条件。这里有两种可能性，第一，理论中所有带电粒子的电荷都是某个最小电荷单位的整数倍，显然，这时候只要最小电荷单位满足狄拉克量子化条件，就必然有任何电荷都满足它。假设记最小电荷单位为 e ，则(3.75)告诉我们，

$$\int_{S^2} \mathbf{B} \cdot d\mathbf{S} = g = \frac{2\pi n}{e}. \quad (3.76)$$

即闭合球面上的磁通(即磁单极子的磁荷)是量子化的，最小的磁通(磁荷)为

$$\Phi_0 = \frac{2\pi}{e}, \quad (3.77)$$

称作磁通量子。

第二种可能性是，理论中带电粒子的电荷互相为无理数比例，比方说，理论中包含电荷为 e 以及电荷为 $\sqrt{3}e$ 的带电粒子。那这时候狄拉克量子化条件是不可能满足的，从而结论就是这样的理论中不可能存在磁单极子。好在，正如上一节讨论过的，现实世界的理论满足的是第一种可能性，因此是允许磁单极子存在的，虽然我们到今天为止都没有发现它。

以上关于狄拉克量子化条件的推导还可以推广到更为一般的情形，比方说球面 S^2 内包围了多个磁单极子的情形。在这种更一般的情形中，庞加

莱引理告诉我们, 依然存在北半球的规范势 \mathbf{A}^N , 以及南半球的规范势 \mathbf{A}^S , 只不过这时候的 \mathbf{A}^N 和 \mathbf{A}^S 不是由(3.68)式和(3.70)式给出。但是在赤道附近的交叠区, 我们依然有

$$\mathbf{A}^N = \mathbf{A}^S + \nabla\epsilon, \quad (3.78)$$

虽然 ϵ 可能不再是 $\frac{q\phi}{2\pi}$, 不过, 它将依然满足(3.74)式。

不妨以 \int_N 表示在北半球面上的积分, 以 \int_S 表示在南半球面上的积分, 以 C 表示赤道回路, 从而我们将有

$$\begin{aligned} \int_{S^2} \mathbf{B} \cdot d\mathbf{S} &= \int_N (\nabla \times \mathbf{A}^N) \cdot d\mathbf{S} + \int_S (\nabla \times \mathbf{A}^S) \cdot d\mathbf{S} \\ &= \oint_C \mathbf{A}^N \cdot d\mathbf{x} - \oint_C \mathbf{A}^S \cdot d\mathbf{x} \\ &= \oint_C \nabla\epsilon \cdot d\mathbf{x} = \epsilon(2\pi) - \epsilon(0), \end{aligned} \quad (3.79)$$

式中我们注意到了回路 C 与北半球的法向成右手关系, 而与南半球的法向成左手关系。代入(3.74)式, 即有

$$\int_{S^2} \mathbf{B} \cdot d\mathbf{S} = \frac{2\pi n}{q}. \quad (3.80)$$

这就是一般的磁通量子化关系, 它实际上对任何两维闭合曲面(不只是球面) Σ 均成立, 因此也写作

$$\int_{\Sigma} \mathbf{B} \cdot d\mathbf{S} = \frac{2\pi n}{q}. \quad (3.81)$$

它告诉我们, 在两维闭合表面上的磁通量必定为磁通量子 Φ_0 的整数倍。

以上考察狄拉克量子化条件的时候, 我们考虑的是单纯的电荷和单纯的磁单极子。但是类似的考察可以推广到既带电荷同时又带磁荷的双荷子(dyons)情形, 假设一个理论既包含一种 (q_1, g_1) 的双荷子, 又包含另一种 (q_2, g_2) 的双荷子, 则它们必定要满足如下Dirac-Zwanziger量子化条件

$$q_1 g_2 - q_2 g_1 \in 2\pi\mathbb{Z}. \quad (3.82)$$

这个条件当然是狄拉克量子化条件的推广。

第四章 粒子源自于场

在物理上，场和粒子似乎完全不同，场的能量是分布在整个空间的，而粒子的能量是集中在一个很小的空间区域上的。看起来似乎是，大自然同时给了我们两种完全不同的物理实体，一种是电子、质子这样的粒子，另一种是电磁场这样的场。然而，现代物理的研究表明，电子、质子同样可以用场来描述，而电磁场也会激发出光子这样的粒子，场和粒子与其说是两种不同的物理实体，不如说是对物理实体的两种相互对偶的不同描述。

4.1 粒子作为场方程的一种解

正确解释粒子与场的关系需要将场量子化，进而将粒子作为场的一种激发态，因此这是量子场论的主题。但是在经典场论的框架下也有可能给粒子提供一种解释(虽然它不能解释所谓的基本粒子)，即将粒子看成是场方程的一种特殊解，通常称作孤立子解，在这种解中，场的能量主要集中在空间上一个很小的局部区域上，并且能够稳定地维持在这样的区域上不弥散，这样从更大的尺度来看，我们就可以把这个小局部区域当成一个粒子，而当这个稳定区域移动时，我们就可以认为是一个粒子在运动。本节就是要以1 + 1维时空(1维空间+1维时间)中一个特殊场论模型为例，解释孤立子解是如何可能的。

这一节的下面一些内容都是在1+1维时空中考察问题。我们以 σ 来表示一维空间坐标，以 $x = (\sigma, t)$ 来表示时空点，从而时空上的场为 $\phi(x) = \phi(\sigma, t)$ 。本节只考虑场本身，不考虑它和其它粒子如何耦合。

4.1.1 场位形的拓扑等价类

假设我们考察的场论模型具有如下作用量(注意时空是两维的, 从而 $\partial_\mu\phi\partial^\mu\phi = -(\partial_t\phi)^2 + (\partial_\sigma\phi)^2$)

$$S = \int d^2x \left[\frac{1}{2} [(\partial_t\phi)^2 - (\partial_\sigma\phi)^2] - \mathcal{U}(\phi) \right], \quad (4.1)$$

式中 $\partial_\sigma\phi = \frac{\partial\phi}{\partial\sigma}$ 。从而假设将作用量写成 $S = \int dt(T - V)$ 的形式, 那么其中场的动能 T 和势能 V 分别为

$$T = \int_{-\infty}^{+\infty} d\sigma \left[\frac{1}{2} (\partial_t\phi)^2 \right], \quad V = \int_{-\infty}^{+\infty} d\sigma \left[\frac{1}{2} (\partial_\sigma\phi)^2 + \mathcal{U}(\phi) \right]. \quad (4.2)$$

根据最小作用量原理, 可以得到场的运动方程为

$$\partial_t^2\phi - \partial_\sigma^2\phi = -\frac{\partial\mathcal{U}}{\partial\phi}. \quad (4.3)$$

特别的, 对于不依赖时间的静态场位形, 场方程的相应解就是使得势能泛函取极值的位形, 这可以通过 $\delta V[\phi(\sigma)] = 0$ 导出方程来验证。

进一步, 我们假设函数 $\mathcal{U}(\phi)$ 有下界(这是为了确保系统的能量有下界), 而由于我们总可以通过给势能密度加上一个常数来调节这个下界的值, 所以我们再假定, $\mathcal{U}(\phi)$ 的下确界为零, 即 $\mathcal{U}(\phi)$ 的最小值为零。

我们称使得 $\mathcal{U}(\phi)$ 取最小值零的场位形为真空场位形, 更严格地说, 我们定义真空场位形的集合 Ω 为

$$\Omega = \{ \phi | \partial_t\phi = \partial_\sigma\phi = 0, \text{ and } \mathcal{U}(\phi) = 0 \}. \quad (4.4)$$

即是说, 真空场位形是一种静态场位形, 它在空间各点取值为常数, 并且这个常数使得函数 $\mathcal{U}(\phi)$ 达到最小。很显然, 真空场位形是场方程(4.3)的解, 称作真空解。同样很显然, 真空场位形的动能和势能都是零, 从而真空场位形的总能量为零。而从动能 T 以及势能 V 的表达式也很容易看出, 其它任何场位形的能量必定大于真空场位形的能量。从而, 真空场位形就是使得场的总能量取最小值(即零)的场位形。

对于本节所要考察的模型来说, 使得 $\mathcal{U}(\phi)$ 取最小值的 ϕ 并不唯一, 从而真空场位形集合 Ω 中的元素多于1个, 即是说, 系统有多个真空。

另外，从 T 和 V 的表达式也可以看出，对于任何场位形，为了使得系统的总能量有限，它在空间无穷远处必须要趋于真空场位形，否则在无穷远区域的动能密度和势能密度就不等于零，积分的话就会贡献无穷大的总能量。我们记

$$\phi_{\pm} = \lim_{\sigma \rightarrow \pm\infty} \phi(\sigma), \quad (4.5)$$

则根据刚才的论述，必有

$$\phi_{\pm} \in \Omega. \quad (4.6)$$

如果 $\phi_+ = \phi_-$ ，则我们可以在保持总能量始终有限的前提下连续地改变场位形，以使得最终 $\phi(x) = \phi_+$ ，即使得最终的场位形处处取同一个真空位形 ϕ_+ 。即是说，如果一个场位形满足 $\phi_+ = \phi_-$ ，那它就可以连续地变形为真空解，我们称这种场位形拓扑上等价于真空位形。

反过来，如果 $\phi_+ \neq \phi_-$ ，那任何将场位形连续变形为真空位形的过程都要至少改变其在 $\pm\infty$ 某一端的边界值，比方说将 $-\infty$ 区域的 ϕ_- 连续变形为 ϕ_+ ，但很显然，这会使得这个边界位形在变形过程中偏离真空，从而需要无穷大的总能量。因此，如果总能量始终有限，那么 $\phi_+ \neq \phi_-$ 的这种场位形就不可能连续变形为真空场位形，我们称它和真空场位形拓扑不等价。假设以无穷远边界位形对 (ϕ_-, ϕ_+) 来标记一个场位形，那么完全类似上面的推理告诉我们，一般来说，两个 (ϕ_-, ϕ_+) 对不同的场位形不可能在保持总能量有限的情况下连续过渡，称之为拓扑不等价¹。所以，一般来说， (ϕ_-, ϕ_+) 对标记的是场位形的拓扑等价类，同一类的就拓扑等价，不同类的就拓扑不等价。

特别的，我们可以将场位形随时间的演化看成是一种对场位形连续变形的过程，那上面的论证就告诉我们，不同拓扑等价类的场位形不可能相互演化，比方说，一个 $\phi_+ \neq \phi_-$ 的场位形不可能演化成一个真空场位形，反过来也一样，一个真空场位形也不可能演化成一个 $\phi_+ \neq \phi_-$ 的场位形。即是说，场位形的拓扑等价类在时间演化之下是稳定不变的。

人们通常称一个 $\phi_- \neq \phi_+$ 的场位形为扭结(kinks)，这个名称来源于，这种场位形从 $-\infty$ 处的 ϕ_- 值过渡到 $+\infty$ 处的 ϕ_+ ，它在中间必定要扭一下。场方

¹除非系统有某种对称性，比如后面sine-Gordon场论的例子讨论的那样，这时候某些 (ϕ_-, ϕ_+) 对不同的场位形可能等价。

程(4.3)的扭结解其实就是一种孤立子, 因为正如我们将要看到的, 对于这种解, 其能量将主要集中在场位形的扭结处, 这通常是一个很小的空间区域。而且, 扭结解拓扑等价类的稳定性正好相应于孤立子的稳定性。

4.1.2 Bogomolny能限和Bogomolny方程

下面我们来考察每一个场位形拓扑等价类的能量下界, 以及使得能量达到下界的场位形。很显然, 为了使得能量 E 尽可能低, 我们只需考察动能为零的场位形, 也就是静态场位形, 这时候 $E = V[\phi(\sigma)] = \int_{-\infty}^{+\infty} d\sigma \left[\frac{1}{2}(\partial_\sigma \phi)^2 + \mathcal{U}(\phi) \right]$ 。本小节下面都是默认考虑静态场位形。

首先, 我们有下面不等式

$$\frac{1}{2} \left(\partial_\sigma \phi \pm \sqrt{2\mathcal{U}(\phi)} \right)^2 \geq 0 \Rightarrow \frac{1}{2}(\partial_\sigma \phi)^2 + \mathcal{U}(\phi) \geq \pm \sqrt{2\mathcal{U}(\phi)} \partial_\sigma \phi. \quad (4.7)$$

将上面的结果对空间坐标 σ 进行积分, 即有

$$\int_{-\infty}^{+\infty} d\sigma \left[\frac{1}{2}(\partial_\sigma \phi)^2 + \mathcal{U}(\phi) \right] \geq \pm \int_{-\infty}^{+\infty} d\sigma \sqrt{2\mathcal{U}(\phi)} \partial_\sigma \phi. \quad (4.8)$$

因此有

$$E \geq \left| \int_{\phi_-}^{\phi_+} \sqrt{2\mathcal{U}(\phi)} d\phi \right|. \quad (4.9)$$

又由于我们假设 $\mathcal{U}(\phi)$ 的最小值为零, 即 $\mathcal{U}(\phi) \geq 0$, 所以可以引入一个 $W(\phi)$, 并令

$$\mathcal{U}(\phi) = \frac{1}{2} \left(\frac{dW}{d\phi} \right)^2. \quad (4.10)$$

代入上面的(4.9)式, 即可以得到

$$E \geq |W(\phi_+) - W(\phi_-)|. \quad (4.11)$$

这就是场位形的能量下界, 称之为Bogomolny能限, 很显然, 这个下限仅仅依赖于场位形的拓扑等价类 (ϕ_-, ϕ_+) 。

从上面的推导过程不难看出, 为了使得场位形的能量尽量低, 以达到Bogomolny能限, 我们不仅要限于考虑静态场位形, 而且此静态场位形还应该满足

$$\partial_\sigma \phi = \pm \sqrt{2\mathcal{U}(\phi)}, \quad (4.12)$$

因为只有这样才能使得推导Bogomolny能限时的出发点不等式 $\frac{1}{2}(\partial_\sigma\phi \pm \sqrt{2\mathcal{U}(\phi)})^2 \geq 0$ 取等于号。方程(4.12)就称为Bogomolny方程。

上面的推导告诉我们，一个静态场位形如果同时还满足Bogomolny方程，那它一定使得势能泛函 $V[\phi(\sigma)]$ 在给定拓扑等价类中取最小值，从而当然也就使得势能泛函 $V[\phi(\sigma)]$ 达到极小值，从而相应的场位形必定满足场方程(4.3)。即，满足Bogomolny方程的静态场位形必定满足场方程(4.3)，这个结论其实也不难直接验证。但是反之则不然，即满足场方程(4.3)的场位形可不一定满足Bogomolny方程。Bogomolny方程相比场方程的巨大优点就在于，它是一个很简单的一阶微分方程，很容易积分出来！

4.1.3 sine-Gordon模型和孤立子

这一节我们考察一个更加具体的场论模型，称之为sine-Gordon模型，这个模型即是在前面两个小节的基础上取 $\mathcal{U}(\phi)$ 为一个三角函数，比如取

$$\mathcal{U}(\phi) = 1 - \cos \phi = 2 \sin^2(\phi/2). \quad (4.13)$$

从而系统的作用量为

$$S = \int d^2x \left[\frac{1}{2} [(\partial_t\phi)^2 - (\partial_\sigma\phi)^2] - (1 - \cos \phi) \right]. \quad (4.14)$$

当这样取时，我们已经假定通过合适地选取长度和能量的单位来使得作用量中的各带量纲参数取1了。

很显然，sine-Gordon模型的真空场位形为， $\phi = 2\pi n$ ，其中 n 为任意整数 $n \in \mathbb{Z}$ 。根据前两个小节，任何有限能量场位形的拓扑等价类一般来说可以由 (ϕ_-, ϕ_+) 来标记， ϕ_\pm 为空间无穷远处的渐近真空场位形，但是，上面的sine-Gordon模型有一种特殊的对称性，即它的作用量在 $\phi \rightarrow \phi \pm 2\pi$ 的变换下保持不变，因此实际上，下面两类场位形是等价的，

$$(\phi_-, \phi_+) \sim (\phi_- + 2\pi n, \phi_+ + 2\pi n). \quad (4.15)$$

因此sine-Gordon模型场位形的拓扑等价类其实由 $\phi_+ - \phi_-$ 分类。人们常常定义场位形的拓扑荷 N 为

$$N = \frac{\phi_+ - \phi_-}{2\pi} = \frac{1}{2\pi} \int_{\phi_-}^{\phi_+} d\phi = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\partial_\sigma\phi) d\sigma, \quad (4.16)$$

式中 $(\partial_\sigma \phi)/(2\pi)$ 称作拓扑荷密度。很显然，拓扑荷 N 是一个整数，而且完全分类了sine-Gordon场位形的拓扑等价类。

sine-Gordon模型的Bogomolny能限为

$$E \geq \left| \int_0^{2\pi N} 2 \sin(\phi/2) d\phi \right| = \int_0^{2\pi|N|} 2 |\sin(\phi/2)| d\phi. \quad (4.17)$$

注意到函数 $|\sin(\phi/2)|$ 的周期为 2π ，从而有

$$E \geq \int_0^{2\pi|N|} 2 |\sin(\phi/2)| d\phi = 4|N| \left[-\cos(\phi/2) \right]_0^{2\pi} = 8|N|. \quad (4.18)$$

能限的等于号仅在静态且满足下面Bogomolny方程的场位形中取到，

$$\partial_\sigma \phi = \pm 2 \sin(\phi/2). \quad (4.19)$$

很容易积分求解Bogomolny方程(4.19)，比方说对于(4.19)式取+号的情形，我们容易积分得出

$$\phi(\sigma) = 4 \tan^{-1} (e^{\sigma-a}), \quad (4.20)$$

式中 a 为积分常数。很显然，这个解满足， $\phi_- = \phi(-\infty) = 0$ ， $\phi_+ = \phi(+\infty) = 2\pi$ ，即这个解的拓扑荷 $N = 1$ 。我们称这种解为静态单扭结解(single kink solution)。类似的，如果我们在(4.19)式取-号，相应的解很明显相当于将静态单扭结解作 $\sigma \rightarrow -\sigma$ 的替换，从而满足 $\phi_- = 2\pi$ ， $\phi_+ = 0$ ，因此是一个拓扑荷为 $N = -1$ 的解，称之为静态反单扭结解(single anti-kink solution)。另外，很显然，Bogomolny方程(4.19)还有 $\phi = 2\pi n$ ， $n \in \mathbb{Z}$ 的真空解，这是方程(4.19)的特解，它相当于在上面的一般解(4.20)中取 $a \rightarrow \pm\infty$ 的极限。这种真空解的拓扑荷当然满足 $N = 0$ 。

下面我们来计算一下静态单扭结解的能量密度 \mathcal{E} ，

$$\mathcal{E} = \frac{1}{2}(\partial_\sigma \phi)^2 + (1 - \cos \phi) = 4 \sin^2(\phi/2). \quad (4.21)$$

代入解(4.20)，可以得到

$$\mathcal{E}(\sigma) = \frac{4}{\cosh^2(\sigma - a)}. \quad (4.22)$$

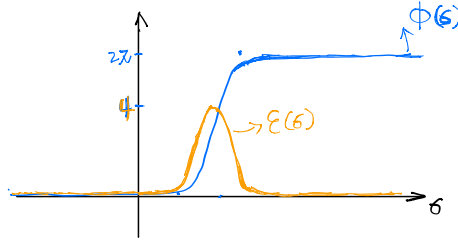


图 4.1: 静态单扭结解的场位形和能量密度分布。

很显然，这个能量密度在 $\sigma = a$ 处达到最大，在 a 点的左右两侧都迅速衰减，很快就衰减为零。并且当然，解的总能量 $E = \int_{-\infty}^{+\infty} \mathcal{E}(\sigma) d\sigma = 8 \cdot |1| = 8$ ，刚好处于Bogomolny能限。我们可以画出静态单扭结解场位形和能量密度分布的示意图，如图(4.1)所示。很显然，静态单扭结解的能量主要集中在 a 点附近的一个小区间内，因此从更大的尺度来看，人们可以等效地认为是在 a 点附近有一个粒子，这就是孤立子(soliton)这一名称的来源。我们常常称这个能量中心位置 a 为静态单扭结解的位置，这当然就是把它当成一个粒子来看待了。

不仅如此，对于一个静态单扭结解，假如我们换一个以速度 $-v$ 运动的参考系来看它，那这个解看起来就是以速度 v 在运动，即它的能量中心在以速度 v 朝前运动，看起来就像是一个以速度 v 运动的粒子。这就说明，sine-Gordon模型不仅存在静态的孤立子解，而且从这些静态解也很容易得到匀速运动的孤立子解。这就加强了将单扭结解和反单扭结解等效地看作粒子的看法。

另外，从前面的讨论可以知道，Bogomolny方程(4.19)只有 $N = 0, \pm 1$ 的解，而没有 N 取其它值的静态多扭结解。这意味着 $N \neq 0, \pm 1$ 时Bogomolny能限是达不到的。而这又意味着，当我们将两个静态单扭结解叠加并做必要的修正，以试图得到场方程(不是Bogomolny方程)的一个 $N = 2$ 的2-扭结解时，这个 $N = 2$ 场位形的总能量一定严格大于 $8 \cdot |2| = 16$ ，即 $E > 8 \cdot |2|$ ，其中多出来的能量 $E - 16$ 可以解释为两个单扭结之间的相互作用势能，它依赖于两个扭结之间的距离 R ，因此不妨记为 $V_I(R)$ 。另一方面，我们也知道，如果相互叠加的这两个单扭结之间的距离趋于无穷，那它们之间的相互影响就可以忽略，这时候场位形的总能量当然就趋于两个静态单扭

结解的能量总和，也就是 $E = 8 \cdot 2$ ，从而 $V_I(\infty) = 0$ 。直观上，这就说明，2-扭结解的相互作用势能是随着两个扭结之间距离的增加而减少的，从而有 $F_I = -\frac{\partial V_I}{\partial R} > 0$ ，这说明2-扭结解的两个扭结之间存在排斥力，在这个排斥力的作用下，这两个扭结不可能保持静止，它们最终一定会相互远离。

以上的分析说明，sine-Gordon模型场方程的2-扭结解必定不是静态解！通常来说，这种2-扭结解应该描述的是两个运动着的单扭结之间的散射，如图(4.2)所示。同样的道理，其它多扭结解也都不是静态解，而是描述多个

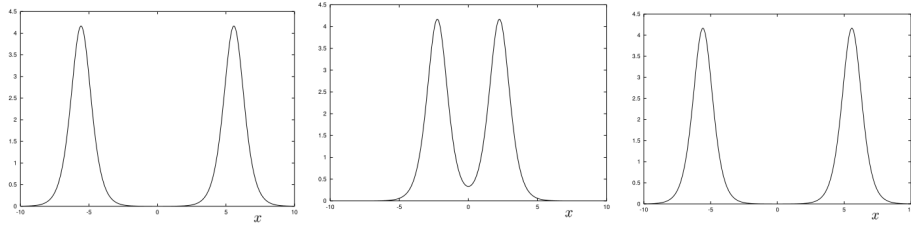


图 4.2: 2-扭结散射的能量密度分布，3幅图的时间依次是 $t = -20, t = 0, t = 20$ 。

单扭结之间的散射。1+1维sine-Gordon模型的特殊之处在于，它是一个可积场论，从而原则上其实可以精确得到它的所有多扭结解！不过，用可积系统的办法处理sine-Gordon模型超出了本书的范围。

4.2 粒子作为场的量子

正如前面所过的，在现代量子场论中，粒子更多是作为场的量子而出现的，比如光子就是电磁场的量子。为了直观理解这个概念，我们考察如图(4.3)所示的两粒子通过中间的场进行相互作用的过程。前面第一章中我们简单讨论过这个过程，但，现在的不同之处在于，我们像图(4.3)右图所示的那样，将中间的场量子化了。所谓的量子化，就是把中间成连续一片的场“剖分”成一些波浪线的集合，这些波浪线就是所谓的场量子。很显然，每一个波浪线都可以看成是由粒子1发出并最终被粒子2接收的中间粒子(场量子)，通过这些中间粒子，粒子1与粒子2发生了相互作用。当然，图中没画的是，粒子2也可以发出中间粒子并被粒子1接收。所以，在对场

进行量子化以后，我们就可以认为粒子1和粒子2是通过交换场量子而发生相互作用的。

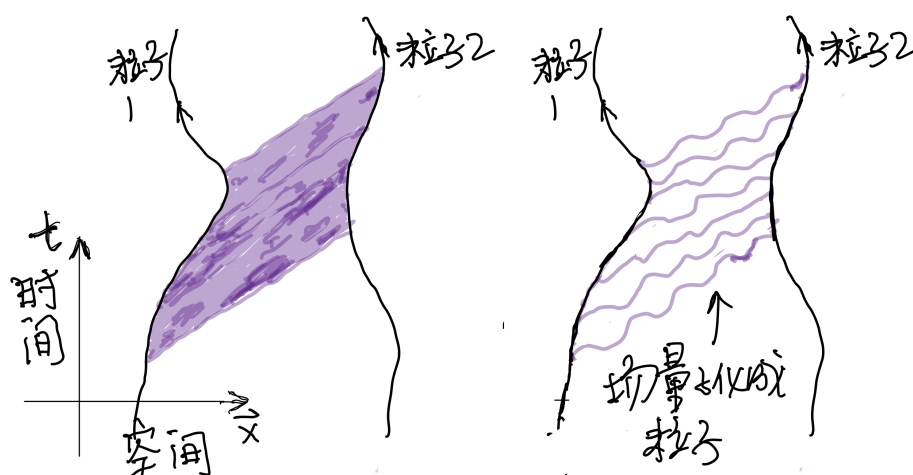


图 4.3: 左图是两粒子通过作为中介物的场相互作用；右图是把中间成连续一片的场“剖分”成波浪线，也就是场量子化。

这里的关键就在于，场的量子自然就是一种粒子。因此，人们不禁要问，可不可能原来的粒子1和粒子2也是某种场的量子呢？回答是肯定的，在量子场论中，一切粒子都可以看成是某种场的量子，粒子间通过交换中间粒子进行相互作用其实就是两种场的相互作用，其中一种场描述原来的粒子，另一种场描述中间粒子。总之，在量子场论中，一切都是场！不仅传递相互作用的光子和胶子是相应电磁场和胶子场的量子，而且电子、夸克等等，都是其相应场的量子。