Physics from Symmetry

1^{st} edition

Translate Version: 2016 年 3 月 13 日

Published by Non

https://github.com/laserroger/Physics-from-Symmetry/

Preface

> 世界上最无法理解的事情是这个世界竟然是可以理解的。
>
> — Albert Einstein

引自 Jon Fripp, Deborah Fripp, and Michael Fripp. *Speaking of Science*. Newnes, 1st edition, 4 2000. ISBN 9781878707512

在物理课上，我像任何一个物理系学生那样熟悉许多基本方程与它们的解，但我不是很清楚它们之间的联系。

当我理解它们中的大多数具有共同的起源 — **对称性 (Symmetry)** 时我十分激动。对我而言搞物理最美的经历莫过于原本费解的东西经过深入探究之后豁然开朗，我因此深爱着对称性。

例如，有一段时间我不能真正理解自旋 — 几乎所有基本粒子都具有的奇特的内禀角动量，后来我学到了原来自旋是一种对称性 (称为 Lorentz 对称性) 的直接结果，于是有关自旋的内容开始充满意义。

本书的目的就是为读者提供这样的经历，在某种意义上，当我开始学物理的时候我就有写这本书的愿望了。[1] 对称性漂亮地解释了其它方面的许多复杂物理现象，这让我们认为可以从对称导出物理学的基本理论，本书正是基于这个锐利而光芒耀眼的信念。[2]

[1]同感! 这就是物理系学生的中二病吗 2333333 — 译者 (SI)

[2]锐 (するど) くギラつかせた希望 (きぼう), もっと本気 (ほんき) 出 (だ) していいよ, あ ～, あ ～, あ ～, あ ～

可以说本书的写作顺序是倒过来的：在我们讨论经典力学与非相对论量子力学之前，我们将利用大自然（据我们所知）精确的对称性导出量子场论的基本方程。不过尽管途径不同，本书的内容仍为标准物理学，我们不涉及仍有争议未经实验验证的内容，而是通过物理学的标准假设导出物理学的标准理论。

根据读者的物理水平，本书可以有两种用途。

- 物理学的初学者[3]可以把这本书当成入门教材。它包含了经典力学、电动力学量子力学、狭义相对论和量子场论的基本理论，在阅读之后，读者可再深入学习各部分内容。各部分有许多更加深入的优秀书籍，在各章的末尾列出了一些延伸阅读的推荐表。如果你觉得你属于这一种，我们建议你在阅读正文之前先从附录的数学补充开始。
- 另一种情况，身经百战见得多的学生可以通过本书将松散的各物理领域紧密联系起来。回顾历史的进程我们会发现许多物理思想可能看起来随意甚至乱来。但从对称性的观点看，它们往往就变得必然而直接。

开始自附录 A。此外，当正文使用一个新的数学概念时，边注中会说明相应的附录目录。

とにかく，不管哪种情况都应该按顺序阅读本书，因为各章之间是递进的。

开始的短章节是关于狭义相对论的，它是之后讨论的所有内容的基础。我们会看到对物理理论最重要的限制之一是它们必须符合狭义相对论。

本书的第二部分引入了描述对称性的数学工具（在物理味道的表述下），大多数工具来自数学的重要分支 — 群论。之后我们介绍拉格朗日形式 (Lagrangian formalism)，它使我们能够在物理体系中直接利用对称性。

第五章和第六章利用之前引入的拉格朗日形式和群论导出了现代物理学的基本方程。

在最后一部分，我们把之前的基本方程加以应用：应用于粒子理论我们可以导出量子力学，应用于场理论我们可以导出量子场论。然后我们研究这些理论在非相对论极限与经典极限的变体，这样又导出了经典力学与电动力学。

每章的开始是本章内容的摘要。如果你发现你在思考 ‘这是在讲些啥’，不妨回到章节开头看看摘要以明白某一部分的目的是什

[3]这个 ‘初学者’ 是 ‘相对而言的’(relatively)... — 译者注

么。书页留有巨大的页边空白[4]~~以防止想出 Fermat 大定理的证明却没地方写~~ 你可以一边阅读一边在页边记下笔记与灵感。

许多页边注的内容是拓展信息与图景。

我希望你读这本书能够像我写这本书的时候一样愉悦。

Karlsruhe, 2015.01

Jakob Schwichtenberg

[4]给译文排版造成了小小的麻烦 — 译者

Acknowledgments

感谢所有帮我编写这本书的人。我特别感激 Fritz Waitz，他的评论、想法与纠正让本书质量大大改善。我十分感谢 Arne Becker 和 Daniel Hilpert，感谢他们无价的建议、意见与细致的校对。感谢 Robert Sadlier 对我英文的帮助以及 Jakob Karalus 的解释。

我还想感谢与我有许多见解深刻的讨论的 Marcel Köpke，感谢 Silvia Schwichtenberg 和 Christian Nawroth 的支持。

最后，我亏欠最多的是我的父母，他们支持着我，教导我知识高于一切。

如果发现文中的错误，我非常希望你能够寄一封短邮件到 errors@jakobschwichtenberg.com。勘误表的地址是 http://physicsfromsymmetry.com/errata。

Contents

I Foundations 基础

II Symmetry Tools 对称性工具

III The Equations of Nature 自然界的方程

Foundations 基础

1. Introduction 简介

1.1 What we Cannot Derive 推不出来的事情

在我们开始讲我们能从对称性里面了解到什么之前，我们首先澄清一下我们需要在我们的理论中人为的加一些什么东西。首先，目前没有任何理论可以得到自然界的常数。这些常数需要从实验中提取出来，比如各种相互作用的耦合常数啊，基本粒子的质量啊这种的。

除了这些，我们还有一些东西解释不了：**数字** 3。这不是术数的那种神秘主义的东西，而是我们不能解释所有的直接与数字 3 相联系的限制。比如：

- 对应三种标准模型描述的基本作用力有三种规范理论[1]。这些力是由分别对应于对称群 $U(1), SU(2)$ 和 $SU(3)$ 的规范理论描述的。为什么没有对应 $SU(4)$ 带来的基本作用力？没人知道！
- 轻子有三代，夸克也有三代。为什么没有第四代？我们只能从实验[2]中知道没有第四代。
- 我们只在拉格朗日量里面包含 Φ 的最低三阶 (Φ^0, Φ^1, Φ^2)，其中 Φ 指代一些描述我们的物理系统的东西，是个通称，而这个拉格朗日量则是被我们用来得到我们的描述自由（= 无相互作用）场/粒子的靠谱的理论的。
- 我们只用三个基本的 Poincare 群双覆盖的表示，分别对应

[1] 如果你不理解这个简介中的某些名词，比如规范理论或者二重覆盖，不需要太过担心。本书将会详尽的解释，在这里提到只是为了完整性。

[2] 比如，现在宇宙中元素的丰度是依赖于代的数量的。更进一步，对撞机实验中有对此的很强的证据。（见 Phys. Rev. Lett. 109, 241802）

自旋 $0, \frac{1}{2}$ 和 1。没有基本粒子的自旋是 $\frac{3}{2}$。

在现代的理论中，这些是我们必须手动增加的假定。我们从实验上知道这些假定是正确的，但是目前为止我们没有更深刻的原理告诉我们为什么我们需要到 3 就停。

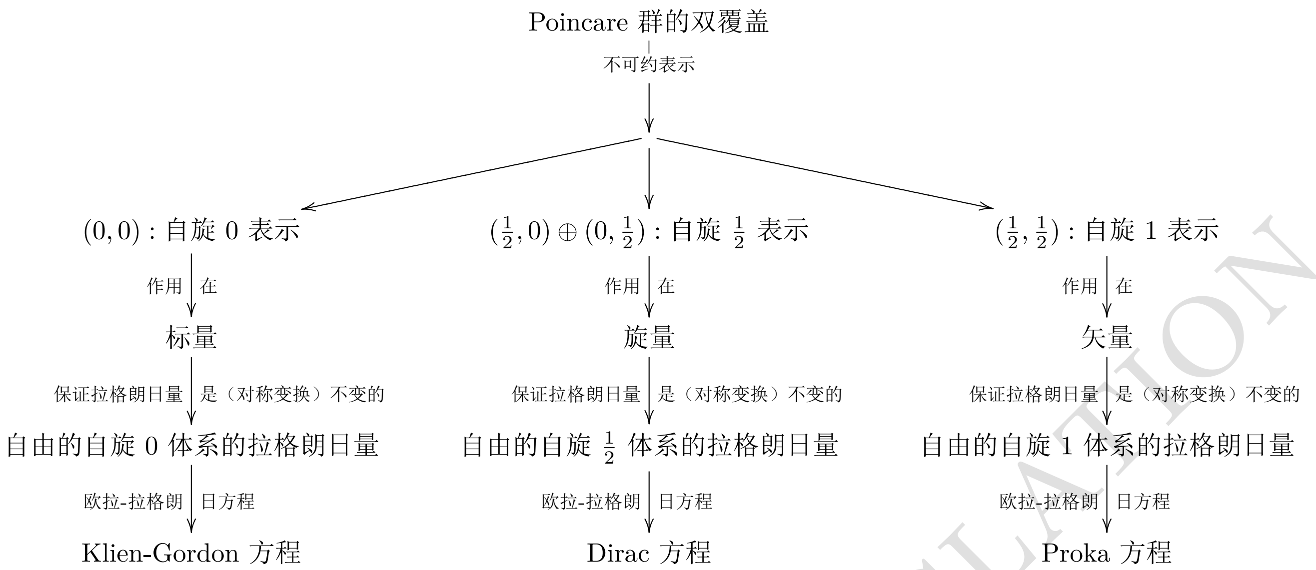
除此之外，还有两件事情我们没法从对称性中得到，但是他们对于一个严谨的理论来说有时必须被考虑到的：

- 我们只允许在拉格朗日量中引入尽可能低阶的非平庸的微分算符 ∂_μ。对于一些理论，我们使用一阶的微分算符 ∂_μ，而另一些理论 Lorentz 不变形禁止了一阶导数，从而二阶导数 $\partial_\mu \partial^\mu$ 是最低阶的可能的非平庸阶项。除此之外我们就再也得不到一个合理的理论了。存在高阶导数项的理论没有下界，这导致能量可以是一个任意大小的负值；因此，这些理论中的态总可以变到能量更低的态，从而永远不会稳定。
- 我们处于类似的原因，我们可以说如果半整数自旋的粒子和整数自旋的粒子拥有完全一样的行为的话，宇宙中就不会有稳定的物质。因此，这两者必然有某些东西不一样，而我们没得选，只有一种可能而且合理的选择[3]是正确的。这引出了半整数自旋粒子的 Fermi-Dirac 统计的概念和整数自旋粒子的 Bose-Einstein 统计的概念。半整数自旋的粒子从而通常被称为 Fermion，它们中永远不存在两个粒子处在完全一样的态上。而与之相反的，这种情况对整数自旋的粒子 — 通常被称为 Boson — 是可能的。

[3] 我们在最开始的量子场论里使用反对易子而不是对易子，从而防止我们的理论变成没有下界的理论。

最后呢，我们提一下剩下的一个我们不能从这本书的其他理论中得到的东西：**引力**。当然，实际上，大名鼎鼎的广义相对论就是优美而准确的描述引力的理论；然而这个理论与其他理论完全不一样，超出了本书的研究范围。而尝试将引力问题划入相同框架下的量子引力理论仍待完善：目前没有人能够成功得出它。除此之外，在最后一章我们会做一些对引力的一些评述。

1.2 Book Overview 全书概览



这本书使用**自然单位制**，也就是说 Planck 常数 $\hbar=1$，光速 $c=1$。这是基本理论中使用的惯例，它免除了很多不必要的笔墨。而对于应用来说呢，这些常数需要被再一次的加上去从而回到标准的 SI 单位制。

狭义相对论的基本假设是我们的起始点；它们是：在所有的惯性参考系 — 一些相互之间的速度保持恒定的参考系 — 中，光的速度不变，为 c；而且物理规律在在所有惯性系中相同。

满足这些对称性的所有的变换构成的集合叫做 **Poincare 群**。为了能够得到它们，我们接下来介绍一些数学知识从而使得我们能够利用好对称性。数学的这一分支叫群论。我们会得到 Poincare 群[4]的不可约表示——你可以当成是组其它所有表示的“地基”。这些表示就是我们后面用来描述粒子和场的不同自旋的表示。自旋一方面是对不同种类的粒子和场的标记，而另一方面也可以视为内禀角动量。

[4] 技术上正确的讲，我们会得到 Poincare 群的双覆盖而不是 Poincare 群本身。“双覆盖” 这个词是由于一个群的双覆盖和这个群的关系是前者的两个元素与后者的一个元素相对应。这在后面的节3.3.1中会讲到。

在这以后，我们要介绍**拉格朗日形式**，它使得我们可以方便的在物理问题里面使用对称性。这里面的核心对象是**拉格朗日量**，我们可以从对不同的物理系统的对称性的考虑出发来得到[1]。而在此之上我们就得到了欧拉-拉格朗日方程，从而我们可以得到给定拉格朗日量的运动方程。利用 Poincare 群的不可约表示，我们实际上得到了场和粒子的基本运动方程。

这里面的中心思想是在 Poincare 群中元素的的变换下，拉格

[1] 实际上，拉格朗日量是人为写出来的，而不是能够推导出来的。它是我们对某类体系『唯相』且经验的刻画---译者注

朗日量必须是不变的。这使得我们在任意的参考系中得到的运动方程都是一样的，就像我们之前说的那样 “物理规律在在所有惯性系中相同”。

之后我们就会发现对于自旋 $\frac{1}{2}$ 场的拉格朗日量的另一个对称性：在 $\mathcal{U}(1)$ 变换下的对称性。类似的，对于自旋 1 的场的**内禀对称性**也可以找到。**局域**的 $\mathcal{U}(1)$ 对称性可以使得我们得到自旋 $\frac{1}{2}$ 场和自旋 1 场之间的耦合。带这样的耦合项的拉格朗日量是**量子电动力学的拉格朗日量**的正确形式。类似的局域 $\mathcal{SU}(2)$ 和 $\mathcal{SU}(3)$ 变换会得到**弱和强相互作用的拉格朗日量**的正确形式。

作为补充，我们会讨论**对称性自发破缺**和 **Higgs 机制**。这些使得我们能够去描述有质量的粒子[5]。

[5] 在对称性自发破缺前，拉格朗日量中的描述质量的项会破坏对称性从而被禁止

之后我们得到 **Nother 定理**，它给我们展示了对称性和守恒量之间深刻的联系。我们会通过鉴别每个物理量和它对应的对称性生成元来展现这种联系。这使得我们看到量子力学中最重要的方程[2]

$$[\hat{x}_i, \hat{p}_j] = i\delta_{ij} \tag{1.1}$$

和量子场论中

$$[\hat{\Phi}(x), \hat{\pi}(y)] = i\delta(x-y) \tag{1.2}$$

我们接下来通过对自旋 0 粒子的运动方程，即 Klein-Gordon 方程，取非相对论性[6]极限，得到了著名的 **Schrödinger 方程**。这一点，同我们认识到的物理量与对称性生成元之间的联系，一起构成了**量子力学**的基石。

[6] 非相对论性指所有物体相比光速来说都运动的十分缓慢，从而狭义相对论的古怪的特性很小不会被观察到。

[7] Kelin-Gordan，Dirac，Proka 和 Maxwell 方程

接下来我们从不同的运动方程[7]的解，和方程(1.2)出发来考察**自由场的量子场论**。这之后我们通过仔细审视包含不同自旋场之间的耦合的拉格朗日量来考虑相互作用。这使得我们可以讨论**散射过程的概率振幅**是如何得到的。

通过得到 **Ehrenfest 定理**，量子和**经典力学**的联系被我们展现出来。更进一步，我们够得了**经典电动力学**的基本方程，包括 **Maxwell 方程组**和 **Lorentz 力定律**。

最后简要的介绍现代**引力**理论，**广义相对论**，的基本结构，并点出一些在构建**引力的量子理论**中的困难。

[2] 原书有 typo

这本书的主要部分是我们需要的处理对称性的数学工具，和被称为**标准模型**的理论的导出。标准模型用量子场论来描述所有已知的基本粒子的行为。直到现在，所有的标准模型的实验预言都是正确的。这里介绍的任何其他的理论都可以看作标准模型下的一个特殊情况，比如宏观物体（经典力学），或者低能的基本粒子（量子力学）。对于那些从来没听说过现今已知的基本粒子和它们之间相互作用的读者，我们在接下来一节会进行一个快速的概述。

1.3 Elementary Particles and Fundamental Forces 基本粒子和基本相互作用力

基本粒子分成两个主要的大类：**bosons** 和 **fermions**。**Pauli 不相容原理**是的不能有两个 fermions 处在完全相同的态上，而 boson 则可以有任意多个粒子处在同一个态上。这种自然界奇怪的事实导致了这些粒子截然不同的性质：

- Fermions 构成物质[3]
- Bosons 构成自然界的力

这说明，比如，原子是由 fermions[8]组成的，但是电磁相互作用力是通过被称为光子的 bosons 传播的。这带来的一个最令人震惊的结果是能有稳定的物质形成了。如果允许有无限多的 fermions 处在相同的状态，根本就不会有稳定的物质，我们会在第??中讨论。

[8] 原子有电子，质子和中子组成，它们都是 fermions。然而注意质子和中子都不是基本的粒子，它们有夸克组成，而夸克也是 fermions。

目前我们知道四种基本作用力

- 电磁相互作用力，通过无质量的**光子**传播
- 弱相互作用力，通过有质量的 $\mathbf{W}^+, \mathbf{W}^-$ 和 **Z - bosons** 传播。
- 强相互作用力通过无质量的**胶子**传播。
- 引力，（或许）通过**引力子**传播。

这里面，一些相互作用对应的 bosons 是无质量的，而另一些不是，这种事情给我们揭露了一些自然界的深刻的事情。我们在做好足够的准备之后会去仔细理解这一点，而现在我们只需要记住每一种相互作用力都与一个对称性相关。弱相互作用力传递的 bosons 是有质量的说明其对应的对称性被破坏了。这种**自发对称性破却**是**所有**基本粒子质量的起源。我们后面会看到这一点通过与另一种基本 boson，**Higgs boson** 的耦合来实现。

[3] 这里，『构成』实际上不是很合适的翻译。直译的话应该叫 Fermions 对物质『负责』。不过大体上就是那么个意思吧。---译者注

[9] 在第??章中我们会讲，所有的荷都有着一个相同的美丽的起源

[10] 通常弱相互作用力的荷会伴随着“弱”这个前缀，所以也被称为弱同位旋；而且还存在一种其它的叫同位旋的概念适用于强相互作用里面的的复合物。不仅如此，这不是一个基本的荷，而本书中会省略“弱”这个前缀。

基本粒子如果携带某种**荷**[9]的时候它可以通过某种力来相互作用。

- 对于电磁力来说，这种荷就是**电荷**，而相应的只有带电粒子能够通过电磁力相互作用。
- 对于弱相互作用力来说，这种荷称为[10]**同位旋**。所有已知的 fermions 都携带同位旋，因此他们之间通过弱作用力来相互作用。
- 强作用力的荷叫**色**没因为它与人对色彩的知觉有某种奇妙的共性。由于这个荷实际上与日常生活中的色彩没有任何关系，所以不要因为这个名字而感到困惑。

基本的 fermions 被分成两个之类：**夸克**，组成质子和中子的基本组成，与**轻子**，如电子和中微子。其中的区别体现着夸克可以通过强作用力来相互作用，这就是说他们带色，而轻子则不是这样。夸克和轻子有三代，每一代有两种粒子：

	第一代	第二代	第三代	电荷	同位旋	色
	上 (u)	粲 (c)	顶 (t)	$\frac{+2}{3}e$	$\frac{1}{2}$	$\surd$
夸克:	下 (d)	奇异 (s)	底 (b)	$\frac{-1}{3}e$	$\frac{-1}{2}$	$\surd$
	电子中微子	μ 子中微子	τ 子中微子	0	$\frac{+1}{2}$	-
轻子:	电子	μ 子	τ 子	$-e$	$\frac{-1}{2}$	-

一般的来说，不同的粒子可以通过**标签**来识别。除了荷和质量之外，它们还有很多其他的十分重要的标签，称为**自旋**，你可以把它理解为内禀的角动量，我们会在第4.5.4节中得到这一点。Bosons 携带整数自旋，fermions 携带半整数自旋。我们上面列出来的基本的 fermions 都只有 $\frac{1}{2}$ 的自旋。基本的 bosons 有着 1 的自旋。除此之外，只有一种基本粒子有 0 自旋：Higgs boson。

[11] 除了质量标签。这点目前正在接受实验的检验，比如瑞士日内瓦的 CERN 进行的 ARGIS，ATRAP 和 ALPHA 实验。

[12] 或许还有中微子，这一点目前正在接受各种关于搜索无中微子的双光子通道衰变实验研究。

每种粒子都有其反粒子，带着大小完全一样的相反[11]的标签。比如电子的反粒子就是正电子，但是一般的话我们不会认为的说一个新的名字，只是在原来的粒子前面加上前缀“反”。比如说，顶夸克的反粒子被称为反顶夸克。有些诸如光子[12]一类的粒子自己就是自己的反粒子。

上面提到的这些想法我们在下面都会更仔细具体的解释。现在，我们或许是时候去尝试带着推导的理解这门能够正确描述粒子大观园中的不同特性是如何相互影响在一起的理论了。而第一步，也就是我们下一章的主题，是 Einstein 的著名的狭义相对论。

2. Special Relativity 狭义相对论

著名的迈克尔逊-莫雷 (Michelson-Morley) 实验告诉我们，光在任何参考系中都具有相同的速度[1]。Albert Einstein 第一个意识到这个结果所蕴含的深刻意义，在这一不寻常的自然真理的基础之上，他建立了狭义相对论。从光速不变原理出发，Einstein 预言了许多有趣而又悖于常理的现象，这些现象最后都被证实是正确的。我们将看到这个原理是多么的强大，但是先让我们澄清有关狭义相对论的一些事实。狭义相对论的两个基本假设是

- **相对性原理：** 在任意惯性系中的物理规律均相同，例如两个相对速度恒定的惯性参考系。
- **光速不变原理：** 在任意惯性系中的光速均为常数 c。

除此之外，我们还假设我们的物理时空均匀且具有各向同性。这意味着不论我们在哪儿 (均匀)，不论朝着什么方向 (各向同性) 做实验，物理规律都是一样的。举个例子，两个物理学家，一个在纽约，一个在东京，做完全一样的两个实验，他们会得到相同 [2] 的物理规律，同样的，就算是在火星上的物理学家也会得到同样的物理规律。

一旦物理规律在数学上清晰地表达出来，那么，对于遵从该规

[1]我们日常生活中所观测到的物体的速度取决于选定的参考系。如果一个观者站在火车站上，测得的火车速度为 $50\mathrm{km\,h^{-1}}$，另一个观者以 $15\mathrm{km\,h^{-1}}$ 的速度与火车一同运动，那么测得的火车速度就应是 $35\mathrm{km\,h^{-1}}$。与此不同的是，光始终以 $1.08\times10^{9}\mathrm{km\,h^{-1}}$ 运动，不论你如何相对于光运动。

[2]这当然要考虑到某些参数的影响，例如重力加速度。

律的物理实验而言，不管你是今天做，明天做，或是换个角度来看实验，它所蕴含的物理规律都应是不变的。此外，狭义相对论的第一条假设告诉我们，在匀速运动的马车上，或是在静止的实验室中，同一物理实验都会有相同的结果。以上论述均与实验相符。举个例子，如果你坐在匀速运动的汽车上，你没办法分辨出你处于运动还是静止。

一旦失去各向同性和均匀性，物理学将变得困难重重：我们从实验中得到的物理定律将仅仅在空间中的某一点对于确定的某一方向成立，这样的物理定律显然是毫无用处的。

唯一有些不直观的就是狭义相对论的第二条假设，毕竟它违背了我们的日常经验。虽然如此，至今为此所有的实验都表明这个假设是正确的。

2.1 The Invariant of Special Relativity 狭义相对论的中的不变量

在接下来的章节中，我们将使用狭义相对论的两个基本假设推导出 Minkowski 度规，有了它，我们就能计算两个物理事件的“距离”。所谓物理事件，是指在 Minkowski 时空中的点，而整个狭义相对论都建立在 Minkowski 时空之上。*任意两个不同惯性参考系之间的变换必须保证 Minkowski 度规不变，通过这一条件我们能到找两惯性系之间所有允许的变换。* 在本书剩下的部分我们将会用到这些关于变换的知识，来寻找在这些变换下都成立的方程。让我们从一个思想实验开始，推出一些从基本假设出发得到的重要结论。 试想坐标系中一观者，在原点处向上发出一个光脉冲，经过一段时间后被垂直镜面反射回原点。如图 2.1 所示

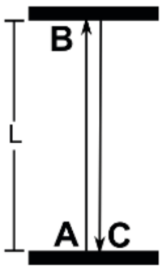


图 2.1: 思想实验图示

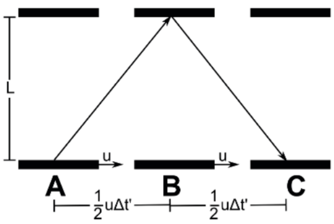


图 2.2: 思想实验图示. 第二位移动观者向左移动，因此第一位观者相对他向右移动

有 3 个重要的事件：

- **A:** 光从原点发出
- **B:** 光在镜面上反射
- **C:** 光回到原点

事件 **AC** 之间的时间间隔为[3]

$$\Delta t = t_C - t_A = \frac{2L}{c} \tag{2.1}$$

式中 L 表示原点与反射点之间的距离。

[3]对于匀速度 v 而言，有 $v = \frac{\Delta s}{\Delta t}$,$\Delta s$ 表示经过的距离，Δt 代表所需时间，因此 $\Delta t = \frac{\Delta s}{v}$

接下来脑补第二位观者，在 t_A 时刻处于他所在的坐标系的原点，并以恒定速度 u 相对于第一个观者[4]向左运动。为简便起见，我们假设第二个参考系的原点在 t_A 时刻与第一位观者的坐标系原点重合。第二位观者所见到的现象就与第一位不一样了。在他的参考系中，光脉冲的起点和终点并不在同一位置 (见图 2.2)。

我们用数学语言表示

$$x'_A = 0 \neq x'_C = u\Delta t' \qquad \rightarrow \qquad \Delta x' = u\Delta t' \tag{2.2}$$

带撇物理量代表第二位观者所测量的量。对于第一位静止系的观者而言有

$$x_A = x_C \qquad \rightarrow \qquad \Delta x = 0 \tag{2.3}$$

我们假定第二位观者运动沿 x 轴，因此

$$y'_A = y'_C \qquad z'_A = z'_C \quad \rightarrow \quad \Delta y' = 0 \qquad \Delta z' = 0 \tag{2.4}$$

那么同样也有

$$y_A = y_C \qquad z_A = z_C \quad \rightarrow \quad \Delta y = 0 \qquad \Delta z = 0 \tag{2.5}$$

接下来的问题是：**第二位观者所测得的时间间隔是多少？**因为我们假定了光速为常数，那么事件 **AC** 对于第二位观者而言将具有不同的间隔！时间间隔 $\Delta t' = t'_C - t'_A$，等于光在第二位观者的参考系中走过的距离 l 除以光速 c。

$$\Delta t' = \frac{l}{c} \tag{2.6}$$

我们可以利用古老的 Pythagoras(毕达哥拉斯) 定理 (见图 2.2) 计算光传播的距离

$$l = 2\sqrt{\left(\frac{1}{2}u\Delta t'\right)^2 + L^2} \tag{2.7}$$

[4]那些能将两相对速度恒定的惯性参考系中的物理量相互转化的变换称为**推动 (Boost)**，后面我们会对此进行详述。

利用式2.6可以得到

$$c\Delta t' = 2\sqrt{\left(\frac{1}{2}u\Delta t'\right)^2 + L^2} \tag{2.8}$$

再利用式2.2中的 $\Delta x' = u\Delta t'$ 可得

$$c\Delta t' = 2\sqrt{\left(\frac{1}{2}u\Delta t'\right)^2 + L^2}$$

$$\to (c\Delta t')^2 = 4\left(\left(\frac{1}{2}u\Delta t'\right)^2 + L^2\right)$$

$$\to (c\Delta t')^2 - (\Delta x')^2 = 4\left(\left(\frac{1}{2}u\Delta t'\right)^2 + L^2\right) - (\Delta x')^2 = 4L^2 \tag{2.9}$$

现在回到式2.1，即 $\Delta t = \frac{2L}{c}$，那么

$$(c\Delta t')^2 - (\Delta x')^2 = 4L^2 = (c\Delta t)^2 = (\Delta t c)^2 - \underbrace{(\Delta x)^2}_{=0\text{由式2.3知}} \tag{2.10}$$

终于，我们得到[5]

$$(c\Delta t')^2 - (\Delta x')^2 - \underbrace{(\Delta y')^2}_{=0} - \underbrace{(\Delta z')^2}_{=0} = (c\Delta t)^2 - \underbrace{(\Delta x)^2}_{=0} - \underbrace{(\Delta y)^2}_{=0} - \underbrace{(\Delta z)^2}_{=0} \tag{2.11}$$

考虑第三个观者，相对于第一个观者以不同的速度运动，用同样的推理可以得到

$$(c\Delta t'')^2 - (\Delta x'')^2 - (\Delta y'')^2 - (\Delta z'')^2 = (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \tag{2.12}$$

[5]注意到我们在此处采用的是得到这个结果的最简方法，因为我们假定 t_A 时刻两坐标的原点重合。尽管如此，我们能够证明任意选择两惯性参考系原点的关系仍然能让结论成立，如果第二位观者的参考系运动方向任意，那就不再有 $\Delta y' = 0$ 和 $\Delta z' = 0$，但能够证明方程仍然是成立的，因为物理定律在任意惯性系中都是一样的，这也给了我们任意选择参考系来简化计算的自由，证明任意情况无非是再多花些功夫罢了。

因此，我们得到了一些对于所有观者都相同的量: 即二次式

$$(\Delta s)^2 \equiv (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \tag{2.13}$$

此外，我们还能看出对于不同观者，$(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$ 或者说 $(c\Delta t)^2$ 是不同的。我们将在下一节讨论不变量所蕴含的物理意义。

2.2 Proper Time 固有时

我们在上一节推导出了狭义相对论中的不变量 Δs^2，即在所有观者眼中都相同的量。接下来我们要讨论这个量的物理意义。为简便起见，我们将问题限制在一维空间中。对于一个相对于观者静止的物体，我们能作出它的时空图 (见图 2.3)。相应的，一个匀速运动的物体能作出如图 2.4 所示的时空图。

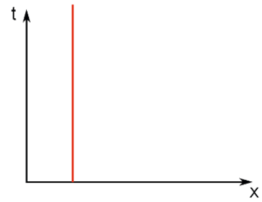


图 2.3: 静止物体的世界线。物体位置不随时间流逝而变化。

我们画来用于确定物体在时空中位置的线称为**世界线 (world line)**。世界线总是依赖于观者的。两不同观者对于同一物体可能会画出完全不同的世界线。若一观者眼中的物体时空图为图 2.4，那么对于速度与物体相同的观者，其时空图将为图 2.5，即对于此观者物体静止。为了解释两位观者给出的不同描述，我们引入 x' 和 t' 表示第二位观者的参考系中的时间和位置。

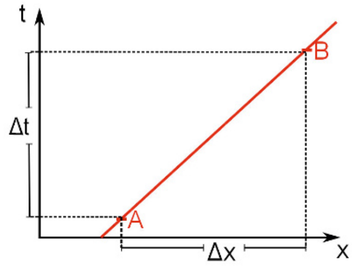


图 2.4: 某匀速直线运动物体的世界线。物体前后经过某两点记为事件 A，B。它们间的空间距离为 Δx，时间间隔为 Δt

我们可以看到，两个观者将对事件 AB 之间的位移持不同观点。对于第一位观者，$\Delta x \neq 0$，但对于第二位观者，$\Delta x' = 0$。两观者都认为事件 A 和 B 之间的时间间隔非零：$\Delta t \neq 0$ 和 $\Delta t' \neq 0$，且认为 Δs^2 也相同 (见上节推导的结论，任意观者都有相同的 Δs^2)。这将导出一个令人惊讶的结论：事件 A 到 B 经历的时间对于两观者不同

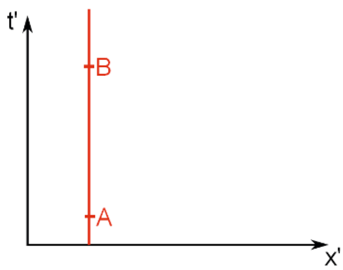


图 2.5: 同一个运动物体的世界线，但观者与该物体速度相同。该观者观测到的事件 AB 之间的空间距离为 $\Delta x' = 0$

$$(\Delta s)^2 = (c\Delta t)^2 - (\Delta x)^2 \tag{2.14}$$

$$(\Delta s')^2 = (c\Delta t')^2 - \underbrace{(\Delta x')^2}_{=0} = (c\Delta t')^2 \tag{2.15}$$

$$(\Delta s)^2 = (\Delta s')^2 \rightarrow (\Delta t')^2 \neq (\Delta t')^2 \quad 因为 (\Delta x)^2 \neq 0 \tag{2.16}$$

这是有关狭义相对论最著名的一个效应，通常称为**钟慢效应 (time-dilation)**。时间间隔依赖于观察者，正如空间距离一样。不同观者时间流逝速度不同，因此两事件经历的时间也就自然不同了。

现在时间将变为一个相对的概念，如果我们能找到一个对于所有观者都相同的时间量，这将会非常有用。在上述例子中，第二位观者与物体有相同的速度，有

$$(\Delta s')^2 = (c\Delta t')^2 \tag{2.17}$$

上式意味着狭义相对论中的不变量恰为观者所观测到的时间间隔乘上光速 c。这让我们有机会能用 $(\Delta s)^2$ 定义一个对于所有观者相同的时间量。我们定义

$$(\Delta s)^2 = (c\Delta \tau)^2 \tag{2.18}$$

τ 称为**固有时 (proper time)**。固有时是所有相对于该物体静止的观者所测量的时间。

当然，现实生活中的物体绝非只能做匀速运动，但我们可以取足够短的时间 (无穷小)，使得运动近似为匀速运动，这样固有时的概念就说得通了。因此数学上我们需将概念过渡到无穷小，即 $\Delta \to d$：

$$(ds)^2 = (cd\tau)^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2 \tag{2.19}$$

因此，就算一个物体到处乱跑，我们也能假设一个与物体一起运动的观者，这个观者自带时钟，且在其看来物体静止。对于这个特殊观者而言，他所测量的时间间隔就是固有时，这一数值对其余观者也是相同的，因为 $(ds)^2 = (cd\tau)^2$ 对于任意观者成立。再次强调，这并不意味着对所有观者都有同样的时间间隔，只是这些观者承认该特殊观者所测量的时间间隔为一不变量罢了。

2.3 Upper Speed Limit 速度上限

在上节中我们对狭义相对论这一不变量有了解释，现在我们可以更进一步，导出另一个狭义相对论的惊人结论。

由于不变量的定义中带有负号，这意味着对于时空中的两个

事件，$(\Delta s)^2$ 的值可能为 0。$(\Delta s)^2$ 的值甚至可能是负的，但这样我们将得到一个复的固有时[6]，同常复的固有时被认为没有物理意义。因此，$(\Delta s)^2 = 0$ 时固有时最小，$\tau = 0$，令

$$\Delta s_{min}^2 = 0 = (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2$$

$$\to (c\Delta t)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

$$\to c^2 = \frac{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}{(\Delta t)^2} \tag{2.20}$$

在上式右侧我们有速度平方 v^2，也就是距离除上时间，对该式取极限，则

$$\to c^2 = \frac{(dx)^2 + (dy)^2 + (dz)^2}{(dt)^2} \tag{2.21}$$

$x(t)$，$y(t)$，$z(t)$ 是描述两事件之间路径的参数方程，即右侧仍为两事件间的运动速度。

因此，当某一观者测得的固有时为 0 时其速度将满足下式

$$\to c^2 = v^2 \tag{2.22}$$

这意味着没有物体能超光速 (否则其固有时为复)！**我们在物理学中有了一个速度上限。**时空中的两事件的联系不能超光速。

这一点满足物理学上的**局域性原理 (principle of locality)**，即物理学中的对象只能被它临近的对象影响。相互作用都是局域的，不存在超距作用，因为相互作用都需要时间。

2.4 The Minkowski Notation Minkowski 记法

> 从现在起，孤立的空间和孤立的时间注定要消失成为影子，只有两者的统一才能保持独立的存在。
>
> **-Hermann Minkowski**[7]

[6] 按固有时定义，$(ds)^2 = (cd\tau)^2$，$(ds)^2 < 0 \to d\tau$ 为复

[7] 出自 Hermann Minkowski 在 Assembly of German Natural Scientists and Physicians(1908.9.21) 上的演讲

写出狭义相对论中不变量

$$ds^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2 \tag{2.23}$$

我们在此处将采用一种新的记法，这种记法初学时可能稍显复杂，不过后文中将经常用到 ~~(所以你就老实学吧)~~:

$$ds^2 = \eta^{\mu\nu} dx_\mu dx_\nu = \eta^{00}(dx_0)^2 + \eta^{11}(dx_1)^2 + \eta^{22}(dx_2)^2 + \eta^{33}(dx_3)^2$$

$$= dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2 \tag{2.24}$$

我们在这儿用到了一些新的约定和符号，这些东西在近代物理中极为常见，所以越早熟悉越好:

- Einstein 求和约定：某指标在同一项内出现两次则代表遍历该指标并求和。例如：$\sum_{i=1}^{3} a_i b_i = a_i b_i = a_1 b_1 + a_2 b_2 + a_3 b_3$，但 $\sum_{i=1}^{3} a_i b_j = a_1 b_j + a_2 b_j + a_3 b_j \neq a_i b_j$
- 像 μ，ν，σ 这样的希腊字母 [8] 作指标时，一般代表从 0 到 3 求和：$x_\mu y_\mu = \sum_{\mu=0}^{3} x_\mu y_\mu$
- 我们约定变量 $x_0 \equiv ct$，$x_1 \equiv x$，$x_2 \equiv y$，$x_3 \equiv z$。这样时间空间地位等同，也便于我们使用 Einstein 求和约定
- 引入 Minkowski 度规 (η 看作一矩阵，η^{ij} 为其矩阵元)：$\eta^{00} = 1$，$\eta^{11} = -1$，$\eta^{22} = -1$，$\eta^{33} = -1$，当 $\mu \neq \nu$ 时 $\eta^{\mu\nu} = 0$ (我们也可简写为 [9] $\eta^{\mu\nu} = diag(1, -1, -1, -1)$)

除此之外，我们还要引入**四维矢量 (four-vector)**，简称 *4 矢*。

[8] 像 i, j, k 这样的罗马字母作指标时一般代表从 1 到 3 求和：$x_i x_i \equiv \sum_i^3 x_i x_i$。后面我们还会用到大写的罗马字母如 A，B，C，它们作指标时代表从 1 到 8 求和。

[9] $$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

$$dx_\mu = \begin{pmatrix} dx_0 \\ dx_1 \\ dx_2 \\ dx_3 \end{pmatrix} \tag{2.25}$$

式2.24可用 4 矢和 Minkowski 度规表出

$$\begin{aligned}(ds)^2 = dx_\mu \eta^{\mu\nu} dx_\nu &= \begin{pmatrix} dx_0 & dx_1 & dx_2 & dx_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} dx_0 \\ dx_1 \\ dx_2 \\ dx_3 \end{pmatrix} \\ &= dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2 \end{aligned} \tag{2.26}$$

采用这种记法能让我们省不少事。对于 ds，我们给出的解释是时空 (Minkowski 时空) 中两事件的“距离”，这个“距离”并不只是空间距离，它还将时间间隔考虑在内。如果我们考虑 3 维 Euclidean 空间[10]中两点间最短距离的平方[11]

$$\begin{aligned}(ds)^2 = dx_i \delta^{ij} dx_j &= \begin{pmatrix} dx_1 & dx_2 & dx_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \\ dx_3 \end{pmatrix} \\ &= (ds)^2 = (dx_1)^2 + (dx_2)^2 + (dx_3)^2 \end{aligned} \tag{2.27}$$

度规 (metric) 能够告诉我们无限临近的两点间的距离。在 Euclidean 空间中度规就是单位矩阵 δ_{ij}。在广义相对论的弯曲时空中将会有更复杂的度规，但在狭义相对论中我们采用的是相对简单的 Minkowski 度规 $\eta^{\mu\nu}$。度规是计算长度的工具，我们可以通过度规定义 **4 矢长度 (length of a four-vector)**，即 4 矢与

[10] 3 维 Euclidean 空间就是经典物理中一般的空间，在此空间中我们将时间和空间分别处理，也就是未将时空考虑成一个整体的几何结构来考虑。但一旦将时间作为新坐标，结合出时空的观念，我们便能将时间空间放在同一坐标下考虑。

[11] 克罗内克函数 (Kronecker delta)δ_{ij} 即是单位矩阵在指标记法下的表示，详见附录 B5.5

自身的标量积[12]

$$x^2 = x \cdot x \equiv x_\mu x_\nu \eta^{\mu\nu}$$

类似地，两任意 4 矢的标量积定义为

$$x \cdot y \equiv x_\mu y_\nu \eta^{\mu\nu} \tag{2.28}$$

对于上下标，也有一些约定能使计算过程更清晰。我们定义带有上指标的 4 矢[13]

$$x^\mu \equiv \eta^{\mu\nu} x_\nu \tag{2.29}$$

或者

$$y^\nu \equiv \eta^{\mu\nu} y_\mu \underbrace{=}_{Minkowski\text{度规具有对称性}\eta^{\mu\nu}=\eta^{\nu\mu}} \eta^{\nu\mu} y_\mu \tag{2.30}$$

因此，标量积可以写为[14]

$$x \cdot y \equiv x_\mu y_\nu \eta^{\mu\nu} = x_\mu y^\mu = x^\nu y_\nu \tag{2.31}$$

上式中，上指标的选择是任意的，这样做只是为了避免公式中总是出现 $\eta^{\mu\nu}$，正如引入 Einstein 求和约定来避免总是出现求和号一样。

2.5 Lorentz Transformations Lorentz 变换

下一步，我们将尝试找出两参考系之间不违背狭义相对论基本假设的变换方式。由上文可知，从狭义相对论的两个基本假设可以推出对于所有惯性参考系均有 $ds^2 = \eta^{\mu\nu} dx'_\mu dx'_\nu$，即

$$ds'^2 = dx'_\mu dx'_\nu \eta^{\mu\nu} = ds^2 = dx_\mu dx_\nu \eta^{\mu\nu} \tag{2.32}$$

---

[12]此定义在 Euclidean 空间中同样适用：由于度规为 $\delta_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$，易得矢量 $\vec{v}$ 的长度 $= \vec{v} \cdot \vec{v} = v_1^2 + v_2^2 + v_3^2$。

[13]带有下指标的 4 矢通常叫做协变 4 矢，带有上指标的 4 矢通常叫做逆变 4 矢。

[14]指标的名称如何选取并不影响最后的值，详见附录 B5.1。

因此，两参考系间所允许的变换要能保证二次式的形式和 Minkowski 时空中的标量积在变换下不变。设变换为 Λ，那么变换后

$$dx_\mu \to dx'_\mu = \Lambda_\mu^\sigma dx_\sigma \tag{2.33}$$

由于 $(ds)^2$ 在变换下不变

$$(ds)^2 = (ds')^2$$

$$\to dx \cdot dx \stackrel{!}{=} dx' \cdot dx' \to dx_\mu dx_\nu \eta^{\mu\nu} \stackrel{!}{=} dx'_\mu dx'_\nu \eta^{\mu\nu} \underbrace{=}_{\text{式}2.33} \Lambda_\mu^\sigma dx_\sigma \Lambda_\nu^\gamma dx_\gamma \eta^{\mu\nu}$$

$$\underbrace{\to}_{\text{重命名哑指标}} dx_\mu dx_\nu \eta^{\mu\nu} \stackrel{!}{=} \Lambda_\sigma^\mu dx_\mu \Lambda_\gamma^\nu dx_\nu \eta^{\sigma\gamma}$$

$$\underbrace{\to}_{dx_\nu\text{任意}} \stackrel{!}{=} \Lambda_\sigma^\mu \Lambda_\gamma^\nu \eta^{\sigma\gamma} \tag{2.34}$$

或者用矩阵形式来写[15]

$$\eta = \Lambda^{\mathrm{T}} \eta \Lambda \tag{2.35}$$

此即变换 Λ_μ^ν 所需满足条件。

这个条件现在看起来可能有些奇怪，但在后文中我们会用更自然的方式导出这个条件 ~~(所以不要慌)~~。在下一章里，我们会看到 Euclidean 空间中的旋转所导出的变换 [16]O 能够保证 Euclidean 空间的标量积不变[17]

$$\vec{a} \cdot \vec{b} \stackrel{!}{=} \vec{a}' \cdot \vec{b}' \underbrace{=}_{\text{注意}(Oa)^{\mathrm{T}}=a^{\mathrm{T}}O^{\mathrm{T}}} \vec{a}^{\mathrm{T}} O^{\mathrm{T}} O \vec{b} \tag{2.36}$$

因此[18] $O^{\mathrm{T}}\mathbf{1}O \stackrel{!}{=} \mathbf{1}$，恰巧 Euclidean 空间的度规也是单位矩阵 $\mathbf{1}$，其地位正如 Minkowski 度规在式2.35 中一样。按旋转原本的定义，

[15]详见附录 C.1。

[16]稍后会讲明 O 的意义。

[17]“$\cdot$”表示的是矢量的点乘，如果我们将矢量写为列向量，那么按矩阵相乘的定义有的 $\vec{a} \cdot \vec{b} = \vec{a}^{\mathrm{T}} \cdot \vec{b}$。有关 $(Oa)^{\mathrm{T}} = a^{\mathrm{T}} O^{\mathrm{T}}$ 这一点详见附录 C.1，式 C.3。

[18]这一条件一般称为**正交性 (orthogonality)** 条件，因此用字母 O 表示。满足 $O^{\mathrm{T}} O = \mathbf{1}$ 的矩阵称为正交矩阵，因为它列与列之间是正交的。换句话说，矩阵的每一列都可以看做一个矢量，正交性条件就是说这些矢量相互正交。

旋转本身不改变矢量的长度，即是保证标量积不变 [19]。此外我们还注意到旋转不改变坐标系的方向[20]，这表明 $\det O \overset{!}{=} 1$，因为保证标量积不变的变换除了旋转变换还有空间反演[21]。

我们定义保 Minkowski 时空中标量积不变的变换为 **Lorentz 变换 (Lorentz transformations)**，这也是为保证狭义相对论的假设所必需的。相应的，每当我们需要得到在 Lorentz 变换下不变的项时，都必须结合上下指标：$x_\mu y^\mu = x_\mu y_\nu \eta^{\mu\nu}$。在我们学会了一些非常优雅的技术来处理这样的情况之后，我们会在下一章构建这些转换具体的矩阵形式。

2.6 Invariance, Symmetry and Covariance 不变性，对称性，协变性

在我们继续之前，有一些重要的概念要事先声明。首先，一个量要称为**不变量 (invariant)**，这个量必须在变换下不变。比如说，我们考虑一个依赖于不同的量 $A, B, C, \dots$ 任意函数 F，$F = F(A, B, C, \dots)$，如果我们将 $A, B, C, \dots$ 变换为 $A', B', C', \dots$，有

$$F(A', B', C', \dots) = F(A, B, C, \dots) \tag{2.37}$$

那么 F 称为变换下的一个不变量。我们可以用对称性来描述。**对称性 (symmetry)** 是指在某一变换下或者某一系列变换下保持不变的性质。举个例子，如果说一个物理系统在任意的旋转变换下不变，那么该系统具有旋转对称性。再比如说，一间室温为常温的房间，房间内各点的温度与位置无关，换言之，如果把所有的点朝着某特定方向平移一段距离，室温不变。因此我们说室温具有平移对称性。

协变性与不变性有共通之处而又不同。如果一个方程在变换下形式不变，那么称这个方程具有协变性。例如下面这个式子

$$E_1 = aA^2 + bBA' + cC^4$$

[19]矢量的长度即是矢量与自身的标量积的平方根。

[20]详见附录 A.5。

[21]空间反演可简单地理解为映射 $\vec{x} \to -\vec{x}$。这种变换满足 $\det O \overset{!}{=}$ 及 $O^{\mathrm{T}} O = \mathbf{1}$。因此，如果我们要将变换限制为旋转变换，就需要加上条件 $\det O \overset{!}{=} \mathbf{1}$。此外，空间反演变换又称宇称变换。

在变换之后这个方程写为

$$E_1' = aA'^2 + bB'A' + cC'^4$$

那么这个方程具有协变性，因其在变换下形式不变。另一方程

$$E_2 = x^2 + 4axy + z$$

若在变换之后写为

$$E_2' = y'^3 + 4az'y' + y'^2 + 8z'x'$$

那么它就不是协变的，因为其形式已经变化。

所有的物理规律都应具有 Lorentz 协变性[22]，因为只有这样的物理规律才能在任意参考系下成立。非协变的物理规律只能在某一特定参考系下成立，这会导致在东京和纽约具有不同的物理规律。显然这个主意糟透了，因为不存在一个特殊的参考系，我们得让物理规律在任意参考系中成立。我们将在后文中学到如何用协变的方法计算物理规律。

[22]即在 Lorentz 变换下具有协变性---译者 (InSight)

II Symmetry Tools 对称性工具

3. Lie Group Theory　Lie 群

本章概述

本章的最终目的是导出 **Poincare 群双覆盖的基本表示**，物理学现在认为 Poincare 群是时空根本的对称性群。这些基本表示是描述所有基本粒子的必要工具，每一种表示对应一种基本粒子，它们揭示了自然界存在何种基本粒子。

我们从两个简单例子引出**群**的定义，然后作为学习 Lie 群理论的第一步，我们讨论描述二维旋转变换的两种方式:

- 2×2 旋转矩阵。
- 单位复数。

接着我们尝试找出描述三维旋转的第二种方法 (像复数那样，当然第一种方法是 3×3 矩阵)，第二法与一种超级重要的群 — $\mathcal{SU}(2)$。[1]有关。之后我们研究 **Lie 代数**，使用简便的 Lie 代数能够深入研究复杂的 Lie 群。不同的群可以有相同的 Lie 代数，但只有其中的一部分是基本的。从上述基础出发就能准确揭示自然的基本对称性群 — Poincare 群的双覆盖。我们将利用已知的变换操作导出 Lie 代数，并利用 Lie 代数得出不同的对称变换表示。这样就能看出我们开始时使用的表示其实只是一种特殊情况。这样我们又能研究 Poincare 群的重要部分 — **Lorentz 群**，我们会看到 Lorentz 群双覆盖的 Lie 代数由两份 $\mathcal{SU}(2)$ Lie 代数所组成，因此我们可以直接利用熟悉的 $\mathcal{SU}(2)$ 群的结论。最后我们将平移

下面的图表是本章结构的示意图。当你迷失方向的时候记得回来看看，初学的时候不太需要看它。~~反正也看不懂~~

$$\begin{tikzcd}[column sep=small] & \text{二维旋转} & \\ \mathcal{U}(1) \arrow[ur] & & \mathcal{SO}(2) \arrow[ul] \end{tikzcd}$$

$$\begin{tikzcd}[column sep=small] & \text{三维旋转} & \\ \mathcal{SU}(2) \arrow[ur] & & \mathcal{SO}(3) \arrow[ul] \end{tikzcd}$$

$$\begin{tikzcd} \text{Lorentz 变换} \arrow[d] \\ \text{Lie 代数} \doteq \mathfrak{su}(2) \oplus \mathfrak{su}(2) \arrow[d] \\ \text{双覆盖表示} \end{tikzcd}$$

$$\begin{tikzcd} \text{Lorentz 变换 + 平移} \arrow[d] \\ \text{Poincare 群} \end{tikzcd}$$

[1]$\mathcal{S}$ 表示特殊 (special)，它的含义为 $\det(M) = 1$。U 表示幺正: $M^\dagger M = 1$，数字 2 表示这个群起初是用 2×2 矩阵定义的。

变换考虑进来，这就是 Poincare 群，Poincare 群就是 Lorentz 群加上平移。完成上述所有之后我们终于可以将 Poincare 群双覆盖的基本表示进行分类，这些在后面的章节中会大用特用，我们将从中导出物理学的基本定律。

3.1 Groups 群

我们需要合适的数学工具描述对称性~~以和民科 (贬义的) 区分开~~。描述对称的数学分支称为**群论**。群论的一个分支 **Lie 理论**~~谎言理论~~描述连续的对称性，物理中经常遇到这种情况。

我们把对称性定义为变换下的不变性，而描述对称的群就定义为某些变换的集合。让我们从两个简单例子开始体会群到底该怎么定义吧。

1. 正方形是一些点的集合 (例如四个顶点是该集合的一部分)，正方形的对称性是在某些变换下 (变换: 将一个点映射到另一个点) 保持不变的性质。
   符合条件的变换有绕中心旋转 $90°, 180°, 270°, 0°$ 等等。这些旋转操作将正方形映射到它自身。我们称这个集合 (正方形点集) 在这样的变换下具有不变性。
   注意不是所有旋转变换都对称。我们关注顶点的变化就能看出来，比如绕中心顺时针旋转 5°，这个变换将顶点映射到原来正方形点集之外，顶点 A 映射到了原来正方形集合之外的点 A'。因此这个旋转变换对正方形不具有对称性。当然，变换后的点集仍然是个正方形，但却是不同的正方形 (即不同点的集合)。绕中心转 90° 是对称的，如图3.3，顶点 A 变换到点 B，等等，原来的正方形点集变换到相同的集合。

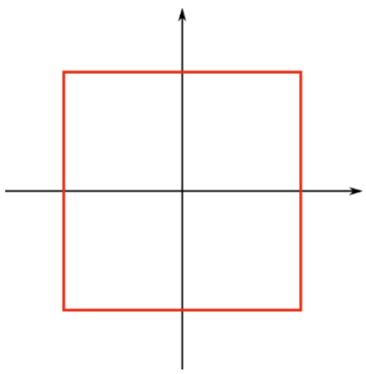


图 3.1: 正方形

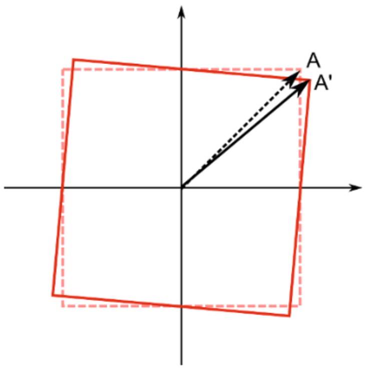


图 3.2: 正方形绕中心顺时针旋转 5°

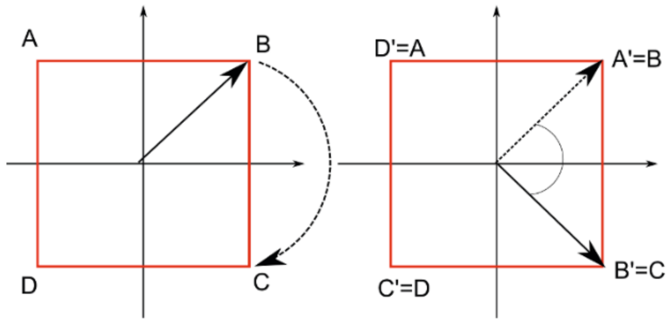


图 3.3: 正方形绕中心旋转 90°

假如你看了一眼原来的正方形，然后闭上眼睛，这时有人把正方形做了一个变换。如果你不能分辨这个正方形是否发生变化，那么这个变换就是个对称变换。

我们把所有使正方形不变的变换构成的集合称为群。变换参数 (本例中就是旋转角度) 不能任意取值 (而是取分立的数)，这个群称为离散群。

2. 另一个例子是使单位圆不变的变换构成的集合。类似地，单位圆还是一些点构成的集合，对称变换把这个集合映射到它自身。

   单位圆绕圆心旋转任意角度都不变。换言之，变换参数 (这里就是旋转角度) 可以取任意值，因此这个群称为连续群。

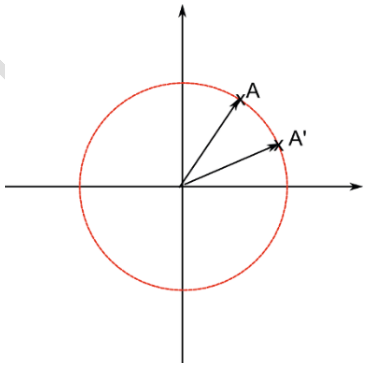


图 3.4: 单位圆绕圆心旋转任意角度都不变

数学中除了几何图形之外还有好多别的都有对称性。例如，对于向量，我们可以考虑所有让向量长度不变的变换构成的集合。看出本节开头对称性定义的普遍性了没？对称性就是变换下的不变性。非常幸运，搞数学的早已建立了群论，它可以研究所有类型的对称性[2]。

[2] 历史上数学家们建立群论最初是为了描述方程的对称性

为了让精确描述对称性的工具 — 群的定义来的自然一点，我们把对称性的定义用数学语言精炼一下:

- 对某物什么也不做 (比如转个 $0°$) 当然也是对称变换 (按照我们的对称性定义)，因此任何群都需要包含一个恒等变换 (恒等元)。在上面的例子里，恒等元就是旋转 $0°$。
- 对某物先做变换再做逆变换的结果必须等于啥也不做。因此对于群中的任意元素 (简称群元)，必须有相应的逆元素。按定义，先做变换再做逆变换等于恒等变换。例如先转 $90°$ 再转 $-90°$(先逆针再顺针转) 与旋转 $0°$ 相同。
- 先做一个对称变换再做一个对称变换，总体效果必须还是对称变换。例如先旋转 $90°$ 再转 $180°$ 等于直接转 $270°$，后者也是对称变换。对称变换的这个性质称为封闭性。
- 对称变换之间必须满足结合律。例如先转 $90°$ 再转 $40°$ 再转 $110°$ 与先转 $130°$ 再转 $110°$ 相同，先转 $90°$ 再转 $150°$ 也一样。用符号表示更加清楚:

$$R(110°)R(40°)R(90°) = R(110°)\Big(R(40°)R(90°)\Big) = R(110°)R(130°) \tag{3.1}$$

$$R(110°)R(40°)R(90°) = \Big(R(110°)R(40°)\Big)R(90°) = R(150°)R(90°)$$

(3.2)

这个性质称为**结合律**。

- 我们要有规定两个群元 (对称变换) 怎样合体的规则，它是个**二元运算** (两个群元合体成一个)，我们称它为**群乘法**。
  在上面的例子里，旋转变换的标准表示方法是用旋转矩阵[3]，两个群元 (两个旋转矩阵，或两个旋转操作) 合体的规则是线性代数里的矩阵乘法。同样的变换经常有不同的表示方法[4]，群论非常系统性地概括了所有形式。群论的分支 — **表示论**就是研究相同变换的不同描述方式的，我们在**??**节学习表示论。

[3] 旋转矩阵见附录 A.2.

[4] 例如，二维平面旋转还可以用单位复数描述，相应的群乘法是复数乘法。稍后就讨论这一点。

我们把上述群与对称变换的特征用严谨的数学语言表达，再将它们提升为公理，满足这些公理的数学结构就是一个群。数学系的群论书可能更喜欢在开头就从天上掉下来这些公理。必须指出满足群公理的结构可能是超级抽象的，但我们现在只关注上面例子里的旋转变换那样的群。~~因为我们是物理系的，而且这是本物理书~~

(我们通过对称变换的性质导出的) 群公理[5]:

一个群就是一个集合 G 加上一个定义在 G 上的二元运算 (群乘法)$\circ$，当然 $(G, \circ)$ 还要满足以下公理:

- 封闭性: 对于任意 $g_1, g_2 \in G, g_1 \circ g_2 \in G$
- 单位元: 存在单位元 $e \in G$ 使得对于所有 $g \in G, g \circ e = g = e \circ g$
- 逆元: 对于任意 $g \in G$，存在相应的逆元 $g^{-1} \in G$ 使得 $g \circ g^{-1} = e = g^{-1} \circ g$
- 结合律: 对于任意 $g_1, g_2, g_3 \in G, g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$

[5] 别担心怎么才能根据这些公理凑出一个群来。搞物理的往往是从某个变换出发，考察它是否符合群公理，如果是~~经常是~~那么就可以应用群论来解决问题。

总结: 某物体在一些变换下保持不变，这些变换组成[1]的集合叫做对称群。对于 Minkowski 时空，Minkowski 度规[6]在变换下保持不变，相应的对称群称为 Poincare 群。

[6] 复习一下，Minkowski 度规就是在 Minkowski 空间中用来计算距离和长度的工具，见第二章。

要注意群的定义完全与变换的物体是啥没有关系。我们可以脱离特定物体而只研究对称变换本身，群的定义将变换从物体中 '提取' 出来了。这是非常有用的，许多不同事物具有同样或同类的对

[1] 我在想这里该用 '构成' 还是 '组成'，后来我觉得无所谓，因为这不是初中化学...(分子构成物质，元素组成物质...) — 译者 (SI)

称性。群论让我们不用管变换的物体 (圆还是正方形)，只研究变换 (例如旋转) 的普遍性质。

3.2 Rotations in two Dimensions　二维旋转

我们从一个简单而重要的例子开始学习群论。考虑那些让二维平面中的向量长度不变的对称变换。符合条件的有旋转和反射[7]。它们也是圆的对称变换，一个单位元旋转或反射之后还是单位圆。可见同一个群（对应一种变换）可以作用于不同类物体：圆（几何图形），或者向量。本节考虑向量，可以用旋转矩阵表示向量的旋转或反射，[8]，将起点位于原点的向量绕原点逆时针旋转 θ 角度的旋转矩阵为：

[7] 当然还有平移。平移的数学描述与前两者有些不同，我们之后再讨论它。

[8] 旋转矩阵的推导见附录 A.2.

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{3.3}$$

关于 x 轴与 y 轴的反射变换用矩阵表示为：

$$P_x = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \quad P_y = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.4}$$

将矩阵乘法作为群乘法 $\circ$, 可以验证 $R(\theta), P_x, P_y$ 与矩阵乘法符合群公理，因而构成一个群，亦即旋转与反射变换构成一个群。

可以用更抽象的方式表达‘所有让二维向量长度不变的变换’，向量长度就是向量与自身点乘的平方根 ($|a| = \sqrt{a \cdot a}$). 向量长度在变换 $a \to a'$ 下不变意味着[2]：

$$a' \cdot a' \stackrel{!}{=} a \cdot a \tag{3.5}$$

将变换矩阵记作 $\mathcal{O}$，变换即为 $a \to a' = \mathcal{O}a$.

$$a \cdot a = a^{\mathrm{T}} a \to a'^{\mathrm{T}} a' = (\mathcal{O}a)^{\mathrm{T}} \mathcal{O}a = a^{\mathrm{T}} \mathcal{O}^{\mathrm{T}} \mathcal{O} a \stackrel{!}{=} a^{\mathrm{T}} a = a \cdot a \tag{3.6}$$

由此可见，使向量长度不变的变换必须满足：

$$\mathcal{O}^{\mathrm{T}} \mathcal{O} = I \tag{3.7}$$

[9] $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

其中 I 表示单位矩阵[9]。前文的旋转和反射矩阵都满足此条

[2] 等号上面的! 只是起强调、提示作用 — 译者 (SI)

件[10] 2×2 矩阵中所有满足3.7式的矩阵构成了 $\mathcal{O}(2)$ 群，即所有 2×2 正交矩阵构成的群[11]。我们可以找出这个群中描述旋转变换的那一部分（构成一个子群）。根据3.7式：

$$\det(\mathcal{O}^{\mathrm{T}}O) = \det(O^{\mathrm{T}})\det(O) \stackrel{!}{=} \det(I) = 1$$
$$(\det(\mathcal{O}))^2 \stackrel{!}{=} 1 \rightarrow \det(\mathcal{O}) \stackrel{!}{=} \pm 1 \tag{3.8}$$

$\det \mathcal{O} = 1$ 的矩阵对应旋转变换[12]

条件

$$\mathcal{O}^{\mathrm{T}}\mathcal{O} = I \tag{3.9}$$
$$\det \mathcal{O} = 1 \tag{3.10}$$

定义了 $\mathcal{SO}(2)$ 群，‘$\mathcal{S}$’ 表示 ‘特殊’ (special)，‘$\mathcal{O}$’ 表示 ‘正交’ (orthogonal)。

$\mathcal{SO}(2)$ 包含的旋转变换保持了坐标系的取向，即一个右手坐标系[13]经旋转变换后还是右手系，而反射变换会改变它的取向。用线性代数的概念来说，我们规定 $\mathcal{SO}(2)$ 中的矩阵行列式必须为 $+1$。

3.2.1 Rotations with Unit Complex Numbers 单位复数表示旋转变换

还可以用单位复数来表示旋转变换：向量绕原点旋转 θ 角可以表示为用单位复数 z 乘此向量（单位复数意为 $z = a + ib$, $|z|^2 = z^*z = 1$）[14]。

所有单位复数构成一个群，称为 $\mathcal{U}(1)$，群乘法即为复数乘法，不难验证它符合群公理。为了看出它与之前引入的 $\mathcal{O}(3), \mathcal{SO}(3)$ 群的关系，我们将 $\mathcal{U}(1)$ 群的条件 — 单位复数表示为[15]，$\forall \mathcal{U} \in \mathcal{U}(1)$：

$$\mathcal{U}^*\mathcal{U} = 1 \tag{3.11}$$

单位复数另一种形式是利用欧拉公式[16]：

$$R_\theta = \mathrm{e}^{i\theta} = \cos\theta + i\sin\theta \tag{3.12}$$

[10] 例如3.3式的矩阵，$R_\theta^{\mathrm{T}} R = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{pmatrix} = \begin{pmatrix}\cos^2\theta+\sin^2\theta & 0\\ 0 & \sin^2\theta+\cos^2\theta\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$

[11] 严谨地说，任意 2×2 正交矩阵可以表示为3.3、3.4式，或者它们的乘积的形式。

[12] 见3.3与3.4式。反射矩阵的行列式等于 -1。

[13] 右手坐标系与左手坐标系见附录 A.5.

[14] 上标 $*$ 表示复数的共轭复数：$z = a + ib \rightarrow z^* = a - ib$

[15] 定义中包含复数乘法的群的普遍性质见??节的附录。

[16] 附录 B.4.2 推导了传说中的欧拉公式。对任意复数 $z = a + ib$, a 称为 z 的实部，记为 $\Re(z) = a$; b 称为虚部，记为 $\Im(z) = b$. 欧拉公式中 R_θ 的实部为 $\cos\theta$, 虚部为 $\sin\theta$.

$R(\theta)$ 的模（长度）为:

$$R_\theta^* R_\theta = \mathrm{e}^{-i\theta}\mathrm{e}^{i\theta} = \big(\cos\theta - i\sin\theta\big)\big(\cos\theta + i\sin\theta\big) = 1 \quad (3.13)$$

例：将向量 $(3,5)$ 旋转 $90°$：

$$z \to z' = \mathrm{e}^{i\,90°} z = \Big(\underbrace{\cos 90°}_{=0} + i\underbrace{\sin 90°}_{=1}\Big)(3+5i) = i(3+5i) = 3i-5 \quad (3.14)$$

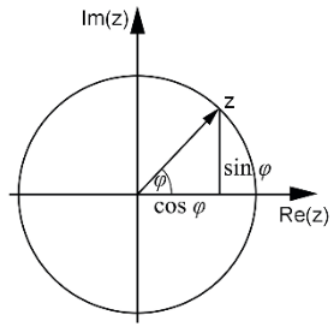


图 3.5: 单位复数在复平面的单位圆上

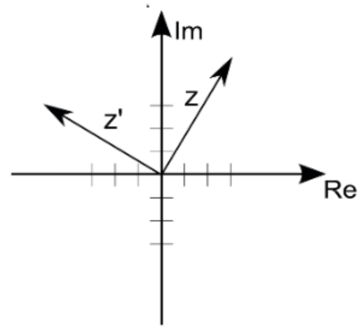


图 3.6: 复数通过乘以单位复数进行旋转

图 3.6 绘制了旋转前后的向量（复数），图中可见复数乘以 $\mathrm{e}^{i\,90°}$ 旋转了 $90°$. 需要注意 $\mathrm{e}^{i\,90°}$ 作用于（乘以）向量对应的复数而非向量本身。描述二维旋转只需要一个参数：旋转角 ϕ. 而复数有两个自由度（实部和虚部），因此我们加上单位复数的限制：$|z| =$ 实部^2 + 虚部$^2 = 1$，只剩一个自由度。

描述二维旋转的两种方式 — 单位复数与 2×2 矩阵（矩阵元都是实数）通过如下方式关联。定义：

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (3.15)$$

不难验证这样定义的 $\mathbf{1}, \mathbf{i}$ 仍然满足:

$$\mathbf{1}^2 = \mathbf{1}, \quad \mathbf{i}^2 = -\mathbf{1}, \quad \mathbf{1i} = \mathbf{i1} = \mathbf{i} \quad (3.16)$$

这样单位复数对应的 2×2 实数矩阵为[3]

$$R_\theta = \cos\theta + i\sin\theta = \cos\theta \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin\theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad (3.17)$$

可见，定义了复单位 $i \to$ 实矩阵 的对应关系后，单位复数就回到了熟悉的旋转矩阵。还有一点要注意：旋转矩阵作用于向量（列矩阵），而我们将 i 与 2×2 矩阵相联系，因此被旋转的向量（与一个复数对应）也将变成一个 2×2 矩阵。

[3]原文此式有误，译文已参照勘误表修改。

例：任意向量 (a,b) 对应的复数对应的 2×2 矩阵为：

$$z = a + ib = a\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} + b\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}a & -b\\ b & a\end{pmatrix} \tag{3.18}$$

将旋转矩阵作用于 z：

$$\begin{aligned} z' &= \begin{pmatrix}a' & -b'\\ b' & a'\end{pmatrix} = R_\theta z = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}a & -b\\ b & a\end{pmatrix} \\ &= \begin{pmatrix}a\cos\theta - b\sin\theta & -b\cos\theta - a\sin\theta\\ a\sin\theta + b\cos\theta & -b\sin\theta + a\cos\theta\end{pmatrix} \end{aligned} \tag{3.19}$$

由上式可得

$$a' = a\cos\theta - b\sin\theta \tag{3.20}$$

$$b' = a\sin\theta + b\cos\theta \tag{3.21}$$

这与旋转矩阵作用于向量（列矩阵形式）相同：

$$\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}a\\ b\end{pmatrix} = \begin{pmatrix}a\cos\theta - b\sin\theta\\ a\sin\theta + b\cos\theta\end{pmatrix} = \begin{pmatrix}a'\\ b'\end{pmatrix} \tag{3.22}$$

我们看到这两种表示方法是一回事儿，用数学术语来说，$\mathcal{SO}(2)$ 与 $\mathcal{U}(1)$ 间有一个同构映射[17]。这一点太重要了，之后的篇幅会不断出现这种思想。

[17] 如果映射 $\Pi: G \to G'$ 是一一映射，并且满足：$\forall g_1, g_2 \in G, \Pi(g_1)\circ\Pi(g_2) = \Pi(g_1\circ g_2)$，则 Π 就是同构映射，并且称群 G, G' 是同构的。

下面研究三维旋转，尝试找出它的两种描述方法。

3.3 Rotations in three Dimensions 三维旋转

三维空间向量旋转变换当然可以用 3×3 旋转矩阵[4]：

$$R_x=\begin{pmatrix}1&0&0\\0&\cos\theta&-\sin\theta\\0&\sin\theta&\cos\theta\end{pmatrix}\quad R_y=\begin{pmatrix}\cos\theta&0&\sin\theta\\0&1&0\\-\sin\theta&0&\cos\theta\end{pmatrix}$$
$$R_z=\begin{pmatrix}\cos\theta&-\sin\theta&0\\\sin\theta&\cos\theta&0\\0&0&1\end{pmatrix}\tag{3.23}$$

类比 $\mathcal{SO}(2)$ 群，上面三个矩阵构成了 $\mathcal{SO}(3)$ 群的一组基[18]。这意味着 $\mathcal{SO}(3)$ 中的任意群元（矩阵）都可以写为 R_x,R_y,R_z 的线性组合，且系数唯一。将向量

$$\vec{v}=\begin{pmatrix}1\\0\\0\end{pmatrix}$$

绕 z 轴旋转[19]就是用旋转矩阵乘以向量：

$$R_z(\theta)\vec{v}=\begin{pmatrix}\cos\theta&-\sin\theta&0\\\sin\theta&\cos\theta&0\\0&0&1\end{pmatrix}\begin{pmatrix}1\\0\\0\end{pmatrix}=\begin{pmatrix}\cos\theta\\\sin\theta\\0\end{pmatrix}\tag{3.24}$$

为找出描述三维旋转的第二法，我们尝试把复数的概念扩展到三维空间。首先试着把 2 维的复数（实部虚部两个自由度）扩展为 3 维复数，但前人探索发现没有 3 维的，但存在 4 维‘复数’，称为**四元数 (quaternions)**。下面就会看到四元数正是描述三维旋转的第二法，并且它还能深刻揭示宇宙 — 四维时空的规律。描述三维旋转需要三个参数（绕 x,y,z 轴的旋转角），而四元数有四个参数，这与二维旋转的情况类似[20]。四元数加上单位长度的限制---单位四元数（三个自由参数）能够描述三维旋转。

[18] 基的概念见附录 A.1.

[19] 三维向量旋转的一般情况的推导见附录 A.2.

[20] 单位复数描述二维旋转。

[4]原文公式有误，译文已按勘误表改正。

3.3.1 Quaternions 四元数

我们类比复数来构造四元数，复数只有一个虚单位 i，而四元数里定义三个虚单位，记作 $\mathbf{i}, \mathbf{j}, \mathbf{k}$，它们仍然满足

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1 \tag{3.25}$$

这样任意一个四元数 q 表示为

$$q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \tag{3.26}$$

其中 $a, b, c, d \in \mathbb{R}$, $\mathbf{1}$ 就是正常的实数 1。[5]

要想定义四元数间的乘法，必须定义虚单位间的乘法规则，比如 $\mathbf{ij} = ?$. 为此再定义

$$\mathbf{ijk} = -1 \tag{3.27}$$

这样虚单位间的乘法就没问题了，例如，为导出 $\mathbf{ij} = \mathbf{k}$, 在3.27式等号两侧右乘 $\mathbf{k}$:

$$\mathbf{ij}\underbrace{\mathbf{kk}}_{=-1} = -\mathbf{k} \rightarrow \mathbf{ij} = \mathbf{k} \tag{3.28}$$

单位四元数 $q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ 满足[21]：

$$q^\dagger q \stackrel{!}{=} 1$$
$$\rightarrow (a\mathbf{1} - b\mathbf{i} - c\mathbf{j} - d\mathbf{k})(a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) = a^2 + b^2 + c^2 + d^2 \stackrel{!}{=} 1 \tag{3.29}$$

[21] 符号 † 称为 ‘dagger’，表示转置加取复共轭，即 $a^\dagger = (a^*)^T$。实数向量间的点乘定义为 $a \cdot b = a^T b$, 向量用列矩阵表示，其转置为行矩阵，于是点乘的定义符合矩阵乘法规则。四元数对应的列矩阵含复数，为使四元数与它自身相乘结果为实数（实数结果可表示 ‘长度’），在转置之外还加上取复共轭。

就像单位复数构成一个群（群乘法为复数乘法）那样，单位四元数也构成一个群（群乘法为四元数乘法）。

类似于我们将二维复数与 2×2 实数矩阵建立联系的做法，四元数也可如此 — 将四个基 $\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{k}$ 与合适的矩阵一一对应。其中

[5] $\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{k}$ 用黑体表示是为了强调它们是四元数的基。---译者 (SI)

一种方法（还有别的方式）是将它们与 2×2 复数矩阵对应：

$$\mathbf{1}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \quad , \quad \mathbf{i}=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$$
$$\mathbf{j}=\begin{pmatrix}0 & -i\\ -i & 0\end{pmatrix} \quad , \quad \mathbf{k}=\begin{pmatrix}i & 0\\ 0 & -i\end{pmatrix} \tag{3.30}$$

不难验证上式满足四元数乘法，比如3.25、3.27式。这样任意一个四元数都可用矩阵表示为：

$$q=a\mathbf{1}+b\mathbf{i}+c\mathbf{j}+d\mathbf{k}=\begin{pmatrix}a+di & -b-ci\\ b-ci & a-di\end{pmatrix} \tag{3.31}$$

由上式可见

$$\det(q)=a^2+b^2+c^2+d^2 \tag{3.32}$$

上式与3.29式比较可以发现单位四元数对应的矩阵的行列式为 1。因此单位四元数对应的 2×2 复数矩阵 $\mathcal{U}$ 满足：

$$\mathcal{U}^\dagger\mathcal{U}=1, \quad 且\ \det(\mathcal{U})=1 \tag{3.33}$$

注意3.30式中的矩阵是线性独立的[22]，它们构成群 $\mathcal{SU}(2)$ 的基。与之前相同，$\mathcal{S}$ 表示特殊 (special)，含义为 $\det(\mathcal{U})=1$；$\mathcal{U}$ 表示幺正 (unitary)[23]，即 $\mathcal{U}^\dagger\mathcal{U}=1$。任意单位四元数都唯一对应 $\mathcal{SU}(2)$ 群中的一个群元。

[22] 线性独立的概念见附录 A.1.

[23] 进一步介绍见本章附录 — ??节。

$\mathcal{SU}(2)$ 群如何与三维旋转联系？事情正在起变化，$\mathcal{SU}(2)$ 群与 $\mathcal{SO}(3)$[24]之间的对应关系并不像二维旋转中 $\mathcal{U}(1)$ 与 $\mathcal{SO}(2)$ 那样简单。

[24] $\mathcal{SO}(3)$ 群就是三维旋转矩阵构成的群。

2 维虚数 $z=a+ib$ 与 2 维向量很容易对应。单位复数保证了矢量长度在旋转变换下不变[25]：$(Rz)^*Rz=z^*R^*Rz=z^*z$。四元数有 4 个参数，它与三维向量的对应关系是并不明显。我们尝试将三维向量 $\vec{v}=(x,y,z)$ 定义为如下四元数[6]：

[25] 注意这里的 R 表示一个单位复数，即 $R^*R=1$, R 与复数 z 相乘就将 z 进行旋转变换。

$$v\equiv x\mathbf{i}+y\mathbf{j}+z\mathbf{k} \tag{3.34}$$

[6]注意下式中的 **ijk** 是四元数的‘基’而非三维直角坐标系的基矢。— 译者

利用前述四元数的矩阵表示可得：

$$\det(v) = x^2 + y^2 + z^2 \tag{3.35}$$

于是保持向量 (x, y, z) 长度不变的变换就对应于保持矩阵行列式不变的矩阵变换。而**单位**四元数对应矩阵都具有**单位**行列式，它与任意矩阵相乘不改变该矩阵的行列式[7]。进展很顺利，但现在有个微妙的状况：继续猜想下去，我们会认为用单位四元数 u 旋转向量（对应的四元数）v 直接用 u 乘以 v 就可以了（按照四元数乘法规则）。但细想起来这却不行，因为根据3.34式，我们把任意三维向量对应的四元数定义在 $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ 的范围内，u 与 v 的四元数乘积会超出这个范围，这样向量旋转后可能不是个向量！这当然不行。

事实上，旋转变换表示为：

$$v' = q^{-1} v q \tag{3.36}$$

其中 v, v' 分别为旋转前后的向量，q 为表示旋转的单位四元数。

这样我们终于实现了描述三维旋转的第二法 — 单位四元数。

例：定义 u 是 $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ 中的某一单位向量，则任意单位四元数 t 可表示为：

$$t = \cos\theta + u \sin\theta \tag{3.37}$$

利用3.34式，任意三维向量 $\vec{v}$ 可表示为：

$$\vec{v} = (v_x, v_y, v_z)^T = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k} \underbrace{=}_{3.31\text{式}} \begin{pmatrix} iv_z & -v_x - iv_y \\ v_x - iv_y & -iv_z \end{pmatrix} \tag{3.38}$$

[7]利用 $\det(AB) = \det(A)\det(B)$，— 译者

简明起见我们举个特例，把向量 $\vec{v}=(1,0,0)^T$ 绕 z 轴旋转。

$$\vec{v}=(1,0,0)^T\rightarrow v=1\mathbf{i}+0\mathbf{j}+0\mathbf{k}=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix} \tag{3.39}$$

$$R_z(\theta)=\cos\theta\mathbf{1}+\sin\theta\mathbf{k}=\begin{pmatrix}\cos\theta+i\sin\theta & 0\\ 0 & \cos\theta-i\sin\theta\end{pmatrix} \tag{3.40}$$

上式也可用欧拉公式写出[26]：

[26] 欧拉公式的推导见附录 B.4.2

$$\mathrm{e}^{ix}=\cos x+i\sin x\Rightarrow R_z(\theta)=\begin{pmatrix}\mathrm{e}^{i\theta} & 0\\ 0 & \mathrm{e}^{-i\theta}\end{pmatrix} \tag{3.41}$$

四元数旋转矩阵的逆矩阵为：

$$R_z^{-1}(\theta)=\begin{pmatrix}\cos\theta-i\sin\theta & 0\\ 0 & \cos\theta+i\sin\theta\end{pmatrix}=\begin{pmatrix}\mathrm{e}^{-i\theta} & 0\\ 0 & \mathrm{e}^{i\theta}\end{pmatrix} \tag{3.42}$$

根据3.36式旋转向量 v；

$$\begin{aligned}v'&=R_z^{-1}(\theta)vR_z(\theta)=\begin{pmatrix}\mathrm{e}^{-i\theta} & 0\\ 0 & \mathrm{e}^{i\theta}\end{pmatrix}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\mathrm{e}^{i\theta} & 0\\ 0 & \mathrm{e}^{-i\theta}\end{pmatrix}\\ &=\begin{pmatrix}0 & -\mathrm{e}^{-i2\theta}\\ \mathrm{e}^{i2\theta} & 0\end{pmatrix}=\begin{pmatrix}0 & -\cos(2\theta)+i\sin(2\theta)\\ \cos(2\theta)+i\sin(2\theta) & 0\end{pmatrix}\end{aligned} \tag{3.43}$$

另一方面，向量 v' 用四元数表示为：

$$v'=\begin{pmatrix}iv'_z & -v'_x-iv'_y\\ v'_x-iv'_y & -iv'_z\end{pmatrix} \tag{3.44}$$

上式与3.43式比较可得：

$$v'_x=\cos(2\theta),\quad v'_y=-\sin(2\theta),\quad v'_z=0 \tag{3.45}$$

所以

$$\rightarrow\vec{v'}=\big(\cos(2\theta),-\sin(2\theta),0\big)^T \tag{3.46}$$

由上式可见从 $\vec{v}$ 到 $\vec{v}'$ 确实进行了旋转[27]，但是要注意，上式表示我们没有把 $\vec{v}$ 旋转 θ 角度，而是旋转了 2θ！因此我们定义 $\phi \equiv 2\theta$，这样 ϕ 表示真正的旋转角，3.37 重写为：

$$t = \cos\left(\frac{\phi}{2}\right) + \sin\left(\frac{\phi}{2}\right) u \tag{3.47}$$

[27] 见3.24式，那里我们用三维旋转矩阵旋转向量。

这样定义的单位四元数与三维旋转矩阵之间的关系并非一一对应的，而是两个单位四元数描述同一个旋转，例如[28]：

$$\begin{tikzcd} t_{\phi=\pi} = u \arrow[dr] & & t_{\phi=3\pi} = -u \arrow[dl] \\ & \text{将向量旋转 } \pi & \end{tikzcd}$$

[28] 将向量旋转 π 与旋转 $3\pi = \pi + 2\pi$ 相同。换言之：两个单位四元数 u 与 $-u$ 对应同一变换：旋转 π.

因此 $\mathcal{SU}(2)$ 群称为 $\mathcal{SO}(3)$ 群的**双覆盖**。一个 $\mathcal{SU}(2)$ 中的群元对应 $\mathcal{SO}(3)$ 的哪个群元是很明确的，反过来就不行，因为 $\mathcal{SO}(3)$ 中的一个群元对应 $\mathcal{SU}(2)$ 中的两个。这并非仅有数学意义，稍后就会看到，在物理上，一个能覆盖别的群的群往往更加深刻[29]。

[29] 剧透：Lorentz 群的双覆盖蕴含**自旋**概念，而 Lorentz 群自身并不具有。自旋是指粒子的某种内禀角动量，它是粒子最重要的特征之一。自旋在小节4.5.4和??详述。

给定某个群，为了找出能覆盖这个群的群，我们需要学习 Lie 代数 — Lie 理论的利器，下一节就讲 Lie 代数。

注意我们把三维向量对应到四元数集 $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ 中，因此四元数的一个参数没有用到，这是相对论的伏笔。比如将时间 t 考虑进来后，四维‘向量’与四元数对应的十分自然，就像二维复数与二维向量之间那样：$v = t\mathbf{1} + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$。纯数学的观点鼓励我们使用 4 维向量。如果我们想描述四维旋转（因为我们目前认为宇宙是 $3+1$ 四维时空），应该怎么办？有两个选择：

- 再去找更高维的‘复数’，或者：
- 再用四元数试试看。

从刚才的遐想可以看出四元数与四维旋转有着暧昧的联系。任意四维旋转需要 6 个参数表示[30]，没有 7 维复数，这样‘单位 7 维复数描述 6 个参数的四维旋转’就行不通了。我们注意到**两个**单位四元数正好有 6 个自由参数。因此两个单位四元数很可能可以描述四维旋转。之后会看到，两个 $\mathcal{SU}(2)$ 群与四维旋转群确实存在千丝万缕的联系。

[30] 四维旋转矩阵 $\mathcal{O}(4\times 4)$ 有 16 个参数，条件 $\mathcal{O}^T\mathcal{O} = 1$ 与 $\det(\mathcal{O}) = 1$ 将 16 个参数限制为 6 个自由参数。

3.4 Lie Algebras Lie 代数

Lie 理论是研究连续对称的理论。连续对称的例子可见本章开头讨论的单位圆的旋转对称性。连续性意味着存在无限接近于恒等变换（恒等变换：什么都不干）的群元。相对的，离散群的群元个数是可数的，不存在无限接近于恒等元的群元。例如把正方形绕中心转 $0.000001°$，这与恒等变换（转 $0°$）非常接近，但它不是正方形的恒等变换。而把圆绕圆心转 $0.000001°$ 就是对称的。圆的对称群（对称变换组成的群）是连续的，因为变换参数（旋转角）可以取任意（连续）的值。用数学符号来表示：把恒等元记作 I，无限接近恒等元的群元 g 可表示为：

$$g(\epsilon) = I + \epsilon X \tag{3.48}$$

其中 ϵ 表示小量（数学总是用 ϵ 表示小量），X 称为生成元，稍后就讨论它。这样一个轻微的小变换作用在物体上几乎啥也不变，有时称 $g(\epsilon)$ 为无穷小变换。把无穷小变换重复许多次就能得到一个有限大小的变换。以旋转为例：许多朝同一方向小旋转等效于一次有限大的旋转。用数学语言来说，我们可以将无穷小变换重复许多次：

$$h(\theta) = \underbrace{(I+\epsilon X)(I+\epsilon X)\dots(I+\epsilon X)}_{k\text{个}(I+\epsilon X)} = (I+\epsilon X)^k \tag{3.49}$$

其中 k 表示无穷小变换重复的次数。如果 θ 表示有限大的旋转角，比如 $50°$ 什么的，然后 N 表示超大的数，则无限接近于恒等变换的旋转变换可表示为：

$$g(\frac{\theta}{N}) = I + \frac{\theta}{N} X \tag{3.50}$$

要使上式表示的变换尽可能小，则让 N 尽可能大，令 $N \to \infty$。为了从这样一个无穷小变换得到一个有限变换，需要把无穷小变换重复无限多次，即：

$$h(\theta) = \lim_{N\to\infty} \left(I + \frac{\theta}{N} X \right)^N \tag{3.51}$$

微积分告诉我们这个极限就是指数函数[31]：

[31] 3.51式经常用作指数函数的定义。指数的级数表示与3.51式等价性的证明见附录 B.4.1. 这在几乎所有的数学分析课本中都能找到。

$$h(\theta) = \lim_{N\to\infty}\left(I + \frac{\theta}{N}X\right)^N = e^{\theta X} \tag{3.52}$$

上式有 X 产生了有限变换 $h(\theta)$ 的感觉，因此 X 成为**生成元**。生成元的定义之后详述，下面先从另一个角度看这个问题。

考虑某个用矩阵表示的连续变换群，对任意群元，在恒等元 I 处做 Taylor 展开[32]。

$$h(\theta) = I + \left.\frac{dh}{d\theta}\right|_{\theta=0}\theta + \frac{1}{2}\left.\frac{d^2h}{d\theta^2}\right|_{\theta=0}\theta^2 + \cdots = \sum_n \frac{1}{n!}\left.\frac{d^nh}{d\theta^n}\right|_{\theta=0}\theta^n \tag{3.53}$$

利用指数函数的级数展开可将上式的级数表示为更紧凑的形式[33]：

$$h(\theta) = \exp\left(\left.\frac{dh}{d\theta}\right|_{\theta=0}\theta\right) \equiv \sum_n \frac{1}{n!}\left.\frac{d^nh}{d\theta^n}\right|_{\theta=0}\theta^n \tag{3.54}$$

这与之前的描述有联系。比较上式与3.52式可得：

$$X = \left.\frac{dh}{d\theta}\right|_{\theta=0} \tag{3.55}$$

蕴含在上述推导中的思想是：通过研究群中重要的无穷小元素 — **生成元**（上面的 X）就能获得群的许多重要信息。

矩阵 Lie 群（记作 G）的 Lie 代数就是下面的集合：

$$\left\{X \middle| 若 e^X \in G\right\}$$

即如果 X 满足 e^X 是 G 的群元，则 X 就是 G 的 Lie 代数（它是一个集合！）中的元素。这个简单定义只对矩阵 Lie 群管用。之后我们引入 Lie 代数的一般定义。

上面的矩阵 Lie 群的 Lie 代数的定义用数学语言表述为[34]：

$n\times n$ 矩阵 Lie 群 G 的 Lie 代数 $\mathfrak{g}$ 是满足如下条件的 $n\times n$ 矩阵 X 的集合：

$$e^{tX} \in G, t \in \mathbb{R}$$

根据群的定义，群无非是一些变换的集合。群乘法 $\circ$ 告诉我们群元之间怎样合体。矩阵 Lie 群的群乘法就是矩阵乘法，我们可能 naive 地认为 Lie 代数的元素之间采用相同的方式合体（$\circ$），但这并不正确！诚然，（矩阵 Lie 群的）Lie 代数的元素都是矩阵[35]，但

[32] 如果你从未听说过 Taylor 展开或 Taylor 级数，作者推荐你看一看附录 B.3. ~~而译者建议你学了微积分再来~~

[33] 附录 B.4.1 中有推导

[34] 群 G 的 Lie 代数通常用哥特体活字（Fraktur）字母 $\mathfrak{g}$ 表示。

[35] Lie 群理论的著名定理 — Ado 定理（Ado's Theorem）告诉我们任意 Lie 代数与矩阵 Lie 代数同构。

两个 Lie 代数的矩阵乘法的结果往往不再是 Lie 代数的元素。Lie 代数元素间有另外的合体规则，当然它与原来群的群乘法直接有关。

Lie 群乘法与 Lie 代数合体法则之间的关系由著名的 Baker-Campbell-Hausdorff 公式（以下简称 BCH 公式）给出[36]：

$$\mathrm{e}^X \circ \mathrm{e}^Y = \mathrm{e}^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}\left[X,[X,Y]\right]-\frac{1}{12}\left[Y,[X,Y]\right]+\dots} \tag{3.56}$$

[36] 我们不讲这个公式的证明，在大多数关于 Lie 理论的教材都能找到，比如 William Fulton and Joe Harris. Representation Theory: A First Course. Springer, 1st corrected edition, 8 1999. ISBN 9780387974958

其中[37]$X, Y \in \mathfrak{g}$, 即 X, Y 是群 G 的生成元。$\mathrm{e}^X, \mathrm{e}^Y$ 是 G 的群元，把它们分别记作 g, h，这样上式写成：

[37] 重复一下，群 G 的 Lie 代数通常用哥特体字母 $\mathfrak{g}$ 表示。

$$\underbrace{g}_{\in G} \circ \underbrace{h}_{\in G} = \mathrm{e}^X \circ \mathrm{e}^Y = \underbrace{\mathrm{e}^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}\left[X,[X,Y]\right]-\frac{1}{12}\left[Y,[X,Y]\right]+\dots}}_{\in G} \tag{3.57}$$

上式等号右侧是群的一个群元，可见两个群元（g, h）相乘可表示为一些 Lie 代数元素的和（再取指数）。上两式出现的新运算 $[,]$ 称为 **Lie 括号**，对于矩阵 Lie 群，$[X,Y] = XY - YX$，$[X,Y]$ 称为 X, Y 的对易子。注意 XY 和 YX 一般不是 Lie 代数的元素，但它们的差一定是[38]！

[38] John Stillwell. Naive Lie Theory. Springer, 1st edition, August 2008a. ISBN 978-0387782140 里面有漂亮的证明。

由 BCH 公式可知 Lie 代数元素间的乘法规则是 Lie 括号 $[,]$，而非最初认为的矩阵乘法。就像群在群乘法 $\circ$ 运算（矩阵群的 $\circ$ 就是矩阵乘法）下封闭那样，我们称 Lie 代数在 Lie 括号运算下封闭。集合在某运算下封闭的含义是集合的任意两个元素进行该运算所得结果仍在此集合内[39]。

[39] 用数学语言表示封闭性：$\forall g, h \in G$, $g \circ h \in G$。对于 $X, Y \in \mathfrak{g}, [X,Y] \in \mathfrak{g}, X \circ Y \notin \mathfrak{g}$

在学习一个 Lie 代数的例子之后，我们将讨论 Lie 代数的现代定义。现代定义是从群的生成元在 Lie 括号运算下的行为来定义的。利用更广泛的现代定义就能看出哪些不同的群具有**相同的** Lie 代数，只从上面的 Lie 代数定义出发看不出这一点来。因此 Lie 代数的新定义能让我们更深刻地认识一种变换的基本特征。同一 Lie 代数可对应许多 Lie 群，Lie 理论的一条重要定理告诉我们这其中有一个**特别的** Lie 群。引入 Lie 群的现代定义之后上面所说的就具体起来了。

下面我们就举一个从给定的群导出相应 Lie 代数的例子。

3.4.1 The Generators and Lie Algebra of $\mathcal{SO}(3)$ $\mathcal{SO}(3)$ 群的生成元与 Lie 代数

$\mathcal{SO}(3)$ 群的定义为（3.10式）：

$$\mathcal{O}^T\mathcal{O} \overset{!}{=} I, \quad \det(\mathcal{O}) \overset{!}{=} 1 \tag{3.58}$$

任意群元 $\mathcal{O}$ 用相应的生成元 J 表示为：

$$\mathcal{O} = \mathrm{e}^{\phi J} \tag{3.59}$$

上式带入群的第一个定义条件得：

$$\mathcal{O}^T\mathcal{O} = \mathrm{e}^{\phi J^T}\mathrm{e}^{\phi J} \overset{!}{=} 1 \quad \to \quad J^T + J \overset{!}{=} 0 \tag{3.60}$$

带入第二个定义条件，再利用等式[40]$\det(\mathrm{e}^A) = \mathrm{e}^{(A)}$ 得：

$$\det(\mathcal{O}) \overset{!}{=} 1 \quad \to \quad \det(\mathrm{e}^{\phi J}) = \mathrm{e}^{\phi(J)} \overset{!}{=} 1$$
$$\to (J) \overset{!}{=} 0 \tag{3.61}$$

[40](A) 表示矩阵 A 的迹，也就是 A 的主对角线上所有元素的和。例如

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

的迹为 $(A) = A_{11} + A_{22}$。

满足??、??式的三个线性无关的矩阵为：

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.62}$$

这三个矩阵构成 $\mathcal{SO}(3)$ 群的生成元的**基**。即群的任意生成元 J 可惟一表示为这三者的线性组合：$J = aJ_1 + bJ_2 + cJ_3$，其中 a, b, c 表示实常数。J_1, J_2, J_3 可用 Levi-Civita 符号更紧凑地表示[41]：

$$(J_i)_{jk} = -\epsilon_{ijk}, \quad i, j, k = 1, 2, 3. \tag{3.63}$$

[41]Levi-Civita 符号的含义见附录 B.5.5.

其中 j,k 表示生成元 J_i 的分量。例：

$$(J_1)_{jk} = -\epsilon_{1jk} \Longleftrightarrow \begin{pmatrix} (J_1)_{11} & (J_1)_{12} & (J_1)_{13} \\ (J_1)_{21} & (J_1)_{22} & (J_1)_{23} \\ (J_1)_{31} & (J_1)_{32} & (J_1)_{33} \end{pmatrix}$$

$$= -\begin{pmatrix} \epsilon_{111} & \epsilon_{112} & \epsilon_{113} \\ \epsilon_{121} & \epsilon_{122} & \epsilon_{123} \\ \epsilon_{131} & \epsilon_{132} & \epsilon_{133} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \tag{3.64}$$

这些生成元可以生成有限大小的变换矩阵。以 J_1 为例，我们可以只关注非零部分 — 右下角的 2×2 矩阵，把它记作 j_1[42]。

$$J_1 = \begin{pmatrix} 0 & \\ & \underbrace{\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}}_{\equiv j_1} \end{pmatrix} \tag{3.65}$$

计算可得[8]

$$j_1^2 = -\mathbf{1} \tag{3.66}$$

接着：

$$j_1^3 = \underbrace{j_1^2}_{=-1} j_1 = -j_1, \quad j_1^4 = +\mathbf{1}, \quad j_1^5 = +j_1 \tag{3.67}$$

一般情况为：

$$j_1^{2n} = (-1)^n I, \quad j_1^{2n+1} = (-1)^n j_1 \tag{3.68}$$

利用上式可计算 j_1 的指数函数的级数展开[43]：

[42] 其实小矩阵 j_1 也可用二维 Levi-Civita 符号表示，$(j_1)_{ij} = \epsilon_{ij}$。见附录 B.5.5. j_1 是二维旋转群 $\mathcal{SO}(2)$ 的生成元。

[43] 推导见附录 B.4.1。相关技巧在 B.4.2. 中进一步解释。正弦和余弦函数的级数展开也在 B.4.1 中。

[8] 以下符号 $\mathbf{1}$、I 均表示单位矩阵。— 译者

$$
\begin{aligned}
R_1 = \mathrm{e}^{\phi j_1} &= \sum_{n=0}^{\infty} \frac{\phi^n j_1^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{\phi^{2n}}{(2n)!} \underbrace{j_1^{2n}}_{(-1)^n I} + \sum_{n=0}^{\infty} \frac{\phi^{2n+1}}{(2n+1)!} \underbrace{j_1^{2n+1}}_{(-1)^n j_1} \\
&= \underbrace{\left(\sum_{n=0}^{\infty} \frac{\phi^{2n}}{(2n)!}(-1)^n\right)}_{=\cos\phi} I + \underbrace{\left(\sum_{n=0}^{\infty} \frac{\phi^{2n+1}}{(2n+1)!}(-1)^n\right)}_{=\sin\phi} j_1 \\
&= \cos\phi \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin\phi \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}
\end{aligned} \tag{3.69}
$$

再利用 $\mathrm{e}^0 = 1$ 计算左上角的 0 对应的元素，得到完整的有限变换矩阵：

$$
R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \tag{3.70}
$$

我们已经得出 $\mathcal{SO}(3)$ 群的生成元显式的矩阵形式（??式），这样就能直接计算生成元之间的 Lie 括号了[44]，不难得到[45]：

$$
[J_i, J_j] = \epsilon_{ijk} J_k \tag{3.71}
$$

其中 ϵ_{ijk} 是 Levi-Civita 符号。

出于物理层面的考虑，在 $\mathcal{SO}(3)$ 生成元 $\mathrm{e}^{\phi J}$ 的幂附加一个虚单位 i，即 $\mathrm{e}^{i\phi J}$，生成元的基变成：

$$
J_1 = i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \quad J_2 = i \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad J_3 = i \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.72}
$$

于是 Lie 代数[46]为：

$$
[J_i, J_j] = i\epsilon_{ijk} J_k \tag{3.73}
$$

[44] 前面说过，Lie 代数元素间的‘乘法’是 Lie 括号。直接计算生成元的基之间的 Lie 括号，就可得到所有生成元之间的 Lie 括号（任意生成元都是基的线性组合）。之后会看到生成元的基在 Lie 括号运算下的行为极其重要。Lie 代数的所有重要信息都蕴含在生成元的基的 Lie 括号中。

[45] 例如：

$$
\begin{aligned}
&[J_1, J_2] = J_1 J_2 - J_2 J_1 \\
&= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} - \\
&\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = \\
&\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \\
&\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \underbrace{\epsilon_{12k}}_{=0,\text{除了}k=3} J_k \\
&= \epsilon_{123} J_3 = J_3
\end{aligned}
$$

[46] 重复一下，我们把生成元的基的 Lie 括号运算结果**等同为** Lie 代数，因为 Lie 代数的所有重要信息都蕴含在那里面。

将生成元加 i 是为了让它们是 Hermitian 生成元[47]，Hermitian 生成元的含义是 $J^\dagger \equiv (J^*)^T \stackrel{!}{=} J$。加 i 是从物理意义考虑的：Hermitian 矩阵的本征值都是实数，这在量子力学里面极其重要，因为生成元的本征值是实验的可观测量，??对此详细讨论。如果不加 i，生成元就是反 Hermitian 的了：$J^\dagger = (J^*)^T = -J$，反 Hermitian 矩阵的本征值是虚数。

[47] 例如 $J_1^* = -i\begin{pmatrix}0&0&0\\0&0&1\\0&-1&0\end{pmatrix}$，而 $J_1^\dagger = (J_1^*)^T = i\begin{pmatrix}0&0&0\\0&0&1\\0&-1&0\end{pmatrix} = J_1$。

还有另一种导出生成元的方法，利用 3.55 式 — $X = \frac{dh}{d\theta}\big|_{\theta=0}$，以及三维旋转矩阵的形式（3.23、??式）可得：

$$J_1 = \frac{dR_1}{d\theta}\bigg|_{\theta=0} = \frac{d}{d\theta}\begin{pmatrix}1&0&0\\0&\cos\theta&-\sin\theta\\0&\sin\theta&\cos\theta\end{pmatrix}\Bigg|_{\theta=0}$$
$$= \begin{pmatrix}0&0&0\\0&-\sin\theta&-cos\theta\\0&\cos\theta&\sin\theta\end{pmatrix}\Bigg|_{\theta=0} = \begin{pmatrix}0&0&0\\0&0&-1\\0&1&0\end{pmatrix} \tag{3.74}$$

这正是3.26式导出的第一个生成元。第一种方法更加普遍，因为像这样已知有限变换矩阵具体形式的好事不太常见。对于之后的 Lorentz 群，我们先根据群的定义导出其生成元的基，然后才得出 Lorentz 变换的显式形式。当然如果已知变换矩阵形式那么利用3.55式就能最简便地导出生成元。

下面开始填作者之前挖的坑吧：我们看看 Lie 代数的现代定义到底是啥。

3.4.2 The Abstract Definition of a Lie Algebra Lie 代数的抽象定义

前面介绍了简化版的 Lie 代数定义。群 G 的 Lie 代数是满足如下条件的 X 的集合：$e^X \in G$。下面我们会看到[9]，群的重要部分 — 群乘法（规定群元之间怎样合体），可以从 Lie 代数元素的 Lie 括号运算中导出。

就像定义 Lie 群那样，我们把 Lie 代数的关键特征表示为公理，这就形成了 Lie 代数的抽象定义：

Lie 代数是一个向量空间 $\mathfrak{g}$ 配备一个二元运算 $[,]:\mathfrak{g}\times\mathfrak{g}\to\mathfrak{g}$，且二元运算 $[,]$ 满足如下公理：

- *双线性：* $\forall a,b\in\mathbb{R}$以及$\forall X,Y,Z\in\mathfrak{g}$, $[aX+bY,Z]=a[X,Z]+b[Y,Z]$，$[Z,aX+bY]=a[Z,X]+b[Z,Y]$。

[9] 作者似乎很喜欢挖这种坑然后慢慢填。— 译者

- 反交换律：$\forall X, Y \in \mathfrak{g}, [X, Y] = -[Y, X]$。
- *Jacobi* 恒等式：$\forall X, Y, Z \in \mathfrak{g}, [X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$。

之前我们把矩阵的对易子作为 Lie 代数的二元运算 $[,]$，不难验证对易子满足上面的要求。其实有许多和对易子丝毫不同的二元运算照样满足这些公理，比如经典力学中著名的 Poisson 括号[10]。

注意 Lie 代数的抽象定义完全不依赖群，这十分重要哟。

在下一小节我们推导 $\mathcal{SU}(2)$ 群的生成元的基（再通过线性组合就能得到所有生成元），并且我们会看到它与 $\mathcal{SO}(3)$ 的生成元满足同样的 Lie 括号关系（??式）。这就是 $\mathcal{SU}(2)$ 与 $\mathcal{SO}(3)$ 群有**相同** Lie 代数的含义。这个超级重要的结论将揭示 $\mathcal{SU}(2)$ 与 $\mathcal{SO}(3)$ 间许多不为人知的关系。

3.4.3 The Generators and Lie Algebra of $\mathcal{SU}(2)$ $\mathcal{SU}(2)$ 的生成元与 Lie 代数

前面我们历经磨难终于找到描述三维旋转的第二法 — $\mathcal{SU}(2)$ 群，并且发现 $\mathcal{SU}(2)$ 是 $\mathcal{SO}(3)$ 的双覆盖[48]。

[48] 复习一下，这意味着两个 $\mathcal{SU}(2)$ 中的两个群元对应 $\mathcal{SO}(3)$ 群的同一元素。

$\mathcal{SU}(2)$ 群的群元是具有单位行列式的幺正的 2×2 矩阵，即：

$$\mathcal{U}^\dagger \mathcal{U} = \mathcal{U}\mathcal{U}^\dagger = \mathbf{1} \tag{3.75}$$

$$\det(\mathcal{U}) = 1 \tag{3.76}$$

从群元满足的条件可导出其生成元满足的条件。设生成元为 $J_i, i = 1, 2, \ldots$[11]，将群元用生成元表示，带入上两式[49]：

[49] 前面说过，为了让生成元是 Hermitian 矩阵，我们在 e 指数幂上加了 i，这样保证量子力学预言的实验值为实数。

$$\mathcal{U}^\dagger \mathcal{U} = (\mathrm{e}^{iJ_i})^\dagger \mathrm{e}^{iJ_i} \stackrel{!}{=} \mathbf{1} \tag{3.77}$$

$$\det(\mathcal{U}) = \det(\mathrm{e}^{iJ_i}) \stackrel{!}{=} 1 \tag{3.78}$$

[10] Poisson 括号可见朗道，力学，高等教育出版社，2007，§42，泊松括号。ISBN：9787040208498。— 译者

[11] 在下面的式子中请注意区分虚数单位 i 和表示数字指标的下标 i，我们对作者采用这样的符号很抱歉。— 译者

结合??式、BCH 公式（3.56式）和 $[J_i, J_i]=0$[12]：

$$\begin{aligned}
&(\mathrm{e}^{iJ_i})^\dagger \mathrm{e}^{iJ_i} = \mathrm{e}^{-iJ_i^\dagger}\mathrm{e}^{iJ_i} \overset{!}{=} \mathbf{1} \\
&\rightarrow \mathrm{e}^{-iJ_i^\dagger + iJ_i + \frac{1}{2}[J_i^\dagger, J_i]+\dots} \overset{!}{=} \mathbf{1} \\
&\qquad\qquad \underbrace{\rightarrow}_{\mathrm{e}^0=1} J_i^\dagger \overset{!}{=} J_i
\end{aligned} \tag{3.79}$$

满足 $J_i^\dagger = J_i$ 的矩阵称为 Hermitian 矩阵，由上可见 $\mathcal{SU}(2)$ 的生成元必须是 Hermitian 的。

利用恒等式 $\det(\mathrm{e}^A) = \mathrm{e}^{(A)}$ 与??式可得：

$$\det(\mathrm{e}^{iJ_i}) = \mathrm{e}^{i(J_i)} = 1 \underbrace{\rightarrow}_{\mathrm{e}^0=1} (J_i) \overset{!}{=} 0 \tag{3.80}$$

综上，$\mathcal{SU}(2)$ 的生成元一定是无迹的 Hermitian 矩阵。2×2 Hermitian 无迹矩阵的一组基为[50]：

[50] 2×2 复数矩阵有 4 个复数元素，因此有 8 个自由度。无迹、Hermitian 性的条件限制了只剩 3 个自由度。

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.81}$$

也就是说任意 2×2 Hermitian 无迹矩阵都可唯一表示为上面三个矩阵的线性组合。这三个矩阵称为 **Pauli 矩阵**。

把 Pauli 矩阵扔进 Lie 括号中进行计算可得：

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k \tag{3.82}$$

其中 ϵ_{ijk} 是 Levi-Civita 符号。上式右侧的 2 很多余，因此通常将 $\mathcal{SU}(2)$ 的生成元的基定义为 $J_i \equiv \frac{1}{2}\sigma_i$，这样 Lie 括号计算结果为：

$$[J_i, J_j] = i\epsilon_{ijk}J_k \tag{3.83}$$

注意这与 $\mathcal{SO}(3)$ 的生成元形式（??式）相同！因此我们说 $\mathcal{SU}(2)$ 与 $\mathcal{SO}(3)$ 有相同的 Lie 代数，因为 Lie 代数是通过 Lie 括号定义的！利用 Lie 代数的抽象定义可以用不同的方式描述 $\mathcal{SU}(2)$ 群对应的变换，即，$\mathcal{SU}(2)$ 群对应的变换可以不用 2×2 矩阵表示。为了做到这一点，我们还需要知道 Lie 群的抽象定义。因为一直以来就是用 2×2 矩阵定义 $\mathcal{SU}(2)$ 的，（$\mathcal{SU}$ ‘2’ 嘛!）换一种表示方

[12] 这由 Lie 括号的反对称性导出：$[J_i, J_i] = -[J_i, J_i] \rightarrow [J_i, J_i] = 0$，— 译者

法（例如用 3×3 矩阵表示）则不知所云。Lie 群的抽象定义能够让我们发现同一变换的不同表示之间的关系。抽象定义将 Lie 群与一种几何结构 — 流形相等同，利用这种抽象结构来定义一个群。这一想法初看十分怪异，但在学习两个例子之后就会发现它是超级有意义的。

3.4.4 The Abstract Definition of a Lie Group　Lie 群的抽象定义

我们讨论过纯良的 $\mathcal{U}(1)$ 群，它由全体单位复数构成。定义为 $z^*z = 1$，设 $z = a + ib$，则：

$$z^*z = (a+ib)^*(a+ib) = (a-ib)(a+ib) = a^2 + b^2 = 1 \tag{3.84}$$

这也是单位圆的定义条件[51]。可见单位复数的集合**正是**复平面上的单位圆。此外，我们知道 $\mathcal{U}(1)$ 与 $\mathcal{SO}(2)$ 群间存在一一映射[52]。因此这两个群可以与几何物体 — 单位圆 — 视为等同。群的抽象定义并非将这个群用 $\mathcal{SO}(2)$，或着是不同的 $\mathcal{U}(1)$ 体现，因为它们都是由特定维数的物体定义的。可以将该群直接利用单位圆定义。即二维旋转变换（对应的 Lie 群）等同于单位圆，然后我们可以用 $\mathcal{SO}(2)$ 群表示这个变换（2×2 矩阵），也可用 $\mathcal{U}(1)$ 群来表示（单位复数）。

下面讨论升级版 — $\mathcal{SO}(3)$ 与 $\mathcal{SU}(2)$。前面讲过在 $\mathcal{SU}(2)$ 与单位四元数之间有一一映射。单位四元数就是满足下列条件的一般四元数 $q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$：

$$a^2 + b^2 + c^2 + d^2 \stackrel{!}{=} 1 \tag{3.85}$$

这与三维超球面 S^3 的定义[53]相同！

[51] 单位圆记作 S^1，它是二维空间中到原点的距离为 1 的所有点构成的集合。用数学语言就是 $S^1 = \{(x_1, x_2) | x_1^2 + x_2^2 = 1\}$

[52] 严谨的说法是：同构映射。数学上把两个物体‘相同’称为它们‘同构’。如果两个物体之间存在一个同构映射，那么它们就是同构的。

[53]

4. The Framework 框架/体系

这一章的基本思路是, 我们要在尽可能少的使用某些东西的前提下，得到正确的关于自然的方程。某些东西是什么？有一件事是确定的：它不应该在 Lorentz 变换下改变，否则我们会在不同的参考系下得到不同的自然规律。在数学意义上，它意味着我们寻找的这个东西是个标量，依照洛伦兹群的 $(0,0)$ 表示作变换。再考虑到自然总依简单而行，我们已经足够导出关于自然的方程了。

从这个想法出发，我们将会引入**拉格朗日形式 (Lagrangian formalism)**。通过极小化理论的中心对象，我们可以得到用以描述问题中的物理系统的运动方程。极小化过程的结果被称作 **Euler-Lagrange 方程**。

通过拉格朗日形式，我们可以得到物理中最重要的定理：Noether 定理。这个定理揭示了对称性和守恒量[1]之间的深刻联系。我们将在下一章中利用它来理解, 理论是如何来描述实验测量量的。

4.1 Lagrangian Formalism 拉格朗日形式

拉格朗日形式是在基础物理中被广泛运用的一个强有力的框架[2]。由于理论的基本对象–**拉格朗日量 (Lagrangian)** 是一个标量[3]，它相对简单。如果你希望从对称性的观点考虑问题，这种形式将会是非常有用的。如果我们要求拉格朗日量的积分，**作用量**

[1] 守恒量指的是不随时间变化的物理量。例如一个给定体系的能量或动量。数学上意味着 $\partial_t Q = 0 \rightarrow Q =$ 常数。

[2] 物理中当然有其他框架,例如以**哈密顿量 (Hamiltonian)** 为中心对象的**哈密顿形式 (Hamiltonian formalism)**。哈密顿量的问题在于它不是洛伦兹不变的，因为它所代表的能量，仅仅是**四动量 (covariant energy-impulse vector)** 的一个分量

[3] 标量指依照洛伦兹群的 $(0,0)$ 表示作变换的对象。这意味着它不在洛伦兹变换下改变

(action)，在某些对称变换下不变，我们即要求体系的动力学遵从该对称性。

4.1.1 Fermat 原理

> Whenever any action occurs in nature, the quantity of action employed by this change is the least possible.
> - Pierre de Maupertius [4]

[4]Recherche des loix du mouvement (1746)

拉格朗日形式的思想源于 Fermat 原理：光在两空间点间传播总依耗时最短的路径 $q(t)$ 而行。数学上来讲，如果我们定义给定路径 $q(t)$ 的作用量为

$$S_{\text{light}}[\mathbf{q}(t)] = \int \mathrm{d}t$$

而我们的任务便是找到一条特定的路径 $q(t)$ 使作用量取极小值[5]为了得到一个给定**函数**的极小值[6]，我们可以求得其导函数并令其为零；而为了找到**泛函**$S[q(t)]$——函数 $q(t)$ 的函数 S——的极小值，就得要一个新的数学工具：**变分法**。

[5] 此处的作用量仅仅是沿给定路径对时间的积分，但一般而言作用量会更加复杂，我们待会儿就能见到

[6] 一般而言，我们希望找到**极值 (extremums)**，即极小值**和**极大值。下一节中的方法足以找到这两者。无论如何，我们将继续谈论极小值

4.1.2 变分法：基本思想

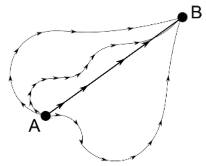


图 4.1: 对于给定初末端点的路径的扰动

在思考如何发展一套能够找到泛函极值的新理论之前，我们需要倒回去想想什么给出一个数学上的极小点。变分法给出的答案是，极小点由极小点邻域的性质决定。例如，让我们尝试寻找一个寻常函数 $f(x) = 3x^2 + x$ 的极小点 $x_{\min}$。我们从一个特定点 $x = a$ 出发，仔细考察其邻域。数学上它意味着 $a + \epsilon$，其中 ϵ 代表无穷小量（可正可负）。我们将 a 的变分代入函数 $f(x)$:

$$f(a+\epsilon) = 3(a+\epsilon)^2 + (a+\epsilon) = 3(a^2 + 2a\epsilon + \epsilon^2) + a + \epsilon。$$

如果 a 是极小点，ϵ 的一阶变分必需为零，否则我们可以取 ϵ 为负 $\epsilon < 0$，这样 $f(a+\epsilon)$ 就会比 $f(a)$ 更小[1]。因此，我们将线性依赖于 ϵ 的项取出并令其为零。

$$3 \cdot 2a\epsilon + \epsilon \stackrel{!}{=} 0 \to 6a + 1 \stackrel{!}{=} 0。$$

[1]译注：此处讨论有误

由此我们找到极小点

$$x_{\min} = a = -\frac{1}{6},$$

它自然和我们求导 $f(x) = 3x^2 + x \leftarrow f'(x) = 6x + 1$ 并令其为零的办法得到的结果一致。对于寻常函数而言，这只是一个用来干同一件事不同方法而已[2]，但是变分法却能找到泛函的极值点。我们马上就能看到，应当如何处理一个一般的作用量泛函。

拉格朗日形式的中心思想在于对于有质量的物体，也存在一个与对光的 Fermat 原理相类似的原理。当然，它不可能直接遵从费马原理，但是我们可以从一个更一般的形式出发

$$S[q(t)] = \int \mathcal{L}\ \mathrm{d}t$$

其中 $\mathcal{L}$ 一般是一个非常数的参量，被称为拉格朗日量。对于光而言，这个参量是个常数。一般的，拉格朗日量依赖于物体的坐标和速度 $\mathcal{L} = \mathcal{L}(q(t), \frac{\partial}{\partial t}q(t))$。在下一节中我们将仔细讨论这件事[7]。在仔细讨论如何对这样一个泛函使用变分法之前，我们需要先讲讲两个小问题。

[7] 我们的任务是找到对于给定拉格朗日量和初始条件有着最小作用量的路径 $q(t)$。在此之前，我们得先找到正确的拉格朗日量，用以描述问题中的物理系统。这是我们在上一章中所讨论的对称性所能发挥作用的地方。通过要求拉格朗日量在洛伦兹群的所有变换下不变，我们就能找到正确的拉格朗日量

4.2 Restrictions 限制

正如我们在1.1节中所提到的，现有理论中有一些限制条件是无法从第一性原理中得到的。我们所能知道是，如果想得到一个有意义的理论，我们必须加上这些限制。

一个重要的限制是，我们仅允许拉格朗日量中出现尽可能低阶的非平凡[3]导数。这里平凡指的是对于系统的动力学，即运动方程，没有影响。某些理论会包含一阶导数，另外一些带二阶导数。一个给定理论所含的最低阶导数由该条件决定：拉格朗日量在 Lorentz 变换下不变[8]，否则我们将在不同的参考系下导出不同的运动方程[4]。对于某些理论，我们无法写出一个仅含一阶导数项的不变量，那么此时二阶导数便成了最低阶导数项。

[8] 实际上，作用量才应该是 Lorentz 不变的。但如果拉格朗日量满足这个条件，那作用量自然也是。

我们并不清楚如何处理包含高阶导数的理论，它们有一些根本的困难[9]。另外，拉格朗日量中的高阶导数会使得运动方程中也包

[9] 这些困难包括 Ostrogradski 不稳定性，即包含高阶导数的系统的能量没有下界，以至于系统中的所有态总会自发衰变到能量更低的态上去。这类系统中找不到稳定的状态。

[2]译注：对于受微元法荼毒的物竞生而言，求导才是用来干这件事的不同方法。
[3]译注：Non-trival. Trival 这个词常包含简单、弱智、无意义之义。
[4]译注：此处讨论疑有误，作者此前并没有说明积分中的时间是坐标时还是固有时

含高阶导数，这样我们必须知道更多的初始条件才能决定物体的运动。

有些人会宣称对理论所包含导数阶数的限制源于我们希望得到一个局域[10]理论，但这个要求仅会排除掉那些无穷阶导数。一个非局域相互作用有如下形式[11]

$$\Phi(x-h)\Phi(x) \tag{4.1}$$

即，时空中距离为 h 的两个点的场相互作用。利用 Taylor 展开我们有[5]

$$\Phi(x-h)=\sum_{k=0}^{\infty}\left(\left(\frac{\partial}{\partial x}\right)^k \Phi(x)|_{x=h}\right)\frac{(x-h)^k}{k!} \tag{4.2}$$

即包含无穷阶导数会导致一个非局域的理论。

另外一个限制是，要得到一个自由（无相互作用）场/粒子的理论，我们只能写到二次项。这意味着我们只会考虑[12]

$$\Phi^0, \Phi^1, \Phi^2$$

这样的项。例如，形如 $\Phi^2\partial_\mu\Phi$ 的项是 Φ 的三阶项，因而不会被包含在我们自由理论的拉格朗日量中。

4.3 Particle Theories vs. Field Theories 粒子理论与场论

我们目前有两套用以描述自然的理论框架。一套是粒子理论，用依赖于时间的粒子位置描述物理系统，即 $\vec{q}=\vec{q}(t)$。由于我们不是非得在使用笛卡尔坐标系[13]，所以用的是字母 q 而不是 x。对于这样的理论，拉格朗日量依赖于坐标 $\vec{q}$，速度 $\partial_t\vec{q}$ 和时间 t：[6]

$$\mathcal{L}=\mathcal{L}(\vec{q},\partial_t\vec{q},t) \tag{4.3}$$

一个著名的例子是这个能导出经典力学的 Newton 运动方程的拉格朗日量 $\mathcal{L}=\frac{1}{2}m\dot{q}^2$。我们将在稍后进行相当详细的讨论。

另一套是场论：使用场而不是独立粒子的坐标来描述自然[14]。在这套理论中，时空构成了场 $\Phi(\vec{x},t)$ 表演的舞台。利用前面提到过的限制，我们得到[15]

[5] 译注：此式有误

[6] 译注：严格说来，依赖于坐标函数 $\vec{q}(t)$，速度函数 $\partial_t\vec{q}(t)$ 和时间 t

[10] 局域性是狭义相对论的基本假定，2.4已作阐述

[11] 我们将会讨论粒子的拉格朗日理论：寻求粒子的路径；也会讨论场的拉格朗日理论：寻求场函数 $\phi(x)$。这将是下一节的主题。

[12] 从另外一个角度来看，这再次表示我们只引入最低阶的非平凡项。我们随后就会看到，含 Φ^0 和 Φ^1 的项是平凡的，因此我们这次用的是 Φ 的最低阶非平凡项

[13] 例如，我们可以用球坐标系

[14] 量子场论有一个特别优美的性质是关于如何将粒子引进的。我们将在??中看到场有能力产生和湮灭粒子

[15] 这里我故意使用了不同的符号 $\mathscr{L}$，因为在场论中大多数时候我们都在和拉格朗日量密度 $\mathscr{L}$ 打交道。两者之间满足关系式 $\mathcal{L}=\int d^3\boldsymbol{x}\ \mathscr{L}$

$$\mathscr{L} = \mathscr{L}(\Phi(\vec{x},t), \partial_\mu \Phi(\vec{x},t), \vec{x}) \tag{4.4}$$

一个著名的例子是我们将用以导出 Klein-Gordon 方程的拉格朗日量 $\mathscr{L} = \frac{1}{2}\left(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2\right)$。

场论的一大优点是它将时空等价的对待。在粒子理论中我们使用空间坐标 $\vec{q}(t)$ 作为时间的函数来描述我们的粒子。尤其是，拉格朗日量中没有类似于 $\partial_{\vec{q}} t$ 的项（如果出现了这类项，我们该怎样理解它呢？）当我们讨论粒子的坐标时，意义是清楚的，但对时间做类似的陈述时，却很难有明确的意义[7]。

讨论了这么多，我们终于可以回到这章开头所说的极小化问题上来了。我们希望找到某些泛函

$$S[q(t)] = \int \mathcal{L}\ \mathrm{d}t$$

的极小值，以得到正确的运动方程。

对于粒子而言，方程的解是使得泛函取极小值的正确路径；而对于场而言，解是正确的场函数。

此刻请不用担心 $\mathcal{L}$ 的具体形式，在下面数章中我们将详细讨论如何对问题中的系统导出正确的拉格朗日量 L。现在，我们将使用之前所介绍过的变分法，对于一个一般的 $\mathcal{L}$ 导出泛函 $S[q(t)]$ 的极小值。极小化过程将给出系统的运动方程。

4.4 Euler-Lagrange equation 欧拉-拉格朗日方程

我们从粒子理论出发，它能让我们搞清楚，在给定初末点后，粒子是如何运动的。数学上，我们需要找到**函数**$q(t)$ 使得作用量[8]

$$S = \int_{t_1}^{t_2} \mathcal{L}\left(q(t), \frac{\mathrm{d}q(t)}{\mathrm{d}t}, t\right)$$

取得其极值（极大值或极小值）。

我们使用记号

$$\dot{q}(t) = \frac{\mathrm{d}q(t)}{\mathrm{d}t}。$$

[7]译注：利用不那么古老的相对论性语言，粒子理论下时空坐标是可以在形式上被同等对待的：用以求导的东西只能是曲线的参数，它可以取作粒子世界线的固有时、某一坐标系的坐标时、或者其他什么东西（只要性质合适，也可以是空间坐标）；而时空坐标则被同样的求导。

[8]此处原文字体有误

与之前的例子类似，取 $q(t)=a(t)$ 并让对这个函数进行一个扰动

$$a(t)+\epsilon(t)$$

其中 ϵ 依旧是一个无穷小量。对于一个粒子而言，我们必须同时改变其速度 $\dot{a}(t)+\dot{\epsilon}(t)$，其中 $\dot{\epsilon}(t)=\frac{\mathrm{d}\epsilon(t)}{\mathrm{d}t}$。

在边界上，变换后的路径应与原路径相同：

$$0=\epsilon(t_1)=\epsilon(t_2) \tag{4.5}$$

这是因为我们寻找的是使作用量积分取极值的**给定初末端点**的路径。

这个扰动使得泛函变为

$$S=\int_{t_1}^{t_2}\mathcal{L}\left(q+\epsilon,\dot{q}+\dot{\epsilon},t\right)\,\mathrm{d}t。$$

与之前的例子类似，当我们寻求极小值时，我们要求线性依赖于扰动 ϵ 的项为零。由于我们处理的是一个一般性的 $\mathcal{L}$，故将其展开成泰勒级数[16]，并令一阶项为零

$$\int_{t_1}^{t_2}\mathrm{d}t\,\left[\epsilon(t)\frac{\partial\mathcal{L}}{\partial q}+\left(\frac{\mathrm{d}}{\mathrm{d}t}\epsilon(t)\right)\frac{\partial\mathcal{L}}{\partial\dot{q}}\right]\stackrel{!}{=}0。 \tag{4.6}$$

对后一项做分部积分[17]得到

$$\int_{t_1}^{t_2}\mathrm{d}t\,\left(\frac{\mathrm{d}}{\mathrm{d}t}\epsilon(t)\right)\frac{\partial\mathcal{L}(q,\dot{q},t)}{\partial\dot{q}}=$$
$$\epsilon(t)\frac{\partial\mathcal{L}(q,\dot{q},t)}{\partial\dot{q}}\bigg|_{t_1}^{t_2}-\int_{t_1}^{t_2}\mathrm{d}t\,\epsilon(t)\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\mathcal{L}(q,\dot{q},t)}{\partial\dot{q}}\right)$$

利用4.5式有

$$\epsilon(t)\frac{\partial\mathcal{L}(q,\dot{q},t)}{\partial\dot{q}}\bigg|_{t_1}^{t_2}=0$$

因此，我们可以将4.6改写为

$$\int_{t_1}^{t_2}\mathrm{d}t\,\epsilon(t)\left[\frac{\partial\mathcal{L}}{\partial q}-\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\mathcal{L}(q,\dot{q},t)}{\partial\dot{q}}\right)\right]\stackrel{!}{=}0$$

易见上式对于任意扰动 $\epsilon(t)$ 为零仅当方括号 [ ] 内的表达式为零。故有[18]

[16] 我们正使用的是这个公式的多变量形式，参见附录??：

$$\mathcal{L}\left(q+\epsilon,\dot{q}+\dot{\epsilon},t\right)=\mathcal{L}(q,\dot{q},t)+\epsilon\frac{\partial\mathcal{L}}{\partial q}+\dot{\epsilon}\frac{\partial\mathcal{L}}{\partial\dot{q}}+\ldots$$

[17] 分部积分是莱布尼兹律的直接推论，详见附录??

[18] 可能你会对于两类不同的导数符号感到困惑。$\frac{\mathrm{d}}{\mathrm{d}t}$ 被称作对 t 的全导数，而 $\frac{\partial}{\partial t}$ 被称为偏导数。全导数给出总改变量，函数 f 的变化率，即偏导数，乘以变量自身的改变，之和。例如，三维空间中的函数 $f(x(t),y(t),z(t))$ 的总变化率为 $\frac{\mathrm{d}f}{\mathrm{d}t}=\frac{\partial f}{\partial x}\frac{\partial x}{\partial t}+\frac{\partial f}{\partial y}\frac{\partial y}{\partial t}+\frac{\partial f}{\partial z}\frac{\partial z}{\partial t}$。即变化率乘以自身的改变量。相对的，偏导数仅给出总改变量的一部分。对于一个不显式的依赖于 t 的函数，其偏导数为零。例如，对于 $f(x(t),y(t))=x^2y+y^3$，我们有 $\frac{\partial f}{\partial t}=0$，但 $\frac{\partial f}{\partial x}=2xy\neq 0, \frac{\partial f}{\partial y}=x^2+3y^2\neq 0$。因此 $\frac{\mathrm{d}f}{\mathrm{d}t}=2xy\frac{\partial x}{\partial t}+(x^2+3y^2)\frac{\partial y}{\partial t}$。作为对比，对于另一个函数 $g(x(t),y(t),t)=x^2t+y$ 我们有 $\frac{\partial g}{\partial t}=x^2$

$$\frac{\partial \mathcal{L}}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}(q,\dot{q},t)}{\partial \dot{q}}\right) = 0 \tag{4.7}$$

这就是著名的**欧拉-拉格朗日方程 (Euler-Lagrange equation)**

我们可以用类似的办法处理场论。首先，注意到我们需要将时空同等对待。因此我们引入拉格朗日量密度：

$$\mathcal{L} = \int \mathrm{d}^3\boldsymbol{x}\ \mathscr{L}(\Phi^i, \partial_\mu \Phi^i) \tag{4.8}$$

并用拉格朗日量密度来表示作用量

$$S = \int \mathrm{d}t\ \mathcal{L} = \int \mathrm{d}^4\boldsymbol{x}\ \mathscr{L}(\Phi^i, \partial_\mu \Phi^i) \tag{4.9}$$

依照上面讲过的步骤，我们可以得到场的运动方程组[19]：

[19] 在这里用“组”字是因为对于每一个场分量 $\Phi_1, \Phi_2, \ldots$ 我们都可以得到一个方程

$$\frac{\partial \mathscr{L}}{\partial \Phi} - \partial_\mu\left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \Phi)}\right) = 0 \tag{4.10}$$

在下一章中我们将用拉格朗日形式导出近代物理学中重要的一个定理。借助其我们能看到对称性与守恒量之间的深刻联系。守恒量是描述自然的合适参量[20]，从这个定理中我们将学会如何和它打交道。

[20] 我们将其视做锚定这个复杂的世界之物。不论万事万物如何改变，守恒量依旧是守恒量。

4.5 Noether's Theorem Noether 定理

Noether 定理表示，拉格朗日量的每一个对称性都对应于一个守恒量。换言之，物理学家用于描述自然（守恒量）的记号和对称性直接相关。这确实是科学史上最美的创见之一。

4.5.1 粒子理论的 Noether 定理

让我们先看看能对粒子理论中的守恒量做点什么。我们将局限于连续对称性，由此我们能使用无穷小变换。如同前面数章所言，我们可以通过重复无穷小变换来得到一个有限的变换。拉格朗日量在无穷小变换[21]$q \to q' = q + \delta q$ 下不变的性质用数学表述如下[9]

[21] 用符号 δ 来表述小扰动，大概不会和用来表述偏导数的 ∂ 相混淆

$$\begin{aligned}\delta \mathcal{L} &= \mathcal{L}(q, \frac{\mathrm{d}q}{\mathrm{d}t}, t) - \mathcal{L}(q + \delta q, \frac{\mathrm{d}(q+\delta q)}{\mathrm{d}t}, t) \\ &= \mathcal{L}(q, \frac{\mathrm{d}q}{\mathrm{d}t}, t) - \mathcal{L}(q + \delta q, \frac{\mathrm{d}q}{\mathrm{d}t} + \frac{\mathrm{d}\delta q}{\mathrm{d}t}, t) \overset{!}{=} 0\end{aligned} \tag{4.11}$$

[9] 译者注：该书中 δS，$\delta \mathcal{L}$ 与平常的约定都差了一个负号。

要求拉格朗日量是变换不变的，这个条件太强了。为了保证动力学一样，不变的应该是作用量而不是拉格朗日量。当然，如果拉格朗日量不变，作用量自然不变：

$$\delta S = \int \mathrm{d}t\ \mathcal{L}(q, \frac{\mathrm{d}q}{\mathrm{d}t}, t) - \int \mathrm{d}t\ \mathcal{L}(q+\delta q, \frac{\mathrm{d}q}{\mathrm{d}t}, t) = \int \mathrm{d}t\ \delta\mathcal{L} \underbrace{=}_{\text{若 }\delta\mathcal{L}=0} 0 \tag{4.12}$$

在什么情况下，拉格朗日量改变而作用量不变呢？答案是，给拉格朗日量加上一个任意函数 G 对时间的全导数

$$\mathcal{L} \to \mathcal{L} + \frac{\mathrm{d}G}{\mathrm{d}t}$$

这时作用量不变[22]

$$\delta S \to \delta S' = \delta S + \int_{t_1}^{t_2} \mathrm{d}t\ \frac{\mathrm{d}}{\mathrm{d}t}\delta G = \delta S + \underbrace{\frac{\partial G}{\partial q}\delta q\Big|_{t_1}^{t_2}}_{=0\ \text{因为}\ \delta q(t_1)=\delta q(t_2)=0}.$$

[22] 用到了 $\delta G = \frac{\partial G}{\partial q}\delta q$。这是因为 $G = G(q)$ 而我们改变了 q，故 G 的扰动等于改变率 $\frac{\partial G}{\partial q}$ 乘上 q 的扰动

在最后一步中，我们用到了扰动 δq 在初末时刻 (t_1, t_2) 为零。由此我们可知，我们并不需要去强求拉格朗日量的扰动 $\delta\mathcal{L}$ 为零，而仅需满足一个更弱的条件

$$\delta\mathcal{L} \stackrel{!}{=} \frac{\mathrm{d}G}{\mathrm{d}t}。\tag{4.13}$$

这意味着拉格朗日量可以在不改变作用量和运动方程的情况下，加减某个函数的全微分 $\frac{\mathrm{d}G}{\mathrm{d}t}$。将4.11式的右端由 0 改作 $\frac{\mathrm{d}G}{\mathrm{d}t}$，我们有

$$\delta\mathcal{L} = \mathcal{L}(q, \frac{\mathrm{d}q}{\mathrm{d}t}, t) - \mathcal{L}(q+\delta q, \frac{\mathrm{d}q}{\mathrm{d}t} + \frac{\mathrm{d}(\delta q)}{\mathrm{d}t}, t) \stackrel{!}{=} \frac{\mathrm{d}G}{\mathrm{d}t} \tag{4.14}$$

将第二项按 Taylor 级数展开到 δq 的一阶项，并使用记号 $\frac{\mathrm{d}q}{\mathrm{d}t} = \dot{q}$

$$\begin{aligned} &\to \delta\mathcal{L} = \mathcal{L} - \mathcal{L} - \frac{\partial\mathcal{L}}{\partial q}\delta q - \frac{\partial\mathcal{L}}{\partial\dot{q}}\delta\dot{q} \stackrel{!}{=} \frac{\mathrm{d}G}{\mathrm{d}t} \\ &\to \delta\mathcal{L} = -\frac{\partial\mathcal{L}}{\partial q}\delta q - \frac{\partial\mathcal{L}}{\partial\dot{q}}\delta\dot{q} \stackrel{!}{=} \frac{\mathrm{d}G}{\mathrm{d}t} \end{aligned} \tag{4.15}$$

利用欧拉-拉格朗日方程[23]重写4.15式，有

[23] 4.7式：$\frac{\partial\mathcal{L}}{\partial q} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}}\right)$

$$\rightarrow \delta\mathcal{L} = -\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}}\right)\delta q - \frac{\partial\mathcal{L}}{\partial\dot{q}}\delta\dot{q} \overset{!}{=} \frac{\mathrm{d}G}{\mathrm{d}t}$$

利用莱布尼兹律[24]，我们有

$$\begin{aligned}&\rightarrow \delta\mathcal{L} = -\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}}\delta q\right) \overset{!}{=} \frac{\mathrm{d}G}{\mathrm{d}t}\\ &\rightarrow \frac{\mathrm{d}}{\mathrm{d}t}\underbrace{\left(\frac{\partial\mathcal{L}}{\partial\dot{q}}\delta q + G\right)}_{\equiv J} \overset{!}{=} 0\end{aligned} \tag{4.16}$$

由此我们得到了一个不随时间变化的物理量 J：

$$J = \frac{\partial\mathcal{L}}{\partial\dot{q}}\delta q + G, \tag{4.17}$$

这是因为

$$\frac{\mathrm{d}J}{\mathrm{d}t} = 0 \rightarrow J = \text{常数。}$$

为了看清楚这一点，我们先借用一下后面章节的结论：常质量自由粒子的 Newton 第二定律是[25]

$$m\ddot{\vec{q}} = 0。 \tag{4.18}$$

用以导出这个著名的运动方程的拉格朗日量是[26]

$$\mathcal{L} = \frac{1}{2}m\dot{\vec{q}}^2 \tag{4.19}$$

你可以将其代入欧拉-拉格朗日方程（4.7式）中去验证它。

让我们从这个拉格朗日量出发，计算不同对称性所分别对应的守恒量。

由于 $\mathcal{L} = \frac{1}{2}m\dot{\vec{q}}^2$ 不依赖于 $\vec{q}$，故我们的拉格朗日量在空间平移变换 $\vec{q} \rightarrow \vec{q} + \vec{a}$ 不变 $(\delta\mathcal{L} = 0)$。这个对应的守恒量满足条件4.13，即 $G = 0$ 的情形[10]

$$J_{\text{trans}} = \frac{\partial\mathcal{L}}{\partial\dot{\vec{q}}}\vec{a} = m\dot{\vec{q}}\vec{a} = \vec{p}\vec{a}, \tag{4.20}$$

其中 $\vec{p} = m\dot{\vec{q}}$ 是经典力学中的**动量 (momentum)**。由于等式 $\frac{\mathrm{d}J}{\mathrm{d}t}$ 对于任意 $\vec{a}$ 都成立，所以说：**拉格朗日量在空间平移下不变导致**

[10] 译注：下式缺少点乘符号

[24] 如果你对莱布尼兹律仍有疑问，请参考??

[25] 如果 q 代表某个物体的位置，则 $\frac{\mathrm{d}q}{\mathrm{d}t} = \dot{q}$ 是这个物体的速度，$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}q}{\mathrm{d}t} = \frac{\mathrm{d}^2q}{\mathrm{d}t^2} = \ddot{q}$ 是加速度。

[26] 我们现在处理的是高于一维的问题，所以使用矢量 $\vec{q}$ 和 $\vec{a}$ 而不是 q 和 a。

了动量守恒。

我们现在把视线转到旋转上来，由此我们需要大于一的维度，这是因为一维上的旋转是没有意义的。我们在这里将不会使用类似 $\vec{q}$ 的矢量符号，而是更方便的指标记号。我们能将所有方程简单的用 $\vec{q} \to q_i$ 改写。让我们来看看无穷小旋转[27]$q_i \to q_i + \epsilon_{ijk} q_j a_k$，由此 $\delta q_i = \epsilon_{ijk} q_j a_k$。仍然是因为由于 $\mathcal{L} = \frac{1}{2} m \dot{\vec{q}}^2$ 不依赖于 $\vec{q}$，我们的拉格朗日量在这样的变换下也不改变，其对应的守恒量是[28]

$$\begin{aligned} J_{\text{rot}} =& \frac{\partial \mathcal{L}(q_i, \dot{q}_i, t)}{\partial \dot{q}_i} \delta q_i = \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \epsilon_{ijk} q_j a_k \\ =& m \dot{q}_i \epsilon_{ijk} q_j a_k = m p_i \epsilon_{ijk} q_j a_k \\ \to J_{\text{rot}} =& (\vec{p} \times \vec{q}) \cdot \vec{a} \equiv \vec{L} \cdot \vec{a} \end{aligned} \tag{4.21}$$

在最后一步中我们将表达式用矢量记号重新写出，其中 $\times$ 称作叉乘，而 $\vec{L}$ 是经典力学中**角动量 (angular momentum)**。即：**旋转下不变导致了角动量守恒。**

接下来让我们看看时间平移不变性[29]。一个无穷小的时间平移 $t \to t' = t + \epsilon$ 会导致

$$\begin{aligned} \delta \mathcal{L} &= \mathcal{L}\left(q(t), \frac{\mathrm{d}q(t)}{\mathrm{d}t}, t\right) - \mathcal{L}\left(q(t+\epsilon), \frac{\mathrm{d}q(t+\epsilon)}{\mathrm{d}t}, t+\epsilon\right) \\ &= -\frac{\partial \mathcal{L}}{\partial q}\frac{\partial q}{\partial t}\epsilon - \frac{\partial \mathcal{L}}{\partial \dot{q}}\frac{\partial \dot{q}}{\partial t}\epsilon - \frac{\partial \mathcal{L}}{\partial t}\epsilon \underbrace{=}_{\text{等式左端正好为全导数}} -\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}t}\epsilon, \end{aligned} \tag{4.22}$$

这告诉我们一般而言 $\delta \mathcal{L} \neq 0$，但有 $G = -\mathcal{L}$。

将其代入4.16式，有[11]

$$\frac{\mathrm{d}}{\mathrm{d}t} \underbrace{\left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \dot{q} - \mathcal{L}\right)}_{\equiv \mathcal{H}} = 0 \tag{4.23}$$

守恒量 $\mathcal{H}$ 被称为**哈密顿量 (Hamiltonian)**，代表了系统的总能量。对于这个例子，我们有

$$\begin{aligned} \mathcal{H} =& \frac{\partial \mathcal{L}}{\partial \dot{q}} \dot{q} - \mathcal{L} = \underbrace{\left(\frac{\partial}{\partial \dot{q}} \frac{1}{2} m \dot{q}^2\right)}_{= m\dot{q}} \dot{q} - \frac{1}{2} m \dot{q}^2 \\ =& m\dot{q}^2 - \frac{1}{2} m \dot{q}^2 = \frac{1}{2} m \dot{q}^2, \end{aligned} \tag{4.24}$$

[27] 这里我们用 Levi-Civita 符号来写旋转生成元。参见??式下的文字。

[28] 下式由4.16式导出，这里对应的依旧是 $G = 0$ 的情况

[29] 这意味着物理不会管我们是在昨天、今天还是五十年前做的实验，给定相同的初始条件，物理定律不会变化。

[11] 译注：原文有字体混乱

正好是系统的动能。由于这里没有势场或外力，故动能就是总能量。能导出在外势场中运动的粒子的 Newton 第二定律 $m\ddot{q} = -\frac{\partial V}{\partial q}$[12]的拉格朗日量是

$$\mathcal{L} = \frac{1}{2}m\dot{q}^2 - V(q)。$$

其哈密顿量为

$$\mathcal{H} = \frac{1}{2}m\dot{q}^2 + V$$

的确是正确的总能量（等于动能加势能）。

由**推动不变性 (boost invariance)** 导致的守恒量略有些奇怪，相应的计算放在了这一章的附录4.6中。其结果为

$$\tilde{J}_{\text{boost}} = \underbrace{pt - \frac{1}{2}mvt}_{\equiv \tilde{p}t} - mq = \tilde{p}t - mq \tag{4.25}$$

我们可以看到这个量依赖于时间零点，可以通过恰当的选取时间零点使其为零。由于这个量是守恒量，这个守恒律告诉我们零始终都是零。[13]

作为粒子理论的总结，我们得到了如下关系：

- 空间平移不变性 $\Rightarrow$ 动量守恒
- 推动不变性[30]$\Rightarrow$ $\tilde{p}t - mq$ 守恒
- 旋转不变性 $\Rightarrow$ 角动量守恒
- 时间平移不变性 $\Rightarrow$ 能量守恒

Noether 定理向我们展示了，为什么这些记号[31]会以这个或那个形式出现在每一个物理理论中。只要我们依旧有通常的时空对称性，动量、能量和角动量就是守恒量。在场论中，我们有两类对称性：一方面，拉格朗日量会在时空变换，如旋转，下不变；另一方面，我们也会有在场自身的变换下的不变性，被称为**内禀对称性 (internal symmetries)**。

[30] 推动的另一个叫法是动量空间的平移，因为变换 $q \to q + vt$ 使动量 $m\dot{q} \to m(\dot{q} + v)$

[31] 除了推动不变性导致的守恒量

4.5.2 场论中的 Noether 定理——时空对称性

对于场，我们得区分在时空变换下两类不同的改变。当观测者 S 看到场 $\Psi(x)$ 时，观测者 S' 看到的则是场 $\Psi'(x')$。这是同一

[12] 原文有误

[13] 译注：这个守恒率没有说的那么平庸，事实上，它等价于质心运动定理；又注：该结果以及4.6的推导有误

个场，只是从不同的视角来看，会有不同的分量值。这两类描述之间由 Lorentz 群中的一个合适的变换相联系。我们现在将会使用在??中介绍过的场表示。（无穷维）微分算符表示 x 变换到 x'。这意味着利用这个表示我们能在不同的时空点或者在一个旋转参考系中计算场分量。Lorentz 群的有限维变换使 Ψ 映射到 Ψ'，即，混合场分量[32]

32 请回忆这个例子，Weyl 旋量有两个分量而矢量场有四个分量。如果我们从一个不同的视角来考察矢量场 $A_\mu = \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix}$，即，从旋转系看，它会形如 $A'_\mu = \begin{pmatrix} A'_0 \\ A'_1 \\ A'_2 \\ A'_3 \end{pmatrix} = \begin{pmatrix} cA_0 \\ -A_2 \\ A_1 \\ A_3 \end{pmatrix}$。$A'_\mu$ 和 A_μ 描述了从两个相差了绕 z 轴旋转 90° 的坐标系分别观察到的同一个场。

对于一个依赖于时空的场的具体变换，需要同时考虑两个部分。我们将分别处理它们，首先来看 x 到 x' 的这一项。对于旋转来说，单由这一个部分推出的守恒量并不会真的守恒，这是因为我们忽略了变换的第二项。仅当将变换的两项 $x \to x'$ 和 $\Psi \to \Psi'$ 分别导致的守恒量加起来之后得到的，才是真正守恒的。

为了使上面这段论证更有说服力，我们考虑一个一般的拉格朗日量密度 $\mathscr{L}(\Phi(x_\mu), \partial_\mu \Phi(x_\mu), x_\mu)$[14]。对称性意味着

$$\mathscr{L}(\Phi(x_\mu), \partial_\mu \Phi(x_\mu), x_\mu) = \mathscr{L}\left(\Phi'(x'_\mu), \partial_\mu \Phi'(x'_\mu), x'_\mu\right)。 \tag{4.26}$$

一般的，当函数自身发生改变且其计算的点也发生改变时，函数的函数的总改变量由下式给出[33]

33 如果你对此不熟悉：这常被称为全导数。总改变量等于变化率，即偏导数，乘以变量自身的改变，之和。例如，三维空间中的函数 $f(x,y,z)$ 的总该变量为 $\frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial z}\delta z$，即变化率乘以自身改变的距离。我们考察的是无穷小变化，故我们可以略去 Taylor 展开的高阶项仅考虑第一项。

$$\delta f(g(x), h(x), \dots) = \frac{\partial f}{\partial g}\delta g + \frac{\partial f}{\partial h}\delta h + \cdots + \frac{\partial f}{\partial x}。 \tag{4.27}$$

运用到拉格朗日量上有

$$\delta \mathscr{L} = \frac{\partial \mathscr{L}}{\partial \Phi}\delta\Phi + \frac{\partial \mathscr{L}}{\partial(\partial_\mu \Phi)}\delta(\partial_\mu \Phi) + \frac{\partial \mathscr{L}}{\partial x_\mu}\delta x_\mu, \tag{4.28}$$

用欧拉-拉格朗日方程[34]将其改写为

34 4.10式：$\frac{\partial \mathscr{L}}{\partial \Phi} = \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \Phi)}\right)$

$$\begin{aligned} \delta \mathscr{L} &= \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \Phi)}\right)\delta\Phi + \frac{\partial \mathscr{L}}{\partial(\partial_\mu \Phi)}\underbrace{\delta(\partial_\mu \Phi)}_{=\partial_\mu \delta\Phi} + \frac{\partial \mathscr{L}}{\partial x_\mu}\delta x_\mu \\ &\underbrace{=}_{\text{莱布尼兹律}} \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \Phi)}\delta\Phi\right) + \frac{\partial \mathscr{L}}{\partial x_\mu}\delta x_\mu \end{aligned} \tag{4.29}$$

其中 $\delta\Phi$ 有两项

$$\delta\Phi = \epsilon_{\mu\nu} S^{\mu\nu}\Phi(x) - \frac{\partial \Phi(x)}{\partial x_\mu}\delta x_\mu, \tag{4.30}$$

其中 $\epsilon_{\mu\nu}$ 为旋转参数，$S_{\mu\nu}$ 为相应的有限维群表示中的旋转算符，

14 译注：原文多出一左括号

额外的负号仅为了方便。$S_{\mu\nu}$ 是由旋转生成元 $S_i = \frac{1}{2}\epsilon_{ijk}S_{jk}$ 和推动生成元 $K_i = S_{i0}$ 构成，类似于??式中 $M_{\mu\nu}$ 的定义。$S_{\mu\nu}$ 的这个定义使得我们同时进行旋转和推动。

第一项仅在旋转和推动时较重要，因为平移不会导致场分量的混合。推动所对应的守恒量依旧不是特别有意思，就像在粒子理论中得到的那样。所以实际上这一项只和旋转相关。

让我们从最简单的场变换出发：时空平移，即

$$x_\mu \to x'_\mu = x_\mu + \delta x_\mu = x_\mu + a_\mu \tag{4.31}$$

将 $\epsilon_{\mu\nu} = 0$（在平移下场分量不会混合）代入方程4.30

$$\delta\Phi = -\frac{\partial\Phi(x)}{\partial x_\mu}\delta x_\mu$$

再由4.29式，考察不变性 $(\delta\mathscr{L} = 0)$ 所导致的守恒量

$$-\partial_\nu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\nu\Phi)}\frac{\partial\Phi(x)}{\partial x_\mu}\delta x_\mu\right) + \frac{\partial\mathscr{L}}{\partial x_\mu}\delta x_\mu \quad = 0 \tag{4.32}$$

$$\to -\partial_\nu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\nu\Phi)}\frac{\partial\Phi(x)}{\partial x^\mu} - \delta^\nu_\mu\mathscr{L}\right)\delta x^\mu \quad = 0 \tag{4.33}$$

由4.31式我们有 $\delta x_\mu = a_\mu$，代入4.33式中可得[15]

$$-\partial_\nu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\nu\Phi)}\frac{\partial\Phi(x)}{\partial x^\mu} - \delta^\nu_\mu\mathscr{L}\right)\delta x^\mu = 0, \tag{4.34}$$

我们定义能动张量

$$T^\nu_\mu := \frac{\partial\mathscr{L}}{\partial(\partial_\nu\Phi)}\frac{\partial\Phi(x)}{\partial x^\mu} - \delta^\nu_\mu\mathscr{L}。 \tag{4.35}$$

由于 a_μ 是任意的，4.34式告诉我们 T^ν_μ 满足连续性方程

$$\partial_\nu T^\nu_\mu = 0 \tag{4.36}$$

$$\to \partial_\nu T^\nu_\mu = \partial_0 T^0_\mu - \partial_i T^i_\mu = 0 \tag{4.37}$$

对于所有 μ 成立。这直接告诉我们有守恒量的存在，例如取 $\mu = 0$ 我们有[35]

[35] 使用了 $\partial_0 = \partial_t$ 和 $\partial^i T^0_i = \nabla\cdot\vec{T}$，以及著名的散度定理 $\int_V \mathrm{d}^3\boldsymbol{x}\ \nabla A = \int_{\delta V}\mathrm{d}^2\boldsymbol{x}\ A$，即能让体积分化作面积分。对于散度定理，在那儿叫 Gauss 定理，有一个非常有启发性的证明，可以在 http://www.feynmanlectures.caltech.edu/II_03.html上的第三章找到。Richard P. Feynman, Robert B. Leighton, and Matthew Sands. The Feynman Lectures on Physics: Volume 2. Addison-Wesley, 1st edition, 2 1977. ISBN 9780201021172（译注：中译本：《费恩曼物理学讲义》）

[15] 原文上下数式的指标不平衡

$$\partial_0 T_0^0 - \partial_i T_0^i = 0 \rightarrow \partial_0 T_0^0 = \partial_i T_0^i$$

$$\partial_t T_0^0 = -\nabla \cdot \vec{T} \underbrace{\rightarrow}_{\text{在某个无限的空间域 } V \text{ 上积分}} \int_V \mathrm{d}^3\boldsymbol{x}\ \partial_t T_0^0 = -\int_V \mathrm{d}^3\boldsymbol{x}\ \nabla \cdot \vec{T}$$

$$\partial_t \int_V \mathrm{d}^3\boldsymbol{x}\ T_0^0 = -\int_V \mathrm{d}^3\boldsymbol{x}\ \nabla \cdot \vec{T} \overbrace{\rightarrow}^{\text{散度定理；}\delta V \text{ 代表 } V \text{ 的表面}} = -\int_{\delta V} \mathrm{d}^2\boldsymbol{x}\ \vec{T} \underbrace{=}_{\text{场在无穷远处为零}} 0 \quad (4.38)$$

$$\rightarrow \partial_t \int_V \mathrm{d}^3\boldsymbol{x}\ T_0^0 = 0 \quad (4.39)$$

最后一步中我们是这样处理的：考虑一个无限大的空间域，如一个有着无限大半径 r 的球体，我们需要在其表面计算积分，即在 $r=\infty$ 处计算场值。在2.3节中，我们发现了物理中任何东西都有一个速度上界。由此无穷远处的场不可能对有限远 x 处的物理有任何影响，故我们说场在无穷远处为零[16]。

最终的结论是：时空平移不变性导致四个守恒量[36]，4.39式告诉我们它们是

[36] 因为对于任意 a_μ 都有 $\partial_0 T_\mu^0 a_\mu = \partial_0 T_0^0 a_0 - \partial_0 T_i^0 a^i = 0$ 成立，所以对于每一个分量我们都能得到一个连续性方程。

$$E = \int \mathrm{d}^3\boldsymbol{x}\ T_0^0 \quad (4.40)$$

$$P_i = \int \mathrm{d}^3\boldsymbol{x}\ T_i^0 \quad (4.41)$$

其中 $i=1,2,3$。它们分别被称为系统的总能量——因时间平移 $x_0 \rightarrow x_0 + a_0$ 不变性而守恒；场的总动量——因空间平移 $x_i \rightarrow x_i + a_i$ 不变性而守恒。

4.5.3 旋转和推动

接下来，我们会仔细研究旋转和推动下的不变性。我们将先从4.30式的第二项出发，随后讨论第一项的含义。我们将从第一项和第二项分别得到一个量，两项之和是守恒的。标量场没有可以混合的分量，故对于它而言第一项始终为零，即 $S_{\mu\nu}=0$；也可以用**??**节的结论解释：我们用标量作为 Lorentz 群一维表示的生成元。故我们在第一部分中导出的将会是对应于标量场真实守恒量。

[16] 译注：这段讨论怪怪的

变换的第二项由下式给出

$$x_\mu \to x'_\mu = x_\mu + \delta x_\mu = x_\mu + M_\mu^\sigma x_\sigma$$

其中 M_μ^σ 是转动和推动的在无穷维表示下的生成元，即在??式中所定义的微分算符。

将 $\delta x_\mu = M_\mu^\sigma x_\sigma$ 代入4.33中得

$$\begin{aligned} \partial_\nu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\nu \Phi)} \frac{\partial \Phi(x)}{\partial x^\mu} - \delta_\mu^\nu \mathscr{L} \right) M^{\mu\sigma} x_\sigma = 0 \\ \underbrace{\to}_{4.35\text{式}} \partial_\nu T_\mu^\nu M^{\mu\sigma} x_\sigma = 0 \end{aligned} \tag{4.42}$$

我们能将它改写成更常见的形式

$$\begin{aligned} &= \frac{1}{2}(\partial^\nu T^\mu_\nu M_{\mu\sigma} x^\sigma + \partial^\nu T^\mu_\nu \underbrace{M_{\mu\sigma}}_{=-M_{\sigma\mu}} x^\sigma) = \frac{1}{2}(\partial^\nu T^\mu_\nu M_{\mu\sigma} x^\sigma + \partial^\nu T^\mu_\nu - M_{\sigma\mu} x^\sigma) \\ &\underbrace{=}_{\text{替换哑指标}} \frac{1}{2}(\partial^\nu T^\mu_\nu M_{\mu\sigma} x^\sigma + \partial^\nu T^\sigma_\nu - M_{\mu\sigma} x^\mu) = \frac{1}{2}(\partial^\nu T^\mu_\nu x^\sigma + \partial^\nu T^\sigma_\nu x^\mu) M_{\mu\sigma} \\ &= \frac{1}{2}\partial_\nu (T^{\mu\nu} x^\sigma + \partial^\nu T^{\sigma\nu} x^\mu) M_{\mu\sigma} = 0 \\ &\to \text{引入}(J^\nu)^{\sigma\mu} \equiv T^{\mu\nu} x^\sigma - \partial^\nu T^{\sigma\nu} x^\mu \to \partial_\nu (J^\nu)^{\sigma\mu} = 0\text{。} \end{aligned} \tag{4.43}$$

我们在最后一步引入的 J^ν 被称为 **Noether 流 (Noether current)**。注意到，由于在??式定义 M 时引入的反对称性 $M_{\mu\sigma} = M_{\sigma\mu}$，我们有 $M_{\mu\mu} = 0$。由此我们找到了六个不同的连续性方程 $\partial_\nu (J^\nu) = 0$，分别对应于每一个非零的 $M_{\mu\sigma}$[17]。使用与4.39式一样的方式，我们可以论证如下的量不随时间改变

$$Q^{\mu\sigma} = \int d^3\boldsymbol{x} \ \left(T^{\mu 0} x^\sigma - T^{\sigma 0} x^\mu\right) \tag{4.44}$$

由此我们得到了对应于转动[37]不变性的守恒量

$$L^i_{\text{orbit}} = \frac{1}{2}\epsilon_{ijk} Q_{jk} = \frac{1}{2}\epsilon_{ijk} \int d^3\boldsymbol{x} \ \left(T^{\mu 0} x^\sigma - \partial^\nu T^{\sigma 0} x^\mu\right), \tag{4.45}$$

其被称为场的轨道角动量[38]。

同样的，我们也有所对应的守恒量[39]

$$Q^{0i} = \int d^3\boldsymbol{x} \ \left(T^{i0} x^0 - \partial^\nu T^{00} x^i\right) \tag{4.46}$$

[17] 译注：排除了零元的同时，也将始终互为相反数的一对元视作一个

[37] 回忆一下??式中提到的转动生成元 J_i 和 $M_{\mu\nu}$ 之间的关系：$J_i = \frac{1}{2}\epsilon_{ijk} M_{jk}$，这里我们使用拉丁字母 $i, j \in \{1, 2, 3\}$ 而不是希腊字母 $\mu, \nu \in \{0, 1, 2, 3\}$，将其限制在空间分量上。

[38] 下标 orbit （轨道）的含义马上就会清楚，因为它实际上只是守恒量的一部分，我们马上就会去考察第二项。

[39] 请回忆推动生成元 K_i 和 $M_{\mu\nu}$ 之间的关系：$K_i = M_{0i}$

将会在附录4.7节中详细讨论。

4.5.4 自旋

接下来我们希望捡回原先在导出守恒量时所丢掉的4.30中的第一项。我们现在考察

$$\delta\Phi = \epsilon_{\mu\nu} S^{\mu\nu} \Phi(x), \tag{4.47}$$

其中 $S^{\mu\nu}$ 是问题中的变换[40]的恰当**有限维表示**表示。这使4.29式多了一项

$$\partial_\rho \left(\frac{\partial \mathscr{L}}{\partial(\partial_\rho \Phi)} \epsilon_{\mu\nu} S^{\mu\nu} \Phi(x) \right) \tag{4.48}$$

同时也使4.45式多了一项。完整的守恒量是

$$L^i = \frac{1}{2}\epsilon_{ijk} Q^{jk} = \frac{1}{2}\epsilon_{ijk} \int \mathrm{d}^3\boldsymbol{x} \left(\frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} S^{jk}\Phi(x) + (T^{k0}x^j - T^{j0}x^k) \right) \tag{4.49}$$

据此写出

$$L^i = L^i_{\text{spin}} + L^i_{\text{orbit}}。 \tag{4.50}$$

第一项是个新东西，但和我们之前考虑的轨道角动量有类似之处，因为这是当我们讨论同一个不变性时出现而加在一起的。标准的观点是这个守恒量是某种内禀[41]角动量。

回忆我们曾用某种也叫做自旋 (spin) 的东西来给 Poincare 群的表示作标签。现在这个记号再一次出现了。我们将在下一章，并在??节中相当仔细地，学习我们如何能测量这类新的角动量，自旋，以及认识到它和我们用以给 Poincare 群的表示作标签的东西，确实是相一致的。

4.5.5 场论中的 Neother 定理——内禀对称性

我们现在来仔细研究内禀对称性[42]。在某些场**自身**的无穷小变换

$$\Phi_i \to \Phi'_i = \Phi_i + \delta\Phi_i \tag{4.51}$$

[40] 场分量的混合是由有限维表示来进行的。例如转动生成元的二维表示 $J_i = \frac{1}{2}\sigma_i$ 混合了外尔旋量的分量。

[41] 我们随后将会看到场产生湮灭粒子。一个自旋 $\frac{1}{2}$ 场产生自旋 $\frac{1}{2}$ 粒子，这是基本粒子的固有属性。所以说我们用“内禀”这个词。轨道角动量是用来描述两个或更多粒子如何相互旋转的一个量。

[42] 毋庸置疑，对于相互作用场论，内禀对称性是相当重要的。除此之外，我们将从最简单的内禀对称性出发开始学习量子场论

下拉格朗日量的不变性 ($\delta\mathscr{L}=0$)，数学上表达为[18]

$$\begin{aligned}\delta\mathscr{L} &= \mathscr{L}(\Phi_i,\partial_\mu\Phi_i)-\mathscr{L}(\Phi_i+\delta\Phi_i,\partial_\mu(\Phi_i+\delta\Phi_i))\\ &= -\frac{\partial\mathscr{L}}{\partial\Phi_i}\delta\Phi_i-\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi_i)}\partial_\mu\delta\Phi_i\overset{!}{=}0,\end{aligned}\tag{4.52}$$

即因变换是无穷小的，故 Taylor 展开到 $\delta\Phi_i$ 第一项。依场的欧拉-拉格朗日方程4.10可知

$$\frac{\partial\mathscr{L}}{\partial\Phi}=\partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi)}\right)。$$

代入4.42式，我们得到

$$\delta\mathscr{L}=-\partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi_i)}\right)\delta\Phi_i-\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi_i)}\partial_\mu\delta\Phi_i=0\tag{4.53}$$

再用上莱布尼兹律，得到

$$\partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi_i)}\delta\Phi_i\right)=0\tag{4.54}$$

这意味着，如果拉格朗日量在变换 $\Phi_i\to\Phi_i'=\Phi_i+\delta\Phi_i$ 下不变，那我们又得到一个被称为了 Noether 流的量

$$J^\mu\equiv\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi)}\delta\Phi_i,\tag{4.55}$$

其满足连续性方程

$$4.54\text{式}\to\partial_\mu J^\mu=0。\tag{4.56}$$

更进一步，用导出4.39式同样的办法，我们能得到一个对时间的守恒量

$$\partial_t\int\mathrm{d}^3\boldsymbol{x}\ J^0=0。\tag{4.57}$$

我们再来看一个重要的例子：在场自身的平移下不变性[43]，即在4.51式中的 $\delta\Phi=-\mathrm{i}\epsilon$：

$$\Phi\to\Phi'=\Phi-\mathrm{i}\epsilon。\tag{4.58}$$

[43] 注意此时我们并没有假定场是实的还是复的。常数 ϵ 是任取的，若我们想让场平移量是实的，我们当然可以令它是虚数。这里用 i 是为了使生成元是厄米的（现在来看还没什么意义），这样其本征值就是实的。现在不要太困扰，把这个额外的 i 当做是一个约定就好，只要通过定义 $\epsilon':=\mathrm{i}\epsilon$，就能把它吸收到常数里去。

[18] 译注：原文上下标混乱

或者对于不止一个分量的场

$$\Phi_i \to \Phi_i' = \Phi_i - \mathrm{i}\epsilon_i。 \tag{4.59}$$

对应的守恒量被称为**共轭 (conjugate)** 动量 Π。方便起见，引入共轭动量密度[44]$\pi \equiv J_0$，通过4.54、4.57式，并利用守恒量中不需要包含 $\delta\Phi = -\mathrm{i}\epsilon$ 这件事。我们得到[45]

$$\partial_t \Pi = 0 \;\to\; \partial_t \Pi = \partial_t \int \mathrm{d}^3\boldsymbol{x}\; J_0 = \partial_t \int \mathrm{d}^3\boldsymbol{x}\; \pi = 0$$
$$\text{用到了 } J_0 = \pi = \frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} \tag{4.60}$$

注意这和源于**空间**平移不变性 $\Phi(x) \to \Phi(x') = \Phi(x+\epsilon)$ **物理的**场动量密度不同，那东西长这样

$$p^i = T^{0i} = \frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} \frac{\partial \Phi}{\partial x_i},$$

即之前在4.41式中导出的东西。下面数章中我们将会看到更多不同的内禀对称性，它们在发展相互作用理论时是无价的。现在来总结一下我们学过的场论中的对称性和守恒量：

- 空间平移不变性 $\Rightarrow$ 物理动量 $P_i = \int \mathrm{d}^3\boldsymbol{x}\; T_i^0$ 守恒
- 推动不变性[46]$\Rightarrow$ 能量中心的速度是常数
- 转动不变性 $\Rightarrow$ 总角动量 $L^i = \frac{1}{2}\epsilon_{ijk} \int \mathrm{d}^3\boldsymbol{x} \left(\frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} S^{ij}\Phi(x) + (T^{k0}x^j - T^{j0}x^k) \right) = L^i_{\text{spin}} + L^i_{\text{orbit}}$，包含自旋和轨道角动量，守恒
- 时间平移不变性 $\Rightarrow$ 能量 $E = \int \mathrm{d}^3\boldsymbol{x}\; T_0^0$ 守恒
- 场自身的平移不变性 $\Rightarrow$ 共轭动量 $\Pi = \int \mathrm{d}^3\boldsymbol{x}\; \frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} = \int \mathrm{d}^3\boldsymbol{x}\; \pi$ 守恒

Further Reading Tips 阅读建议

- **Cornelius Lanczos - The Variational Principles of Mechanics**[47]是一本关于拉格朗日体系在经典力学中的应用的绝妙书。
- **Herbert Goldstein - Classical Mechanics**[48]是经典力学的标准教材，其中有许多对拉格朗日体系非常好的解释。

[44] 这个相当概括的量是量子场论最重要的记号，我们将在第??章中看到这一点

[45] 4.55式中的 $\delta\Phi$ 成为一个常数 ϵ，因为这个量对于任何 ϵ 都守恒，所以我们在定义守恒量时不需要它。

[46] 推动又被称为在动量的空间平移

[47] Cornelius Lanczos. The Variational Principles of Mechanics. Dover Publications, 4th edition, 3 1986. ISBN 9780486650678

[48] Herbert Goldstein, Charles P. Poole Jr., and John L. Safko. Classical Mechanics. Addison-Wesley, 3rd edition, 6 2001. ISBN 9780201657029

4.6 4.6 Appendix: Conserved Quantity from Boost Invariance for Particle Theories 附录：粒子理论中的推动不变性与其守恒量

4.7 Appendix: Conserved Quantity from Boost Invariance for Field Theories 附录：场论中的推动不变性与其守恒量

III The Equations of Nature 自然界的方程

5. Measuring Nature 测量

我们现在已经发现了对称性与守恒量之间的关系了，那么我们接下来就可以利用这种联系。用更专业的话说，Noether 定理建立了对称变换的生成元与一个守恒量之间的联系。我们本章将利用这种联系。

守恒量常常被物理学家们用来描述物理系统，因为无论这个系统经历了怎样复杂的变化，守恒量都是一样的。比如，物理学家们为了描述一个实验会使用动量，能量或角动量。Noether 定理提示我们了一个方向和一个至关重要的想法：我们在挑选哪些量我们用来描述自然的时候，我们实际上在看伴随着其对应的生成元：

$$\text{物理量} \Rightarrow \text{对应对称性的生成元} \tag{5.1}$$

我们会看到，这种选择会自然地引导我们给出给出量子理论。

5.1 Operators of Quantum Mechanics 量子力学中的算符

惯例上用一个尖帽子 $\hat{O}$ 来表示一个算符

拉格朗日量在空间平移操作的生成元的作用下不变导致我们得出动量守恒。因此，我们可以得到

$$\text{动量}\hat{p}_i \to \text{空间平移操作的生成元} - i\partial_i$$

类似的，时间平移的生成元带来的不变形给出我们能量守恒，也就是

$$\text{能量}\hat{E} \to \text{时间平移操作的生成元} - i\partial_0$$

没有关于“位置守恒”的对称性，所以位置并没有伴随生成元，我们有[1]

$$\text{位置}\hat{x}_i \to x_i$$

我们现在利用算符给出了我们在描述自然的理论中使用的物理量。接下来很合理的我们就需要问：这些算符作用在什么上面？它们是如何与实验中我们做的测量联系起来的呢？我们下一章将会仔细的讨论这件事情，在这里我们只需要知道我们的物理量，也就是算符，是作用在一些东西上的。让我们暂时继续用抽象的概念理解算符作用在什么伤。我们叫它 Ψ，我们后面会探究这个一些东西是什么鬼。

现在，我们可以去得到一些非常重要的[2]量子力学的方程了。就像上面已经解释过的那样，我们假定我们的算符作用在抽象的 Ψ 上。于是我们就有

$$\begin{aligned}[\hat{p}_i, \hat{x}_k]\Psi &= (\hat{p}_x \hat{x}_j)\Psi = (\partial_i \hat{x}_j - \hat{x}_j \partial_i)\Psi \\ &\underbrace{=}_{\text{莱布尼兹律}} (\partial_i \hat{x}_j)\Psi + \cancel{\hat{x}_j(\partial_i \Psi)} - \cancel{\hat{x}_j(\partial_i \Psi)} \underbrace{=}_{\text{因为}\partial_i \hat{x}_j = \frac{\partial x_j}{\partial x_i}} i\delta_{ij}\Psi\end{aligned} \tag{5.2}$$

这个方程对任意的 Ψ 都成立，因为我们没有对 Ψ 做任何假设，因此我们可以不考虑它[3]而直接的写出

$$[\hat{p}_i, \hat{x}_j] = i\delta_{ij} \tag{5.3}$$

5.1.1 Spin and Angular Momentum 自旋与角动量

上一章的第4.5.4节中我们看到，只是对标量理论来说，旋转不变形带来的守恒量有两个部分，而这个第二部分则对应了轨道角动量，我们把它称为**无限**维的轨道角动量生成元的表示。$\hat{L}_i \to$ 转动生成元（无限维表示）$i\frac{1}{2}\epsilon_{ijk}(x^j\partial^k - x^k\partial^j)$。类似的我们把第一部分称作自旋，对应**有限**维度的生成元[4]的表示。

[1] 或者说，我们可以观察守恒量与对应的推动下的不变性。注意到我们在4.6节中得到的非相对论性 Galilei 推动下的守恒量是从非相对论性的拉格朗日量得出来的。统一，我们可以对相对论性拉格朗日量做一样的事情，得到守恒量 $i\vec{p} - \vec{x}E$。相对论性能量由 $E = \sqrt{m^2 + p^2}$ 给出。在相对论性极限 $c \to \infty$ 下，我们有 $E \approx m$，并且，Lorentz 推动下的守恒量回到我们的道德 Galilei 推动的结果。因此，粒子理论的守恒量为 $M_i = (tp_i - x_iE)$。推动的生成元（见(??)，其中 $K_i = M_{0i}$）为 $K_i = i(x^0\partial_i - x_i\partial_0)$。比较两者，其中 $x_0 = t$，给出 $M_i = (tp_i - x_iE) \leftrightarrow K_i = i(t\partial_i - x_i\partial_0)$。因此，利用我们之前的选择，很直观的我们有位置：$\hat{x}_i \to x_i$

[2] 如果你完全不了解量子力学，为什么这些简单地式子会如此重要对你来说会看起来比较奇怪。你也许听说过 Heisenberg 不确定性原理。在第??节，我们会更进一步的研究量子力学的形式与结构，然后我们就可以看到这个方程告诉我们我们不能够一任意精度同时测量一个粒子的动量和坐标。我们的物理量的解释是算符的测量，而这个方程告诉我们先测量动量再测量位置与先测量位置再测量动量是完全不一样的。

[3] 如果这对你来说比较罕见，注意我们可以把**每**一个矢量方程都写成只包含数的形式。比如，牛顿第二定律 $\vec{F} = m\ddot{\vec{x}}$，可以对于任意矢量 $\vec{C}$ 写出 $\vec{F}\vec{C} = m\ddot{\vec{x}}\vec{C}$。不仅如此，如果对于任意的 $\vec{C}$ 都成立，那么一直把这个 $\vec{C}$ 写着也就没什么意义了。

[4] 记得这就是从场的分量的混杂的不变性带来的守恒量，因此是有限维的表示。

自旋$\hat{S}_i \to$ 转动生成元（有限维表示）S_i

如前面(4.30)解释过的那样，$S_{\mu\nu}$ 和转动算符生成元 S_i 的关系为 $S_i = \frac{1}{2}\epsilon_{ijk}S_{jk}$。

比如，当我们考虑自旋 $\frac{1}{2}$ 场的时候，我们需要用到我们在节**??**中得到的二维表示：

$$\hat{S}_i = i\frac{\sigma_i}{2} \tag{5.4}$$

其中 σ_i 是 Pauli 矩阵。我们会在节**??**学习了如何利用我们本章得到的这些算符之后回到这一有趣而奇怪的类型的角动量。时刻注意，只有这两部分的**和**是守恒的。

5.2 The Operators of Quantum Field Theory 量子场论的算符

场论的最重要的客体当然就是**作为时空位置的函数**[5]的场了 $\Phi = \Phi(x)$。我们接下来希望描述时空上的一个点发生的相互作用，从而我们需要考虑动力学变量 $\pi = \pi(x)$ 的密度，而不是其其总量 $\Pi = \int d^3x\pi(x) \neq \Pi(x)$。

[5] 这里，$x = x_0, x_1, x_2, x_3$，包含时间 $x_0 = t$。

我们上一章发现了在奇异不变性下场自己 $\Phi \to \Phi - i\epsilon$ 产生了一种新的守恒量，称为共轭动量 Π。就像我们上一章做的那样，我们用对应的生成元来表示共轭动量密度(**??**)。

$$\text{共轭动量密度}\pi(x) \to \text{关于场自己的平移的生成元：} -i\frac{\partial}{\partial\Phi(x)}$$

我们后面会看到量子场论与这里处理的稍稍不同，而且我们这里的处理现在已经足够了。

就像我们上一节讨论的那样，我们需要给我们的算符一些作用上去的东西。我们这里同样适用抽象的 Ψ，我们后面的章节会具体的说明。我们同样可以得到一个十分重要的方程，这次是针对量子场论[6]

$$[\Phi(x),\pi(y)]\Psi = \left[\Phi(x), -i\frac{\partial}{\partial\Phi(y)}\right]\Psi$$
$$\underbrace{=}_{\text{莱布尼兹律}} \cancel{-i\Phi(x)\frac{\Psi}{\partial\Phi(y)}} \cancel{+i\Phi(x)\frac{\Psi}{\partial\Phi(y)}} + i\left(\frac{\partial\Phi(x)}{\partial\Phi(y)}\right)\Psi = i\delta(x-y)\Psi \tag{5.5}$$

[6] 最后一步的时候我们类似 $\frac{\partial x_i}{\partial x_j} = \delta_{ij}$ 给出 delta 分布 $\frac{\partial f(x_i)}{\partial f(x_j)} = \delta(x_i - x_j)$，它实际上可以严格的说明的。更具体的内容请看附录 D.2。

同样的，由于这个式子对任意的 Ψ 都成立，我们可以写

$$[\Phi(x), \pi(y)] = i\delta(x - y) \tag{5.6}$$

类似的，对于场有多于一个分量的情况，我们有

$$[\Phi_i(x), \pi_j(y)] = i\delta(x - y)\delta_{ij} \tag{5.7}$$

在后面我们会卡到，机会所有的量子场论中的东西都要从这些简单的方程出发。