

Physics from Symmetry

1st edition

Translate Version: 2016 年 4 月 8 日

DRAFT TRANSLATION

Copyright © 2016 Chaoli-Translating-Group

PUBLISHED BY NON

[HTTPS://GITHUB.COM/LASERROGER/PHYSICS-FROM-SYMMETRY/](https://github.com/LASERROGER/PHYSICS-FROM-SYMMETRY/)

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the “License”). You may not use this file except in compliance with the License. You may obtain a copy of the License at <http://creativecommons.org/licenses/by-nc/3.0>. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

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

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

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

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。

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

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。

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。

我希望你读这本书能够像我写这本书的时候一样愉悦。

许多页边注的内容是拓展
信息与图景。

Karlsruhe, 2015.01

Jakob Schwichtenberg

DRAFT TRANSLATION

⁴给译文排版造成了小小的麻烦 — 译者

DRAFT TRANSLATION

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>。

DRAFT TRANSLATION

Contents

I	Foundations 基础	
1	简介	3
1.1	推不出来的事情	3
1.2	全书概览	5
1.3	基本粒子和基本相互作用力	7
2	狭义相对论	9
2.1	狭义相对论的中的不变量	10
2.2	固有时	12
2.3	速度上限	14
2.4	Minkowski 记法	15
2.5	Lorentz 变换	18
2.6	不变性，对称性，协变性	19

3	Lie 群	23
3.1	群	24
3.2	二维旋转	27
3.2.1	单位复数表示旋转变换	28
3.3	三维旋转	31
3.3.1	四元数	32
3.4	Lie 代数	37
3.4.1	$SO(3)$ 群的生成元与 Lie 代数	40
3.4.2	Lie 代数的抽象定义	43
3.4.3	$SU(2)$ 的生成元与 Lie 代数	44
3.4.4	Lie 群的抽象定义	46
3.5	表示论	48
3.6	$SU(2)$	51
3.6.1	The Finite-dimensional Irreducible Representations of $SU(2)$ $SU(2)$ 群的有限维不可约表示	52
3.6.2	$SU(2)$ 的 Casimir 算符	54
3.6.3	$SU(2)$ 的一维表示	55
3.6.4	$SU(2)$ 的二维表示	55
3.6.5	$SU(2)$ 的三维表示	57
3.7	Lorentz 群	57
3.7.1	Lorentz 群的 1 表示	60
3.7.2	Lorentz 群其他部分的生成元	63
3.7.3	$O(1,3)^\uparrow$ 群的 Lie 代数	66
3.7.4	$(0,0)$ 表示	67
3.7.5	$(\frac{1}{2}, 0)$ 表示	67
3.7.6	$(0, \frac{1}{2})$ 表示	69
3.7.7	Van der Waerden 符号	70
3.7.8	$(\frac{1}{2}, \frac{1}{2})$ 表示	75
3.7.9	旋量和宇称	79
4	框架/体系	81
4.1	拉格朗日形式	81
4.1.1	Fermat 原理	82

	11
4.1.2 变分法：基本思想	82
4.2 限制	83
4.3 粒子理论与场论	84
4.4 欧拉-拉格朗日方程	85
4.5 Noether 定理	87
4.5.1 粒子理论的 Noether 定理	87
4.5.2 场论中的 Noether 定理——时空对称性	91
4.5.3 旋转和推动	94
4.5.4 自旋	96
4.5.5 场论中的 Noether 定理——内禀对称性	96
4.6 附录：粒子理论中的推动不变性与其守恒量	99
4.7 附录：场论中的推动不变性与其守恒量	100

III The Equations of Nature 自然界的方程

5 测量	103
5.1 量子力学中的算符	103
5.1.1 自旋与角动量	104
5.2 量子场论的算符	105
6 自由场理论	107
6.1 Lorentz 协变性与不变量	107
7 相互作用理论	109
7.1 $U(1)$ 相互作用	111
7.1.1 自旋 $\frac{1}{2}$ 自由场的内禀对称性	111
7.1.2 自旋为 1 自由场的内禀对称性	113
7.1.3 把奇怪的东西堆一起	114
7.1.4 非齐次麦克斯韦方程和最小耦合	116
7.1.5 再一次电荷共轭变换	118
7.1.6 内禀 $U(1)$ 对称性的 Noether 定理	119
7.1.7 零自旋有质量场的相互作用	121
7.1.8 自旋为 1 有质量场的相互作用	121
7.2 $SU(2)$ 相互作用	122
7.3 质量项、 $U(1)$ 和 $SU(2)$ 的统一	122

7.4	宇称破坏	122
7.5	轻子质量项	122
7.6	夸克质量项	122
7.7	同位旋	122
7.7.1	Labelling 态	122
7.8	$SU(3)$ 相互作用	122
7.8.1	色	122
7.8.2	夸克描述	122
7.9	Bosons 和 Fermions 间的相互影响	122



Foundations 基础

1	简介	3
1.1	推不出来的事情	
1.2	全书概览	
1.3	基本粒子和基本相互作用力	
2	狭义相对论	9
2.1	狭义相对论的中的不变量	
2.2	固有时	
2.3	速度上限	
2.4	Minkowski 记法	
2.5	Lorentz 变换	
2.6	不变性，对称性，协变性	

DRAFT TRANSLATION

1. Introduction 简介

1.1 What we Cannot Derive 推不出来的事情

在开始讲我们能从对称性里面了解到什么之前，我们首先说明需要在理论中人为加一些什么东西。首先，目前没有任何理论可以得到自然界的常数。这些常数需要实验测定，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，而我们用拉格朗日量描述自由（即无相互作用）的场/粒子。
- 我们只用三个 Poincare 群双覆盖的基本表示，分别对应自旋 $0, \frac{1}{2}$ 和 1 。没有基本粒子的自旋是 $\frac{3}{2}$ 。

这是现代理论必须人为引入的假设。我们从实验上知道这些假

¹ 如果你不理解这个简介中的某些名词，比如规范理论或者双覆盖，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。

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）

设是正确的，但是目前为止没有更深刻的原理告诉我们为什么到 3 就停。

除此之外，还有两件事情没法从对称性中得到，但一个严谨的理论有时必须考虑它们：

- 我们只允许在拉格朗日量中引入尽可能低阶的非平庸微分算符 ∂_μ 。一些理论使用一阶的微分算符 ∂_μ ，而另一些理论中，Lorentz 不变性禁止了一阶导数，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。
- 类似的，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 Fermion¹，它们中永远不存在两个粒子处在完全一样的态上。相反，这种情况对整数自旋的粒子 — 通常被称为 Boson² — 是可能的。

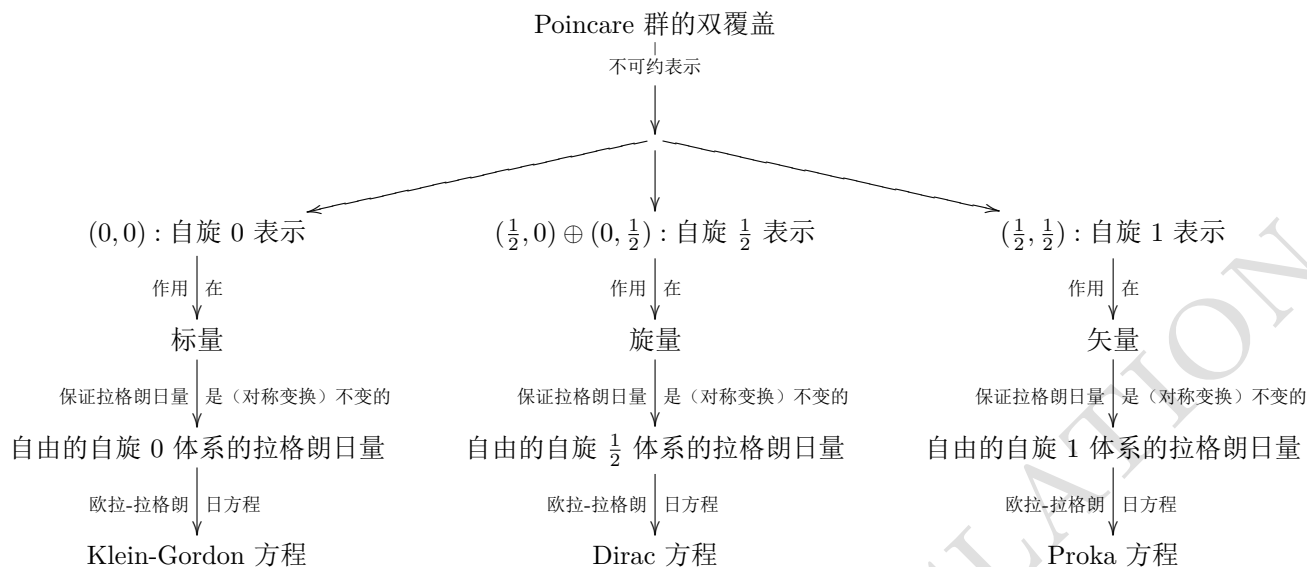
最后呢，本书不能导出引力理论。大名鼎鼎的广义相对论就是优美而准确的描述引力的理论；然而这个理论与其他理论完全不一样，超出了本书的研究范围。而尝试将引力问题划入相同框架下的量子引力理论仍待完善：目前没有人能够成功得出它。但最后一章我们会对引力做一些评述。

³ 量子场论里一开始使用反对易子而不是对易子，从而防止理论没有下界。

¹常译为费米子

²常译为玻色子

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——你可以当成是组成其它所有表示的“地基”。这些表示就是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。

之后介绍**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**，通过考虑不同物理系统的对称性来得到它³。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 群中元素的变换下，拉格

³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『唯象』且经验的刻画 — 译者注

⁴ 严格来说，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。

朗日量必须是不变的。这使得在任意的参考系中得到的运动方程都是一样的，就像之前说的那样，“物理规律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 $U(1)$ 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 $U(1)$ 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 $SU(2)$ 和 $SU(3)$ 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，我们会讨论自发对称破缺和 Higgs 机制。这些使得我们能够去描述有质量的粒子⁵。

之后我们得到 Noether 定理，它给我们展示了对称性和守恒量之间深刻的联系。我们将物理量和相应对称性的生成元相等同。这就导出量子力学中最重要方程⁴。

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

和量子场论中

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

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。

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 and 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，并点出一些构建引力的量子理论的困难。

这本书的主要部分处理对称性的数学工具，以及导出被称为

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

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 Klein-Gordon, Dirac, Proca 和 Maxwell 方程

⁴原书有 typo

“标准模型”的理论。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，下一小节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**说明不能有两个 **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⁵
- 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6中讨论。

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 W^+ , 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，现在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**自发对称破缺**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。

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。

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**。

⁵ 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7章中我们会讲，所有的荷都有着有一个相同的美丽起源

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。

- 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，因此它们之间通过弱作用力来相互作用。
- 强作用力的荷叫**色**，因为它与人对色彩的知觉有某种奇妙的共性。这个荷实际上与日常生活中的色彩没有任何关系，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
夸克：	上 (u)	粲 (c)	顶 (t)	$+\frac{2}{3}e$	$\frac{1}{2}$	√
	下 (d)	奇异 (s)	底 (b)	$-\frac{1}{3}e$	$-\frac{1}{2}$	√
轻子：	电子中微子	μ 子中微子	τ 子中微子	0	$+\frac{1}{2}$	-
	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之后都会更具体的解释。现在是时候去尝试推导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 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，在这一奇特实验事实的基础之上，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，这些最后都证实是正确的。我们将看到这个原理的强大，但是我们首先阐明狭义相对论的基础。狭义相对论的两个基本假设是：

- **相对性原理：** 任意惯性系，即任意两个相对速度恒定的参考系的物理规律相同。
- **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，我们还假设时空均匀且各向同性。这意味着不论在哪儿 (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，就算是在火星也一样。

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。

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 50km h^{-1} ，另一个观者以 15km h^{-1} 的速度与火车一同运动，那么测得的火车速度就应是 35km h^{-1} 。与此不同的是，光始终以 $1.08 \times 10^9\text{km h}^{-1}$ 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。

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找到两惯性系之间所有允许的变换。在本章其余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

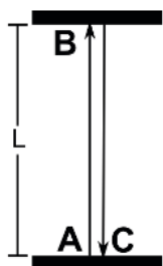


图 2.1: 思想实验图示

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

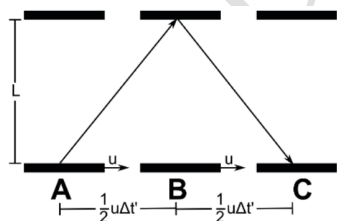


图 2.2: 思想实验图示. 第二位移动观者向左移动，因此第一位观者相对他向右移动

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。

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

$$\Delta t = t_C - t_A = \frac{2L}{c} \quad (2.1)$$

式中 L 表示原点与反射点之间的距离。

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

$$x'_A = 0 \neq x'_C = u\Delta t' \quad \rightarrow \quad \Delta x' = u\Delta t' \quad (2.2)$$

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

$$x_A = x_C \quad \rightarrow \quad \Delta x = 0 \quad (2.3)$$

我们假定第二位观者运动沿 x 轴，因此

$$y'_A = y'_C \quad z'_A = z'_C \quad \rightarrow \quad \Delta y' = 0 \quad \Delta z' = 0 \quad (2.4)$$

那么同样也有

$$y_A = y_C \quad z_A = z_C \quad \rightarrow \quad \Delta y = 0 \quad \Delta z = 0 \quad (2.5)$$

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 。

$$\Delta t' = \frac{l}{c} \quad (2.6)$$

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

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

利用式2.6可以得到

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

再利用式2.2中的 $\Delta x' = u\Delta t'$ 可得

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

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

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

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

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

终于，我们得到⁵

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

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

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

因此，我们得到了一些对于所有观者都相同的量：即二次式

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

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 Δs^2 所蕴含的物理意义。

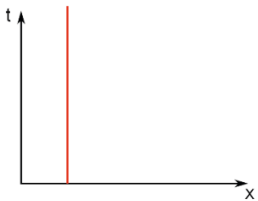


图 2.3: 静止物体的世界线。物体位置不随时间流逝而变化。

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。

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。

我们可以看到, 两个观者将对事件 AB 之间的变化持不同观点。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

$$(\Delta s)^2 = (c\Delta t)^2 - (\Delta x)^2 \quad (2.14)$$

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

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

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

$$(\Delta s')^2 = (c\Delta t')^2 \quad (2.17)$$

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

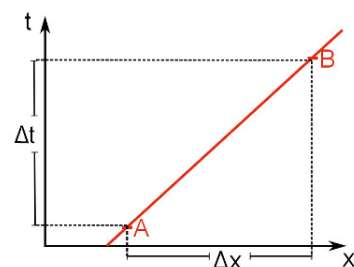


图 2.4: 某匀速直线运动物体的世界线。物体前后经过某两点记为事件 A, B 。它们间的空间距离为 Δx , 时间间隔为 Δt

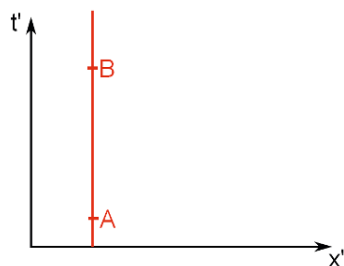


图 2.5: 同一个运动物体的世界线, 但观者与该物体速度相同。该观者观测到的事件 AB 之间的空间距离为 $\Delta x' = 0$

者相同的时间量。我们定义

$$(\Delta s)^2 = (c\Delta\tau)^2 \quad (2.18)$$

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 $\Delta \rightarrow d$:

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

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，且总与该物体相对静止。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，通常复的固有时没有物理意义。因此， $(\Delta s)^2 = 0$ 时固有时最小， $\tau = 0$ ，令

$$\begin{aligned} \Delta s_{min}^2 = 0 &= (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \\ \rightarrow (c\Delta t)^2 &= (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \\ \rightarrow c^2 &= \frac{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}{(\Delta t)^2} \end{aligned} \quad (2.20)$$

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

⁶ 按固有时定义, $(ds)^2 = (cd\tau)^2$, $(ds)^2 < 0 \rightarrow d\tau$ 为复数。

极限，则

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

$x(t)$, $y(t)$, $z(t)$ 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

$$\rightarrow c^2 = v^2 \quad (2.22)$$

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邻近的对象影响。相互作用都是局域的, 不存在超距作用, 物理效应的传播需要时间。

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 Minkowski⁷

写出狭义相对论中不变量

$$ds^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2 \quad (2.23)$$

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:

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

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

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:

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

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。

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

- 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 $\eta^{\mu\nu} = 0$ (也可简写为 $\eta^{\mu\nu} = \text{diag}(1, -1, -1, -1)$)

除此之外, 我们还要引入四维矢量 (four-vector), 简称 4 矢。

$$dx_\mu = \begin{pmatrix} dx_0 \\ dx_1 \\ dx_2 \\ dx_3 \end{pmatrix} \quad (2.25)$$

式 2.24 可用 4 矢和 Minkowski 度规表出

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

$$= dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2 \quad (2.26)$$

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 Kronecker 函数 δ_{ij} 即是单位矩阵在指标法下的表示, 详见附录 B5.5

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

$$y^\nu \equiv \eta^{\mu\nu} y_\mu \underbrace{=} \eta^{\nu\mu} y_\mu \tag{2.30}$$

Minkowski 度规具有对称性 $\eta^{\mu\nu} = \eta^{\nu\mu}$

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。

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。

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

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

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

$$dx_\mu \rightarrow dx'_\mu = \Lambda_\mu^\sigma dx_\sigma \quad (2.33)$$

由于 $(ds)^2$ 在变换下不变

$$\begin{aligned} (ds)^2 &= (ds')^2 \\ \rightarrow dx \cdot dx &\stackrel{!}{=} dx' \cdot dx' \rightarrow dx_\mu dx_\nu \eta^{\mu\nu} \stackrel{!}{=} dx'_\mu dx'_\nu \eta^{\mu\nu} \underbrace{=}_{\text{式2.33}} \Lambda_\mu^\sigma dx_\sigma \Lambda_\nu^\gamma dx_\gamma \eta^{\mu\nu} \\ &\underbrace{\rightarrow}_{\text{重命名哑指标}} dx_\mu dx_\nu \eta^{\mu\nu} \stackrel{!}{=} \Lambda_\sigma^\mu dx_\mu \Lambda_\gamma^\nu dx_\nu \eta^{\sigma\gamma} \\ &\underbrace{\rightarrow}_{dx_\nu \text{任意}} \stackrel{!}{=} \Lambda_\sigma^\mu \Lambda_\gamma^\nu \eta^{\sigma\gamma} \end{aligned} \quad (2.34)$$

¹⁵ 详见附录 C.1。

或者用矩阵形式来写¹⁵

$$\eta = \Lambda^T \eta \Lambda \quad (2.35)$$

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。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

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

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 $\vec{a} \cdot \vec{b} = \vec{a}^T \cdot \vec{b}$ 。有关 $(Oa)^T = a^T O^T$ 这一点详见附录 C.1，式 C.3。

因此¹⁸

$O^T \mathbf{1} O \stackrel{!}{=} \mathbf{1}$, 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, 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 $\det(O) \stackrel{!}{=} 1$, 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。

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 $O^T O = \mathbf{1}$ 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。

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 $\vec{x} \rightarrow -\vec{x}$ 。这种变换满足 $\det(O) \stackrel{!}{=} -1$ 及 $O^T O = \mathbf{1}$ 。因此, 如果我们要将变换限制为旋转变换, 就需要加上条件 $\det(O) \stackrel{!}{=} 1$ 。此外, 空间反演变换又称宇称变换。

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

$$F(A', B', C', \dots) = F(A, B, C, \dots) \quad (2.37)$$

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

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

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¹，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。

¹即在 Lorentz 变换下具有协变性---译者 (InSight)

Symmetry Tools 对称性工具

3 Lie 群 23

- 3.1 群
- 3.2 二维旋转
 - 单位复数表示旋转变换
- 3.3 三维旋转
 - 四元数
- 3.4 Lie 代数
 - $SO(3)$ 群的生成元与 Lie 代数
 - Lie 代数的抽象定义
 - $SU(2)$ 的生成元与 Lie 代数
 - Lie 群的抽象定义
- 3.5 表示论
- 3.6 $SU(2)$
 - The Finite-dimensional Irreducible Representations of $SU(2)$ $SU(2)$ 群的有限维不可约表示
 - $SU(2)$ 的 Casimir 算符
 - $SU(2)$ 的一维表示
 - $SU(2)$ 的二维表示
 - $SU(2)$ 的三维表示
- 3.7 Lorentz 群
 - Lorentz 群的 1 表示
 - Lorentz 群其他部分的生成元
 - $O(1,3)^\uparrow$ 群的 Lie 代数
 - $(0,0)$ 表示
 - $(\frac{1}{2},0)$ 表示
 - $(0,\frac{1}{2})$ 表示
 - Van der Waerden 符号
 - $(\frac{1}{2},\frac{1}{2})$ 表示
 - 旋量和宇称

4 框架/体系 81

- 4.1 拉格朗日形式
 - Fermat 原理
 - 变分法：基本思想
- 4.2 限制
- 4.3 粒子理论与场论
- 4.4 欧拉-拉格朗日方程
- 4.5 Noether 定理
 - 粒子理论的 Noether 定理
 - 场论中的 Noether 定理——时空对称性
 - 旋转和推动
 - 自旋
 - 场论中的 Noether 定理——内禀对称性
- 4.6 附录：粒子理论中的推动不变性与其守恒量
- 4.7 附录：场论中的推动不变性与其守恒量

DRAFT TRANSLATION

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。

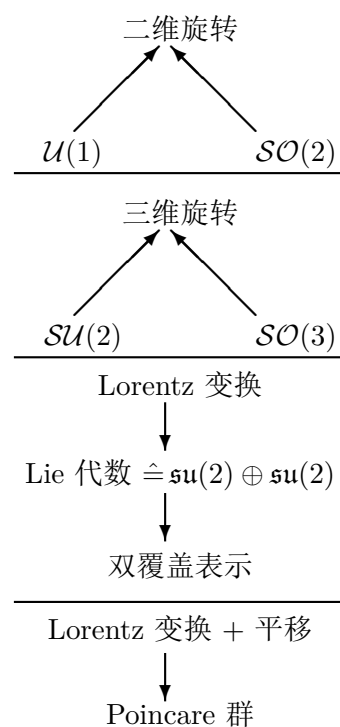
我们从两个简单例子引出群的定义，然后作为学习 Lie 群理论的第一步，我们讨论描述二维旋转变换的两种方式：

- 2×2 旋转矩阵。
- 单位复数¹。

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 — $SU(2)$ ¹ 有关。之后我们研究 **Lie 代数**，使用相对简单的 Lie 代数就能深入研究复杂的 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 代数得出对称变换的不同表示。这样就能看出开始时使用的表示其实只是一种特殊情况。于是又能研究 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 $SU(2)$ Lie 代数所组成，因此可以直

¹指模长为 1 的复数。

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



¹ 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。

接利用熟悉的 $SU(2)$ 群的结论。最后将平移变换考虑进来, 就得到了 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, 物理中经常遇到这类对称性。

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 $90^\circ, 180^\circ, 270^\circ, 0^\circ$ 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

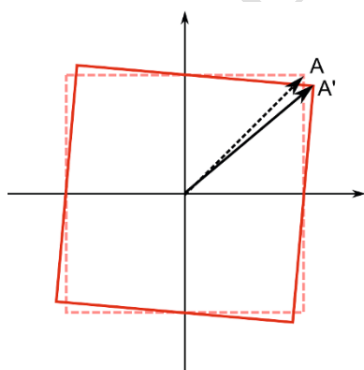


图 3.2: 正方形绕中心顺时针旋转 5°

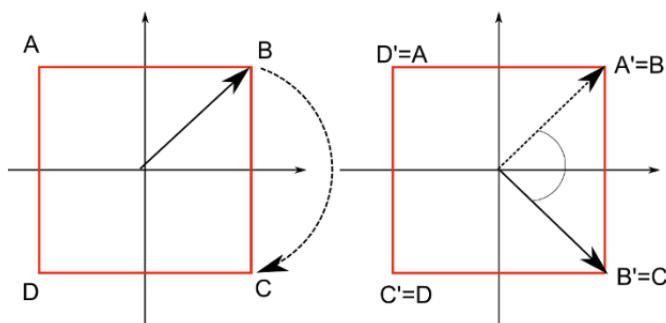


图 3.3: 正方形绕中心旋转 90°

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。

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。

为了让描述对称的工具 — 群 — 有确切的含义，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° (先逆时针再顺时针转) 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³。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：

$$R(110^\circ)R(40^\circ)R(90^\circ) = R(110^\circ)\left(R(40^\circ)R(90^\circ)\right) = R(110^\circ)R(130^\circ) \quad (3.1)$$

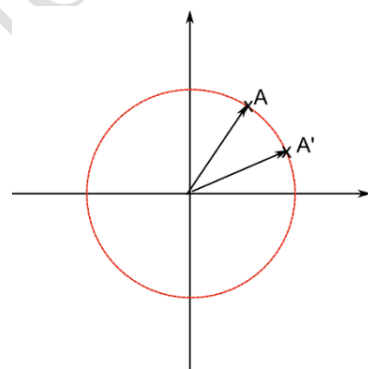


图 3.4: 单位圆绕圆心旋转任意角度都不变

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

³ 它们不一定满足交换律！例如绕不同轴的旋转变换之间不能交换顺序，比如 $R_x(\theta)R_z(\phi) \neq R_z(\phi)R_x(\theta)$

$$R(110^\circ)R(40^\circ)R(90^\circ) = \left(R(110^\circ)R(40^\circ) \right) R(90^\circ) = R(150^\circ)R(90^\circ) \quad (3.2)$$

这个性质称为**结合律**。

- 我们要有关于两个群元 (对称变换) 怎样结合的规则, 它是个**二元运算**², 我们称它为**群乘法**。

⁴ 旋转矩阵见附录 A.2.

⁵ 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。

在上面的例子里, 旋转变换的标准表示方法是用旋转矩阵⁴, 两个群元 (两个旋转矩阵, 或两个旋转操作) 结合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⁵, 群论非常系统地概括了所有形式。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3.5节学习表示论。

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。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中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

⁶ 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on应用群论来解决问题。

(我们通过对称变换的性质导出的) 群公理⁶:

一个群是一个集合 G 加上一个定义在 G 上的二元运算 (群乘法) \circ , 若 (G, \circ) 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 $g_1, g_2, g_3 \in G, g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$

总结: 某对象在一些变换下保持不变, 这些变换组成³的集合叫做对称群。对于 Minkowski 时空, Minkowski 度规⁷在变换下保持不变, 相应的对称群称为 Poincare 群。

要注意群的定义完全与变换的对象是啥没有关系。我们可以脱

⁷ 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。

²两个群元结合成一个

³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)

离特定对象而只研究对称变换本身，群的定义将变换从对象中‘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称性。群论让我们不用管变换的对象（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⁸。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）可以作用于不同类对象：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⁹，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：

$$R_\theta = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad (3.3)$$

关于 x 轴与 y 轴的反射变换用矩阵表示为：

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

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 $a \rightarrow a'$ 下不变意味着⁴：

$$a' \cdot a' \stackrel{!}{=} a \cdot a \quad (3.5)$$

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

$$a \cdot a = a^T a \rightarrow a'^T a' = (\mathcal{O}a)^T \mathcal{O}a = a^T \mathcal{O}^T \mathcal{O}a \stackrel{!}{=} a^T a = a \cdot a \quad (3.6)$$

⁸ 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。

⁹ 旋转矩阵的推导见附录 A.2.

⁴等号上面的! 只是起强调、提示作用 — 译者 (SI)

由此可见，使向量长度不变的变换必须满足：

$$\mathcal{O}^T \mathcal{O} = I \quad (3.7)$$

其中 I 表示单位矩阵¹⁰。前文的旋转和反射矩阵都满足此条件¹¹ 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¹²。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：

$$\begin{aligned} \det(\mathcal{O}^T \mathcal{O}) &= \det(\mathcal{O}^T) \det(\mathcal{O}) \stackrel{!}{=} \det(I) = 1 \\ (\det(\mathcal{O}))^2 &\stackrel{!}{=} 1 \rightarrow \det(\mathcal{O}) \stackrel{!}{=} \pm 1 \end{aligned} \quad (3.8)$$

$\det \mathcal{O} = 1$ 的矩阵对应旋转变换¹³。

条件

$$\mathcal{O}^T \mathcal{O} = I \quad (3.9)$$

$$\det \mathcal{O} = 1 \quad (3.10)$$

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¹⁴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：绕原点旋转 θ 角可以表示为乘以单位复数 z （单位复数意为 $z = a + ib$, $|z|^2 = z^* z = 1$ ）¹⁵。

所有单位复数构成一个群，称为¹⁶ $\mathcal{U}(1)$ ，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¹⁷, $\forall \mathcal{U} \in \mathcal{U}(1)$:

$$\mathcal{U}^* \mathcal{U} = 1 \quad (3.11)$$

单位复数另一种写法是¹⁸:

¹⁰

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

¹¹ 例如 3.3 式的矩阵, $R_\theta^T R = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} \cos^2 \theta + \sin^2 \theta & 0 \\ 0 & \sin^2 \theta + \cos^2 \theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

¹² 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。

¹³ 见 3.3 与 3.4 式。反射矩阵的行列式等于 -1。

¹⁴ 右手坐标系与左手坐标系见附录 A.5。

¹⁵ 上标 * 表示复数的共轭复数： $z = a + ib \rightarrow z^* = a - ib$

¹⁶ \mathcal{U} 表示么正性，它的含义为 $\mathcal{U}^* \mathcal{U} = 1$ 。

¹⁷ 定义中包含复数乘法的群的普遍性质见??节的附录。

¹⁸ 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 $\text{Re}(z) = a$; b 称为虚部，记为 $\text{Im}(z) = b$ 。欧拉公式中 R_θ 的实部为 $\cos \theta$, 虚部为 $\sin \theta$ 。

$$R_\theta = e^{i\theta} = \cos \theta + i \sin \theta \quad (3.12)$$

$R(\theta)$ 的模（长度）为:

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

例：将向量 $(3, 5)$ 对应的复数 $z = 3 + 5i$ 旋转 90° :

$$z \rightarrow z' = e^{i90^\circ} z = \left(\underbrace{\cos 90^\circ}_{=0} + i \underbrace{\sin 90^\circ}_{=1} \right) (3 + 5i) = i(3 + 5i) = 3i - 5 \quad (3.14)$$

图 3.6 绘制了旋转前后的向量（复数），图中可见复数乘以 e^{i90° 旋转了 90° 。需要注意 e^{i90° 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： $|z| = \text{实部}^2 + \text{虚部}^2 = 1$ ，刚好只剩一个自由度。

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）通过如下方式相联系。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：

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

这样单位复数对应的 2×2 实数矩阵为⁵

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

可见，定义了复单位 $i \rightarrow$ 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。

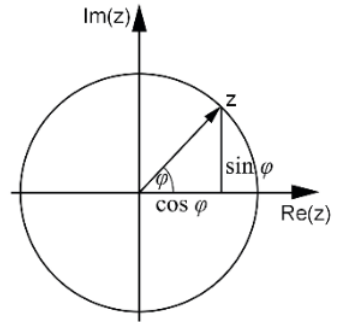


图 3.5: 单位复数在复平面的单位圆上

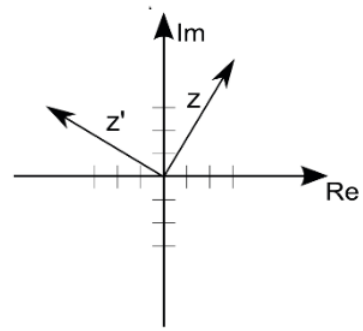


图 3.6: 复数通过乘以单位复数进行旋转

⁵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：

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

将旋转矩阵作用于 z ：

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

由上式可得

$$a' = a \cos \theta - b \sin \theta \quad (3.20)$$

$$b' = a \sin \theta + b \cos \theta \quad (3.21)$$

这与旋转矩阵作用于向量（列矩阵形式）相同：

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

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¹⁹。这一点太重要了，之后的篇幅会不断出现这种思想。

¹⁹ 如果映射 $\Pi : G \rightarrow G'$ 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。

²⁰ 其中蕴含着十分硬却失挺 (interesting) 的内容... 与二维旋转类似，我们将找出三维旋转的第二法，这个第二法能够揭示大自然更深刻的内容。

下面研究三维旋转，尝试找出它的两种描述方法²⁰。

3.3 Rotations in three Dimensions 三维旋转

三维向量旋转变换当然可以用 3×3 旋转矩阵⁶:

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

²¹ 基的概念见附录 A.1.

类比 $\mathcal{SO}(2)$ 群, 上面三个矩阵构成了 $\mathcal{SO}(3)$ 群的一组基²¹。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

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

²² 三维向量旋转的一般情况的推导见附录 A.2.

绕 z 轴旋转²²就是用旋转矩阵乘以向量:

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

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⁷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²³。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。

²³ 单位复数描述二维旋转。

⁶原文公式有误, 译文已按勘误表改正。

⁷实部虚部两个自由度

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

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1 \quad (3.25)$$

这样任意一个四元数 q 表示为

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

其中 $a, b, c, d \in \mathbb{R}$, $\mathbf{1}$ 就是正常的实数 1。⁸

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

$$\mathbf{ijk} = -1 \quad (3.27)$$

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} ：

$$\mathbf{ij} \underbrace{\mathbf{kk}}_{=-1} = -\mathbf{k} \rightarrow \mathbf{ij} = \mathbf{k} \quad (3.28)$$

单位四元数 $q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ 满足²⁴：

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

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

²⁴ 符号 \dagger 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。

⁸ $\mathbf{1}, \mathbf{i}, \mathbf{j}, \mathbf{k}$ 用黑体表示是为了强调它们是四元数的基。---译者 (SI)

一种方式（还有别的方式）是将它们与 2×2 复数矩阵对应：

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

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：

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

由上式可见

$$\det(q) = a^2 + b^2 + c^2 + d^2 \quad (3.32)$$

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：

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

注意3.30式中的矩阵是线性独立的²⁵，它们构成群 $SU(2)$ 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)²⁶，即 $\mathcal{U}^\dagger \mathcal{U} = 1$ 。任意单位四元数都唯一对应 $SU(2)$ 群中的一个群元。

$SU(2)$ 群如何与三维旋转联系？事情正在起变化， $SU(2)$ 群与 $SO(3)$ ²⁷ 之间的对应关系并不像二维旋转中 $U(1)$ 与 $SO(2)$ 那样简单。

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²⁸： $(Rz)^* R z = z^* R^* R z = z^* z$ 。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⁹：

$$v \equiv x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \quad (3.34)$$

²⁵ 线性独立的概念见附录 A.1.

²⁶ 进一步介绍见本章附录 —— ??节。

²⁷ $SO(3)$ 群就是三维旋转矩阵构成的群。

²⁸ 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。

⁹注意下式中的 \mathbf{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：

$$\det(v) = x^2 + y^2 + z^2 \quad (3.35)$$

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¹⁰。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：

$$v' = q^{-1}vq \quad (3.36)$$

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：

$$t = \cos \theta + u \sin \theta \quad (3.37)$$

利用3.34式，任意三维向量 \vec{v} 可表示为：

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

¹⁰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。

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

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

上式也可用欧拉公式写出²⁹：

²⁹ 欧拉公式的推导见附录 B.4.2

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

四元数旋转矩阵的逆矩阵为：

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

根据3.36式旋转向量 v ；

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

另一方面，向量 v' 用四元数表示为：

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

上式与3.43式比较可得：

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

所以

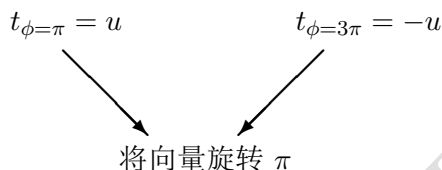
$$\rightarrow \vec{v}' = (\cos(2\theta), -\sin(2\theta), 0)^T \quad (3.46)$$

³⁰ 见3.24式，那里我们用三维旋转矩阵旋转向量。

由上式可见从 \vec{v} 到 \vec{v}' 确实进行了旋转³⁰，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：

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

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³¹：



³¹ 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 $SU(2)$ 群称为 $SO(3)$ 群的双覆盖。一个 $SU(2)$ 中的群元对应 $SO(3)$ 的哪个群元是很明确的，反过来就不行，因为 $SO(3)$ 中的一个群元对应 $SU(2)$ 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³²。

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³³，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 $SU(2)$ 群与四维旋转群确实存在千丝万缕的联系。

³² 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。

³³ 四维旋转矩阵 $O(4 \times 4)$ 有 16 个参数，条件 $O^T O = 1$ 与 $\det(O) = 1$ 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：

$$g(\epsilon) = I + \epsilon X \quad (3.48)$$

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对象上几乎啥也不变，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：

$$h(\theta) = \underbrace{(I + \epsilon X)(I + \epsilon X) \dots (I + \epsilon X)}_{k \uparrow (I + \epsilon X)} = (I + \epsilon X)^k \quad (3.49)$$

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：

$$g\left(\frac{\theta}{N}\right) = I + \frac{\theta}{N} X \quad (3.50)$$

要使上式表示的变换尽可能小，则让 N 尽可能大，令 $N \rightarrow \infty$ 。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：

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

微积分告诉我们这个极限就是指数函数³⁴：

³⁴3.51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。

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

上式有 X 生成了有限变换 $h(\theta)$ 的感觉, 因此 X 成为**生成元**。生成元的严格定义之后详述, 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³⁵。

$$h(\theta) = I + \frac{dh}{d\theta} \Big|_{\theta=0} \theta + \frac{1}{2} \frac{d^2 h}{d\theta^2} \Big|_{\theta=0} \theta^2 + \cdots = \sum_n \frac{1}{n!} \frac{d^n h}{d\theta^n} \Big|_{\theta=0} \theta^n \quad (3.53)$$

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³⁶:

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

这与之前的描述有联系。比较上式与3.52式可得:

$$X = \frac{dh}{d\theta} \Big|_{\theta=0} \quad (3.55)$$

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**¹¹ (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:

$$\left\{ X \mid \text{若 } e^X \in G \right\}$$

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³⁷:

$n \times n$ 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。告诉我们群元之间怎样结合。矩阵 Lie 群的群乘法就是矩阵乘法, 我们可能

¹¹注意生成元 X 并非群元, $I + \epsilon X$ 才是。

³⁵ 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.

³⁶ 附录 B.4.1 中有推导

³⁷ 群 G 的 Lie 代数通常用哥特体活字 (Fraktur) 字母 \mathfrak{g} 表示。

naive 地认为 Lie 代数的元素之间采用相同的方式结合 (\circ), 但这并不正确! 诚然, (矩阵 Lie 群的) Lie 代数的元素都是矩阵³⁸, 但两个 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结合规则, 当然它与原来群的群乘法直接有关。

Lie 群乘法与 Lie 代数结合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³⁹:

$$e^X \circ e^Y = e^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]-\frac{1}{12}[Y,[X,Y]]+\dots} \quad (3.56)$$

其中⁴⁰ $X, Y \in \mathfrak{g}$, 即 X, Y 是群 G 的生成元。 e^X, e^Y 是 G 的群元, 把它们分别记作 g, h , 这样上式写成:

$$\underbrace{g}_{\in G} \circ \underbrace{h}_{\in G} = e^X \circ e^Y = \underbrace{e^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]-\frac{1}{12}[Y,[X,Y]]+\dots}}_{\in G} \quad (3.57)$$

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 $[\cdot, \cdot]$ 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⁴¹!

由 BCH 公式可知 Lie 代数元素间的乘法规则是 Lie 括号 $[\cdot, \cdot]$, 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⁴²。

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。

³⁸ 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。

³⁹ 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

⁴⁰ 重复一下, 群 G 的 Lie 代数通常用哥特体字母 \mathfrak{g} 表示。

⁴¹ John Stillwell. Naive Lie Theory. Springer, 1st edition, August 2008a. ISBN 978-0387782140 里面有漂亮的证明。

⁴² 用数学语言表示封闭性: $\forall g, h \in G, g \circ h \in G$ 。对于 $X, Y \in \mathfrak{g}, [X, Y] \in \mathfrak{g}, X \circ Y \notin \mathfrak{g}$

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

$SO(3)$ 群的定义为 (3.10式):

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

任意群元 \mathcal{O} 用相应的生成元 J 表示为:

$$\mathcal{O} = e^{\phi J} \quad (3.59)$$

⁴³ $\text{tr}(A)$ 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 $\text{tr}(A) = A_{11} + A_{22}$ 。

⁴⁴ 线性无关的概念见附录 A.1.

上式带入群的第一个定义条件得:

$$\mathcal{O}^T \mathcal{O} = e^{\phi J^T} e^{\phi J} \stackrel{!}{=} 1 \quad \rightarrow \quad J^T + J \stackrel{!}{=} 0 \quad (3.60)$$

带入第二个定义条件, 再利用等式⁴³ $\det(e^A) = e^{\text{tr}(A)}$ 得:

$$\begin{aligned} \det(\mathcal{O}) \stackrel{!}{=} 1 &\quad \rightarrow \quad \det(e^{\phi J}) = e^{\phi \text{tr}(J)} \stackrel{!}{=} 1 \\ &\quad \rightarrow \text{tr}(J) \stackrel{!}{=} 0 \end{aligned} \quad (3.61)$$

满足3.60、3.61式的三个线性无关的⁴⁴矩阵为:

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

这三个矩阵构成 $SO(3)$ 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⁴⁵:

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

⁴⁵Levi-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：

$$\begin{aligned} (J_1)_{jk} = -\epsilon_{1jk} &\longleftrightarrow \begin{pmatrix} (J_1)_{11} & (J_1)_{12} & (J_1)_{13} \\ (J_1)_{21} & (J_1)_{22} & (J_1)_{23} \\ (J_1)_{31} & (J_1)_{32} & (J_1)_{33} \end{pmatrix} \\ &= - \begin{pmatrix} \epsilon_{111} & \epsilon_{112} & \epsilon_{113} \\ \epsilon_{121} & \epsilon_{122} & \epsilon_{123} \\ \epsilon_{131} & \epsilon_{132} & \epsilon_{133} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad (3.64) \end{aligned}$$

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 ⁴⁶。

⁴⁶ 其实子矩阵 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。

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

计算可得¹²

$$j_1^2 = -\mathbf{1} \quad (3.66)$$

接着：

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

一般情况为：

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

利用上式可计算 j_1 的指数函数的级数展开⁴⁷：

⁴⁷ 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。

¹²以下符号 $\mathbf{1}$ 、 I 均表示单位矩阵。— 译者

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

再利用 $e^0 = 1$ 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 $SO(3)$ 群的生成元显式的矩阵形式 (3.62式)，这样就能直接计算生成元之间的 Lie 括号了⁴⁸，不难得到⁴⁹：

$$[J_1, J_2] = J_1 J_2 - J_2 J_1 \quad [J_i, J_j] = \epsilon_{ijk} J_k \tag{3.71}$$

其中 ϵ_{ijk} 是 Levi-Civita 符号。

出于物理层面的考虑，在 $SO(3)$ 生成元 $e^{\phi J}$ 的幂附加一个虚单位 i ，即 $e^{i\phi J}$ ，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⁵⁰为：

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

⁴⁸ 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。

⁴⁹ 例如：

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

⁵⁰ 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。

将生成元加 i 是为了让它们是厄米生成元⁵¹，厄米生成元的含义是 $J^\dagger \equiv (J^*)^T \stackrel{!}{=} J$ 。加 i 是从物理意义考虑的：厄米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 ，生成元就是反厄米的了： $J^\dagger = (J^*)^T = -J$ ，反厄米矩阵的本征值是虚数。

还有另一种导出生成元的方法，利用 3.55 式 — $X = \frac{dh}{d\theta}|_{\theta=0}$ ，以及三维旋转矩阵的形式 (3.23、3.97 式) 可得：

$$\begin{aligned} J_1 &= \left. \frac{dR_1}{d\theta} \right|_{\theta=0} = \left. \frac{d}{d\theta} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix} \right|_{\theta=0} \\ &= \left. \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\sin \theta & -\cos \theta \\ 0 & \cos \theta & \sin \theta \end{pmatrix} \right|_{\theta=0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad (3.74) \end{aligned}$$

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¹³，群的重要部分 — 群乘法（规定群元之间怎样结合），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 $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ ，且二元运算 $[\cdot, \cdot]$ 满足如下公理：

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

¹³作者似乎很喜欢挖这种坑然后慢慢填。 — 译者

$$\begin{aligned} \text{例如 } J_1^* &= -i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \\ \text{而 } J_1^\dagger &= (J_1^*)^T = i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = J_1. \end{aligned}$$

- *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 $[\cdot, \cdot]$, 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¹⁴。

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 $SU(2)$ 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 $SO(3)$ 的生成元满足同样的 Lie 括号关系 (3.73 式)。这就是 $SU(2)$ 与 $SO(3)$ 群有相同 Lie 代数的含义。这个超级重要的结论将揭示 $SU(2)$ 与 $SO(3)$ 间许多不为人知的关系。

3.4.3 The Generators and Lie Algebra of $SU(2)$ $SU(2)$ 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 — $SU(2)$ 群, 并且发现 $SU(2)$ 是 $SO(3)$ 的双覆盖⁵²。

$SU(2)$ 群的群元是具有单位行列式⁵³的幺正的 2×2 矩阵, 即:

$$\mathcal{U}^\dagger \mathcal{U} = \mathcal{U} \mathcal{U}^\dagger = \mathbf{1} \quad (3.75)$$

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

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_1, J_2, \dots ¹⁵, 将群元用生成元表示, 带入上两式⁵⁴:

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

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

结合 3.77 式、BCH 公式 (3.56 式) 和 $[J_i, J_i] = 0$ ¹⁶:

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

¹⁴Poisson 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

¹⁵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

¹⁶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

⁵² 复习一下, 这意味着两个 $SU(2)$ 中的两个群元对应 $SO(3)$ 群中的同一元素。

⁵³ 具有单位行列式正是 ‘特殊’ (S) 的含义。

⁵⁴ 前面说过, 为了让生成元是厄米矩阵, 我们在 e 指数上加了 i , 这样可以确保量子力学预言的实验值为实数。

满足 $J_i^\dagger = J_i$ 的矩阵称为厄米矩阵，由上可见 $SU(2)$ 的生成元必须是厄米的。

利用恒等式 $\det(e^A) = e^{\text{tr}(A)}$ 与 3.78 式可得：

$$\det(e^{iJ_i}) = e^{i\text{tr}(J_i)} = 1 \xrightarrow[e^0=1]{\text{tr}(J_i)} \stackrel{!}{=} 0 \quad (3.80)$$

综上， $SU(2)$ 的生成元一定是无迹的厄米矩阵。 2×2 无迹厄米矩阵的一组基为⁵⁵：

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

也就是说任意 2×2 无迹厄米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：

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

其中 ϵ_{ijk} 是 Levi-Civita 符号。上式右侧的 2 很多余，因此通常将 $SU(2)$ 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：

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad (3.83)$$

注意这与 $SO(3)$ 的生成元形式 (3.73 式) 相同！因此我们说 $SU(2)$ 与 $SO(3)$ 有相同的 Lie 代数，因为 Lie 代数是通 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 $SU(2)$ 群对应的变换，即， $SU(2)$ 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 $SU(2)$ 的，(SU ‘2’ 嘛!) 换一种表示方法（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。

⁵⁵ 2×2 复数矩阵有 4 个复数元素，因此有 8 个自由度。无迹、厄米性的条件限制了只剩 3 个自由度。

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 $U(1)$ 群，它由全体单位复数构成。定义为 $z^*z = 1$ ，设 $z = a + ib$ ，则：

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

⁵⁶ 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 $S^1 = \{(x_1, x_2) | x_1^2 + x_2^2 = 1\}$

⁵⁷ 严谨的说法是：同构映射。数学上把两个对象‘相同’称为它们‘同构’。如果两个对象之间存在一个同构映射，那么它们就是同构的。

这也是单位圆的定义条件⁵⁶。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 $U(1)$ 与 $SO(2)$ 群间存在一一映射⁵⁷。因此这两个群可以与几何对象——单位圆——视为等同。群的抽象定义并非将这个群用 $SO(2)$ ，或是不同的 $U(1)$ 体现，因为它们都是由特定维数的对象定义的。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 $SO(2)$ 群表示这个变换（ 2×2 矩阵），也可用 $U(1)$ 群来表示（单位复数）。

下面讨论升级版—— $SO(3)$ 与 $SU(2)$ 。前面讲过在 $SU(2)$ 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 $q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ ：

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

⁵⁸ 单位圆 S^1 的定义是 $\{(x_1, x_2) | x_1^2 + x_2^2 = 1\}$ 。二维球面 S^2 的定义是 $\{(x_1, x_2, x_3) | x_1^2 + x_2^2 + x_3^2 = 1\}$ ，它是三维球体的表面。三维超球面 S^3 的定义是 $\{(x_1, x_2, x_3, x_4) | x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1\}$ 。 S 的上标表示维数。在二维空间中加上单位长度的条件之后得到了一维对象 S^1 ，相同地，在四维空间加上单位长度条件 $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$ 后得到三维对象 S^3 ，它是四维超球体的‘表面’，注意区分 S^3 与三维球体。

⁵⁹ 下面定义里的数学细节不必在意，本书用不着那些内容。但必须记住现代定义的灵魂：Lie 群 = 流形。

⁶⁰ 流形是满足某些条件的点集，例如球面就是一个流形，它局部看起来就像平坦 Euclidean 空间 \mathbb{R}^n 。或者， n 维流形就是满足如下条件的点的集合：任意一点可以对应 n 个独立坐标。本章末的??节有流形的详细说明。

这与三维超球面 S^3 的定义⁵⁸相同！上式给出了一个 $SU(2)$ 与 S^3 之间的映射，该映射是同构映射（是单射也是满射），因此我们可以把 $SU(2)$ 与三维超球面 S^3 等同。

结合上面的例子，Lie 群的现代定义就不难理解了⁵⁹：

Lie 群是一个群，也是一个微分流形⁶⁰。当然，这个流形要满足如下条件：

群乘法 \circ 诱导出的从流形到流形自身的映射必须是可微的¹⁷。这称为相容性条件，它保证了群定义与流形定义的兼容。例如，群 G 的任意群元 a 诱导出了从 G 到 G 的映射：给定 $\forall b \in G$ ， $c \equiv a \circ b \in G$ ，映射 $a : G \rightarrow G$ ¹⁸ 必须是可微的。

流形中的任意点都有相应的坐标，可以用坐标的语言表示上述内容： $c = ab$ 对应的坐标必须是 b 的坐标的可微函数。

从 Lie 代数的抽象定义可以看出 $SO(3)$ 与 $SU(2)$ 有相同的 Lie

¹⁷把函数看做一个映射，则映射的可微性就是函数的可微性。流形之间映射的可微性的一般定义可以见 微分几何入门与广义相对论，梁灿彬，上册，第二版，科学出版社，第二章定义 3。——译者

¹⁸或者记作 $a : b \in G \mapsto c \in G$ 。

代数 (3.83 式), 现在有了 Lie 群的抽象定义, 我们可以填在 3.4 节结尾处开的一个坑了:

“...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。”

为啥有一个特别的群? 因为这个特别的群是单连通的。单连通性是流形的性质, 它的含义是流形上的任意闭合曲线可以平滑地收缩为一点⁶¹。只有通过现代定义将群与流形联系起来才能找出特别的群。

把上面的重点重复一遍⁶²:

任一 Lie 代数仅对应一个单连通 Lie 群。

给定 Lie 代数, 它可对应许多 Lie 群, 而其中唯一的单连通 Lie 群可以视为其他群的‘母体’, 因为总存在从这个单连通群到其他群的映射, 反之则不一定。何妨把这个单连通群称为母群呢? 数学家们给它起了一个不太文艺的名字: 覆盖群。并且称该群覆盖了同一 Lie 代数对应的其他群。我们已经遭遇过这种情况: $SU(2)$ 群是 $SO(3)$ 群的双覆盖, 即两个 $SU(2)$ 元素对应一个 $SO(3)$ 的元素。

$SU(2)$ 就是三维超球面, 它正是一个单连通流形。因此 Lie 代数 3.83 式对应的覆盖群就是 $SU(2)$ 。从 $SU(2)$ 出发可以构造出其他所有对应的群。

$SO(3)$ 是什么流形? 已知 $SU(2)$ 到 $SO(3)$ 的映射将 $SU(2)$ 的两个点映射为 $SO(3)$ 中的一个。因此 $SO(3)$ 是单位超球面的一半。

因此从流形的角度看, $SU(2)$ 群比 $SO(3)$ 更完整, $SO(3)$ 只是其中的一部分。

本书的指导思想之一就是用最基本的群描述大自然最基本的对称性。具体到三维旋转的情况, 我们采用 $SU(2)$ 而非 $SO(3)$ 。之后其他的对称性我们都这么处理。

世界似乎就是这么设计的!¹⁹ 比如必须用 Poincare 群的双覆盖表示来描述基本粒子, 而常用却不完整的 4-向量表示就不行。要从最深刻的层面描述自然, 必须使用覆盖群, 而非从覆盖群导出的其他群。

给定一个 Lie 代数, 我们可以通过 Lie 代数的表示来导出它的覆盖群的表示⁶³, 然后把 Lie 代数元素 (生成元) 的矩阵表示扔进

¹⁹ 先不管设计者有没有、到底是谁吧

⁶¹ 本书不讨论这些抽象内容。感兴趣的读者可以阅读本章末尾的推荐教材。在这里我们知道一个 Lie 代数总是有一个特别的群就够了。

⁶² 这个命题的证明可见 Michael Spivak. *A Comprehensive Introduction to Differential Geometry*, Vol.1, 3rd Edition. Publish or Perish, 3rd edition, 1 1999. ISBN 9780914098706

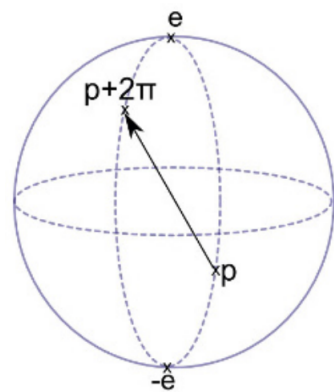


图 3.7: 三维超球面 S^3 的一个二维切面 (S^3 自身是四维超球的‘表面’, 愚蠢的人类没法直接把它画出来), 图中可见球面的上半部分对应 $SO(3)$, 因为从 $SU(2)$ 到 $SO(3)$ 的映射将前者中的 p 点与 $p + 2\pi$ 映射到后者中的同一点。

⁶³ ‘表示’的具体含义下一节就讲。

e 指数里面就能得到任意群元的矩阵表示。

这就是 Lie 理论的威力所在。人类通过纯数学就能揭示自然的某些深刻规律。狭义相对论的标准对称群隐藏了一些真相⁶⁴，因为它（尽管常用）不是这一对称性对应的最基本的覆盖群。Poincare 群⁶⁵的覆盖群才是最基本的，我们用它描述自然。

总结一下⁶⁶：

- $S^1 \cong U(1) \xleftrightarrow{\text{一一对应}} SO(2)$
- $S^3 \cong SU(2) \xleftrightarrow{\text{二对一}} SO(3) \cong S^3$ 的一半
 $\Rightarrow SU(2)$ 是 Lie 代数 $[J_i, J_j] = i\epsilon_{ijk} J_k$ (3.83式) 的覆盖群，
 因为 S^3 是单连通流形。

下一节介绍 Lie 理论的重要分支 — 表示论。表示论使我们能够从任意 Lie 群导出相应基本群。

3.5 Representation Theory 表示论

群论可以描述变换本身而不必联系现实中的特定对象，这是群论的重要特征。

理论上经常从抽象观点来考虑一个群。这意味着用流形结构与群乘法定义群。例如 $SU(2)$ 就是 S^3 (超球面)，群元就是流形中的点，群乘法规定两个群元 a, b 怎样结合，且满足群公理。搞物理的更关心一个群到底能做什么，即群作用。

同一个群能作用于许多不同种类的东西⁶⁷，这一重要事实使我们定义了表示的概念：群 G 的一个表示就是从 G 到某个向量空间 V 上全体线性变换组成的集合⁶⁸的映射，将表示记作 R ，映射

$$R : \underbrace{g}_{\in G} \mapsto R(g) \quad (3.86)$$

其中 $R(g)$ 表示 V 上的某个线性变换。映射 R 必须满足如下条件⁶⁹：

- $R(e) = I$, 恒等元对应的线性变换是恒等变换。
- $R(g^{-1}) = (R(g))^{-1}$, 逆元对应的变换是逆变换。
- $R(g) \circ R(h) = R(g \circ h)$, 群元 g, h 结合对应的变换等于他们分别对应变换的结合。

群的表示⁷⁰将抽象的群元（群流形上的点）与向量空间中的线性变换相联系。尽管按定义来讲，表示是个映射，但大多数情况把

⁶⁴ 学过量子力学的人可能听说过：标准对称群不能告诉我们自旋！

⁶⁵ 简明起见，我们省略了“双覆盖”或“覆盖群”的说法，我们利用 Poincare 群的一个表示导出相应的 Lie 代数，然后利用此 Lie 代数导出相应覆盖群的表示，也就是 Poincare 双覆盖群的表示。

⁶⁶ S^2 未出现，没有与 S^2 对应的 Lie 群，因为没有三维复数。我们从二维复数 i 直接跳到了四维的四元数 i, j, k 。

⁶⁷ 这话的含义稍后会更加清楚。

⁶⁸ 本书中 V 上的线性变换其实就是指矩阵，即表示 R 把群元映射为矩阵，矩阵通过矩阵乘法作用于向量空间中的向量。

⁶⁹ 满足这些条件的映射称为同态映射，同构映射当然是同态的，并且比同态性还多了一一映射的特征。

⁷⁰ 表示的定义可以进行推广，不一定非得是向量空间（别的空间也行）上的线性变换（非线性也可能），拓展后的映射叫 **realization**。物理主要考虑向量空间中的线性变换（例如量子力学的 Heilbert 空间、狭义相对论的 Minkowski 空间），因此对物理而言表示比更广泛的 realization 更重要。

表示看做许多矩阵的集合就行了。例如，向量空间⁷¹ \mathbb{R}^3 中的旋转矩阵组成的集合就是 $\mathcal{SO}(3)$ 群的一个表示，旋转矩阵就是 \mathbb{R}^3 上的线性变换。注意 $\mathcal{SO}(3)$ 群对应的变换也可以作用到其他向量空间。

表示论让我们能系统研究同一个群对不同向量空间的作用，这让群论变得精彩多了。

$SU(2)$ 是物理使用最多的群之一。它可以作用到许多向量空间上，比如作用于一维复数向量空间 C^1 ，稍后会看到结果是很平凡的。再比如 C^2 ，下面我们会详细探究， C^2 的元素是二维复数向量，因此 $SU(2)$ 的群元以 2×2 复数矩阵的形式出现，这是 $SU(2)$ 最常用的表示，之前一直采用此表示。 $SU(2)$ 还可作用到 C^3 ，前人已经对此有完备的研究， $SU(3)$ 以 3×3 矩阵出现，作用于 3 维向量。剧透： $SU(3)$ 的 C^3 版本的生成元是⁷²：

$$J_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad J_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.87)$$

把生成元扔进 e 指数里就可得到 $SU(2)$ 的群元在这个表示之下对应的矩阵的形式。

上面的路可以一直走到 C^n 。因此作者纳闷人们为啥叫把这个群叫 $SU(2)$ 而非 S^3 。⁷³但 $SU(2)$ 通常定义为满足如下条件的 2×2 (是 2 哟!) 复数矩阵集合 (即 3.33 式)

$$U^\dagger U = \mathbf{1}, \quad \det(U) = 1 \quad (3.88)$$

下面会把 $SU(2)$ 的群元写成 3×3 矩阵的形式，因此要牢记当谈论一般情况 $SU(2)$ 或其他群群元的时候，我们指的是抽象的群流形中的点，而非只是 2×2 矩阵或者别的。

本章前面都是从群的某个表示出发定义一个群的。比如用 2×2 矩阵定义 $SU(2)$ ，此方法的优点是群的含义十分具体。在入门之后，把群视为抽象流形的观点⁷⁴更有效，因为这样可以更方便地找出其他有用的群表示。

下面定义几个之后要用的概念，它们与表示的‘等级’有关 (不是所有表示都是基本的)。

⁷¹ \mathbb{R}^3 表示三维 Euclidean 空间，它的元素就是三维向量，附录 A.1 里面有很多例子。

⁷² 下一节就推倒 (无误) 它，这里只是为了说明就算是 $SU(2)$ 也有不得不守护的东西，也可以作用于三维向量。

⁷³ 在作者撰写本书手稿时就把这个群叫 S^3 ，考虑到这个非标准名字会对初学者造成困扰所以改了，因为主流教科书都用 $SU(2)$ 。

⁷⁴ 例如把 $SU(2)$ 视为 S^3

⁷⁵ 矩阵 S 可逆的含义是存在矩阵 T 使得 $ST = TS = \mathbf{1}$, S 的逆矩阵一般记作 S^{-1} 。

相似变换。给定任意矩阵 D 与可逆⁷⁵矩阵 S 。则可逆变换的形式为：

$$R \rightarrow R' = S^{-1}RS \quad (3.89)$$

可逆变换与表示的关系在于，给定群 G ，群的表示 $R(G)$ 以及任意可逆矩阵 S ，则 $S^{-1}RS$ 也是群表示。这可以从群的定义直接看出来： $\forall g_1, g_2 \in G$ ，表示 R 把群元映射为向量空间 V 上的线性变换 $R(g_1), R(g_2)$ 。群的定义要求：

$$R(g_1)R(g_2) = R(g_1g_2) \quad (3.90)$$

现在把表示进行相似变换：

$$S^{-1}R(g_1)\underbrace{SS^{-1}}_{=1}R(g_2)S = S^{-1}R(g_1)R(g_2)S = S^{-1}R(g_1g_2)S \quad (3.91)$$

上式可得 $S^{-1}RS$ 也是一个表示。如果我们有一个表示，那么就能利用任意可逆矩阵变出好多其他表示，新表示的形式可以更漂亮（比如让它是 diagonal 矩阵）。

第二个概念：不变子空间。给定群 G 在向量空间 V 上的表示 R ，如果 $V' \subset V$ 满足：

$$\forall v \in V', g \in G, \text{ 都有 } R(g)v \in V'$$

则 V' 称为 V 的不变子空间⁷⁶。

任意群元作用于 V' 中的任意向量所得向量仍然在 V' 内，这就可以在 V' 上定义群 G 的表示 R' 了：

$$R'(g)v = R(g)v, \quad \forall v \in V', g \in G \quad (3.92)$$

这种情况下，表示 R 就有一种不太称职的感觉。我们称原来的表示 R 不是原来的向量空间 V 的基本表示， R 称为子表示。

定义了上述概念之后就能引入超级重要的不可约表示。群 G 在向量空间 V 上的不可约表示 R 意味着除了 0 空间和 V 自身之外再无不变子空间。不可约表示是基本的，它们不能再分解为子表示。从一个不可约表示可导出其他所有表示。

⁷⁶ 自然 $v \in V$ ，子空间 $V' \subset V$ ，换言之 V' 的任意元素都是 V 的元素。

还可以从另一个角度看不可约表示：不可约表示不能通过相似变换变成分块对角矩阵形式。而可约表示能被相似变换变成分块对角形式。表示的可约性很关键，比如描述基本粒子要用不可约表示⁷⁷，基本粒子在变换下的行为由相应对称群的不可约表示描述。

一个群有许多表示⁷⁸，怎么知道哪个是描述自然的基本表示？Casimir 元可以给我们答案。Casimir 元（记作 C ）从 Lie 代数的元素 — 生成元导出，它的定义是，对任意生成元 X 都有

$$[C, X] = 0 \quad (3.93)$$

著名的 Schur 引理⁷⁹告诉我们，给定一个不可约群 $R: \mathfrak{g} \rightarrow GL(V)$ ²⁰，如果某个线性变换 $T: V \rightarrow V$ （即 $T \in GL(V)$ ）与所有 $R(g), \forall g \in G$ 对易²¹，则 T 必为恒等变换的常数倍。因此 Casimir 元是群表示的‘常数’，这些‘常数’的值自然地为群表示‘编号’⁸⁰。之后我们就从 Casimir 元数值最低的不可约表示开始。

群表示的定义与向量空间 V 有关。一个表示就是将抽象群到 V 的线性变换的映射。线性代数告诉我们以任意矩阵的本征向量为基可以张成一个向量空间。任意 Lie 群的生成元至少有一个可以通过相似变成对角矩阵⁸¹，然后这个对角生成元的本征向量就可作为向量空间的基。

下一节导出 $SU(2)$ 群的不可约表示。研究 $SU(2)$ 是因为 Lorentz 群的 Lie 代数是两份 $SU(2)$ 的 Lie 代数。Lorentz 群是 Poincare 群的一部分，本章由浅入深对他们逐一讨论。

3.6 $SU(2)$ 特殊么正群 $SU(2)$

我们在 3.4.3 小节中用具体的矩阵形式进行计算得到了 $SU(2)$ 生成元的 Lie 括号⁸²。通过这一点便能得到 $SU(2)$ 群的进一步表示。回到开头我们对 $SU(2)$ 群的表示，即 2×2 的么正矩阵，其行列式的值为 1，但这只是一个特例罢了。在进一步处理这个问题之前，我们可以想一想大概能得到什么样的表示。

⁷⁷ 要不用啥呢？

⁷⁸ 比如我们前面讨论了二维旋转变换的两种方法 — 单位复数、 2×2 矩阵。它们都是群流形 S^1 的表示。

⁷⁹ Schur 引理是群论的基本结论之一，它在任何一本群论书都有。

⁸⁰ 具体内容见下一节的例子。

⁸¹ 对角生成元的集合称为 Cartan 子代数，相应的集合元素称为 Cartan 生成元。它们在量子场论中大放异彩，因为 Cartan 生成元的本征值可以给基本粒子‘编号’。例如量子色动力学使用 $SU(3)$ 群，它有两个 Cartan 生成元。因此色动力学描述的粒子都有两个‘编号’。通常不用数字表示编号，而是用‘色’（红、蓝、绿...）来标注。再比如描述弱相互作用的 $SU(2)$ 群只有一个 Cartan 生成元，因此每个粒子只有一个标签 — Cartan 生成元的本征值。

⁸² 想当年我们就是用这东西抽象地定义了群的 Lie 代数。具体结果见 3.83 式

²⁰ $GL(V)$ 表示向量空间 V 上所有线性变换组成的集合。

²¹ 即 $\forall g \in G, [R(g), T] = 0$ 。

3.6.1 The Finite-dimensional Irreducible Representations of $SU(2)$

$SU(2)$ 群的有限维不可约表示

为了知道 $SU(2)$ 群有限维不可约表示的一些东西，我们用在 3.4.3 小节中出现的算符的线性组合来定义一个新的算符⁸³：

$$J_+ = \frac{1}{\sqrt{2}}(J_1 + iJ_2) \quad (3.94)$$

$$J_- = \frac{1}{\sqrt{2}}(J_1 - iJ_2) \quad (3.95)$$

用 3.83 式可以验证新定义的算符满足如下对易关系：

$$[J_3, J_{\pm}] = \pm J_{\pm} \quad (3.96)$$

$$[J_+, J_-] = J_3 \quad (3.97)$$

若 v 为 J_3 本征矢，本征值为 b ，用新的算符作用在 v 上会得到很有意思的结果：

$$\begin{aligned} J_3(J_{\pm}v) &= J_3(J_{\pm}v) + \underbrace{J_{\pm}J_3v - J_3J_{\pm}v}_{=0} \\ &= \underbrace{J_{\pm}J_3v}_{=J_{\pm}bv} + \underbrace{J_3J_{\pm}v - J_{\pm}J_3v}_{=[J_3, J_{\pm}]v} \\ &\stackrel{\text{3.96式}}{=} (b \pm 1)J_{\pm}v \end{aligned} \quad (3.98)$$

⁸³ 即 $J_3v = bv$ ，详见附录 C.4。

可见 $J_{\pm}v$ 也是 J_3 的本征矢，本征值为 $(b \pm 1)$ ⁸⁴，记该本征矢为 w ，则：

$$J_3w = (b \pm 1)w \quad \text{以及} \quad w = J_{\pm}v \quad (3.99)$$

算符 J_- 和 J_+ 通常称为降算符和升算符，也称**阶梯算符 (ladder operators)**。我们可以通过重复利用阶梯算符 J_{\pm} 来得到关于 J_3 的本征矢，但这个过程不能无休止地进行，因为不同本征值的本征矢之间是线性独立的，而我们处理的是有限维的表示。这意味着相应的矢量空间是有限维的，我们只能得到有限个独立的矢量。

我们推断一定有一本征矢对应着最大的本征值 v_{\max} 。若将 J_+ 作用到 v N 次以后，得到本征值最大的本征矢 v_{\max}

$$v_{\max} = J_+^N v \quad (3.100)$$

我们有

$$J_+ v_{max} = 0 \quad (3.101)$$

因为 v_{max} 已经是有最大本征值的本征矢了。称最大本征值为 $j := b + N$ 。相应地有对应着最小本征值的本征矢 v_{min} ，满足

$$J_- v_{min} = 0 \quad (3.102)$$

将 J_- 作用到 v_{max} M 次以后，得到本征值最小的本征矢 v_{min}

$$v_{min} = J_-^M v_{max} \quad (3.103)$$

因此， v_{min} 对应的本征值为 $j - M$ 。为进一步讨论，我们需要知道 J_{\pm} 作用到本征矢上的具体形式。上述的计算说明， $J_- v_k$ 应等于 v_{k-1} 乘上一常数：

$$J_- v_k = \alpha_k v_{k-1} \quad (3.104)$$

如果非要深究 J_- 作用到 v_{max} 上的细节的话，那我只能告诉你⁸⁵ α 满足

$$\alpha_{j-k} = \frac{1}{\sqrt{2}} \sqrt{(2j-k)(k+1)} \quad (3.105)$$

注意到当 $k = 2j$ 时 α 为 0，因此，经过 $2j$ 步之后，我们从梯子的顶端到达了底端。 v_{min} 的本征值即是 $j - 2j = -j$ 。总的来说，我们得到了 $2j + 1$ 个本征态的本征值

$$-j, -j+1, \dots, j-1, j \quad (3.106)$$

此式仅当 j 为整数或半整数⁸⁶时成立。因为我们有 $2j + 1$ 个线性独立的本征矢，故矢量空间应为 $2j + 1$ 维⁸⁷。这些 J_3 的本征矢构成了完备的矢量空间，因为 J_1 和 J_2 能够被 J_+ 和 J_- 线性表示，就算是一个任意的线性组合 $\sum_i a_i v_i$ 也能变换成另一种线性组合 $\sum_i b_i v_i$ ，其中 a_i, b_i 是常数。因此， J_3 的本征矢构成了 V 的一个非平庸不变子空间，而我们也正在寻找能构成完备矢量空间 V 的不可约表示。

我们可以利用上面定义的 $2j + 1$ 维的矢量空间及 J_3 的本征矢

⁸⁵ 见 Matthew Robinson. Symmetry and the Standard Model. 第 90 页, Springer, 1st edition, August 2011. ISBN 978-1-4419-8267-4.

⁸⁶ 不信的话你就用分数试试看！

⁸⁷ 见此书 189 页: Nadir Jeevanjee. An Introduction to Tensors and Group Theory for Physicists. Birkhaeuser, 1st edition, August 2011. ISBN 978-0817647148

⁸⁸ 例如此书 190 页: Nadir Jeevanjee. An Introduction to Tensors and Group Theory for Physicists. Birkhaeuser, 1st edition, August 2011. ISBN 978- 0817647148

⁸⁹ 回想一下, Casimir 算符指的是那些与群中每个生成元都对易的群元, 即 $[C, X] = 0$ 。

v_k 来给出 $SU(2)$ 的一个表示。进一步说, 我们能证明 $SU(2)$ 的每一种不可约表示都等价于这一种⁸⁸。

3.6.2 The Casimir Operator of $SU(2)$ $SU(2)$ 的 Casimir 算符

如 3.5 节所述, 我们可以用 Casimir 算符来标记不同的群表示。 $SU(2)$ 恰好有一个 Casimir 算符⁸⁹:

$$J^2 := (J_1)^2 + (J_2)^2 + (J_3)^2 \quad (3.107)$$

满足定义的条件:

$$[J^2, J_i] = 0 \quad (3.108)$$

我们可以用 3.95 式和 3.94 式定义的 J_{\pm} 来重新表示 J^2 :

$$\begin{aligned} J^2 &= J_+ J_- + J_- J_+ + (J_3)^2 \\ &= \frac{1}{2}(J_1 + iJ_2)(J_1 - iJ_2) + \frac{1}{2}(J_1 - iJ_2)(J_1 + iJ_2) + (J_3)^2 \\ &= \frac{1}{2}((J_1)^2 - iJ_1 J_2 + iJ_2 J_1 + (J_2)^2) + \frac{1}{2}((J_1)^2 + iJ_1 J_2 - iJ_2 J_1 + (J_2)^2) \\ &\quad + (J_3)^2 \\ &= (J_1)^2 + (J_2)^2 + (J_3)^2 \end{aligned} \quad (3.109)$$

带入⁹⁰

$$J_+ v_k = \frac{1}{\sqrt{2}} \sqrt{(j+k+1)(j-k)} v_{k+1} \quad (3.110)$$

以及

$$J_- v_k = \frac{1}{\sqrt{2}} \sqrt{(j+k)(j-k+1)} v_{k-1} \quad (3.111)$$

⁹⁰ 这些东西只是一些归一化常数, 如果我们将 J_{\pm} 作用到归一态上, 得到的结果一般都不是归一化的。然而, 学物理的都喜欢用归一态, 详细的原因以后再谈。这些常数的推导有些无聊, 大体思路是先假设 c 是个归一化常数, 然后从 $J_{\pm} v_k = c v_{k\pm 1}$ 开始。计算过程能在大多数的量子力学教材中找到 (讲角动量和角动量算符那块儿)。如果你不知道也没关系, 因为这节讲的对接下来要讲的东西不是太重要。

我们就能计算对于每个表示都成立的系数：

$$\begin{aligned}
 J^2 v_k &= (J_+ J_- + J_- J_+ + (J_3)^2) v_k \\
 &= J_+ J_- v_k + J_- J_+ v_k + (J_3)^2 v_k \\
 &= J_+ \frac{1}{\sqrt{2}} \sqrt{(j+k)(j-k+1)} v_{k-1} + J_- \frac{1}{\sqrt{2}} \sqrt{(j+k+1)(j-k)} v_{k+1} + k^2 v_k \\
 &= \frac{1}{\sqrt{2}} \sqrt{(j+k)(j-k+1)} J_+ v_{k-1} + \frac{1}{\sqrt{2}} \sqrt{(j+k+1)(j-k)} J_- v_{k+1} + k^2 v_k \\
 &= \frac{1}{\sqrt{2}} \sqrt{(j+k)(j-k+1)} \frac{1}{\sqrt{2}} \sqrt{(j+(k-1)+1)(j-(k-1))} v_k \\
 &\quad + \frac{1}{\sqrt{2}} \sqrt{(j+k+1)(j-k)} \frac{1}{\sqrt{2}} \sqrt{(j+(k+1)+1)(j-(k+1)+1)} v_k + k^2 v_k \\
 &= \frac{1}{2} (j+k)(j-k+1) v_k + \frac{1}{2} (j-k)(j+k+1) v_k + k^2 v_k \\
 &= (j^2 + j) v_k = j(j+1) v_k
 \end{aligned} \tag{3.112}$$

现在会看这种表示下的几个特例，当然，从最简单的低维数情况开始。

3.6.3 The Representation of $SU(2)$ in one Dimension $SU(2)$ 的一维表示

j 的最小可能值为 0，这样群的表示是一维的。可以看出这个表示是平庸的，因为只有 1×1 矩阵满足 $SU(2)$ 李代数的对易关系 $[J_l, J_m] = i\epsilon_{lmn} J_n = 0$ 。而生成元 0 的指数映射得到的是变换 $U = e^0 = 1$ ，即群的恒等元。

3.6.4 The Representation of $SU(2)$ in two Dimension $SU(2)$ 的二维表示

接下来我们讨论 $j = \frac{1}{2}$ 的情况，这个表示是二维的。群的生成元 J_3 的本征值为 $\frac{1}{2}$ 和 $\frac{1}{2} - 1 = -\frac{1}{2}$ ，正如从 3.106 式中所见，因此 J_3 可表示为

$$J_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.113}$$

此处我们选择 J_3 为对角化的生成元⁹¹。对应其本征值 $+\frac{1}{2}$ 和 $-\frac{1}{2}$ 的本征矢为：

$$v_{\frac{1}{2}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{以及} \quad v_{-\frac{1}{2}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{3.114}$$

⁹¹ 由于对易关系， $SU(2)$ 群仅有一个对角化的生成元。还有就是记住我们可以用相似变换把一个不对角的生成元变成对角的。

可以用在这一基矢下的阶梯算符来得到 $SU(2)$ 另外两个生成元的矩阵形式

$$J_1 = \frac{1}{\sqrt{2}}(J_- + J_+) \quad (3.115)$$

$$J_2 = \frac{i}{\sqrt{2}}(J_- - J_+) \quad (3.116)$$

这一点可以直接从3.95式和3.94式关于 J_{\pm} 的定义的得到。回忆一下，表示空间中的基矢是由 J_3 的本征矢给出的，而我们也基矢给出用 $J_1 J_2$ 。换句话说， $J_1 J_2$ 是由它们作用在 J_3 的本征矢上的效果来定义的。例如

$$J_1 v_{\frac{1}{2}} = \frac{1}{\sqrt{2}}(J_- + J_+)v_{\frac{1}{2}} = \frac{1}{\sqrt{2}}(J_- v_{\frac{1}{2}} + \underbrace{J_+ v_{\frac{1}{2}}}_{=0}) = \frac{1}{\sqrt{2}}J_- v_{\frac{1}{2}} = \frac{1}{2}v_{-\frac{1}{2}} \quad (3.117)$$

式中用到 $J_+ v_{\frac{1}{2}} = 0$ ，因为 $\frac{1}{2}$ 是最大的本征值。最后结果的系数 $\frac{1}{2}$ 是从3.105式得到的。同理可得

$$J_1 v_{-\frac{1}{2}} = \frac{1}{\sqrt{2}}(J_- + J_+)v_{-\frac{1}{2}} = \frac{1}{2}v_{-\frac{1}{2}} \quad (3.118)$$

而我们之前得到的基矢是 $v_{\frac{1}{2}} = (1, 0)^T$ 和 $v_{-\frac{1}{2}} = (0, 1)^T$ ，因此 J_1 的矩阵为：

$$J_1 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (3.119)$$

⁹² 推导出 的3.117式为 $J_1 v_{\frac{1}{2}} = \frac{1}{2}v_{-\frac{1}{2}}$ 。带入这里具体的矩阵形式可得 $J_1 v_{\frac{1}{2}} =$

$$\frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2}v_{-\frac{1}{2}} \quad \checkmark。$$

⁹³ 再次说明， $SU(2)$ 最开始的定义是全体行列式值为 1 的 2×2 么正矩阵的集合，而在这儿我们采用抽象的定义，即其对应的流形 S^3 ，接下来还要讨论这个群的更高维表示，比如 3×3 矩阵的表示。其实叫成别的名字可能更合适（比如说对应的流形 S^3 ），但按照传统，它就叫 $SU(2)$ 。

可以验证上式满足我们所说的 J_1 作用到基矢上的效果⁹²。同理可得

$$J_2 = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad (3.120)$$

正如我们在这一章开头寻找 $SU(2)$ 群 Lie 代数时所发现的一样 (3.81式)，这些生成元 $J_i = \frac{1}{2}\sigma_i$ 正是 Pauli 矩阵。可以看出，现在使用表示的是二维表示。当然除此之外还有其他的表示形式，例如我们下节会看到的三维空间中的表示⁹³。

3.6.5 The Representation of $SU(2)$ in two Dimension $SU(2)$ 的三维表示

重复与二维情况相似的操作⁹⁴ 可以得到:

$$J_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad J_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad J_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.121)$$

这就是 $SU(2)$ 的三维表示。如果你有兴趣的话, 还可以推出三维表示下 $SU(2)$ 群元的形式 (把生成元扔进 e 指数里。)。介于现在所学已足够理解最重要的 Lorentz 群的表示, 我们将不会推导更高维的 $SU(2)$ 表示。

⁹⁴ 仍然是从对角化的生成元 J_3 出发。学过的人都知道 J_3 的本征值为 $(1, 0, -1)$, 因此立刻就能得到其矩阵形式。对于 $J_1 J_2$ 而言, 仍然是从其作用结果出发, 即把他们写成 J_{\pm} 的线性组合然后作用到 J_3 的本征矢上。

3.7 The Lorentz Group $\mathcal{O}(1, 3)$ Lorentz 群 $\mathcal{O}(1, 3)$

抽象总是从具体的现实中产生的 ... 你必须从某件事物出发, 最后再把所有现实的痕迹抹除掉。

— Pablo Picasso⁹⁵

⁹⁵ 引自 Robert S. Root-Bernstein and Michele M. Root-Bernstein. Sparks of Genius. Mariner Books, 1st edition, 8 2001. ISBN 9780618127450

本节从 Lorentz 群已知的一种表示出发导出相应的 Lie 代数。推导过程与之前 $SU(2)$ 的情形相同 — 用 $SU(2)$ 的 2×2 矩阵表示导出了 Lie 代数。我们会发现 Lorentz 群的 Lie 代数可以用两份 $SU(2)$ Lie 代数表示, 这会导出 Lorentz 群的更多表示, 著名的向量表示 — 即 Lorentz 群群元为 4×4 矩阵, 作用于 4 维向量 — 仅仅是众多表示中的一种。新表示让我们能发现物理系统中向量表示不能描述的许多现象。这就是 Lie 理论的威力所在, 它使我们找出对称性隐藏的抽象结构, 从而可以描述自然最基本的层面。

首先导出 Lorentz 群及其子群的特性。Lorentz 群定义为 Minkowski 空间中所有保证内积不变的变换组成的集合⁹⁶:

$$x^\mu x_\mu = x^\mu \eta_{\mu\nu} x^\nu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \quad (3.122)$$

⁹⁶ 下式的推导在第 2 章。重复一下, Lorentz 群的定义借鉴了 Euclidean 空间中旋转、反射变换保证 Euclidean 空间中内积不变的特性。

其中 $\eta_{\mu\nu}$ 表示 Minkowski 度规:

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (3.123)$$

这个定义的条件对变换 (矩阵) 有什么限制? Lorentz 群的群元 — Lorentz 变换通常记作 Λ , 考虑变换 $x^\mu \rightarrow x'^\mu = \Lambda^\mu_\nu x^\nu$, 变换前后的内积为:

$$x^\mu \eta_{\mu\nu} x^\nu \rightarrow x'^\sigma \eta_{\sigma\rho} x'^\rho = (x^\mu \Lambda^\sigma_\mu) \eta_{\sigma\rho} (\Lambda^\rho_\nu x^\nu) \stackrel{!}{=} x^\mu \eta_{\mu\nu} x^\nu \quad (3.124)$$

由 x^μ 的任意性可得:

$$\Lambda^\sigma_\mu \eta_{\sigma\rho} \Lambda^\rho_\nu \stackrel{!}{=} \eta_{\mu\nu} \quad (3.125)$$

⁹⁷ 复习: 两个向量的内积用矩阵乘法表示是, 左侧的向量转置成行矩阵。因此下式左侧有 Λ^T 。

上式写成矩阵形式⁹⁷:

$$\Lambda^T \eta \Lambda \stackrel{!}{=} \eta \quad (3.126)$$

上式就是 Lorentz 变换 Λ 定义的数学形式! 对上式取行列式, 利用 $\det(AB) = \det(A) \det(B)$ 可得:

$$\det(\Lambda) \underbrace{\det(\eta)}_{=-1} \det(\Lambda) = \underbrace{\det(\eta)}_{=-1} \rightarrow \det(\Lambda)^2 \stackrel{!}{=} 1 \quad (3.127)$$

$$\rightarrow \det(\Lambda) \stackrel{!}{=} \pm 1 \quad (3.128)$$

⁹⁸ 稍后就会看到这有什么用。

取3.125式的 $\mu = \nu = 0$ 分量⁹⁸:

$$\begin{aligned} \Lambda^\sigma_0 \eta_{\sigma\rho} \Lambda^\rho_0 &\stackrel{!}{=} \underbrace{\eta_{00}}_{=-1} \\ \rightarrow \Lambda^\sigma_0 \eta_{\sigma\rho} \Lambda^\rho_0 &= (\Lambda^0_0)^2 - \sum_i (\Lambda^i_0)^2 \stackrel{!}{=} 1 \end{aligned} \quad (3.129)$$

由此可得:

$$\Lambda^0_0 \stackrel{!}{=} \pm \sqrt{1 + \sum_i (\Lambda^i_0)^2} \quad (3.130)$$

根据3.128与3.130式的正负可以把 Lorentz 群分成 4 分支。保证坐标系取向⁹⁹不变的两个分支满足 $\det(\Lambda) = +1$ 。如果想保证时间轴的方向不变,则需要条件 $\Lambda_0^0 \geq 0$, 因为

$$x^0 = t \rightarrow x'^0 = t' = \Lambda_0^0 t + \Lambda_1^0 x^1 + \Lambda_2^0 x^2 + \Lambda_3^0 x^3 \quad (3.131)$$

如果 $\Lambda_0^0 \geq 0$, 则 t' 与 t 同号。

满足 $\det(\Lambda) \geq 0$ 与 $\Lambda_0^0 \geq 0$ 的分支称为 $\mathcal{SO}(1,3)^\uparrow$ 。我们主要研究这个子群, 它有漂亮的名字 — 正规¹⁰⁰或正时¹⁰¹Lorentz 群。Lorentz 群的这四分支是分离的, 因为从一个分支出发, 仅用该分支的 Lorentz 变换无法得到其他任何分支的元素。但引入两个新变换之后就可以从 $\mathcal{SO}(1,3)^\uparrow$ 出发得到其他分支¹⁰²:

$$\Lambda_P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (3.132)$$

$$\Lambda_T = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.133)$$

Λ_P 称为宇称算符, 相应的宇称变换其实就是镜面反射。 Λ_T 称为时间反演算符。

于是完整的 Lorentz 群 $\mathcal{O}(1,3)$ 可以表示为如下集合:

$$\mathcal{O}(1,3) = \{\mathcal{SO}(1,3)^\uparrow, \Lambda_P \mathcal{SO}(1,3)^\uparrow, \Lambda_T \mathcal{SO}(1,3)^\uparrow, \Lambda_P \Lambda_T \mathcal{SO}(1,3)^\uparrow\} \quad (3.134)$$

这样, 研究 Lorentz 群的表示就可简化为研究 $\mathcal{SO}(1,3)^\uparrow$ 的表示, 因为只要再找出 Λ_P 、 Λ_T 的表示, Lorentz 群的其他分支就都有啦。

⁹⁹ 这意味着右手坐标系在变换后仍为右手系, 左手坐标系在变换后仍为左手系。左手、右手坐标系的定义见附录 A.5.

¹⁰⁰ ‘正规’ (proper) 意思是 $\mathcal{SO}(1,3)^\uparrow$ 子群的群元都是保证坐标系取向和宇称不变的变换。

¹⁰¹ ‘正时’ (orthochronous) 是指子群的群元保持时间轴的方向不变。

¹⁰² 在现在的群表示下这两个算符的矩阵长这样。后面换用不同的表示, 它们会大不一样。

3.7.1 One Representation of the Lorentz Group Lorentz 群的 1 表示

¹⁰³ 通常狭义相对论涉及的向量空间是四维实 Minkowski 空间 $R^{(1,3)}$ 。Lorentz 群就是定义在这个空间的 (Lorentz 变换就是保证 4×4 度规不变的变换), 因此首先研究群在 $R^{(1,3)}$ 上的表示。类似地, $SU(2)$ 群起初是用 2×2 复数矩阵表示的, 从那里出发可以得到 $SU(2)$ 好多其他的表示。

¹⁰⁴ 自行验证 $\Lambda^T \eta \Lambda = \eta$ 对 Λ 的限制条件是 10 个。留做习题

下面利用 Lorentz 群的定义 (3.125 式) 构造 Lorentz 变换的显式矩阵表示。首先考虑矩阵的形式。对 Lorentz 群, 它的群元作用于 4 向量¹⁰³, 因此群元的形式为 4×4 实数矩阵。矩阵元必须是实数, 因为 Minkowski 空间 $R^{(1,3)}$ 是实数空间。一般的 4×4 实数矩阵有 16 个参数, 而 Lorentz 群的定义规定了 10 个条件¹⁰⁴, 这样还剩 6 个自由参数。也就是说描述最一般的 Lorentz 变换需要 6 个参数。因此只要集齐 6 个线性无关的生成元, 就可以召唤 Lorentz 群的 Lie 代数。这 6 个生成元~~一个~~基构成 Lie 代数的一组基, 即任意生成元可唯一地表示为它们的线性组合。然后就能计算这些基在 Lie 括号下的运算结果, 从而导出 Lie 代数的抽象定义。

注意三维 Euclidean 空间中的旋转矩阵 (它们只变换空间坐标而不管时间) 满足定义 3.125 式。这是因为 Minkowski 度规的空间部分¹⁰⁵与 3×3 单位矩阵成比例¹⁰⁶。因此对于只涉及空间坐标的变换, 由 3.125 可得:

$$\begin{aligned} -R^T I_{3 \times 3} R &= -R^T R \stackrel{!}{=} -I_{3 \times 3} \\ \rightarrow R^T I_{3 \times 3} R &= R^T R \stackrel{!}{=} I_{3 \times 3} \end{aligned}$$

上式正是 $O(3)$ 群的定义式。再加上条件

$$\det(\Lambda) \stackrel{!}{=} 1$$

就是 $SO(3)$ 群的定义。只涉及空间坐标的 Lorentz 变换的形式为:

$$\Lambda_{\text{rot}} = \begin{pmatrix} 1 & \\ & R_{3 \times 3} \end{pmatrix}$$

其中旋转矩阵 $R_{3 \times 3}$ 的形式见 3.23 式, 相关推导见 3.4.1 节。相应的生成元的推导与 3.4.1 节相似, 结果为:

$$J_i = \begin{pmatrix} 0 & \\ & J_i^{3\text{dim}} \end{pmatrix}, \quad i = 1, 2, 3. \quad (3.135)$$

¹⁰⁵ 空间部分是指分量 $\mu = 1, 2, 3$ 对应的部分。空间部分通常记作 η_{ij} , 人们习惯用拉丁字母 (例如 i, j) 表示从 1 到 3, 而希腊字母 (例如 μ, ν) 表示从 0 到 3。

¹⁰⁶ $\eta_{11} = \eta_{22} = \eta_{33} = -1, \eta_{ij} = 0, \text{若 } i \neq j$

比如, 由3.65式可得:

$$J_1 = \begin{pmatrix} 0 & & & \\ & J_1^{3\text{dim}} & & \\ & & & \\ & & & \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (3.136)$$

下面包含时空坐标的变换。按 Lie 理论的套路, 考虑无穷小变换¹⁰⁷:

$$\Lambda_\rho^\mu \approx \delta_\rho^\mu + \epsilon K_\rho^\mu \quad (3.137)$$

将上式带入 Λ 的定义条件3.125式:

$$\begin{aligned} \Lambda_\rho^\mu \eta_{\mu\nu} \Lambda_\sigma^\nu &\stackrel{!}{=} \eta_{\rho\sigma} \\ \rightarrow (\delta_\rho^\mu + \epsilon K_\rho^\mu) \eta_{\mu\nu} (\delta_\sigma^\nu + \epsilon K_\sigma^\nu) &\stackrel{!}{=} \eta_{\rho\sigma} \\ \rightarrow \cancel{\eta_{\rho\sigma}} + \epsilon K_\rho^\mu \eta_{\mu\sigma} + \epsilon K_\sigma^\nu \eta_{\rho\nu} + \underbrace{\epsilon^2 K_\rho^\mu \eta_{\mu\nu} K_\sigma^\nu}_{\text{略去高阶小量}\epsilon^2\text{项}} &= \cancel{\eta_{\rho\sigma}} \\ \rightarrow K_\rho^\mu \eta_{\mu\sigma} + K_\sigma^\nu \eta_{\rho\nu} &= 0 \end{aligned} \quad (3.138)$$

写成矩阵形式¹⁰⁸:

$$K^T \eta = -\eta K \quad (3.139)$$

上面导出了时空变换的生成元 K 需满足的条件。这些生成元生成的变换称为**推动**。推动意味着变换到相对于初始坐标系匀速直线运动的另一坐标系。比如从一个相对于观者静止的坐标系变换到相对于观者匀速直线运动的坐标系。下面考虑2.1节中的例子: 沿 x 轴的推动变换。已知 $y' = y, z' = z$, 因此生成元的形式为:

$$K_x = \begin{pmatrix} \underbrace{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}_{\equiv K_x} & \\ & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix} \quad (3.140)$$

¹⁰⁷ Kronecker δ 符号定义为

$$\delta_\rho^\mu = \begin{cases} 1, & \mu = \rho \\ 0, & \mu \neq \rho \end{cases}$$

将 Kronecker δ 符号写成矩阵形式就是恒等矩阵 I 。

¹⁰⁸ 第一个下标表示行指标, 第二个是列指标。不过指标为一上一下的时候会混淆, 这时我们把上下标错开表示, 比如 $K_\rho^\mu \equiv K_\rho^\mu \rightarrow (K^T)^\mu_\rho = K_\rho^\mu$ 。矩阵乘法的定义是用行乘以列, 因此 $K_\sigma^\nu \eta_{\rho\nu} = \eta_{\rho\nu} K_\sigma^\nu$ 的含义为, 矩阵 η 的第 ρ 行乘以 K 的第 σ 列, 用矩阵表示为 ηK 。此外, $K_\rho^\mu \eta_{\mu\sigma} = K_\rho^\mu \eta_{\mu\sigma} = (K^T)_\rho^\mu \eta_{\mu\sigma}$, 为将上式表示为矩阵乘法而引入矩阵 K 的转置, 只有这样才能表示为行乘以列的形式: K^T 的第 ρ 行乘以 η 的第 σ 列, 上式用矩阵表示为 $K^T \eta$ 。用指标记号可以自由移项, 比如 K_ρ^μ 只是矩阵 K 得一个元素, 即一个数。

只需要考虑非零的 2×2 矩阵部分就行了, 3.139式简化为:

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = - \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

不难解得¹⁰⁹:

$$k_x = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\begin{aligned} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \\ & \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \text{以 及} \\ & - \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \\ & - \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \end{aligned}$$

因此沿 x 轴的推动变换的生成元为:

$$K_x = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (3.141)$$

同理可以求出沿 y 、 z 轴推动变换的生成元:

$$K_y = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad K_z = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad (3.142)$$

¹¹⁰ 生成元 K_x, K_y, K_z 都是厄米的, 即 $K_i^\dagger = K_i$ 。因此 e 指数里不加虚数单位 i , 因为加了之后就变成反厄米的了。

¹¹¹ 不难算出 $k_x^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, 然后 $k_x^4 = \mathbf{1}$ 等等等等。对于奇次幂有 $k_x^3 = k_x, k_x^5 = k_x, \dots$

根据 Lie 理论, 把生成元扔进 e 指数里就能得到有限大小的变换¹¹⁰:

$$\Lambda_x(\phi) = e^{\phi K_x}$$

还以 K_x 为例, 简明起见, 只考虑其中 exciting 的部分 — 2×2 矩阵 k_x (见3.140式)。将它的指数函数用级数展开式计算, 并利用¹¹¹ $k_x^2 = \mathbf{1}$:

$$\begin{aligned}
\Lambda_x(\phi) &= e^{\phi k_x} = \sum_{n=0}^{\infty} \frac{\phi^n k_x^n}{n!} = \sum_{n=0}^{\infty} \frac{\phi^{2n}}{(2n)!} \underbrace{k_x^{2n}}_{=1} + \sum_{n=0}^{\infty} \frac{\phi^{2n+1}}{(2n+1)!} \underbrace{k_x^{2n+1}}_{=k_x} \\
&= \left(\sum_{n=0}^{\infty} \frac{\phi^{2n}}{(2n)!} \right) I + \left(\sum_{n=0}^{\infty} \frac{\phi^{2n+1}}{(2n+1)!} \right) k_x = \cosh(\phi) I + \sinh(\phi) k_x \\
&= \begin{pmatrix} \cosh(\phi) & 0 \\ 0 & \cosh(\phi) \end{pmatrix} + \begin{pmatrix} 0 & \sinh(\phi) \\ \sinh(\phi) & 0 \end{pmatrix} = \begin{pmatrix} \cosh(\phi) & \sinh(\phi) \\ \sinh(\phi) & \cosh(\phi) \end{pmatrix}
\end{aligned} \tag{3.143}$$

计算过程与3.4.1的内容相似。这里的求和式里没出现 $(-1)^n$ ，因此结果不是 $\sin(\phi)$ 和 $\cos(\phi)$ ，而是双曲正弦 $\sinh(\phi)$ 和双曲余弦 $\cosh(\phi)$ 。沿 x 轴推动变换的完整 4×4 矩阵为：

$$\Lambda_x = \begin{pmatrix} \cosh(\phi) & \sinh(\phi) & 0 & 0 \\ \sinh(\phi) & \cosh(\phi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.144}$$

同理可导出沿 y, z 轴的推动变换矩阵：

$$\Lambda_y = \begin{pmatrix} \cosh(\phi) & 0 & \sinh(\phi) & 0 \\ 0 & 1 & 0 & 0 \\ \sinh(\phi) & 0 & \cosh(\phi) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{3.145}$$

$$\Lambda_z = \begin{pmatrix} \cosh(\phi) & 0 & 0 & \sinh(\phi) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh(\phi) & 0 & 0 & \cosh(\phi) \end{pmatrix} \tag{3.146}$$

任意的推动变换都可以分解为上述三个变换矩阵的乘积。

3.7.2 Generators of the Other Components of the Lorentz Group

Lorentz 群其他部分的生成元

为得到 Lorentz 群其他部分¹¹²生成元的形式，只需将宇称算符 Λ_P 和时间反演算符 Λ_T 作用于刚刚导出的生成元 J_i, K_i 上，用矩阵指标记法写为¹¹³：

¹¹² 复习：Lorentz 群可以分成 $\mathcal{O}(1, 3) = \{\mathcal{SO}(1, 3)^\uparrow, \Lambda_P \mathcal{SO}(1, 3)^\uparrow, \Lambda_T \mathcal{SO}(1, 3)^\uparrow, \Lambda_P \Lambda_T \mathcal{SO}(1, 3)^\uparrow\}$ ，上一节导出了 $\mathcal{SO}(1, 3)^\uparrow$ 的生成元。

¹¹³ 对 J, K 变换需要两个 Λ_P ，因为 J, K 都是二维矩阵，一维一个 Λ_P 。这就是算符矩阵在坐标变换下的一般形式，喵。

$$(\Lambda_P)_{\alpha'}^{\alpha}(\Lambda_P)_{\beta'}^{\beta}(J_i)^{\alpha'\beta'} \underbrace{\hat{=}}_{\text{变成矩阵形式}} \Lambda_P J_i (\Lambda_P)^T = J_i \hat{=} (J_i)^{\alpha\beta} \quad (3.147)$$

$$(\Lambda_P)_{\alpha'}^{\alpha}(\Lambda_P)_{\beta'}^{\beta}(K_i)^{\alpha'\beta'} \underbrace{\hat{=}}_{\text{变成矩阵形式}} \Lambda_P K_i (\Lambda_P)^T = -K_i \hat{=} -(K_i)^{\alpha\beta} \quad (3.148)$$

上两式可以用上节导出的 J, K 显式矩阵形式暴力计算，例如：

$$J_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \rightarrow J'_x = \Lambda_P J_x (\Lambda_P)^T = J_x \quad (3.149)$$

$$\begin{aligned} \Lambda_P J_x (\Lambda_P)^T &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}^T \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{aligned} \quad (3.150)$$

同理，

$$K_x = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow K'_x = \Lambda_P K_x (\Lambda_P)^T = -K_x \quad (3.151)$$

$$\begin{aligned} \Lambda_P K_x (\Lambda_P)^T &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}^T \\ &= - \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (3.152) \end{aligned}$$

综上，在宇称变换下：

$$J_i \xrightarrow[\Lambda_P]{} J_i, \quad K_i \xrightarrow[\Lambda_P]{} -K_i \quad (3.153)$$

之后会用得到它们，因为宇称算符在其他表示之下的形式可能不向现在的向量表示这样简单。时间反演算符 Λ_T 的推导过程相似，由于 Λ_T 只变换第一个分量（时间分量），因此它对 J_i 无影响，而只使得 K_i 有变化：

$$(\Lambda_T)_{\alpha'}^{\alpha} (\Lambda_T)_{\beta'}^{\beta} (J_i)^{\alpha' \beta'} \underbrace{\hat{=} \Lambda_T J_i (\Lambda_T)^T}_{\text{变成矩阵形式}} = J_i \hat{=} (J_i)^{\alpha \beta} \quad (3.154)$$

$$(\Lambda_T)_{\alpha'}^{\alpha} (\Lambda_T)_{\beta'}^{\beta} (K_i)^{\alpha' \beta'} \underbrace{\hat{=} \Lambda_T K_i (\Lambda_T)^T}_{\text{变成矩阵形式}} = -K_i \hat{=} - (K_i)^{\alpha \beta} \quad (3.155)$$

或者简写为：

$$J_i \xrightarrow[\Lambda_T]{} J_i, \quad K_i \xrightarrow[\Lambda_T]{} -K_i \quad (3.156)$$

3.7.3 The Lie Algebra of the Proper Orthochronous Lorentz Group

$\mathcal{O}(1,3)^\uparrow$ 群的 Lie 代数

利用上节导出的 $\mathcal{O}(1,3)^\uparrow$ 生成元的矩阵形式¹¹⁴可以直接计算相应的 Lie 代数¹¹⁵:

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad (3.157)$$

$$[J_i, K_j] = i\epsilon_{ijk}K_k \quad (3.158)$$

$$[K_i, K_j] = -i\epsilon_{ijk}J_k \quad (3.159)$$

其中 J_i 是旋转变换的生成元, K_i 是推动的生成元。一般形式的 Lorentz 变换可以表示为:

$$\Lambda = e^{iJ\theta + iK\Phi} \quad (3.160)$$

3.158式表明两种生成元 (J_i 与 K_i) 之间不对易。旋转生成元在对易运算下封闭¹¹⁶, 而推动生成元则不是这样¹¹⁷。从 J_i, K_i 出发可以定义在对易运算下封闭并且相互对易的新生成元²²:

$$N_i^\pm \equiv \frac{1}{2}(J_i \pm iK_i) \quad (3.161)$$

不难算出对易关系为:

$$[N_i^+, N_j^+] = i\epsilon_{ijk}N_k^+ \quad (3.162)$$

$$[N_i^-, N_j^-] = i\epsilon_{ijk}N_k^- \quad (3.163)$$

$$[N_i^+, N_j^-] = 0 \quad (3.164)$$

这正是 $SU(2)$ Lie 代数的形式! 我们发现了 $\mathcal{SO}(1,3)^\uparrow$ 群的 Lie 代数包含两份 $SU(2)$ 的 Lie 代数。

我们搞了个大新闻, 因为前面已经构造了 $SU(2)$ 的所有不可约表示。然而就像 $\mathcal{SO}(3)$ 那样, Lorentz 群不是单连通群¹¹⁸, Lie 理论告诉我们, 对于非单连通群, 不存在 Lie 代数的不可约表示和群的表示之间的一一映射¹¹⁹。通过推导 Lorentz 群 Lie 代数的不可约表示, 可以导出 Lorentz 群的覆盖群的不可约表示, 只要把相应的生成元扔进 e 指数里就行了。这样导出的一部分表示是 Lorentz 群的表示, 还有一些不是。这些额外的表示很有用处,

²²注意区分之后式子中的虚数单位 i 与指标 i , 我们对作者采用这样的符号很抱歉。

¹¹⁴ 推动变换的生成元见 3.141 式, 旋转变换的生成元见 3.62 式。

¹¹⁵Levi-Civita 符号 ϵ_{ijk} 的定义见附录 B.5.5.

¹¹⁶ 在对易运算下封闭意味着对易子 $[J_i, J_j] = J_i J_j - J_j J_i$ 还是一个旋转变换的生成元。3.157 体现了这一点。

¹¹⁷3.159 表明两个推动生成元 K_i, K_j 的对易子并非推动生成元, 而是个转动生成元。

¹¹⁸ 本书不证明这个命题, 证明过程很复杂并且与物理关系不大。

¹¹⁹ 这话理解起来有点难。前面提过, 一个 Lie 代数总对应一个特殊的群, 这个群的特殊之处在于它是单连通的。导出一个 Lie 代数的不可约表示之后, 再将 Lie 代数的元素 (即生成元) 扔进 e 指数里就能得到单连通群 (覆盖群) 的表示。只有对于单连通群, 前面说的一一映射才存在。

它们可以用来描述某些基本粒子。

简明起见，我们把之后要推导的 Lorentz 群 Lie 代数的表示，或者 Lorentz 群双覆盖的表示，全部简称为 Lorentz 群的表示。

$SU(2)$ 的每一不可约表示都可以用 $SU(2)$ 的 Casimir 算符对应的标量 j 来标记。因此 Lorentz 群覆盖群¹²⁰的不可约表示可以用两个数 j_1, j_2 （它们都是整数或半整数）来标记。下面依次研究 Lorentz 群的 (j_1, j_2) 表示，并利用前面导出的 $SU(2)$ 群的 $j_1, j_2 = 0, \frac{1}{2}, 1, \dots$ 表示的形式。

Lorentz 群的 Lie 代数可以用 $M_{\mu\nu}$ 表示为更紧凑的形式， $M_{\mu\nu}$ 的定义为：

$$J_i = \frac{1}{2}\epsilon_{ijk}M_{jk} \quad (3.165)$$

$$K_i = M_{0i} \quad (3.166)$$

这样 Lie 代数可表示为：

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\mu\rho}M_{\nu\sigma} - \eta_{\mu\sigma}M_{\nu\rho} - \eta_{\nu\rho}M_{\mu\sigma} + \eta_{\nu\sigma}M_{\mu\rho}) \quad (3.167)$$

下面从 Lorentz 群的 Lie 代数构造不可约表示，さ、いくぞ！

¹²⁰Lorentz 群的覆盖群记作 $SL(2, C)$ ，它是具有单位行列式（行列式为 1）的 2×2 复数矩阵的集合。 $SL(2, C) \rightarrow SO(1, 3)$ 之间的关系与之前讨论的 $SU(2) \rightarrow SO(3)$ 之间的关系类似。

3.7.4 The $(0, 0)$ Representation $(0, 0)$ 表示

$SU(2)$ 的最低阶表示是平凡的，因为 Lie 代数对应是 1 维的向量空间，因此生成元必须是 1×1 矩阵，满足对易关系的只有平凡的 0：

$$N_i^+ = N_i^- = 0 \rightarrow e^{N_i^+} = e^{N_i^-} = e^0 = 1 \quad (3.168)$$

因此 Lorentz 群的 $(0, 0)$ 表示作用于在 Lorentz 变换下不变的对象。 $(0, 0)$ 表示又称为 Lorentz 标量表示。

3.7.5 The $(\frac{1}{2}, 0)$ Representation $(\frac{1}{2}, 0)$ 表示

$(\frac{1}{2}, 0)$ 表示就是对 $SU(2)$ 的一份 Lie 代数 N_i^+ 使用 2 维表示¹²¹，另一份 Lie 代数 N_i^- 使用 1 维表示，即 $N_i^- = 0$ 。由 N_i^-

¹²¹ 复习：表示对应的向量空间的维数是 $2j + 1$ ，因此 $j = 1/2$ 表示对应 $2 \cdot \frac{1}{2} + 1 = 2$ 维向量空间。

的定义式3.161可得:

$$N_i^- = \frac{1}{2}(J_i - iK_i) = 0 \quad (3.169)$$

$$\rightarrow J_i = iK_i \quad (3.170)$$

此外, 根据3.6.4节中导出的 $SU(2)$ 的 2 维表示的形式:

$$N_i^+ = \frac{\sigma_i}{2} \quad (3.171)$$

其中 σ_i 表示 Pauli 矩阵, 其定义见3.81式。

另一方面:

$$N_i^+ \stackrel{\text{3.161式}}{=} \frac{1}{2}(J_i + iK_i) \stackrel{\text{3.170式}}{=} \frac{1}{2}(iK_i + iK_i) = iK_i \quad (3.172)$$

比较3.171与3.172式可得:

$$iK_i = \frac{\sigma_i}{2} \rightarrow K_i = \frac{\sigma_i}{2i} = \frac{i\sigma_i}{2i^2} = -\frac{i}{2}\sigma_i \quad (3.173)$$

$$\text{3.170式} \rightarrow J_i = iK_i = -\frac{i^2}{2}\sigma_i = \frac{1}{2}\sigma_i \quad (3.174)$$

于是在这一表示下, Lorentz 旋转变换为:

$$R_\theta = e^{i\vec{\theta}\vec{J}} = e^{j\vec{\theta}\frac{\vec{\sigma}}{2}} \quad (3.175)$$

Lorentz 推动变换为²³:

$$B_{\vec{\phi}} = e^{i\vec{\phi}\vec{K}} = e^{\vec{\phi}\frac{\vec{\sigma}}{2}} \quad (3.176)$$

将生成元扔进 e 指数函数里, 再将 e 指数函数用级数展开, 就可以算出 Lorentz 群元的表示。例如, 绕 x 轴的旋转变换为:

$$R_x(\theta) = e^{i\theta J_1} = e^{i\theta \frac{1}{2}\sigma_1} = 1 + \frac{i}{2}\theta\sigma_1 + \frac{1}{2!}\left(\frac{i}{2}\theta\sigma_1\right)^2 + \cdots + \frac{1}{n!}\left(\frac{i}{2}\theta\sigma_1\right)^n + \cdots \quad (3.177)$$

¹²² 具体过程与3.4.1中的推导十分相似。

将 σ_1 的显式形式 (3.81式) 代入, 并利用 $\sigma_1^2 = 1$ 可得¹²²:

²³原文 B_θ 在译文中修改为 $B_{\vec{\phi}}$ — 译者

$$\begin{aligned}
R_x(\theta) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{i}{2}\theta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \frac{1}{2}\left(\frac{\theta}{2}\right)^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \dots \\
&= \begin{pmatrix} \cos \frac{\theta}{2} & i \sin \frac{\theta}{2} \\ i \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix}
\end{aligned} \tag{3.178}$$

关于其他轴的旋转及推动变换可类似地导出。注意 Lorentz 变换在这儿用 2×2 复矩阵表示，它们当然不是作用在 Minkowski 空间的（Minkowski 空间的元素是 4 维向量）。Lorentz 变换在 $(\frac{1}{2}, 0)$ 表示下作用的 2 分量¹²³对象称为左手旋量¹²⁴：

$$\chi_L = \begin{pmatrix} (\chi_L)_1 \\ (\chi_L)_2 \end{pmatrix} \tag{3.179}$$

旋量在这里指的是 2 分量对象。左手旋量的另一种定义是在 Lorentz 变换下根据 Lorentz 群的 $(\frac{1}{2}, 0)$ 表示变换的对象。注意旋量并不只是 2 分量的向量，旋量有很多新的性质。比如 e 指数中的 $\frac{1}{2}$ 因子，这个因子使得旋量¹²⁵在旋转 2π 之后不再和原来相同，而多了一个负号。这个性质十分奇特，因为日常遇到的对象在转 2π 之后总是和原来一样。

“旋量可以说是 Lorentz 变换最基本的数学对象。”

— A. M. Steane¹²⁶

3.7.6 The $(0, \frac{1}{2})$ Representation $(0, \frac{1}{2})$ 表示

$(0, \frac{1}{2})$ 表示的构造过程与之前的 $(\frac{1}{2}, 0)$ 相似。这里 N_i^+ 采用 1 维表示，即 $N_i^+ = 0$ ， N_i^- 采用 2 维表示，即 $N_i^- = \frac{1}{2}\sigma_i$ 。容易认为 $(0, \frac{1}{2})$ 表示和 $(\frac{1}{2}, 0)$ 表示差不多，但并非如此！根据 3.161 式可得：

$$N_i^+ = \frac{1}{2}(J_i + iK_i) = 0 \tag{3.180}$$

$$\rightarrow J_i = -iK_i \tag{3.181}$$

注意上面的负号。 N_i^- 采用 $SU(2)$ 的二维表示（在 3.6.4 节推导）：

$$N_i^- = \frac{1}{2}\sigma_i = \frac{1}{2}(J_i - iK_i) \underbrace{=}_{\text{3.181式}} \frac{1}{2}(-iK_i - iK_i) = -iK_i \tag{3.182}$$

¹²³ 之后会看到这些二分量对象与自旋向上和自旋向下的状态有关。

¹²⁴ 这个名字在引入右手旋量之后的含义会更清楚。我们会看到宇称变换将左手旋量变换为右手旋量，反之亦然。旋量通常分为左手和右手的，它们的起源与螺旋性有关 — 螺旋性与手性不完全相同。重复一遍宇称算符的效果：将左手坐标系变换为右手坐标系，反之亦然。左手、右手旋量的名字就是从那儿来的。

¹²⁵ 旋量的更多性质可见，chapter 3.2 in J. J. Sakurai. Modern Quantum Mechanics. Addison Wesley, 1st edition, 9 1993. ISBN 9780201539295

¹²⁶ Andrew M. Steane. An introduction to spinors. ArXiv e-prints, December 2013

可以得到 $(0, \frac{1}{2})$ 表示的生成元为:

$$-iK_i = \frac{1}{2}\sigma_i \rightarrow K_i = -\frac{1}{2i}\sigma_i = \frac{i}{2}\sigma_i \quad (3.183)$$

再结合3.181式:

$$J_i = -iK_i = \frac{1}{2}\sigma_i \quad (3.184)$$

于是, 在这一表示之下, Lorentz 旋转变换为:

$$R_\theta = e^{i\theta\vec{J}} = e^{i\theta\frac{\vec{\sigma}}{2}} \quad (3.185)$$

Lorentz 推动变换为²⁴:

$$B_{\vec{\phi}} = e^{i\vec{\phi}\vec{K}} = e^{-\vec{\phi}\frac{\vec{\sigma}}{2}} \quad (3.186)$$

可见 $(0, \frac{1}{2})$ 表示的旋转变换与 $(\frac{1}{2}, 0)$ 一样, 但推动变换的 e 指数幂差一个负号。我们推测这两个表示作用的对象相似但不完全相同。Lorentz 群的 $(0, \frac{1}{2})$ 表示作用的对象称为**右手旋量**:

$$\chi_R = \begin{pmatrix} (\chi_R)^1 \\ (\chi_R)^2 \end{pmatrix} \quad (3.187)$$

左手/右手旋量统称为 Weyl 旋量。

3.7.7 Van der Waerden Notation Van der Waerden 符号

本节引入便于处理旋量的标记法。上面导出了按不同方式变换的两种旋量, 如何将它们区分? 它们确实不同, 但也并非毫无关系。实际上左手旋量 (按 $(\frac{1}{2}, 0)$ 表示变换) 与右手旋量 (按 $(0, \frac{1}{2})$ 表示变换) 之间存在联系。引入带点与不带点的指标来表示不同旋量 — 这一记法以发明者命名, 称为 Van der Waerden 符号。这能清楚地表示何种对象按何种方式变换。在完整的体系建立起来之后 Van der Waerden 符号的优点会更清楚。

左手旋量 χ_L 用不带点的下指标表示:

$$\chi_L = \chi_a \quad (3.188)$$

²⁴原文 B_θ 在译文中修改为 $B_{\vec{\phi}}$ — 译者

右手旋量用带点的上指标：

$$\chi_R = \chi^{\dot{a}} \quad (3.189)$$

下面引入“旋量度规”的概念。旋量度规使得右手旋量可以变换为左手旋量，反之亦然。旋量度规定义为¹²⁷：

$$\epsilon^{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (3.190)$$

不难验证它满足上述性质。再定义¹²⁸：

$$\chi_L^C \equiv \epsilon \chi_L^* \quad (3.191)$$

其中 $*$ 表示取复共轭。下面考虑变换得到的 χ_L^C 怎样在 Lorentz 变换下变换，之后会发现它正是一个右手旋量。右手旋量的主要特征是它在 Lorentz 变换下怎样变换，因此认为 χ_L^C 正是一个右手旋量。下面让推动变换作用到 χ_L^C ，利用：

$$(-\epsilon)(\epsilon) = \mathbf{1} \quad (3.192)$$

$$(\epsilon)\sigma_i^*(-\epsilon) = -\sigma_i \quad (3.193)$$

上式不难用 Pauli 矩阵 σ_i 的显式形式直接验证。变换过程为¹²⁹：

$$\begin{aligned} \chi_L^C &\rightarrow (\chi')_L^C = \epsilon(\chi')_L^* \\ &= \epsilon(e^{\frac{\vec{\phi}}{2}\vec{\sigma}}\chi_L)^* \\ &= \epsilon(e^{\frac{\vec{\phi}}{2}\vec{\sigma}} \underbrace{(-\epsilon)(\epsilon)}_{=1, \text{见3.192式}} \chi_L)^* \\ &= \underbrace{\epsilon(e^{\frac{\vec{\phi}}{2}\vec{\sigma}*}(-\epsilon))}_{\text{3.193式} \rightarrow e^{-\frac{\vec{\phi}}{2}\vec{\sigma}}} (\epsilon)\chi_L^* \\ &= e^{-\frac{\vec{\phi}}{2}\vec{\sigma}} \underbrace{\epsilon\chi_L^*}_{=\chi_L^C} \\ &= e^{-\frac{\vec{\phi}}{2}\vec{\sigma}} \chi_L^C \end{aligned} \quad (3.194)$$

由此可见 χ_L^C 按右手旋量的形式变换¹³⁰。上式第 5 行利用了 $e^{\frac{\vec{\phi}}{2}\vec{\sigma}}$ 的 Taylor 级数展开，并对每一项利用了 3.193 式。不难验证 $\epsilon(\chi_L^*)$ 在旋转变换下仍与原来的变换方式相同，理应如此，因此 χ_L 与

¹²⁷ 注意下式就是二维 Levi-Civita 符号，其定义见附录 B.5.5.

¹²⁸ 这里对 χ_L^C 做一简介。上标 C 表示电荷共轭性，?? 节具体说明它的含义。我们看到 ϵ 的作用结果是把对象的‘标签’翻转——将左手旋量变换为右手旋量。之后会发现 ϵ 可以翻转所有‘标签’，比如，电荷的正负。

¹²⁹ 下面 $\vec{\phi}\vec{\sigma} \equiv \sum_i \sigma_i \phi_i \underbrace{=}_{\text{求和约定}} \sigma_i \phi_i$ ， $\vec{\sigma}$ 并非表示“向量”，而只是为了方便而采用的缩写。

¹³⁰ 右手旋量在推动变换下的形式见 3.186 式： $B_{\vec{\phi}} = e^{i\vec{\phi}\vec{K}} = e^{-\vec{\phi}\frac{\vec{\sigma}}{2}}$ 。注意它与左手旋量的变换 (3.176 式) 的不同： $B_{\vec{\phi}} = e^{i\vec{\phi}\vec{K}} = e^{\vec{\phi}\frac{\vec{\sigma}}{2}}$

χ_R 在旋转变换形式相同:

$$\chi_L^C \rightarrow (\chi')_L^C = \epsilon(\chi')_L^* = \epsilon(e^{\frac{i\vec{\theta}}{2}\vec{\sigma}}\chi)_L^* = e^{\frac{i\vec{\theta}}{2}\vec{\sigma}}\epsilon(\chi_L)^* \quad (3.195)$$

还可以验证以下结果: ϵ 在任意 Lorentz 变换下不变; 将右手旋量变换为左手旋量需要用 $(-\epsilon)$ 。

仿照狭义相对论中的张量记号, 定义“旋量度规”可以升降旋量的指标:

$$\epsilon\chi_L \quad \underbrace{\quad}_{\text{用指标记号表示}} \quad \epsilon^{ac}\chi_c = \chi^a \quad (3.196)$$

其中重复的上下标表示求和 (即 Einstein 求和约定)。此外, 从上面可知如果要把 χ_R 变换为 χ_L 还需要取复共轭:

$$\chi_R = \epsilon\chi_L^* \quad (3.197)$$

用 Van der Waerden 符号, $\epsilon +$ 取复共轭 得到的右手旋量用带点的上指标表示:

$$\chi_R = \epsilon\chi_L^* = \chi^{\dot{a}} \quad (3.198)$$

再与指标记号相比较, 可以发现 χ_L 的复共轭可以用带点的下指标表示:

$$\chi_L^* = \chi_a^* = \chi_{\dot{a}} \quad (3.199)$$

同理, 右手旋量 χ_R 的复共轭可用不带点的上指标表示:

$$\chi_R^* = (\chi^{\dot{a}})^* = \chi^a \quad (3.200)$$

下面考虑 $\chi_{\dot{a}}$ 与 χ^a 在 Lorentz 变换下如何改变, 这非常重要, $\chi_{\dot{a}}$ 与 χ^a 可以构造在 Lorentz 变换下不变的某种旋量的组合形式¹³¹。

从左手旋量在 Lorentz 变换下的行为出发:

$$\chi_L = \chi_a \xrightarrow{\text{Lorentz 变换}} \chi'_a = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^b \chi_b \quad (3.201)$$

¹³¹ 这些旋量构造出的不变量极为重要, 因为我们要导出在所有惯性系下都成立的物理定律。这些不变量的重要性之后会不断体现。

考虑旋量 $\chi_{\dot{a}}$ 的变换:

$$\begin{aligned}\chi_L^* = \chi_a^* = \chi_{\dot{a}} \xrightarrow{\text{Lorentz 变换}} \chi'_{\dot{a}} = (\chi'_a)^* &= \left(\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^b \right)^* \chi_b^* \\ &= \left(e^{-i\vec{\theta}\frac{\vec{\sigma}^*}{2} + \vec{\phi}\frac{\vec{\sigma}^*}{2}} \right)_{\dot{a}}^{\dot{b}} \chi_{\dot{b}}\end{aligned}\quad (3.202)$$

类似地, 利用右手旋量的变换形式:

$$\chi_R \xrightarrow{\text{Lorentz 变换}} \chi'_R = \chi'^{\dot{a}} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2}} \right)^{\dot{a}}_{\dot{b}} \chi^{\dot{b}} \quad (3.203)$$

可以导出旋量 χ^a 的变换:

$$\begin{aligned}\chi_R^* = (\chi^{\dot{a}})^* = \chi^a \xrightarrow{\text{Lorentz 变换}} \chi'^a = (\chi'^{\dot{a}})^* &= \left(\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}} \right)^{\dot{a}}_{\dot{b}} \right)^* (\chi^{\dot{b}})^* \\ &= \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}} \right)^a_b \chi^b\end{aligned}\quad (3.204)$$

为了组合出神秘的 Lorentz 不变量, 还需要一件事: 考虑两个向量标量积的定义: $\vec{a} \cdot \vec{b} = \vec{a}^T \vec{b}$ 。仿照这个我们在旋量的标量积定义里也引入转置操作。注意上面新旋量变换式的 e 指数中出现了 Pauli 矩阵的复共轭 σ_i^* (例如 $e^{-i\vec{\theta}\frac{\vec{\sigma}^*}{2}}$), 再加上转置操作就变成取厄米共轭的形式: $\sigma_i^\dagger = (\sigma_i^*)^T$, 其中前面已经解释过符号 \dagger (称为 dagger) 的含义。所有 Pauli 矩阵的厄米共轭仍是 Pauli 矩阵自身, 即:

$$\sigma_i^\dagger = (\sigma_i^*)^T = \sigma_i \quad (3.205)$$

这从 Pauli 矩阵的形式 3.81 式容易导出。

比较 3.201 与 3.204 式, 再结合 3.205 式, 可以看出旋量 χ_a 的转置与旋量 χ^a 分别对应的变换形式正好抵消, 这意味着 $(\chi^a)^T \chi_a$

是 Lorentz 变换下的不变量（即在 Lorentz 变换下不变），因为：

$$\begin{aligned}
 (\chi^a)^T \chi_a &\xrightarrow{\text{Lorentz 变换}} \\
 (\chi'^a)^T \chi'_a &= \left(\left(e^{-i\vec{\theta}\frac{\vec{\sigma}^*}{2} - \vec{\phi}\frac{\vec{\sigma}^*}{2}} \right)_b^a \chi^b \right)^T \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c \chi_c \\
 &= (\chi^b)^T \left(\exp \left(-i\vec{\theta}\frac{(\vec{\sigma}^*)^T}{2} - \vec{\phi}\frac{(\vec{\sigma}^*)^T}{2} \right) \right)_b^a \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c \chi_c \\
 &\stackrel{\text{3.205式}}{=} (\chi^b)^T \underbrace{\left(e^{-i\vec{\theta}\frac{\vec{\sigma}^*}{2} - \vec{\phi}\frac{\vec{\sigma}^*}{2}} \right)_b^a \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c}_{=\delta_b^c} \chi_c \\
 &= (\chi^c)^T \chi_c \tag{3.206}
 \end{aligned}$$

类似地，将 $\chi_{\dot{a}}$ 与 $\chi^{\dot{a}}$ 适当组合，利用3.202、3.203式可以得出另一个不变量 $(\chi_{\dot{a}})^T \chi^{\dot{a}}$ 。但是形如 $(\chi^{\dot{a}})^T \chi_L$ 的项不是 Lorentz 变换下的不变量，因为：

$$\chi_R^T \chi_L = (\chi^{\dot{a}})^T \chi_a \rightarrow (\chi'^{\dot{a}})^T \chi'_a = \chi^{\dot{b}} \underbrace{\left(e^{i\vec{\theta}\frac{\vec{\sigma}^T}{2} - \vec{\phi}\frac{\vec{\sigma}^T}{2}} \right)_b^{\dot{a}} \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c}_{\neq \delta_b^c} \chi_c \tag{3.207}$$

可见这样的左手旋量与右手旋量结合的项并非 Lorentz 不变量。综上，Lorentz 不变量必须是一对同类上下标组合成的量¹³²。或者，不变量必须是右手旋量的厄米共轭左乘左手旋量： $\chi_R^\dagger \chi_L = (\chi_R^*)^T \chi_L \hat{=} (\chi^a)^T \chi_a$ ，或者左手旋量的厄米共轭左乘右手旋量： $\chi_L^\dagger \chi_R = (\chi_L^*)^T \chi_R \hat{=} (\chi_{\dot{a}})^T \chi^{\dot{a}}$ 。Lorentz 不变量是之后推导的自然界的普适定律的关键。

在上述基础上， ϵ^{ab} 的名字“旋量度规”就变得合情合理了，因为旋量间的内积（在 Lorentz 变换下不变，见3.206式）可以写为：

$$\chi_a^T \chi^a \stackrel{\text{3.196式}}{=} \chi_a^T \epsilon^{ab} \chi_b \tag{3.208}$$

与第二章 Minkowski 空间的内积（它是 Lorentz 变换的不变量）的定义式??相对比：

$$x_\mu y^\mu = x_\mu \eta^{\mu\nu} y_\nu \tag{3.209}$$

其中 $\eta^{\mu\nu}$ 是 Minkowski 度规。可见旋量度规与四维向量的 Minkowski

¹³² 即上下标都带点 \dot{a} 或都不带点 a

度规地位相同¹³³。

在定义了上述符号之后，不难写出旋量度规的下标形式：

$$\epsilon_{ab} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (3.210)$$

因为 $(-\epsilon)(e^{i\vec{\theta}\frac{\vec{\sigma}}{2}-\vec{\phi}\frac{\vec{\sigma}}{2}})_a^b \chi_R = \chi_L$ 。¹³⁴ 一般的 Lorentz 变换（转动 + 推动）通常用 Λ 表示，当 Λ 的指标带点，说明它作用于右手旋量，并由此推断相应的具体形式：

$$\chi_R \rightarrow \chi'_R = \chi'^{\dot{a}} = \Lambda_{\dot{b}}^{\dot{a}} \chi^{\dot{b}} = (e^{i\vec{\theta}\frac{\vec{\sigma}}{2}-\vec{\phi}\frac{\vec{\sigma}}{2}})_{\dot{b}}^{\dot{a}} \chi^{\dot{b}} \quad (3.211)$$

同理，左手旋量情形：

$$\chi_L \rightarrow \chi'_L = \chi'_a = \Lambda_a^b \chi_b = (e^{i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}})_a^b \chi_b \quad (3.212)$$

这样 $(\frac{1}{2}, 0)$ 与 $(0, \frac{1}{2})$ 表示下的 Lorentz 变换记作：

$$\Lambda_{(\frac{1}{2}, 0)} = (e^{i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}})_a^b \triangleq \Lambda_a^b \quad (3.213)$$

$$\Lambda_{(0, \frac{1}{2})} = (e^{i\vec{\theta}\frac{\vec{\sigma}}{2}-\vec{\phi}\frac{\vec{\sigma}}{2}})_a^b \triangleq \Lambda_{\dot{b}}^{\dot{a}} \quad (3.214)$$

Van der Waerden 符号非常便捷，它有效地体现了 χ_L 与 χ_R 之间的联系。左右旋量之间可以互相变换，因此必然需要一套能将它们联系起来的符号体系。

之后考虑下一个不可约表示，有一位故人正在那里久候。

3.7.8 The $(\frac{1}{2}, \frac{1}{2})$ Representation $(\frac{1}{2}, \frac{1}{2})$ 表示

在 $(\frac{1}{2}, \frac{1}{2})$ 表示之下，两份 $SU(2)$ Lie 代数 N_i^+, N_i^- 都采用 2 维表示。首先考虑此表示下的 Lorentz 变换作用的对象是什么。两份 Lie 代数之间无相互作用，因为 N_i^+, N_i^- 是对易的，即 $[N_i^+, N_i^-] = 0$ 。因此作用的对象（记作 v ）在两份 Lie 代数下各自独立变换。 v 有两个指标 $v_a^{\dot{b}}$ ，每个指标在独立的 $SU(2)$ Lie 代数下变换。 $v_a^{\dot{b}}$ 即采用了上节介绍的 Van der Waerden 符号。

$v_a^{\dot{b}}$ 的每个指标有两个取值 $(\frac{1}{2} \text{ 和 } -\frac{1}{2})$ ，²⁵ 因为每个表示都是 2 维的， v 共有 4 个分量。因此 v 可表示为 2×2 矩阵，但 v 也可以表示为四维向量，稍后会看到¹³⁵。

²⁵似乎指标应该为 1, 2? — 译者

¹³³ 为啥上式 — 四维向量的内积的定义式中既没有转置操作？其实上式既可以做向量方程也可视为分量方程，符号 x_μ 既代表向量本身又代表向量的分量，不幸的是这有时会造成混乱，比如现在。将上式看做分量方程，则无需转置，这对旋量内积也一样。然而转置运算不可或缺，为了避免遗忘，我们在旋量内积的定义式中加上上标 T 表示强调。对于常用的三维向量有更好的符号避免上述混乱：向量自身用 \vec{a} ，向量分量用 a_i 表示。

¹³⁴ 读者可自行检验，但这对之后的内容关系不大。

¹³⁵ 在之前的二维旋转中遇到过类似情况，二维旋转操作可以用单位复数作用到向量对应的复数表示，在定义了复数与实数矩阵间的映射后，旋转也可以用实矩阵作用于列向量来表示。

首先考虑复数矩阵，一个一般的 2×2 复数矩阵有 4 个复数元素，因此有 8 个自由参数。从上述可见只需要 4 个自由参数。任意复矩阵 M 可表示为厄米矩阵 $H (H^\dagger = H)$ 与反厄米矩阵 $A (A^\dagger = -A)$ 之和： $M = H + A$ 。 2×2 厄米矩阵和反厄米矩阵都有 4 个自由参数。此外，下面会证明厄米矩阵经 $(\frac{1}{2}, \frac{1}{2})$ 表示下的 Lorentz 变换后仍为厄米矩阵，反厄米矩阵也一样。这意味着厄米矩阵与反厄米矩阵分别组成的集合都是不变子空间（见 3.5 节），亦即，一般的 2×2 复矩阵上的表示是可约的。综上所述，下面研究 2×2 厄米矩阵（组成的向量空间）上的不可约表示。

¹³⁶ 它的含义是任意 2×2 厄米矩阵可以唯一地表示为基的线性组合： $a_0 \mathbf{1} + a_i \sigma_i$ 。

2×2 厄米矩阵的一组基¹³⁶是 Pauli 矩阵 $\sigma_i, i = 1, 2, 3$ 加上单位矩阵。

首先考虑 v_{ab} ，而非 v_a^b ，这样可以利用 3.81 式的 Pauli 矩阵的显式形式。注意 v_a^b, v_{ab} 之间可用旋量度规 ϵ^{bc} 相互变换，因此容易从 v_a^b 回到 v_{ab} 。

$$\text{记 } \sigma^0 = I_{2 \times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad ^{26}$$

$$v_{ab} = v_\nu \sigma_{ab}^\nu = v_0 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + v_1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + v_3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (3.215)$$

¹³⁷ 这只是基的选择问题，这里采用了 Pauli 矩阵 asflkjldk-fjslf（翻译无力。。。）This is really just a basis choice and here we choose the basis that gives us with our definition of the Pauli matrices, the transformation behaviour we derived earlier for vectors.

上面解释过，我们也可以使用¹³⁷ $v_a^b = v_\mu \sigma_{ac}^\mu \epsilon^{bc}$ ，这样的话选用的基就是 $(\tilde{\sigma}^b)^\mu = \sigma_{ac}^\mu \epsilon^{bc}$ 。一般的厄米矩阵表示为：

$$v_{ab} = \begin{pmatrix} v_0 + v_3 & v_1 - iv_2 \\ v_1 + iv_2 & v_0 - v_3 \end{pmatrix} \quad (3.216)$$

上节讲过，不同类的指标按不同方式变换，具体来说：带点的下标与不带点的下标变换方式不同。

下面研究 v_{ab} 在 Lorentz 变换下如何表现，利用不带点的下标

²⁶ 原文下式等号右侧的 v 应为下指标，译文已改正。

的变换式^{3.202}:

$$\begin{aligned}
 v \xrightarrow{\text{Lorentz 变换}} v' &= v'_{ab} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}^*}{2} + \vec{\phi}\frac{\vec{\sigma}^*}{2}} \right)_b^d \Big)^T \\
 &= \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}^\dagger}{2} + \vec{\phi}\frac{\vec{\sigma}^\dagger}{2}} \right)_b^d \\
 &\stackrel{\sigma^\dagger = \sigma}{=} \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^d
 \end{aligned} \tag{3.217}$$

前面说过, 厄米矩阵经 Lorentz 变换后仍为厄米的, 下面来验证¹³⁸:

$$\begin{aligned}
 \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^d &\rightarrow \left(\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^d \right)^\dagger \\
 &\stackrel{(ABC)^\dagger = C^\dagger B^\dagger A^\dagger}{=} \left(\left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^d \right)^\dagger v_{cd}^\dagger \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c \Big)^\dagger \\
 &= \left(e^{i\vec{\theta}\frac{\vec{\sigma}^\dagger}{2} + \vec{\phi}\frac{\vec{\sigma}^\dagger}{2}} \right)_b^d v_{cd}^\dagger \left(e^{-i\vec{\theta}\frac{\vec{\sigma}^\dagger}{2} + \vec{\phi}\frac{\vec{\sigma}^\dagger}{2}} \right)_a^c \\
 &\stackrel{\text{若 } v_{cd}^\dagger = v_{cd}}{=} \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^d v_{cd} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c
 \end{aligned} \tag{3.218}$$

一般形式变换的计算过程冗长无聊, 这里只推导一个特例¹³⁹, 考虑将 v 沿 z 轴做推动变换¹⁴⁰:

$$\begin{aligned}
 v_{ab} \xrightarrow{\text{沿 } z \text{ 轴推动}} v'_{ab} &= \left(e^{\phi\frac{\sigma_3}{2}} \right)_a^c v_{cd} \left(e^{\phi\frac{\sigma_3}{2}} \right)_b^d \\
 &= \begin{pmatrix} e^{\frac{\phi}{2}} & 0 \\ 0 & e^{-\frac{\phi}{2}} \end{pmatrix} \begin{pmatrix} v_0 + v_3 & v_1 - iv_2 \\ v_1 + iv_2 & v_0 - v_3 \end{pmatrix} \begin{pmatrix} e^{\frac{\phi}{2}} & 0 \\ 0 & e^{-\frac{\phi}{2}} \end{pmatrix} \\
 &= \begin{pmatrix} e^\phi(v_0 + v_3) & v_1 - iv_2 \\ v_1 + iv_2 & e^{-\phi}(v_0 - v_3) \end{pmatrix}
 \end{aligned} \tag{3.219}$$

其中利用了 σ_3 是对角矩阵¹⁴¹以及对角矩阵 e 指数的计算公式: $\begin{pmatrix} e^{A_{11}} & 0 \\ 0 & e^{A_{22}} \end{pmatrix}$ 。将变换后的 v' 用一般形式写出, 并与^{3.219}式进行

¹³⁸ 反厄米矩阵经 Lorentz 变换仍为反厄米矩阵的证明过程同理。只要把下面最后一步 $v_{cd}^\dagger = v_{cd}$ 改成 $v_{cd}^\dagger = -v_{cd}$ 。

¹³⁹ 一般过程的推导可见 Matthew Robinson. Symmetry and the Standard Model. Springer, 1st edition, August 2011. ISBN 978-1-4419-8267-4, 128 页。

¹⁴⁰ 沿 z 轴意味着 $\vec{\phi} = (0, 0, \phi)^T$ 。沿 z 轴的推动变换形式最简单, 因为只有 σ_3 是对角矩阵, 其他方向的推动变换的 e 指数计算较复杂。

¹⁴¹ $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

比较:

$$v'_{ab} = \begin{pmatrix} v'_0 + v'_3 & v'_1 - iv'_2 \\ v'_1 + iv'_2 & v'_0 - v'_3 \end{pmatrix} = \begin{pmatrix} e^\phi(v_0 + v_3) & v_1 - iv_2 \\ v_1 + iv_2 & e^{-\phi}(v_0 - v_3) \end{pmatrix}$$

¹⁴² 下式利用了双曲函数的定义, 即由 $\cosh \phi \equiv \frac{1}{2}(e^\phi + e^{-\phi})$, $\sinh \phi \equiv \frac{1}{2}(e^\phi - e^{-\phi})$ 得出 $e^{-\phi} = (\cosh \phi - \sinh \phi)$, $e^\phi = (\cosh \phi + \sinh \phi)$ 。当然不这么做也可以, 不过用双曲函数使得形式更加紧凑。

比较可得对象 v 的分量在沿 z 轴的推动变换下的形式为¹⁴²:

$$\begin{aligned} \rightarrow v'_0 + v'_3 &= e^\phi(v_0 + v_3) = (\cosh \phi + \sinh \phi)(v_0 + v_3) \\ \rightarrow v'_0 - v'_3 &= e^{-\phi}(v_0 - v_3) = (\cosh \phi - \sinh \phi)(v_0 - v_3) \end{aligned}$$

解得:

$$\begin{aligned} \rightarrow v'_0 &= \cosh(\phi)v_0 + \sinh(\phi)v_3 \\ \rightarrow v'_3 &= \sinh(\phi)v_0 + \cosh(\phi)v_3 \end{aligned} \quad (3.220)$$

¹⁴³ 见 3.146 式。

上面正是四维向量的 Lorentz 变换式¹⁴³:

$$\begin{aligned} \begin{pmatrix} v'_0 \\ v'_1 \\ v'_2 \\ v'_3 \end{pmatrix} &= \begin{pmatrix} \cosh \phi & 0 & 0 & \sinh \phi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh \phi & 0 & 0 & \cosh \phi \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \end{pmatrix} \\ &= \begin{pmatrix} \cosh(\phi)v_0 + \sinh(\phi)v_3 \\ v_1 \\ v_2 \\ \sinh(\phi)v_0 + \cosh(\phi)v_3 \end{pmatrix} \end{aligned} \quad (3.221)$$

上述结果对任意 Lorentz 变换都成立, 其他情形 (如转动) 由读者自行计算验证。由此可见 $(\frac{1}{2}, \frac{1}{2})$ 表示正是 4 向量表示。用向量形式可以简化计算, 因为用变换矩阵乘以向量比三个矩阵相乘要简单。人们熟悉的 4 向量其实与更基本的旋量有关, 一个 4 向量是二阶旋量, 即具有两个分量, 根据 Lorentz 群的 $(\frac{1}{2}, \frac{1}{2})$ 表示变换的量。此外, 4 向量并非描述最基本物理体系的量, 它们不够基本, 有一些系统它们无法描述。

¹⁴⁴ 不妨把二阶张量简单看成一个二阶矩阵 $M_{\mu\nu}$

有人说“旋量是向量的平方根”, 这话与“向量是二阶张量¹⁴⁴的平方根”相似, 一个二阶张量有两个向量指标, 一个向量有两个旋量指标。因此 Lorentz 变换作用的最基本的对象其实是旋量。

前面讲过 Lorentz 群有四部分。不同部分之间通过宇称算符与

时间反演算符联系¹⁴⁵，因此为了描述所有保证光速不变的变换，还需要考虑宇称和时间反演算符在各个表示下的形式。本书只考虑宇称变换，因为实验表明自然在宇称变换下并不总是对称的（之后的章节讨论），而且下一节有关宇称变换表示的推导也可应用于时间反演变换。

¹⁴⁵ 见3.134式。

3.7.9 Spinors and Parity 旋量和宇称

DRAFT TRANSLATION

DRAFT TRANSLATION

4. The Framework 框架/体系

这一章的基本思路是，我们要通过使某些东西取极小值，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 Lorentz 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)。通过使理论的中心对象取极小值，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欧拉-拉格朗日方程 (Euler-Lagrange equations)。

通过拉格朗日形式，可以得到物理中最重要的定理：Noether 定理 (Noether's Theorem)。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。要求拉格朗日量的积分，作用量 (action)，

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 $dQ/dt = 0 \rightarrow Q = \text{常数}$ 。

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，仅仅是协变能动矢量 (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

在某些对称变换下不变，即保证了体系的动力学遵从该对称性。

4.1.1 Fermat 原理

自然界中不论发生什么过程，由此造成作用量的改变总尽可能的小。

- Pierre de Maupertius ⁴

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

$$S_{\text{light}}[q(t)] = \int dt$$

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。

⁴Recherche des loix du mouvement (1746)

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

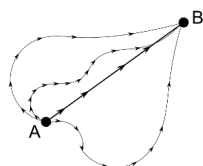


图 4.1: 对于给定初末端点的路径的扰动

4.1.2 变分法：基本思想

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 的一阶变分必须为零，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。

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

¹译注：此处讨论有误。可改作“否则若 ϵ 的一阶变分为正，我们可以...，反之亦然。”

由此我们找到极小点

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

它自然和我们求导 $f(x) = 3x^2 + x \rightarrow f'(x) = 6x + 1$ 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

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

其中 \mathcal{L} 一般是一个非常数的参量，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 $\mathcal{L} = \mathcal{L}(q(t), \frac{\partial}{\partial t}q(t))$ 。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。

4.2 Restrictions 限制

正如1.1节所提到的，现有理论中有些限制条件无法从第一性原理中得到。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。

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

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

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。

⁹ 这些困难称为 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-trivial. Trivial 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，这样就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

$$\Phi(x-h)\Phi(x) \quad (4.1)$$

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

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

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

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

一个著名的拉格朗日量是 $\mathcal{L} = \frac{1}{2} m \dot{\vec{q}}^2$ ，它能导出经典力学中的 Newton 运动方程。在稍后将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，已更正

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6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 \mathcal{L} ，因为在场论中大多数时候我们都在和拉格朗日量密度 \mathcal{L} 打交道。两者之间满足关系式 $\mathcal{L} = \int d^3x \mathcal{L}$

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

这里一个著名的拉格朗日量密度 $\mathcal{L} = \frac{1}{2} (\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2)$ ，它可以导出 Klein-Gordon 方程。

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

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

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，下面数章将详细讨论如何对问题中的系统导出正确的拉格朗日量 \mathcal{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，在给定初末端点后，粒子是如何在两点间运动的。数学上，即寻找函数 $q(t)$ 使得作用量⁸

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

取极值（极大值或极小值）。

使用记号

$$\dot{q}(t) = \frac{dq(t)}{dt}。$$

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译注：此处原文字体有误，已更正

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 $\dot{\epsilon}(t) = \frac{d\epsilon(t)}{dt}$ 。

在边界上, 扰动后的路径应与原路径相同:

$$0 = \epsilon(t_1) = \epsilon(t_2) \quad (4.5)$$

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

$$S = \int_{t_1}^{t_2} \mathcal{L}(q + \epsilon, \dot{q} + \dot{\epsilon}, t) dt.$$

与之前的例子类似, 极小值要求线性依赖于扰动 ϵ 的项为零。由于我们处理的是一般的 \mathcal{L} , 故将其展开成 Taylor 级数¹⁶, 并令一阶项为零

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

对后一项做分部积分¹⁷得到

$$\begin{aligned} & \int_{t_1}^{t_2} dt \left(\frac{d}{dt} \epsilon(t) \right) \frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} = \\ & \epsilon(t) \frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} dt \epsilon(t) \frac{d}{dt} \left(\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \right) \end{aligned}$$

利用4.5式有

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

因此, 我们可以将4.6改写为

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

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 我们使用的是这个公式的多变量形式, 参见附录??:

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

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。 $\frac{d}{dt}$ 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 $\frac{df}{dt} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial t} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial f}{\partial z} \frac{\partial z}{\partial t}$, 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 $f(x(t), y(t)) = x^2 y + y^3$, 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 $\frac{df}{dt} = 2xy \frac{\partial x}{\partial t} + (x^2 + 3y^2) \frac{\partial y}{\partial t}$ 。作为对比, 对于另一个函数 $g(x(t), y(t), t) = x^2 t + y$ 我们有 $\frac{\partial g}{\partial t} = x^2$

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

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, 注意到场论将时空同等对待。因此引入拉格朗日量密度:

$$\mathcal{L} = \int d^3\mathbf{x} \mathcal{L}(\Phi^i, \partial_\mu \Phi^i) \quad (4.8)$$

并用拉格朗日量密度来表示作用量

$$S = \int dt \mathcal{L} = \int d^4\mathbf{x} \mathcal{L}(\Phi^i, \partial_\mu \Phi^i) \quad (4.9)$$

依照上面讲过的步骤, 我们可以得到场的运动方程组¹⁹:

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

下一节中将在拉格朗日形式下导出近代物理学中重要的一个定理。从中可以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, 这个定理将教会我们如何和它打交道。

¹⁹ 在这里用“组”字是因为对于每一个场分量 Φ_1, Φ_2, \dots 我们都可以得到一个方程

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直接对应一个守恒量。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丽的创见之一。

4.5.1 粒子理论的 Noether 定理

先看看能对粒子理论中的守恒量做点什么。我们限于讨论连续对称性, 这样才能使用无穷小变换。如同上一章所言, 重复进行无穷小变换可以得到一个有限变换。拉格朗日量在无穷小变换²¹ $q \rightarrow q' = q + \delta q$ 下不变的性质用数学表述如下⁹

$$\begin{aligned} \delta \mathcal{L} &= \mathcal{L}(q, \frac{dq}{dt}, t) - \mathcal{L}(q + \delta q, \frac{d(q + \delta q)}{dt}, t) \\ &= \mathcal{L}(q, \frac{dq}{dt}, t) - \mathcal{L}(q + \delta q, \frac{dq}{dt} + \frac{d\delta q}{dt}, t) \stackrel{!}{=} 0 \end{aligned} \quad (4.11)$$

要求拉格朗日量是变换不变的, 这个限制太强了。为了保证动

²¹ 用符号 δ 来表述小扰动, 希望不会和用来表述偏导数的 ∂ 相混淆。

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

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：

$$\delta S = \int dt \mathcal{L}(q, \frac{dq}{dt}, t) - \int dt \mathcal{L}(q + \delta q, \frac{dq}{dt}, t) = \int dt \delta \mathcal{L} \stackrel{\text{若 } \delta \mathcal{L}=0}{=} 0 \quad (4.12)$$

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

$$\mathcal{L} \rightarrow \mathcal{L} + \frac{dG}{dt}$$

这时作用量不变²²

$$\delta S \rightarrow \delta S' = \delta S + \int_{t_1}^{t_2} dt \frac{d}{dt} \delta G = \delta S + \underbrace{\left. \frac{\partial G}{\partial q} \delta q \right|_{t_1}^{t_2}}_{=0 \text{ 因为 } \delta q(t_1)=\delta q(t_2)=0}.$$

最后一步中用到了扰动 δq 在初末时刻¹⁰ t_1, t_2 为零。由此可知，拉格朗日量的扰动 $\delta \mathcal{L}$ 不必非要为零，而仅需满足一个更弱的条件

$$\delta \mathcal{L} \stackrel{!}{=} \frac{dG}{dt}. \quad (4.13)$$

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 $\frac{dG}{dt}$ 。将4.11式的右端由 0 改作 $\frac{dG}{dt}$ ，我们有

$$\delta \mathcal{L} = \mathcal{L}(q, \frac{dq}{dt}, t) - \mathcal{L}(q + \delta q, \frac{dq}{dt} + \frac{d(\delta q)}{dt}, t) \stackrel{!}{=} \frac{dG}{dt} \quad (4.14)$$

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

$$\begin{aligned} \rightarrow \delta \mathcal{L} &= \mathcal{L} - \mathcal{L} - \frac{\partial \mathcal{L}}{\partial q} \delta q - \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} \stackrel{!}{=} \frac{dG}{dt} \\ \rightarrow \delta \mathcal{L} &= -\frac{\partial \mathcal{L}}{\partial q} \delta q - \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} \stackrel{!}{=} \frac{dG}{dt} \end{aligned} \quad (4.15)$$

²² 4.7式: $\frac{\partial \mathcal{L}}{\partial q} = \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right)$

利用欧拉-拉格朗日方程²³重写4.15式，有

$$\rightarrow \delta \mathcal{L} = -\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \delta q - \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} \stackrel{!}{=} \frac{dG}{dt}$$

¹⁰ 译注：右侧删去了原文的一对括号

利用莱布尼兹律²⁴，我们有

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

由此得到了一个不随时间变化的物理量 J ：

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

这是因为

$$\frac{dJ}{dt} = 0 \rightarrow J = \text{常数}.$$

为了看清楚这一点，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

$$m\ddot{q} = 0. \quad (4.18)$$

相应的拉格朗日量是²⁶

$$\mathcal{L} = \frac{1}{2} m \dot{q}^2 \quad (4.19)$$

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 $\mathcal{L} = \frac{1}{2} m \dot{q}^2$ 不依赖于 q ，故此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¹¹

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

其中 $\vec{p} = m \dot{q}$ 是经典力学中的动量 (momentum)。由于等式 $\frac{dJ}{dt}$ 对于任意 \vec{a} 都成立，所以说：拉格朗日量的空间平移不变性导致了动量守恒。

现在考虑旋转变换，这得需要大于一的维度，因为一维上的旋

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 $\frac{dq}{dt} = \dot{q}$ 是这个物体的速度， $\frac{d}{dt} \frac{dq}{dt} = \frac{d^2 q}{dt^2} = \ddot{q}$ 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 。

¹¹ 译注：求和约定，见第3章边注 129

转是没有意义的。下面将不会使用类似 \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 $\vec{q} \rightarrow q_i$ 改写。考虑无穷小旋转²⁷ $q_i \rightarrow q_i + \epsilon_{ijk} q_j a_k$ ，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

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

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：旋转不变性导致了角动量守恒。

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 $t \rightarrow t' = t + \epsilon$ 会导致

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

这表明一般而言 $\delta \mathcal{L} \neq 0$ ，但有 $G = -\mathcal{L}$ 。

将其代入4.16式，有¹²

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

守恒量 \mathcal{H} 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

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

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

²⁷ 这里我们用 Levi-Civita 符号来写旋生成元。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、今天还是五十年后做的实验，给定相同的初始条件，物理定律不会变化。

¹²译注：原文有字体混乱，已更正

$m\ddot{q} = -\frac{\partial V}{\partial q}$ ¹³ 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

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

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。¹⁴

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形形色色的物理理论中。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，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，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 $q \rightarrow q + vt$ 使动量 $m\dot{q} \rightarrow m(\dot{q} + v)$

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

¹³ 原文有误，已更正

¹⁴ 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，参见译注21

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 $A'_\mu = \begin{pmatrix} A'_0 \\ A'_1 \\ A'_2 \\ A'_3 \end{pmatrix} = \begin{pmatrix} A_0 \\ -A_2 \\ A_1 \\ A_3 \end{pmatrix}$ 。 A'_μ 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。

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。

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

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。

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 $x \rightarrow x'$ 和 $\Psi \rightarrow \Psi'$ 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 $\mathcal{L}(\Phi(x_\mu), \partial_\mu \Phi(x_\mu), x_\mu)$ ¹⁵。对称性意味着

$$\mathcal{L}(\Phi(x_\mu), \partial_\mu \Phi(x_\mu), x_\mu) = \mathcal{L}(\Phi'(x'_\mu), \partial_\mu \Phi'(x'_\mu), x'_\mu)。 \quad (4.26)$$

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

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

运用到拉格朗日量上有

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

用欧拉-拉格朗日方程³⁴将其改写为

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

其中 $\delta \Phi$ 有两项

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

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，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

¹⁵ 译注：原文多出一左括号

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

$$x_\mu \rightarrow x'_\mu = x_\mu + \delta x_\mu = x_\mu + a_\mu \quad (4.31)$$

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 ($\delta\mathcal{L} = 0$) 所导致的守恒量

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

$$\rightarrow -\partial_\nu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\nu\Phi)} \frac{\partial\Phi(x)}{\partial x^\mu} - \delta_\mu^\nu \mathcal{L} \right) \delta x^\mu = 0 \quad (4.33)$$

由4.31式我们有 $\delta x_\mu = a_\mu$ ，代入4.33式中可得¹⁶

$$-\partial_\nu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\nu\Phi)} \frac{\partial\Phi(x)}{\partial x^\mu} - \delta_\mu^\nu \mathcal{L} \right) \delta x^\mu = 0, \quad (4.34)$$

我们定义能动张量

$$T_\mu^\nu := \frac{\partial\mathcal{L}}{\partial(\partial_\nu\Phi)} \frac{\partial\Phi(x)}{\partial x^\mu} - \delta_\mu^\nu \mathcal{L}. \quad (4.35)$$

由于 a_μ 是任意的，4.34式告诉我们 T_μ^ν 满足连续性方程

$$\partial_\nu T_\mu^\nu = 0 \quad (4.36)$$

$$\rightarrow \partial_\nu T_\mu^\nu = \partial_0 T_\mu^0 - \partial_i T_\mu^i = 0 \quad (4.37)$$

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

$$\begin{aligned} \partial_0 T_0^0 - \partial_i T_0^i &= 0 \rightarrow \partial_0 T_0^0 = \partial_i T_0^i \\ \partial_t T_0^0 &= -\nabla \cdot \vec{T} \xrightarrow{\text{在某个无限的空间域 } V \text{ 上积分}} \int_V d^3\mathbf{x} \partial_t T_0^0 = - \int_V d^3\mathbf{x} \nabla \cdot \vec{T} \end{aligned}$$

¹⁶原文上下数式的指标不平衡

³⁵ 使用了 $\partial_0 = \partial_t$ 和 $\partial^i T_i^0 = \nabla \cdot \vec{T}$ ，以及著名的散度定理 $\int_V d^3\mathbf{x} \nabla \cdot \mathbf{A} = \int_{\partial V} d^2\mathbf{x} A$ ，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.feynmanlecture.s.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》)

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

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

最后一步中是这样处理的：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零¹⁷。

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

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

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

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。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3.7.4节的结论解释：我们用标量作为 Lorentz 群一维表示的生成元。故在第一部分中导出的将会是对应于标量场真实守恒量。

变换的第二项由下式给出

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

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。

¹⁷译注：这段讨论怪怪的

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。

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

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

我们能将它改写成更常见的形式

$$\begin{aligned} &= \frac{1}{2} (\partial^\nu T_\nu^\mu M_{\mu\sigma} x^\sigma + \partial^\nu T_\nu^\mu \underbrace{M_{\mu\sigma} x^\sigma}_{=-M_{\sigma\mu}}) = \frac{1}{2} (\partial^\nu T_\nu^\mu M_{\mu\sigma} x^\sigma + \partial^\nu T_\nu^\mu - M_{\sigma\mu} x^\sigma) \\ &\stackrel{\text{替换哑指标}}{=} \frac{1}{2} (\partial^\nu T_\nu^\mu M_{\mu\sigma} x^\sigma + \partial^\nu T_\nu^\sigma - M_{\mu\sigma} x^\mu) = \frac{1}{2} (\partial^\nu T_\nu^\mu x^\sigma + \partial^\nu T_\nu^\sigma x^\mu) M_{\mu\sigma} \\ &= \frac{1}{2} \partial_\nu (T^{\mu\nu} x^\sigma + \partial^\nu T^{\sigma\nu} x^\mu) M_{\mu\sigma} = 0 \\ &\rightarrow \text{引入 } (J^\nu)^{\sigma\mu} \equiv T^{\mu\nu} x^\sigma - \partial^\nu T^{\sigma\nu} x^\mu \rightarrow \partial_\nu (J^\nu)^{\sigma\mu} = 0. \end{aligned} \quad (4.43)$$

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 $M_{\mu\sigma} = -M_{\sigma\mu}$, 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 $M_{\mu\sigma}$ ¹⁸。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

$$Q^{\mu\sigma} = \int d^3\mathbf{x} (T^{\mu 0} x^\sigma - T^{\sigma 0} x^\mu) \quad (4.44)$$

由此我们得到了对应于转动³⁷不变性的守恒量

$$L_{\text{orbit}}^i = \frac{1}{2} \epsilon_{ijk} Q_{jk} = \frac{1}{2} \epsilon_{ijk} \int d^3\mathbf{x} (T^{\mu 0} x^\sigma - \partial^\nu T^{\sigma\nu} x^\mu), \quad (4.45)$$

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

$$Q^{0i} = \int d^3\mathbf{x} (T^{i0} x^0 - T^{00} x^i) \quad (4.46)$$

将会在附录 4.7 节中详细讨论。

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: $K_i = M_{0i}$

¹⁸ 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

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。现在考察

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

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

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

同时也使4.45式多了一项。完整的守恒量是

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

据此写出

$$L^i = L_{\text{spin}}^i + L_{\text{orbit}}^i. \quad (4.50)$$

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。

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

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

$$\Phi_i \rightarrow \Phi'_i = \Phi_i + \delta\Phi_i \quad (4.51)$$

下拉格朗日量的不变性 ($\delta\mathcal{L} = 0$), 数学上表达为¹⁹

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

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

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

代入 4.42 式, 我们得到

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

再用上莱布尼兹律, 得到

$$\partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\Phi_i)} \delta\Phi_i \right) = 0 \quad (4.53)$$

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

$$J^\mu \equiv \frac{\partial\mathcal{L}}{\partial(\partial_\mu\Phi)} \delta\Phi, \quad (4.54)$$

其满足连续性方程

$$\text{4.53 式} \rightarrow \partial_\mu J^\mu = 0. \quad (4.55)$$

更进一步, 用导出 4.39 式同样的办法, 能得到一个对时间的守恒量

$$\partial_t \int d^3x J^0 = 0. \quad (4.56)$$

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 $\delta\Phi = -i\epsilon$:

$$\Phi \rightarrow \Phi' = \Phi - i\epsilon. \quad (4.57)$$

¹⁹译注: 原文上下标混乱

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 $\epsilon' := i\epsilon$, 就能把它吸收到常数里去。

或者对于不止一个分量的场

$$\Phi_i \rightarrow \Phi'_i = \Phi_i - i\epsilon_i. \quad (4.58)$$

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4.53、4.56式, 并利用守恒量中不需要包含 $\delta\Phi = -i\epsilon$ 这件事。我们得到⁴⁵

$$\begin{aligned} \partial_t \Pi = 0 \rightarrow \partial_t \Pi = \partial_t \int d^3\mathbf{x} J_0 = \partial_t \int d^3\mathbf{x} \pi = 0 \\ \text{用到了 } J_0 = \pi = \frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} \end{aligned} \quad (4.59)$$

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

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

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 $P_i = \int d^3\mathbf{x} T_i^0$ 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 $L^i = \frac{1}{2}\epsilon_{ijk} \int d^3\mathbf{x} \left(\frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} S^{jk} \Phi(x) + (T^{k0} x^j - T^{j0} x^k) \right) = L_{\text{spin}}^i + L_{\text{orbit}}^i$, 包含自旋和轨道角动量, 守恒
- 时间平移不变性 \Rightarrow 能量 $E = \int d^3\mathbf{x} T_0^0$ 守恒
- 场自身的平移不变性 \Rightarrow 共轭动量 $\Pi = \int d^3\mathbf{x} \frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} = \int d^3\mathbf{x} \pi$ 守恒

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6章中看到这一点

⁴⁵ 4.54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

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的书。
- **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。

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

此附录中我们想找到对于示例拉格朗日量 $\mathcal{L} = \frac{1}{2}m\dot{q}^2$ 其来自于推动不变性的守恒量。一个推动变换即 $q \rightarrow q' = q + \delta q = q + vt$ ，其中 v 代表一个常值速率。该拉格朗日量在推动下并不是不变的，但正如之前就看到过的，这并不是一个问题：

为了使运动方程在变换下不变，拉格朗日量至多允许变化一个任意函数的全导数。这个条件在推动变换 $q \rightarrow q' = q + vt \Rightarrow \dot{q}' = \dot{q} + v$ $v=\text{常数!}$

下是显然满足的，即我们的拉格朗日量

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

的改变量为

$$\begin{aligned}\delta\mathcal{L} &= \mathcal{L}(q, \dot{q}, t) - \mathcal{L}(q', \dot{q}', t) = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}m\dot{q}'^2 \\ &= \frac{1}{2}m\dot{q}^2 - \frac{1}{2}m(\dot{q} + v)^2 = -m\dot{q}v - \frac{1}{2}mv^2,\end{aligned}\quad (4.60)$$

等于下面这个函数的全导数⁴⁹

$$G \equiv -m\dot{q}v - \frac{1}{2}mv^2t.$$

我们断言运动方程⁵⁰不变⁵¹，故我们找到了一个守恒量。

对于推动，守恒量是（4.16式）²⁰

$$J_{\text{boost}} = \underbrace{\frac{\partial\mathcal{L}}{\partial\dot{q}}}_{\equiv p} \underbrace{\delta q}_{=vt} + G = pvt - m\dot{q}v - \frac{1}{2}mv^2t,$$

由于 4.16 式右端为零，故我们能除去一个 v 的公因子²¹。由此有

$$\tilde{J}_{\text{boost}} = \underbrace{pt - \frac{1}{2}mv^2t}_{\equiv \tilde{p}t} - m\dot{q}v = \tilde{p}t - m\dot{q}v \quad (4.61)$$

这是一个奇怪的守恒量，因为它的值依赖于时间零点的选取。我

²⁰ 译注：原文中下式有符号错误

²¹ 译注：由于之前关于守恒量的诸多推导都使用了变换是无穷小变换这一条件，故这里 v 必须取无穷小量，即略去 J_{boost} 中 v 的二阶项，最后得到的才是守恒量。否则再利用其它条件我们会得到 $t \equiv 0$ 的荒谬结论。

⁴⁹ $\frac{dG}{dt} = \frac{d}{dt}(-m\dot{q}v - \frac{1}{2}mv^2t) = -m\dot{q}v - \frac{1}{2}mv^2$ ，因为 v 和 m 都是常数。

⁵⁰ 看到这一点的另一个方法是直接看运动方程： $m\ddot{q} = 0$ 。我们有变换 $q \rightarrow q' = q + \delta q = q + vt$ ，其中 v 是常数，且有 $\dot{q} \rightarrow \dot{q}' = \dot{q} + v$ 以及 $\ddot{q} \rightarrow \ddot{q}' = \ddot{q}$ 。

⁵¹ 拉格朗日量改变了，即 $\delta\mathcal{L} \neq 0$ ，但是作用量值最多改变一个全导数，故我们得到了相同的运动方程。

们能选择一个合适的时间零点使得这个量为零。由于它是守恒量，故这个零始终是零。

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

由场论中的推动不变性，我们有如下守恒量

$$Q^{0i} = \int d^3\mathbf{x} (T^{i0}x^0 - T^{00}x^i), \quad (4.62)$$

即在4.46式中所导出的，利用了 $K_i = M_{0i}$ 。

我们怎样诠释这个守恒量呢？ Q^{0i} 是守恒的，意味着

$$\begin{aligned} 0 = \frac{\partial Q^{0i}}{\partial t} &= \underbrace{\int d^3\mathbf{x} \frac{\partial T^{i0}}{\partial t} x^0}_{=x^0 \frac{\partial}{\partial t}} + \int d^3\mathbf{x} T^{i0} \underbrace{\frac{\partial x^0}{\partial t}}_{=1 \text{ 因为 } x^0=t} - \frac{\partial}{\partial t} \int d^3\mathbf{x} T^{00} x^i \\ &= t \frac{\partial P^i}{\partial t} + P^i - \frac{\partial}{\partial t} \int d^3\mathbf{x} T^{00} x^i \end{aligned}$$

由4.41式²²可知 P^i 是守恒的 $\frac{\partial P^i}{\partial t} = 0 \rightarrow P^i = \text{常量}$ ，故我们断言

$$\frac{\partial}{\partial t} \int d^3\mathbf{x} T^{00} x^i = P^i = \text{常数}。 \quad (4.63)$$

⁵² 不要太过在意这一项的意义，因为它并不像其他守恒量那样有用。

故这个守恒率告诉我们⁵²能量中心以一个恒定的速度运动。或者从另一个视角来看：动量等于总能量乘以质心速度。

不论如何，4.46式中导出的守恒量还是显得很奇怪。它能够通过恰当选取起始时刻而取作零。我们能选择一个合适的时间零点使得这个量为零。由于它是守恒量，故这个零始终是零。²³

²² 译注：原文公式引用有误

²³ 译注：很是不能理解作者的意思。毕竟我们可以选取合适的能量零点使能量为零、合适的惯性系使动量为零、以及合适的匀速转动系使角动量为零。

The Equations of Nature 自然界的方程

5	测量	103
5.1	量子力学中的算符 自旋与角动量	
5.2	量子场论的算符	
6	自由场理论	107
6.1	Lorentz 协变性与不变量	
7	相互作用理论	109
7.1	$U(1)$ 相互作用 自旋 $\frac{1}{2}$ 自由场的内禀对称性 自旋为 1 自由场的内禀对称性 把奇怪的东西堆一起 非齐次麦克斯韦方程和最小耦合 再一次电荷共轭变换 内禀 $U(1)$ 对称性的 Noether 定理 零自旋有质量场的相互作用 自旋为 1 有质量场的相互作用	
7.2	$SU(2)$ 相互作用	
7.3	质量项、 $U(1)$ 和 $SU(2)$ 的统一	
7.4	宇称破坏	
7.5	轻子质量项	
7.6	夸克质量项	
7.7	同位旋 Labelling 态	
7.8	$SU(3)$ 相互作用 色 夸克描述	
7.9	Bosons 和 Fermions 间的相互影响	

DRAFT TRANSLATION

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：

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

我们会看到，这种选择会自然地引导我们给出量子理论。

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

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

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

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

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

$$\text{位置 } \hat{x}_i \rightarrow x_i$$

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上。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

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

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

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

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。 $\hat{L}_i \rightarrow$ 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 $c \rightarrow \infty$ 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 $M_i = (tp_i - x_i E)$ 。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 $M_i = (tp_i - x_i E) \leftrightarrow K_i = i(t\partial_i - x_i \partial_0)$ 。因此，利用我们之前的选择，很直观的我们有位置： $\hat{x}_i \rightarrow x_i$

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 $\vec{F} = m\vec{\ddot{x}}$ ，可以对于任意矢量 \vec{C} 写出 $\vec{F}\vec{C} = m\vec{\ddot{x}}\vec{C}$ 。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 $\hat{S}_i \rightarrow$ 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3.7.5中得到的二维表示:

$$\hat{S}_i = i\frac{\sigma_i}{2} \quad (5.4)$$

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

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

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, 而不是其总量 $\Pi = \int d^3x \pi(x) \neq \Pi(x)$ 。

我们上一章发现了在奇异不变性下场自己 $\Phi \rightarrow \Phi - i\epsilon$ 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(??)。

共轭动量密度 $\pi(x) \rightarrow$ 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

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。

$$\begin{aligned} [\Phi(x), \pi(y)]\Psi &= \left[\Phi(x), -i\frac{\partial}{\partial\Phi(y)} \right] \Psi \\ &\stackrel{\text{莱布尼兹律}}{=} -i\Phi(x)\frac{\partial\Psi}{\partial\Phi(y)} + i\Phi(x)\frac{\partial\Psi}{\partial\Phi(y)} + i\left(\frac{\partial\Phi(x)}{\partial\Phi(y)}\right)\Psi = i\delta(x-y)\Psi \end{aligned} \quad (5.5)$$

同样的，由于这个式子对任意的 Ψ 都成立，我们可以写

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

类似的，对于场有多于一个分量的情况，我们有

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

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。

6. Free Theory 自由场理论

在这章中，我们建立自由的体系的物理理论，也就是说无相互作用的情况下的对称性给出的场¹。我们会

- 利用 Lorentz 群的 $(0,0)$ 表示给出 Klein-Gordon 方程
- 利用 Lorentz 群的 $(\frac{1}{2},0) \oplus (0,\frac{1}{2})$ 表示给出 Dirac 方程
- 利用 Lorentz 群的矢量 $(\frac{1}{2},\frac{1}{2})$ 表示给出 Proca 方程，在无质量极限下退化到著名的 Maxwell 方程组

¹ 尽管我们这里针对的是场，后面会看到这里得到的方程同样可以用来处理粒子

6.1 Lorentz Covariance and Invariance Lorentz 协变性与不变量

在下面的章节中，我们会得到粒子物理的标准模型- 这是我们拥有的最好的物理理论 - 中的运动方程。我们期待这些房产在所有的惯性系中看起来都一样，因为如果不这样的话，我们就会对每一个可能的参考系都有各自的方程。狭义相对论指出并没有一个特别的参考系，所以这一点是没有意义的。这个术语被称为 **Lorentz 协变性**。一个 Lorentz 协变的物体是说它按照某一种给定的 Lorentz 群来进行变换。比如说，矢量 A_μ ，通过 $(\frac{1}{2},\frac{1}{2})$ 表示来进行变换，因此是 Lorentz 协变的；这其实是说， $A_\mu \rightarrow A'_\mu$ ，但实际上这两个指代同一个东西，并不是完全不同的。另一方面，比如说， $A_1 + A_3$ 就不是 Lorentz 协变的，因为它不按照 Lorentz 群的某种表示来变换。但这并不是说我们不知道它是如何变换的，

因为它的变换规律从 A_μ 的变换规律里面可以简单地得出，但是在不同的惯性系中它看起来完全不一样。在一个“推动”的参考系中它看起来或许像 $A_2 + A_4$ 。只涉及到 Lorentz 协变的東西的方程被称为 Lorentz 协变方程，比如

$$A_\mu + 7B_\mu + C_\nu A^\nu D_\mu = 0$$

就是一个 Lorentz 协变方程，因为在另一个参考系中，它看起来就是

$$A_\mu + 7B_\mu + \underbrace{C_\nu A^\nu}_{\text{一个 Lorentz 标量, 根据 } (0,0) \text{ 表示来进行变换}} D_\mu = A'_\mu + 7B'_\mu + C_\nu A^\nu D'_\mu = 0$$

7. Interaction Theory 相互作用理论

Summary 总结

在这一章中我们将导出不同的场之间的相互作用。这使得我们能够，例如，描述电子是如何和光子作用的¹。

内禀对称性，或者这里叫做规范² 对称性，能指引我们得到拉格朗日量的正确形式。我们将从局域³ $\mathcal{U}(1)$ 对称性出发来得到拉格朗日量

$$\mathcal{L} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_{\mu}\partial^{\mu}\Psi + A_{\mu}\bar{\Psi}\gamma^{\mu}\Psi + \partial^{\mu}A^{\nu}\partial_{\mu}A_{\nu} - \partial^{\mu}A^{\nu}\partial_{\nu}A_{\mu},$$

即量子电动力学 (quantum electrodynamics) 的拉格朗日量。这个拉格朗日量描述了有电荷有质量的场和无质量自旋为 1 的场 (光子场) 之间的相互作用。在排除了形如 $mA_{\mu}A^{\mu}$ 的“质量项”后 (这和用 A_{μ} 来描述的光子是无质量的实验事实相符)，拉格朗日量只有这样才是局域 $\mathcal{U}(1)$ 不变的⁴。利用 Noether 定理，我们能从 $\mathcal{U}(1)$ 对称性中导出一个新的守恒量，它常被解释为电荷 (electric charge)。

接下来是 $SU(2)$ 对称性。引入一个二分量的二重态 (doublet)

$$\bar{\Psi} := \begin{pmatrix} \bar{\psi}_1 & \bar{\psi}_2 \end{pmatrix},$$

这样一个二重态场中包含了两个自旋 $\frac{1}{2}$ 场，例如，电子和电子中

¹ 从另外一个视角看：电子 (自旋 $\frac{1}{2}$ 且有质量) 场如何和光子 (自旋为 1) 场作用。

² 马上就会解释这个奇怪的名字

³ 这意味着不仅仅是一个 $e^{i\alpha}$ ，而是在每一个时空点上作用一个不同的因子，数学上写作 $e^{i\alpha(x)}$ 。或者这样说：变换参数 $\alpha = \alpha(x)$ 现在是 x 的函数，在不同的时空点上有不同的值。

⁴ 译注：应该为“只有这样才是且仅是”。

微子场在 $SU(2)$ 的“旋转”下相互转换。

⁵ $W_j^{\mu\nu}$ 这玩意会通过三个 W_j^μ 定义, 就像从 $U(1)$ 规范场 A^μ 定义 $F_{\mu\nu}$ 一样。

我们可以使用二重态记号写下局域 $SU(2)$ 不变的拉格朗日量⁵

$$\mathcal{L} = i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + \bar{\Psi}\gamma_\mu\sigma_j W_j^\mu\Psi - \frac{1}{2}(W_{\mu\nu})_i(W^{\mu\nu})_i,$$

其中包含了三个自旋为 1 的场 W_j^μ 。由于 $SU(2)$ 群的生成元有三个基 $J_i = \sigma_i/2$, 所以我们需要三个场来保证拉格朗日量是局域 $SU(2)$ 不变的。我们将看到局域 $SU(2)$ 对称性仅当形如 $m\bar{\Psi}\Psi$ 、 $mW_\mu W^\mu$ 的质量项 (其中质量 m 是任意给定矩阵) 不存在时才有可能实现, 这是因为 Ψ 现在是二分对象了。所以此时不仅自旋为 1 的场 W_j^μ 得是无质量的了, 自旋 $\frac{1}{2}$ 场也一样。另外一种可能是两个自旋 $\frac{1}{2}$ 等质量场, 但是它被实验否决了: 电子质量远大于电子中微子质量。除此之外, 我们从实验知道三个自旋为 1 的场 W_j^μ 不是无质量的。这常解释成 $SU(2)$ 对称性被破缺了。

这是随后被引入的 **Higgs 机制 (Higgs formalism)** 的想法来源。这个机制能使我们得到一个包含质量项的 $SU(2)$ 不变的拉格朗日量。它通过引入一个与零自旋场, 即 Higgs 场, 的相互作用来达成目标。相同的机制使我们能给自旋 $\frac{1}{2}$ 场加上任意的质量项以与实验相符。最后的相互作用拉格朗日量描述了一种新的相互作用: **弱相互作用 (weak interaction)**, 由三个⁶有质量自旋为 1 的场: W^+ 、 W^- 和 Z , 作为传播媒介。利用 Noether 定理, 我们能从 $SU(2)$ 对称性得到一个新的守恒量: **同位旋 (isospin)**, 它相当于于电磁相互作用中的电荷, 是弱相互作用中的荷。

最后, 我们将考虑内禀 $SU(3)$ 对称, 它将会将我们能描述一类新的相互作用, **强相互作用 (strong interaction)**, 的拉格朗日量。为此我们将引入一个三重态

$$Q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix},$$

在 $SU(3)$ 下变换, 包含三个自旋 $\frac{1}{2}$ 场。这三个自旋 $\frac{1}{2}$ 场被解释为有不同颜色 (color) 的夸克 (quarks), 这是电磁作用中的电荷、或者弱作用中的同位旋在强相互作用中的对应物。一样, 质量项是被禁止的, 但这次和实验结论一致: 8 个相应的 bosons⁷, 称为胶子, 是无质量的⁸。根据实验我们知道只有夸克 (自旋 $\frac{1}{2}$) 和胶

⁶ 准确的讲: 我们从 $U(1) \otimes SU(2)$ 对称破缺到 $U(1)$ 。这个过程产生了三个有质量矢量 bosons: W^+ 、 W^- 、 Z , 和一个无质量矢量 bosons: γ 光子。所以说为什么常讲电磁学 (来自 $U(1)$ 对称性) 和弱相互作用理论 (来自 $SU(2)$) 是统一的。一开始的 $U(1)$ 对称性和最后留下来的 $U(1)$ 对称性不是一个东西。光子和 Z Bosons 可以看做是两个矢量 Bosons 的线性组合, 常记做 B 和 W^3 , 用以使拉格朗日量具有局域 $U(1)$ (B 波色子) 和 $SU(2)$ (W^1 、 W^2 、 W^3 -bosons) 不变性。

⁷ 数字 8 源于 $SU(3)$ 的生成元有 8 个基。

⁸ 作为补充, 我们从实验知道 $SU(3)$ 的三重态中的场有相同的质量。这是个好消息, 因为局域 $SU(3)$ 对称性禁止形如 $m\bar{Q}Q$ 这样带有任意质量矩阵 m 的项存在, 但允许 $\bar{Q} \begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix} Q$ 这样一项, 这代表三重态里的项有相同的质量。由此局域 $SU(3)$ 不变性不会在拉格朗日量的质量项上造成新的障碍, $SU(3)$ 对称也没有被破缺。

子（自旋为 1）携带颜色。最后，拉格朗日量的结果是

$$\mathcal{L} = -\frac{1}{4}F_{\alpha\beta}^A F_A^{\alpha\beta} + \bar{Q}(iD_\mu\gamma^\mu - m)Q,$$

咱仅仅引用一下装个逼，因为推导太繁琐了，并且和我们之前做过的完全类似。

给总结来总结一下：

$U(1)$	\rightarrow	1 个规范场	\rightarrow	无质量光子	\rightarrow	电荷
$SU(2)$	\rightarrow	3 个规范场	\rightarrow	有质量 W- 和 Z-bosons （需要 Higgs）	\rightarrow	同位旋
$SU(3)$	\rightarrow	8 个规范场	\rightarrow	无质量胶子	\rightarrow	色荷

7.1 $U(1)$ Interaction $U(1)$ 相互作用

为了得到拉格朗日量中正确的相互作用项，我们得使用内禀对称性，或者称作规范对称性 (gauge symmetries)。规范对称这个叫法有一些历史上的原因，和我们现在要讨论的东西之间关系不是太大。Weyl 曾尝试将电磁学⁹ “作为时空对称性的结果，特别是在局域尺度变换下的对称性”。将其称作规范对称性是有原因的，因为它意味着，例如，我们能任意改变用于定义长度标准的铂金尺（用来规范¹实验中测长的仪器），而不改变物理。这个尝试没有成功，但稍晚些时候，Weyl 找到了能正确导出电磁学的对称性，并仍用这个词来命名。

⁹Frank Wilczek. Riemann-einstein structure from volume and gauge symmetry. Phys. Rev. Lett., 80:48514854, Jun 1998. doi: 10.1103/Phys-RevLett.80.4851

7.1.1 自旋 $\frac{1}{2}$ 自由场的内禀对称性

再来看一眼用来导出自由自旋 $\frac{1}{2}$ 理论的拉格朗日量（??式）

$$\mathcal{L}_{\text{Dirac}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi = \bar{\Psi}(i\gamma_\mu\partial^\mu - m)\Psi \quad (7.1)$$

它是在 Lorentz 对称性的要求下导出的，但如果我们仔细观察它，能发现这个拉格朗日量中蕴含着另一个对称性。拉格朗日量在场 Ψ 的如下变换下不变

$$\begin{aligned} \Psi &\rightarrow \Psi' = e^{ia}\Psi \\ \Rightarrow \bar{\Psi} &\rightarrow \bar{\Psi}' = \Psi'^\dagger\gamma_0 = (e^{ia}\Psi)^\dagger\gamma_0 = \bar{\Psi}e^{-ia}, \end{aligned} \quad (7.2)$$

¹译注：没有规矩，不成方圆。线长是某个线性空间中的某种范数，将改变线长的定义的变换，称作规范变换，应当是一个合理的称呼。关于 gauge 这个词的命名以及为何翻译成“规范”，可以参见曹则贤. Norm and Gauge[J]. 物理, 2013, 42(11): 815-819.

¹⁰ 别忘了 $\bar{\Psi} = \Psi^\dagger \gamma_0$

其中负号来源于复共轭¹⁰，而 a 是一个任给实数。为了看清这一点，我们将变换后的场的拉格朗日量明确写出

$$\begin{aligned}\mathcal{L}'_{\text{Dirac}} &= -m\bar{\Psi}'\Psi' + i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' \\ &= -m(\bar{\Psi}e^{-ia})(e^{ia}\Psi) + i(\bar{\Psi}e^{-ia})\gamma_\mu\partial^\mu\Psi(e^{ia}\Psi) \\ &= -m\bar{\Psi}\Psi \underbrace{e^{-ia}e^{ia}}_{=1} + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi \underbrace{e^{-ia}e^{ia}}_{=1} \\ &= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi = \mathcal{L}_{\text{Dirac}}\end{aligned}\quad (7.3)$$

¹¹ 技术上讲：一个复数和所有矩阵对易，例如 γ_μ 。

其中我们利用了 e^{ia} 仅是一个复数这一点来将它自由前后挪动¹¹。回忆我们在第3章中曾讲过的，任何单位复数（能写作 e^{ia} 的形式）构成了一个 $\mathcal{U}(1)$ 群，由此可以将我们刚才发现的事实用数学语言表达出来：拉格朗日量是 $\mathcal{U}(1)$ 不变的。由于它丝毫不涉及时空的变换而只改变场自身，所以说这是一个内禀对称性。猛地一看，这玩意还挺萌的，但好像没什么用。不要走开，稍后的节目会告诉你它有多精彩！

现在咱再来钻深一点。我们刚展示了，给我们的场乘一个任意单位复数什么都不会改变。这个对称变换 $\Psi \rightarrow \Psi' = e^{ia}\Psi$ 叫做**全局 (global)** 变换，因为我们在每一个点 x 给场 $\Psi = \Psi(x)$ 乘上了一个相同的因子 e^{ia} 。

所以说，为什么一个时空点的相位因子会和另一个时空点的相关呢？一个点的选取不应该瞬间改变整个宇宙的情况。这是挺奇怪的，因为正如2.4节中所述，狭义相对论告诉我们没有信息能比光传播得更快。而全局对称性选择会瞬间应用到在整个宇宙中的每一个点上²。

让我们来检查一下，如果给每一个时空点变换一个不同的因子 $a = a(x)$ ，拉格朗日量是不是还是不变的。这叫做**局域 (local)** 变换。

我们作变换

$$\begin{aligned}\Psi &\rightarrow \Psi' = e^{ia(x)}\Psi \\ \Rightarrow \bar{\Psi} &\rightarrow \bar{\Psi}' = \Psi'^\dagger \gamma_0 = (e^{ia(x)}\Psi)^\dagger \gamma_0 = \bar{\Psi}e^{-ia(x)},\end{aligned}\quad (7.4)$$

现在 $a = a(x)$ 依赖于坐标了，我们得到变换后的拉格朗日量¹²

²译注：引入局域规范的另一种讲法是，我们不能保证外星人里物理学家和我们选取同一套规范（这里的 a 是可以随意选取的，但是必须要选取一个才能得到完整的理论。 a 的选取被称为规范的选取），但是物理定律不应该有区别。

¹² 可能你会好奇包含全部可能导数项的拉格朗日量（由简洁性的要求被忽略）是不是局域 $\mathcal{U}(1)$ 不变的： $\mathcal{L} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + i(\partial^\mu\Psi)\gamma_\mu\Psi$ 。你可以检查一下，这个拉格朗日量自然是局域 $\mathcal{U}(1)$ 不变的，但注意第二项和第三项之和为零： $i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + i(\partial^\mu\Psi)\gamma_\mu\Psi \xrightarrow{\text{分部积分}} i(\partial^\mu\Psi)\gamma_\mu\Psi - i(\partial^\mu\Psi)\gamma_\mu\Psi = 0$ 。包含了全部可能导数的正确拉格朗日量的两导数项间必须有一个相对负号： $\mathcal{L}_{\text{Dirac}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - i(\partial^\mu\Psi)\gamma_\mu\Psi$ ，故它并不是局域 $\mathcal{U}(1)$ 不变的。

$$\begin{aligned}
\mathcal{L}'_{\text{Dirac}} &= -m\bar{\Psi}'\Psi' + i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' \\
&= -m(\underbrace{\bar{\Psi}e^{-ia}}_{=1})(e^{ia}\Psi) + i(\bar{\Psi}e^{-ia(x)})\gamma_\mu\partial^\mu\Psi(e^{ia(x)}\Psi) \\
&= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi \underbrace{e^{-ia}e^{ia}}_{=1 \text{ 莱布尼兹律}} + i(e^{-ia(x)}\bar{\Psi})\gamma_\mu\Psi(\partial^\mu e^{ia(x)}) \\
&= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + i^2(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi \neq \mathcal{L}_{\text{Dirac}} \quad (7.5)
\end{aligned}$$

由于莱布尼兹律导致了一个额外的项，所以说我们的拉格朗日量在局域 $\mathcal{U}(1)$ 对称变换下不是不变的。但按照上面所讨论的，拉格朗日量应该是局域不变的，在这里却不是。这里有一些事情可做，但我们得先研究另一种对称性。

¹³ 对比??式，注意我们为了简明起见而丢掉的 $\frac{1}{2}$ 因子

7.1.2 自旋为 1 自由场的内禀对称性

接下来，来看看自旋为一的粒子¹³

$$\mathcal{L}_{\text{Proca}} = \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu + m^2 A_\mu A^\mu. \quad (7.6)$$

这儿也能找到一个全局内禀对称性。如果我们作变换

$$A_\mu \rightarrow A'_\mu = A_\mu + a_\mu \quad (7.7)$$

其中 a_μ 为任意常数，拉格朗日量变为

$$\begin{aligned}
\mathcal{L}'_{\text{Proca}} &= \partial^\mu A'^\nu \partial_\mu A'_\nu - \partial^\mu A'^\nu \partial_\nu A'_\mu + m^2 A'_\mu A'^\mu \\
&= \partial^\mu (A^\mu + a^\mu) \partial_\mu (A_\nu + a_\nu) - \partial^\mu (A^\nu + a^\nu) \partial_\nu (A_\mu + a_\mu) + m^2 (A_\mu + a_\mu) (A^\mu + a^\mu) \\
&= \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu + m^2 (A_\mu + a_\mu) (A^\mu + a^\mu) \quad (7.8)
\end{aligned}$$

由此我们作出结论，对于无质量场 ($m = 0$)，这个变换是个全局对称的变换。

那局域对称性又会怎么样呢？作变换

$$A_\mu \rightarrow A'_\mu = A_\mu + a_\mu(x) \quad (7.9)$$

无质量的拉格朗日量变成

$$\begin{aligned}
 \mathcal{L}'_{\text{Maxwell}} &= \partial^\mu A'_\nu \partial_\mu A'_\nu - \partial^\mu A'_\nu \partial_\nu A'_\mu \\
 &= \partial^\mu (A^\mu + a^\nu(x)) \partial_\mu (A_\nu + a_\nu(x)) - \partial^\mu (A^\nu + a^\nu(x)) \partial_\nu (A_\mu + a_\mu(x)) \\
 &= \partial^\mu A^\nu \partial_\mu A_\nu + \partial^\mu a^\nu(x) \partial_\mu A_\nu + \partial^\mu A^\nu \partial_\mu a_\nu(x) + \partial^\mu a^\nu(x) \partial_\mu a_\nu(x) \\
 &\quad - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\mu A^\nu \partial_\nu a_\mu(x) - \partial^\mu a^\nu(x) \partial_\nu A_\mu - \partial^\mu a^\nu(x) \partial_\nu a_\mu(x)
 \end{aligned}$$

即没有内禀对称性。

然而，如果我们这样作变换 $A_\mu \rightarrow A'_\mu + \partial_\mu a(x)$ ，我们能弄出一个局域内禀对称性来。这意味着我们加的是一个任意函数的导数，而不是一个任意函数。这会导致³

$$\begin{aligned}
 \mathcal{L}'_{\text{Maxwell}} &= \partial^\mu A'_\nu \partial_\mu A'_\nu - \partial^\mu A'_\nu \partial_\nu A'_\mu \\
 &= \partial^\mu (A^\mu + \partial^\nu a(x)) \partial_\mu (A_\nu + \partial_\nu a(x)) - \partial^\mu (A^\nu + \partial^\nu a(x)) \partial_\nu (A_\mu + \partial_\mu a(x)) \\
 &= \partial^\mu A^\nu \partial_\mu A_\nu + \partial^\mu \partial^\nu a(x) \partial_\mu A_\nu + \partial^\mu A^\nu \partial_\mu \partial_\nu a(x) + \partial^\mu \partial^\nu a(x) \partial_\mu \partial_\nu a(x) \\
 &\quad - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\mu A^\nu \partial_\nu \partial_\mu a(x) - \partial^\mu \partial^\nu a(x) \partial_\nu A_\mu - \partial^\mu \partial^\nu a(x) \partial_\nu \partial_\mu a(x) \\
 &\stackrel{\text{重命名哑指标}}{=} \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu = \mathcal{L}_{\text{Maxwell}}
 \end{aligned}$$

$\partial_\nu \partial_\mu = \partial_\mu \partial_\nu$ 以及重命名哑指标

所以说这里的确有一个局域内禀对称变换。这看上去也像是一个技术细节。好的，我们找到了一些局域的、内禀的对称性，然后呢？

7.1.3 把奇怪的东西堆一起

总结一下咱现在都干了啥

- 我们发现自旋 $\frac{1}{2}$ 自由场的拉格朗日量有一个全局的内禀对称性 $\Psi \rightarrow \Psi' = e^{ia} \Psi$ 。或者说：自旋 $\frac{1}{2}$ 自由场的拉格朗日量在全局 $\mathcal{U}(1)$ 变换下不变。
- 这个对称性不是局域的（尽管它应当是），因为在 $a = a(x)$ 的情况下，我们在拉格朗日量中得到如下形式的额外项（7.5式）

$$-(\partial_\mu a(x)) \bar{\Psi} \gamma^\mu \Psi. \quad (7.12)$$

换言之：拉格朗日量不是局域 $\mathcal{U}(1)$ 不变的。

- 上一节中我们发现了无质量自旋为一的场的一个局域内禀对

³译注：原文缺少一个 (x)

称性

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu a(x), \quad (7.13)$$

它仅当我们加上的是一个任意函数的导数 $\partial_\mu a(x)$ ，而不是一个任意函数 $a_\mu(x)$ 时，才是一个局域对称性。

这两坨奇怪的东西看上去真该加在一起：拉格朗日量中的一个额外项 $A_\mu \bar{\Psi} \gamma^\mu \Psi$ 会变换作

$$A_\mu \bar{\Psi} \gamma^\mu \Psi \rightarrow (A_\mu + \partial_\mu a(x)) \bar{\Psi} \gamma^\mu \Psi = A_\mu \bar{\Psi} \gamma^\mu \Psi + \partial_\mu a(x) \bar{\Psi} \gamma^\mu \Psi \quad (7.14)$$

比较 7.12 式中的第二项⁴。拉格朗日量中这个将 Ψ 、 $\bar{\Psi}$ 和 A_μ 耦合在一起的新项，正好能干掉那个不自量力阻挡自旋 $\frac{1}{2}$ 自由场拥有局域 $\mathcal{U}(1)$ 不变性的家伙。换言之：**通过添加这个新项使得拉格朗日量局域 $\mathcal{U}(1)$ 不变。**

再来仔细研究一下这件事。首先习惯上都会在局域 $\mathcal{U}(1)$ 变换的指数上添一个常数因子 g ： $e^{ig a(x)}$ 。这样额外项变成

$$-(\partial_\mu a(x)) \bar{\Psi} \gamma^\mu \Psi \rightarrow -g(\partial_\mu a(x)) \bar{\Psi} \gamma^\mu \Psi. \quad (7.15)$$

正如我们下面将看到的，这个额外的因子 g 代表一个任意的耦合常数¹⁴ g 。将这个新项

$$g A_\mu \bar{\Psi} \gamma^\mu \Psi$$

加到自旋 $\frac{1}{2}$ 自由场的拉格朗日量上，其中 γ^μ 用来保证这一项的 Lorentz 不变性，否则带有一个自由指标 μ 的 A_μ 和拉格朗日量密度这个标量无法匹配，另外也插入了**耦合常数 (coupling constant)**¹⁵ g 。这样拉格朗日量

$$\mathcal{L}_{\text{Dirac}+ \text{额外项}} = -m \bar{\Psi} \Psi + i \bar{\Psi} \gamma_\mu \partial^\mu \Psi + g A_\mu \bar{\Psi} \gamma^\mu \Psi.$$

依据 Ψ 、 $\bar{\Psi}$ 和 A_μ 局域变换的规则来变换这个拉格朗日量，得到¹⁶

¹⁴ 耦合常数总能告诉我们相互作用的强度是多少。这里讨论的是电磁相互作用， g 代表着它的强度。

¹⁵ 我们能看到这里 g 决定了 Ψ 、 $\bar{\Psi}$ 和 A_μ 耦合在一起的强度。

¹⁶ Ψ 、 $\bar{\Psi}$ 和 A_μ 的联合变换被称作 $\mathcal{U}(1)$ 规范变换。

⁴ 译注：此处疑引用有误

$$\begin{aligned}
\mathcal{L}'_{\text{Dirac+ 额外项}} &= -m\bar{\Psi}'\Psi' + i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' + gA'_\mu\bar{\Psi}'\gamma^\mu\Psi' \\
&\stackrel{\text{见 7.5 式}}{=} -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - g(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi + gA'_\mu\bar{\Psi}\gamma^\mu\Psi' \\
&= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - g(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi + g(A_\mu + \partial_\mu a(x))(e^{-iga}\bar{\Psi}\gamma^\mu\Psi) \\
&= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - \underline{g(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi} + gA_\mu\bar{\Psi}\gamma^\mu\Psi + \underline{g(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi} \\
&= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi = \mathcal{L}_{\text{Dirac+ 额外项}}.
\end{aligned}$$

由此，在加上一个额外项后，我们得到了，正如所期待的，一个局域 $U(1)$ 不变的拉格朗日量。为了描述一个包含了包含有质量自旋 $\frac{1}{2}$ 场和无质量自旋为一的场的系统，我们必须将无质量自旋为一的自由场的拉格朗日量也加到拉格朗日量中来。这样就得到了完整的拉格朗日量

$$\mathcal{L}_{\text{Dirac+ 额外项 + Maxwell}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi + \partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu. \quad (7.17)$$

为了方便起见，引入记号

$$D_\mu \equiv i\partial_\mu + gA_\mu, \quad (7.18)$$

称作协变导数 (covariant derivative)⁵。这样拉格朗日量写作

$$\begin{aligned}
\mathcal{L}_{\text{Dirac+ 额外项 + Maxwell}} &= -m\bar{\Psi}\Psi + \bar{\Psi}\gamma_\mu(i\partial^\mu + gA^\mu)\Psi + \partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu \\
&= -m\bar{\Psi}\Psi + \bar{\Psi}\gamma_\mu D^\mu\Psi + \partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu \quad (7.19)
\end{aligned}$$

这即为电磁场的量子场论，常被称为量子电动力学，的正确拉格朗日量。我们通过观察自旋 $\frac{1}{2}$ 自由场和自旋为一的自由场的内禀对称性，简单的得到了它。

我们将要回答的下一个问题是：这个拉格朗日量会导致什么样的运动方程？

7.1.4 非齐次麦克斯韦方程和最小耦合

谜底揭晓：这个拉格朗日量给出有电流情况下的非齐次 Maxwell 方程组。

⁵译注：常见的记号似乎与其差了一个虚数单位

推导是很直白的：将拉格朗日量⁶

$$\mathcal{L}_{\text{Dirac} + \text{额外项} + \text{Maxwell}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA^\mu\bar{\Psi}\gamma_\mu\Psi + \partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu$$

代入每一个场的拉格朗日方程中

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\Psi} - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\Psi)}\right) &= 0 \\ \frac{\partial\mathcal{L}}{\partial\bar{\Psi}} - \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\bar{\Psi})}\right) &= 0 \\ \frac{\partial\mathcal{L}}{\partial A_\rho} - \partial_\sigma\left(\frac{\partial\mathcal{L}}{\partial(\partial_\sigma A_\rho)}\right) &= 0.\end{aligned}$$

得到

$$\bar{\Psi}(i\gamma_\mu\partial^\mu + m) + gA_\mu\bar{\Psi}\gamma^\mu = 0 \quad (7.20)$$

$$(i\gamma_\mu\partial^\mu - m)\Psi - gA_\mu\gamma^\mu\Psi = 0 \quad (7.21)$$

$$\partial_\sigma(\partial^\sigma A^\rho - \partial^\rho A^\sigma) + g\bar{\Psi}\gamma^\rho\Psi = 0 \quad (7.22)$$

头两个方程描述了自旋 $\frac{1}{2}$ 粒子/场在外加电磁场下的行为。在许多书里这些方程是利用一个叫做**最小耦合 (minimal coupling)** 的办法得到的，即当外场存在时，导数 ∂_μ 应当换成协变导数⁷

$$\partial_\mu \rightarrow D_\mu = i\partial_\mu + A_\mu \quad (7.23)$$

来导出正确的方程。“最小”这个词的意思是，这里只用了一个规范场 A_μ 的四个分量⁸。

现在我们有用来描述 Dirac 旋量在外加电磁场中的行为的方程 (7.21 式)，所以说可以来兑现在??节中许下的承诺了。在那儿我们宣称电荷共轭变换，带有一个暗示性的名字，改变其描述对象的电荷。换言之，如果 Ψ 描述带电荷 $+e$ 的某些东西，而电荷共轭旋量 Ψ^C 描述带电荷 $-e$ 的某些东西。电荷决定了自旋 $\frac{1}{2}$ 粒子/场和外加自旋为一的场的耦合强度，我们将考虑 Ψ^C 的运动方程是什么。之后我们将谈谈第三个方程，7.22 式。

⁶译注：原文下式标记有误，已更正。

⁷译注：此处虚单位 i 的位置似有不妥，同后文

⁸译注：否则，我们可以使用张量场 $\bar{\Psi}\sigma_{\mu\nu}F^{\mu\nu}\Psi$

7.1.5 再一次电荷共轭变换

为了得到对应的方程，我们需要找到 Dirac 旋量的电荷共轭变换算符确切的形式。我们在??节中得到了变换（??式）

$$\Psi = \begin{pmatrix} \chi_L \\ \tilde{\xi}_R \end{pmatrix} \rightarrow \Psi^C = \begin{pmatrix} \tilde{\xi}_L \\ \chi_R \end{pmatrix}. \quad (7.24)$$

这个变换可以简单地用一个 γ_μ 矩阵来描述。利用??式中 γ_2 的定义，我们有

$$\Psi^C = i\gamma_2 \Psi^* = i \begin{pmatrix} 0 & \sigma_2 \\ -\sigma_2 & 0 \end{pmatrix} \begin{pmatrix} \chi_L^* \\ \tilde{\xi}_R^* \end{pmatrix}, \quad (7.25)$$

利用 $i\sigma_2 = \epsilon$ 正好就是旋量度规，改写作

$$\begin{pmatrix} 0 & \epsilon \\ -\epsilon & 0 \end{pmatrix} \begin{pmatrix} \chi_L^* \\ \tilde{\xi}_R^* \end{pmatrix} = \begin{pmatrix} \epsilon \tilde{\xi}_R^* \\ \epsilon \chi_L^* \end{pmatrix}. \quad (7.26)$$

等价于

$$= \begin{pmatrix} \tilde{\xi}_L \\ \chi_R \end{pmatrix}, \quad (7.27)$$

正如3.7.7节，特别是??式中展示的一样。故我们从7.21式开始：

$$(i\gamma_\mu \partial^\mu - m)\Psi - gA_\mu \gamma^\mu \Psi = (\gamma_\mu (i\partial^\mu - gA^\mu) - m)\Psi = 0, \quad (7.28)$$

并取其复共轭，作为得到 Ψ^C 的方程的第一步：

$$\rightarrow (\gamma_\mu^* (-i\partial^\mu - gA^\mu) - m) \Psi^* = 0. \quad (7.29)$$

然后左乘一个 γ_2 ，并在 Ψ 前插入一个 $1 = \gamma_2^{-1} \gamma_2$ ：

$$\rightarrow \gamma_2 (\gamma_\mu^* (-i\partial^\mu - gA^\mu) - m) \underbrace{\gamma_2^{-1} \gamma_2}_{=1} \Psi^* = 0. \quad (7.30)$$

$$\rightarrow \underbrace{(\gamma_2 \gamma_\mu^* \gamma_2^{-1})}_{=-\gamma_\mu} (-i\partial^\mu - gA^\mu) - m \gamma_2 \gamma_2^{-1} \gamma_2 \Psi^* = 0. \quad (7.31)$$

$$\underbrace{\rightarrow}_{\text{方程乘上 } i} (-\gamma_\mu (-i\partial^\mu - gA^\mu) - m) \underbrace{i \gamma_2 \Psi^*}_{=\Psi^C \text{ 见 7.25 式}} = 0. \quad (7.32)$$

$$\rightarrow (-\gamma_\mu (-i\partial^\mu - gA^\mu) - m) \Psi^C = 0, \quad (7.33)$$

这和 Ψ 的运动方程只在耦合常数上差了个负号 $g \rightarrow -g$ 。这与其名字，电荷共轭，相符¹⁷。

¹⁷ 注意 $\Psi^C \neq \bar{\Psi}$ 。 $\Psi^C = i\gamma_2\Psi$ 而 $\bar{\Psi} = \Psi^\dagger\gamma_0 = (\Psi^*)^T\gamma_0$ 。电荷共轭变换让我们能用反粒子来解释一些东西，稍后便会详细讨论。

接下来，我们转向上一节中的第三个式子，7.22式，称为有电流情况下的非齐次麦克斯韦方程 (inhomogeneous Maxwell equation in the presence of an electric current)。为了讲清楚这个名字，我们又需要 Noether 定理了。

7.1.6 内禀 $\mathcal{U}(1)$ 对称性的 Noether 定理

在4.5.5节中我们得知 Noether 定理将所有内禀对称性和一个守恒量联系了起来。和我们刚才发现的 $\mathcal{U}(1)$ 对称性相联系的守恒量是什么？Noether 定理告诉我们，有如下形式的变换

$$\Psi \rightarrow \Psi' = \Psi + \delta\Psi$$

能导出 Noether 流

$$J^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \Psi)} \delta\Psi$$

其满足连续性方程

$$\partial_\mu J^\mu = 0. \quad (7.34)$$

一个全局¹⁸ $\mathcal{U}(1)$ 变换是

$$\Psi \rightarrow \Psi' = e^{iga} = (1 + iga + \dots)\Psi$$

如通常所做的那样，我们仅将指数函数展开到第一项，因为 $\mathcal{U}(1)$ 是一个 Lie 群，而任意变换都可以由无穷小的来生成。一个无穷小的变换写作

$$\Psi \rightarrow \Psi' = \Psi + iga\Psi$$

¹⁸ 记住，自旋 $\frac{1}{2}$ 自由场的拉格朗日量仅仅是全局 $\mathcal{U}(1)$ 不变的。而上一节最终得到的拉格朗日量是全局 $\mathcal{U}(1)$ 不变的。全局对称性是局域对称性在 $a = \text{常数}$ 时的特殊情况。所以说，如果我们有一个局域 $\mathcal{U}(1)$ 不变的拉格朗日量，它自动是全局 $\mathcal{U}(1)$ 不变的。这里考虑的全局 $\mathcal{U}(1)$ 对称性将导出一个对自由场和相互作用场都守恒的量

故我们有 $\delta\Psi = i g a \Psi$ 。就如4.5.5节中所做的一样，导出 Noether 流

$$\begin{aligned} J^\mu &= \frac{\partial \mathcal{L}}{\partial(\partial_\mu \Psi)} \delta\Psi \\ &= \frac{\partial(-m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi)}{\partial(\partial_\mu \Psi)} i g a \Psi \\ &= -\bar{\Psi}\gamma^\mu g a \Psi = -g a \bar{\Psi}\gamma^\mu\Psi. \end{aligned} \quad (7.35)$$

¹⁹ 保留常数 g ，它的值由实验决定，而不是任意选取的。

由于连续性方程对于任意 a 都成立，故略去¹⁹任意的常数因子 a 。故，我们定义

$$J^\mu = -g\bar{\Psi}\gamma^\mu\Psi \quad (7.36)$$

这常被称作**四电流 (electric four-current)**。零分量正是电荷密度，其空间积分对时间守恒²⁰

²⁰ 类似的推导参见4.39式。

$$Q = \int d^3\mathbf{x} \underbrace{\rho}_{\text{电荷密度}} = \int d^3\mathbf{x} J^0 = -g \int d^3\mathbf{x} \bar{\Psi}\gamma^0\Psi \quad (7.37)$$

在量子框架下， Ψ 和概率幅相关，有要求 $\int d^3\mathbf{x} \bar{\Psi}\gamma^0\Psi = 1$ ，即总概率必须为 $100\% = 1$ 。所以说守恒量实际就是耦合强度 g ，在电磁作用意义下正比于电荷。所以，全局 $\mathcal{U}(1)$ 对称性导致了电荷守恒。

注意7.22式，我们可以借助7.36式将其改写成

$$\begin{aligned} \partial_\sigma(\partial^\sigma A^\rho - \partial^\rho A^\sigma) + \underbrace{g\bar{\Psi}\gamma^\rho\Psi}_{=-J^\rho} &= 0 \\ \rightarrow \partial_\sigma(\partial^\sigma A^\rho - \partial^\rho A^\sigma) &= J^\rho. \end{aligned} \quad (7.38)$$

借助??式中定义的电磁张量，改写作

$$\partial_\sigma F^{\rho\sigma} = J^\rho. \quad (7.39)$$

²¹ “组”，因为对于每一个分量都有一个方程

²² 在??章中将会见到

这即是在有电流情况下的非齐次 Maxwell 方程组。这些方程组²¹，和非齐次 Maxwell 方程组（可以由 $F^{\rho\sigma}$ 的定义直接得到²²）一起，是经典电动力学的基础。

接下来我们将快速的过一遍有质量自旋为一的场和零自旋场的相互作用。

7.1.7 零自旋有质量场的相互作用

注意我们导出的零自旋场的拉格朗日量

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2)$$

不是 $U(1)$ 不变的, 做变换 $\Phi \rightarrow \Phi' = e^{ia}\Phi$ 看一下就知道了。

但是, 复标量理论

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi^* \partial^\mu \Phi - m^2 \Phi^* \Phi) \quad (7.40)$$

具有 $U(1)$ 对称性, 因为有 $\Phi \rightarrow \Phi' = e^{ia}\Phi$ 及 $\Phi^* \rightarrow (\Phi^*)' = e^{-ia}\Phi^*$ 。

所以我们会像 7.1 节中对自旋 $\frac{1}{2}$ 理论所做的那样, 导出一个相互作用理论。推导完全是类似的²³, 结果是

$$\mathcal{L} = \frac{1}{2}(((\partial_\mu - iqA)\Phi^*)((\partial^\mu + iqA^\mu)\Phi) - m^2 \Phi^* \Phi) \quad (7.41)$$

代入拉格朗日方程

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right) &= 0 \\ \frac{\partial \mathcal{L}}{\partial \Phi^*} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi^*)} \right) &= 0, \end{aligned}$$

最终得到相应的运动方程⁹

$$(\partial_\mu + iqA_\mu)(\partial^\mu - iqA^\mu)\Phi^* - m^2 \Phi^* = 0 \quad (7.42)$$

$$(\partial_\mu - iqA_\mu)(\partial^\mu + iqA^\mu)\Phi - m^2 \Phi = 0, \quad (7.43)$$

描述了每一个零自旋场和自旋为一的无质量场相耦合。

7.1.8 自旋为 1 有质量场的相互作用

自旋为 1 有质量场和自旋为 1 无质量场的相互作用也可以简单的由对称性得到。自旋为 1 无质量场的拉格朗日量为

$$L_{\text{Maxwell}} = \frac{1}{4} F^{\mu\nu} F_{\mu\nu}.$$

⁹译注: 下式符号有误, 已更正

²³ 正确的拉格朗日量可以仅通过做替换 $\partial_\mu \rightarrow D_\mu = i\partial_\mu + A_\mu$ 来得到

为了区分有质量场和无质量场，我们将有质量场命名为 B^μ ，并定义

$$G^{\mu\nu} := \partial^\mu B^\nu - \partial^\nu B^\mu$$

自旋为 1 的有质量场的拉格朗日量写作 (??式)

$$G_{\text{Proca}} = \frac{1}{2} G^{\mu\nu} G_{\mu\nu} + m^2 B^\mu B_\mu$$

Lorentz 对称性说拉格朗日量中的相互作用项一定具有形式

$$\mathcal{L}_{\text{Proca-相互作用}} = C G_{\mu\nu} F^{\mu\nu}$$

其中耦合常数 C 需要由实验来定。感兴趣的读者可以借助欧拉-拉格朗日方程自行推导相应的运动方程。

7.2 $SU(2)$ Interaction $SU(2)$ 相互作用

7.3 Mass Terms and Unification of $U(1)$ and $SU(2)$ 质量项、 $U(1)$ 和 $SU(2)$ 的统一

7.4 Parity Violation 宇称破坏

7.5 Lepton Mass Terms 轻子质量项

7.6 Quark Mass Terms 夸克质量项

7.7 Isospin 同位旋

7.7.1 Labelling States Labelling 态

7.8 $SU(2)$ Interactions $SU(2)$ 相互作用

7.8.1 Color 色

7.8.2 Quark Description 夸克描述

7.9 The Interplay Between Fermions and Bosons Bosons 和 Fermions 间的相互影响