

系统科学基础

吴金闪

2015 年 4 月 21 日

目录

第一部分 系统科学导论	19
第一章 引言：系统科学与科学	21
1.1 抛一块砖：系统科学的思想、目标和定位	21
1.2 关于科学方法	22
1.3 与系统科学有关的一般的科学研究的思想	24
1.3.1 物理学的受力分析：按照研究目的来划分系统	24
1.3.2 最优编码与信息熵：概念是从现实世界里抽象出来的	24
1.3.3 力学思想无处不在：状态、状态变化以及状态变化的原因	25
1.3.4 普适性：一个原理或者模型尽可能多地描述现象	25
第二章 一些具有系统科学特色的研究实例	27
2.1 热寂说、传统物理系统的相变与开放系统结构的产生	27
2.2 整体运动的激发模式	27
2.3 相变与临界性的例子	27
2.3.1 传统相互作用与相变、临界性：Ising 模型	27
2.3.2 涌现、自组织临界：Bak-Tang-Wiesenfeld 沙堆	27
2.4 相互作用，周期分叉与混沌的例子	28
2.5 相互作用与网络科学的例子：汉字研究	28
2.6 相互作用与投入产出分析：经济学、科学学、数学	28
2.7 博弈与相互作用的例子	28
2.8 演化与定态的例子	28
2.9 约束下定态与动态过程的优化的例子	28
2.9.1 约束下定态优化的例子：项目管理	28

2.9.2	动态过程优化的例子：银行存款	29
第三章	概念地图与系统图示法	31
3.1	如何描述一个系统：图示举例	31
3.2	几种特殊的图示法	31
3.3	概念地图图示法	31
第四章	为什么要学习数学物理基础	33
4.1	概念、数学模型的来源：前面的例子再阐述	33
4.2	处理方法的来源：例子	33
4.3	研究对象的来源：例子	33
第二部分	系统科学的数学物理基础	35
第五章	线性代数	39
5.1	集合与映射的语言	39
5.2	从群到线性空间	41
5.2.1	群、半群	42
5.2.2	线性空间的矢量	44
5.2.3	矢量内积	48
5.3	线性算符	52
5.3.1	共轭算符	53
5.3.2	本征值与本征向量	55
5.3.3	线性变换与表象理论	57
5.3.4	矩阵的奇异值分解	57
5.4	数值线性代数	59
5.4.1	矩阵相乘的 Strassen 算法	59
5.4.2	本征值与本征向量	59
5.4.3	线性系统的解	59
第六章	概率论	61
6.1	古典概型：离散概率与几何概率	62
6.2	现代概率三元体	62
6.3	条件概率与 Bayesian 公式	64

目录	5
6.4 典型概率分布函数和基本统计量	65
6.5 大数定律与中心极限定理	66
6.6 概率论的 Dirac 符号形式	67
6.7 随机过程初步	69
6.8 Monte Carlo 方法	75
6.9 随机变量的测量	75
第七章 力学	77
7.1 力学思想	77
7.2 Newton 力学的基本概念	81
7.3 势函数、保守力与能量守恒	84
7.4 分析力学的技术	85
7.4.1 Lagrangian 力学	86
7.4.2 Hamiltonian 力学	86
7.4.3 Lagrangian 力学与 Hamiltonian 力学等价以及最小作用量原理	87
7.4.4 从单体到多体的分析力学形式	89
7.5 关于什么是科学的进一步讨论	90
7.6 本章小结	93
第八章 统计力学	95
8.1 从力学到统计力学	95
8.2 统计力学的基本理论：系综理论	98
8.3 相变与临界现象	100
8.4 Metropolis 方法	100
8.5 在物理学以及其他学科中的熵	103
8.6 从熵最大推导正则分布	104
8.7 本章小结	105
8.8 本章技术性附录	105
第九章 量子力学	107
9.1 量子系统的实验	107
9.2 量子系统的基本理论	110
9.2.1 量子系统的静态描述	110

9.2.2 量子系统的演化	114
9.3 量子与经典密度矩阵的区别	115
9.4 正则量子化与表象理论初步	117
9.5 Schrödinger 绘景与 Heisenberg 绘景	121
9.6 再论表象理论：相干态表象	122
9.7 写在量子力学后面的话	124
第十章 随机过程	127
10.1 Langevin 方程：随机力	129
10.2 Master 方程与 Fokker-Planck 方程	129
10.2.1 Markov 过程的 Master 方程	130
10.2.2 从 Master 方程到 Fokker-Planck 方程	130
10.2.3 几个典型的随机过程	130
10.3 Langevin 方程与 Fokker-Planck 方程的关系	130
10.4 平稳性	130
10.5 随机过程的模拟与计算	130
10.6 本章小结	130
第三部分 系统科学的基本理论	133
第十一章 线性与非线性动力学简论	135
11.1 线性系统的演化	136
11.2 非线性系统的线性化：定态与线性稳定性	137
11.3 混沌与分支	140
11.4 更多的具体系统上的研究实例	141
第十二章 相变、临界性与自组织临界性	143
12.1 相变的研究方法	144
12.2 临界行为	151
12.3 自组织临界性	151
12.4 更多的具体系统上的研究实例	151
第十三章 复杂网络概论	153
13.1 复杂网络的描述方法	154

目录	7
13.2 基本静态量与中心性	156
13.3 网络演化	156
13.4 网络上的动力学	156
13.5 网络科学应用	156
13.5.1 用网络研究方法增进人类交往，提高创造性和生产力	156
13.5.2 用网络研究方法发现容易发生脑退化的人	157
13.6 本章小结	157
第十四章 博弈论与演化博弈概论	159
14.1 博弈、博弈论与 Nash 均衡	161
14.2 理论与实际的对比：Nash 均衡的描述能力	161
14.3 演化博弈：均衡实现的问题	161
14.4 新的均衡的问题	161
14.5 更多的具体系统上的研究实例	161
第十五章 优化理论	163
15.1 运筹学概论	163
15.1.1 基本问题及其数学描述	163
15.1.2 问题的解决方法举例	163
15.1.3 更多的具体系统上的研究实例	163
15.2 控制论概论	163
15.2.1 基本问题及其数学描述	163
15.2.2 问题的解决方法举例	163
15.2.3 更多的具体系统上的研究实例	163
第四部分 附录：计算技术与学习方法	165
第十六章 计算机语言与工具的学习	167
16.1 Linux 系统中安装、编辑、编译、执行程序	167
16.2 C 语言、Python 与编程风格	172
16.3 BLAS、Lapack 及其实现	174
16.4 并行计算初步与 BLAS、Lapack、Petsc、Slepc	174
16.5 xmds2	178
16.6 sage	178

16.7 NetLogo	178
16.8 Swarm	178
16.9 C 与 R 中的 igraph	178
第十七章 概念地图与理解型学习方法	179
名词索引	181
人名与常用翻译	183
插图目录	185
举例目录	188

前言

本书的特点：

1. 来自于研究论文的例子
2. 简约内容，保留最核心的概念以及掌握这些概念必须的基础概念与分析计算技术
3. 注重通过概念与概念之间的联系来学习概念，以及概念背后的动机
4. 例子、概念、逻辑框架、动机、计算分析技术，这几个方面分开，并结合的处理方式
5. 概念地图学习方法和思维方式的应用
6. 通过例子来讲解基本概念，逻辑框架和举例并重
7. 数学物理基础和基本的系统理论的概念与方法并重
8. 每一章的前言部分交代这一章的主要思想、主要学习任务、核心概念和技术、推荐阅读材料
9. 附有计算技术附录
10. 包含一定量的例题和习题，其中大部分来自于研究论文或者模型化之前的实际问题（一定程度上训练读者从实际问题提炼模型的能力）

系统科学还不是一个非常成熟的学科，这意味着概念体系尚未完备，概念之间也还没有达到数学、物理学等领域的紧密相联的程度。但是，系统科学这个领域确实已经形成一些有自身特点的概念、研究方法和研究对象。因此，本书编撰的指导思想是通过具体的理论研究工作以及少量的应用性

实例来反映这个学科的概念、方法和对象。后者通常做为对前者的观察、评论和总结来体现，同时也是前者编排顺序、材料选择的隐藏线索。

同时，一个没有非常成熟的学科还决定了在这个学科里，概念与方法还没有完全压倒思想和思辨。在本书中，我们尽量把思想和思辨与概念和方法区分开来，重点放在后面。尽管思想和思辨是很重要的，但是我们采取的方式是：只有当思想和思辨对于理解概念和方法有非常大的好处的时候，我们介绍一下思想和思辨。系统科学的基本的思想源流有几个方面，在合适的地方在具体的例子当中，我们会做进一步的阐述：系统的演化与结构的产生（与经典热力学图景的矛盾）、整体运动与激发模式、临界行为与普适性、网络科学与一般的相互作用。我们会尽量把思想和思辨层次的内容与具体的例子结合，最好能够通过计算体现出来。所有的这些思想有一个共同点就是探讨复杂现象背后的机制，这样的机制有可能是简单的，同时把各种研究方法系统化，形成相互协调的理论体系。借用 Anderson[?] 和 Kadanoff[?] 的两句话来表达就是：More is different, more is the same。多个个体通过相互作用形成整体运动的复杂的模式和行为，这样的整体运动甚至有可能独立于原始的系统，同时系统科学的研究方法追求用一般的普适的分析方法来研究各种多个个体相互作用的系统。对于对思想和思辨的兴趣大过概念、方法和计算的读者，尽管我们强烈推荐这样的读者也熟悉一下具体的例子的分析计算，我们在这本书的基础上，采用了更简单的例子（也因此失去了很多有意义的细节）编撰了《系统科学概论》。有需要可以和本书一起阅读。在具体的例子之中，我们可以很容易地理解系统性思考的角度、整体性思考的角度是什么，当离开具体的例子的时候，我们就很难做一般的讨论。因此，我们在这里再一次强调，系统科学的基本思想的理解和学习不能离开具体的研究工作的例子。这样一门学科的探索者们必须与街头卖药的区别开来。完全兜售思想和思辨就会沦为伪科学，或者至少被严肃的科学家鄙视的学科。

另外，在这里我们希望对系统理论与系统工程做一个区分。系统理论主要是基础性的科学研究，同时当然也做一些应用性研究，系统工程则是利用来自于数学的、物理学的、系统理论的概念与方法解决实际工程与管理领域的问题，是应用性的研究甚至直接就是应用。控制论、信息论、运筹学等属于系统工程的基础性学科，但是却属于系统理论的应用性学科。当然，这样的应用性学科本身也存在这基础研究的问题，这部分问题就会进入到系统理论的范畴。因此，在这里，我们不把系统工程相关学科作为重点，但是

我们会讨论和学习这些学科的基本问题的部分。

作为教材，尤其是旨在促进一个学科的发展的教材，应该尽可能的来回答这个领域独立成为一个学科的理由是什么，研究对象的共同特点是什么，最基本的概念、方法与核心公式是什么。这也是本书的任务。但是，我们没有把握来回答好这个问题。在这里我们仅仅来抛一块砖。

系统科学的基本研究目标有两个层面。第一、如何处理在各种各样的系统之中的多个个体之间的相互作用，讨论其对系统的性质和功能的影响；第二、如何把对一种系统的研究方法抽象出来应用于更多的系统。两者都是方法论层次的目标，前者比后者稍微具体一些。如果我们把相互作用局限在四种基本相互作用（引力、电磁、强、弱）的框架内，那么我们讨论的就是统计物理学与场论的基本对象和目标。系统科学可以包含更广义的相互作用，例如人群之间的意见形成与传播，其相互作用形式就不能直接还原成为四种基本相互作用。沿着这个思路。在每一个层次，有这个层次自己的相互作用，这种形式的相互作用如何从更底层更基本的相互作用中涌现出来，也是系统科学研究的一个研究对象。后者其实是所有的自然科学（以及部分社会科学）的研究目标，想办法处理一个系统，接着抽象一般概念与方法来处理更多的系统。那么为什么要单独提出来作为系统科学的一个有特点的研究目标呢？在这里，我们强调，传统的学科是按照研究对象的某种共性来区分的，而系统科学的研究对象可以没有任何共性，可以来自于任何一个传统学科，只要而且希望，这样的跨学科的研究对象的研究方法上存在这共性。例如相变与临界现象原来是来自于物理学，尤其是统计物理学，的研究对象，后来人们发现在地理学、生态学、社会学甚至神经科学中都大量存在这多个体组成的系统整体状态出现定性变化，也就是相变这样的现象，因此从物理学发展起来的临界现象的研究方法（例如序参量、相变点附近的长程关联、重整化群、Metropolis 模拟方法）自然也就在以上这些学科的相应研究中的到了应用。通过本书的例子，多大程度上我们能够回答这个问题，多大程度上我们能够按照这个思路来回答这个问题，只能够留待本书使用者的检验了。

由于前面提到的本学科基本概念与方法之间的联系紧密和一致性程度还不高，我们又希望通过本书能够给出一个比较清晰的系统科学的图像，在本书中我们尝试使用了一个比较新的学习技术：概念地图。概念地图就是把概念与概念之间的连接画成一张图，用来整理制作者的思路，指引学习者对概念的理解。

本书的逻辑体系、内容选取都仅仅是一个尝试，而且是与通常的教材不那么一样的一个尝试。我们希望本书具体内容发挥一定作用的同时，也希望它尽早地被逻辑体系和具体内容都更好的书取代。这样才能表明本学科的成熟度的到了提高，学科取得了发展。另一方面，我们也希望本书的编撰原则和特点，它们是内容来自于研究论文、大量的来自于研究课题的习题、学习方法方面的一定的考虑、技术和概念并重压缩对思想的讨论，得到延续。

在各个章节具体学科的内容选择上，我们企图选择最核心的和最基础的。我们希望做到学习这些基础与核心的东西之后，读者可以独立地进一步学习，同时看到这个学科的基本图景（现状和未来，以及应用）。为了完成这个目的，实际上，每一门课程，我们都准备了这门课程的概念地图，然后综合考虑这些个课程的概念地图组合而成的大的概念地图来决定内容的取舍。我们希望我们的选择是合理的，但是我们更希望其他人看得见我们利用概念地图来选择这些核心和基础内容的方法的合理性，并加以发展和应用。另一方面，我们也看到，这样的比较简约的集中在核心内容上的学习方式对于学生真正掌握这些学科来说，是不够的。一定程度上的重复劳动对于掌握内容来说是必须的。我们提供的练习题和例题的数量是远远不够的。如果读者将进一步掌握好这些内容，我们建议从其他教材中选择一定数量的习题来完成。我们也准备在将来给每一个这样的具体学科出一本相同风格（简约和注重核心，保证理解核心部分的最低要求的基础，尽量让学习者明白每一个概念每一个定理的动机，注重概念与概念之间的联系，体现概念地图的学习方法，包含导论、习题、例题而且导论、习题和例题有一大部分直接来自于研究论文），但是更详细的教材：包含的核心内容稍微多一点点，解释和计算的细节增加一点点，包含的习题的数量多很多很多。

阅读本书的读者，数学方面最好有微积分、线性代数、概率论的基础，了解一点近世代数会很有帮助；物理学方面最好有力学的基础，了解统计物理学、量子力学也会提高对内容的理解；计算机方面最好有一门语言的编程基础，如果有 Linux 使用基础和科学计算的初步了解也会有帮助；系统科学方面最好什么都不会，尤其是不能看过很多系统科学的哲学书籍。在实际上课过程中，由于有反馈的存在，这些必要基础的问题都可以解决：大部分内容是自足的，如果学生的理解力和悟性足够好，都可以明白。少部分内容必要的时候可以上课补充或者课后自己阅读相关材料。但是，在此，我们还是想指出来，这本书不是写给所有人看的，很多人不适合看这本书，很多人第一遍看这本书的话，我们保证是看不太懂的。但是，我们相信，就算你不

是最合适的读者，就算你很多地方没有看懂，你看了者们书之后也会有很大的收获，尤其是如果你能够坚持多看上几遍。例如，每一章的前言都是这一章的基本思想的总结，对于初次学习这一章内容的读者，这些总结都是天书，但是如果你是第二遍以及以上的读者，你可能就会有更好的体会。如果你看了本书之后，有体会、有感想、有意见、有建议，我们都希望你能够来这本书的网站 (<http://systemsci.org/books>) 参加讨论。同时，我们也尽我们的能力做到这本书的电子版本都会在她自己的网站上免费提供。

本书的每一个小节每一章都附有概念地图，还有概念地图的解说和阅读指引。我们推荐读者通过对比概念地图和正文的内容来熟悉概念地图，熟悉之后，自己来制作章节的概念地图，再来看本书附带的概念地图。如果需要学习概念地图和理解型学习，可以参考吴金闪的《概念地图学习与教学方法》[?]

学习本书的大部分内容需要：微积分（数学分析），高中阶段的牛顿力学；有帮助但是不必要的知识：线性代数，概率论，大学阶段的牛顿力学，一点点微分方程的知识。

致谢

感谢大家：导师（杨展如、Mona Berciu）、老师（裴寿镛、梁灿彬、喀兴林、漆安慎）、学生、同事（狄增如、袁强）、家人，合作研究者。

献给

黄云弟、冯倩、吴立心、吴逸兮。

第一部分
系统科学导论

第一章 引言：系统科学与科学

这一章我们企图给系统科学做一个定位，讨论其基本研究对象、学科的基本任务、基本的研究方法和思想这些问题。然后，我们会用整本书的例子来让大家进一步体会我们对这个学科的这样一个定位。

引言中的很多部分都会后面的章节中展开阐述，其中我们也会提到很多的应用性研究的例子。实际上，所有的应用性研究的例子，我们都可以把它们放到引言里面来，起到开阔大家眼界的作用。有些例子我们选择放到具体章节中，有可能是它们在技术细节和细节概念上要求更多，不太适宜放在读者学习各个章节的具体概念和技术之前来做一般的讨论。这个基本上就是引言部分的例子和其它章节内部的例子的唯一的区别。所以，从这个意义上说，引言部分就是本书的主题思想所在，本书的核心就是引言部分。当然，不理解后续章节，有可能很难真正理解引言部分。我们推荐本书的读者在浏览完本书之后，来决定怎么看；我们也鼓励看完整本书的读者回过头来，再来看一遍引言部分；读者们也可以把本书的引言部分当作系统科学的普及读物，仅仅阅读本章，然后在后续章节中选择少量合适的例子做进一步的了解。

1.1 抛一块砖：系统科学的思想、目标和定位

任何一门学科，要成熟就必须有自己的研究对象，自己这个学科的目的，核心的概念和分析方法。我们已经提到系统科学还没有成熟到能够把这些内容成体系地整理出来。在这里，我们尝试对这些问题给一个答案，不求精确，不求永恒，但求对这个学科的成熟有一定促进作用。我们认为所谓系统科学，就是把来自于属于具体科学领域，例如物理学、化学、生物学、信息科学、计算机科学等等，的思想和方法抽象和提炼出来——通常这个抽象和提炼的结果是一个数学结构，然后把这些思想和方法应用于更加广泛的

其他领域的问题的研究。也就是说，系统科学是一个来自于具体系统，同时以具体系统为最终的研究对象，但是其基本理论又不在具体系统的层次上的科学。这个定位使得这个学科非常象数学。但是，两者不是完全一样：数学，尽管本质上也来自于现实世界，只要逻辑上自洽是可以不接受实践的检验的¹；系统科学是科学，而科学最重要的特征是来源于现实世界，并接受实践的检验。因此，哲学的以及完全从心智来构造的系统科学的理论是不存在的。当然，倒过来，系统科学的哲学思考，当系统科学本身已经比较明确的时候，是可以存在的。

因此，我们把系统科学的基本任务和研究对象、目标总结为下面的列表。

目标：从具体系统中来，提炼一般概念与方法，到其他具体系统中去

定位：融合学科领域，方法与概念的迁移、创新

思想：系统元素之间的相互关系决定系统元素的内涵

主要概念与方法（系统科学目前已有的）：相互作用与网络科学，相变与临界性（也包含涌现、自组织临界），博弈与相互作用，演化与定态（也包含非线性动力学），约束下静态与动态过程的优化（也包含运筹学、控制的基本理论研究）

1.2 关于科学方法

既然系统科学首先是科学，那么我们就需要了解科学是什么，最主要的特征、思想、方法、技术、概念是什么。科学是运用科学方法来回答和解决现实世界中的问题的思想、方法和结果。这样的科学的研究的结果就是每一个具体科学领域的理论体系，或者简单称为具体科学的知识。粗略地说，针对现实世界通过科学方法得到的就是科学。既然如此，那么，这里最关键的就成了什么是科学方法。在人类探索对现实世界的问题的解决的过程中，人们形成了一些比较有共识的科学方法和科学思想。关于这个科学方法和科学思想已经有人做了一个比较全面的整理和批判（中性词，critical thinking 的意思）²。粗略来说，科学方法就是通过观察和实验对现象的特征做出确

¹关于数学与现实世界的关系，可以阅读 Timothy Gowers 的《Mathematics: A Very Short Introduction》[?]

²关于这个主题的带有强烈哲学意味的探讨，可以从 Karl Popper 的《The Logic of Scientific Discovery》找到。

定和整理，然后提出关于这个现象发生的原因以及条件等等的猜想，接着运用进一步的实验来检验这些猜想，并在得到验证的猜想的基础上通过逻辑上的推演来构造进一步的理论，然后把进一步的理论再放在观察和实验中检验的这样一个用来回答和解决现实世界的问题的方法。其中包含的观察、实验、猜想、验证、构造理论等等这些步骤，在实际运用中，往往是反复的，没有特定的顺序的。以后我们提到科学方法就是指这个通过观察、实验、猜想、验证、构造理论来回答问题的方式。人们希望通过这个科学方法得到的问题的答案是客观的真理。至于科学方法得到的答案是不是就是真理的问题，在此我们不展开讨论（实际上，Karl Popper 提出科学知识的本质特征不是通常认为的可以得到验证的真理，而是科学知识的可证伪性：也就是原则上存在被证明是错的可能性的，迄今为止又没有被证明是错的知识）。

既然系统科学首先是科学，那么科学方法就是我们学习系统科学首先需要学习的东西。有的时候，与系统科学的研究相结合，我们也把科学方法总结成一个过程：提出问题（来自于实际系统的经验），完整地描述问题（数学模型），解决问题（解析或数值计算，借助于其他学科），实践检验，提出理论，进一步解决可能是其它学科的问题。实际上，这个过程的任何步骤都可以打乱次序，有反复，不是一个线性的过程。具体的科学研究工作也可以在任何一个步骤上展开，而不一定完成整个过程。了解了这个科学方法的内涵之后，在本书的例子中，我们会尽量让大家通过具体的研究工作的例子来体会这个科学方法。

除了观察、实验、猜想、验证、构造理论，更大范围上的检验这样的具体的科学方法，科学方法还有一个非常重要的思维的层次的内涵：批判性思维³。永远不要相信任何东西，直到你搞清楚为什么你可以相信它。结论是不可信的，就算是看起来通过科学方法获得的结论：如果证据和论证过程你不能理解，那么你不应该相信它。就算是教科书中的数学定理，在你自己能够证明它们之前，也是不可信的。怀疑一切，明确逻辑和事实的起点，还有得到结论的逻辑过程，是所有的科学的最基本的要求。因此，当你给其他人呈现你自己的科学研究工作的时候，也需要面对其他人的批判性思考，尽可能完善地呈现你的证据和论证，考虑到可能的逻辑上的漏洞，收集足够有说服力的证据，而且这些证据和论证是通过科学方法获得的。学会怀疑，学会批判，学会问为什么，是每一个科学家首先需要做的事情。

³推荐 M.N. Browne 和 S. S. M. Keeley 的《Asking the right question: A guide to critical thinking》
[?]

1.3 与系统科学有关的一般的科学研究的思想

除了科学方法，还有一些从科学研究，尤其是物理学的研究，中体现出来的一般思想也是我们学习系统科学需要学习的。以下四个小节，我们举例讨论四个对于系统科学来说非常重要的科学的一般思想：如何划分系统、概念是现实的抽象、力学思想的重要性，以及普适性。

1.3.1 物理学的受力分析：按照研究目的来划分系统

系统科学很重要的很基础的一个概念就是系统。系统粗略地说就是一个研究对象的集合。系统是一个可以在具体问题的研究中指代非常明确，但是却很难给一个一般定义的概念。把系统认为是一个包含了研究对象的集合，除了哲学上的讨论，在实际研究工作中我们认为就够用了。那么，如何划分一个系统呢？一个物体的状态显然可以是和另外一个物体的状态相关的，于是整个世界才是一个有天然的定义的。那么，是不是所有的问题的研究都需要把整个世界都包含进来呢？如果是这样的话，我们就没有学科之间的区分了。学科之间的边界，尽管这个边界会不断地变化，就是研究者通过把自身所感兴趣的研究对象从整个世界“隔离”出来形成的。在任何实际问题的研究中，这样的隔离都是必须的。那么，既然这样的隔离时必须的，在实际问题的研究中，又如何划分呢？在这一点上，物理学中，受力分析的训练可以给我们很大的启发。针对同一个情境，有的时候我们会按一种方式来划分系统，有的时候我们又会按照另一种方式。这个经验在将来更一般的系统中的讨论也是有意义的。因此，我们把这个受力分析的训练作为划分系统的一种体验。

1.3.2 最优编码与信息熵：概念是从现实世界里抽象出来的

除了实用主义的角度，在那里科学就是一个可复现的解决问题的方式，很多时候，我们把科学也作为现实世界的心智模型。也就是说，我们认为科学就是对现实世界如何运作的理解。换句话说，如果给科学家们一个明确的情境，那么通过在头脑里面做演算（实际上可以通过计算机等工具），我们就可以对这个情境下的系统的行为有一个完全的把握（当然，有的时候这个完全的把握是概率性的，例如量子力学）。很多科学家，包括我自己，都对科学持这个态度：科学是现实世界的心智模型。在这个图景下面，科学最重要的特征就是对现实的抽象。所有的概念和思想，原则上，都应该来自

于对现实世界的抽象。因此，抽象——有的时候通俗地说“透过现象看本质”——的过程就是一个研究者最重要的科学研究的能力。因此，在下面的小节里面，我们也选取了一些例子来让大家体会这个抽象的过程。

1.3.3 力学思想无处不在：状态、状态变化以及状态变化的原因

尽管，科学的任务，也就是问题的解决，可以面对各种各样的问题，实际上，典型的问题就是对现象的理解、预测和控制（或者说干预可能更好）。对于理解、预测和控制这些个目的来说，最重要的信息就是关于这个现象中的主体的状态的描述，状态的变化，以及探索状态变化的原因。在物理学里面，通常称这样的问题为一个动力学问题，或者说演化的问题。从更一般的意义上说，这个能够把大多数问题归结为一个动力学问题的情况是一个非常让人惊奇的事情。在这里，我们希望给大家展示典型力学问题，以及从力学问题开始到更加一般的动力学问题的这样一个推广。

1.3.4 普适性：一个原理或者模型尽可能多地描述现象

科学还有一个潜在的要求：希望用更少的原理和模型来解释更多的现象。有的时候这个也被称为普适性的要求。例如，关于打雷、下雨、闪电原则上我们可以存在一个雷公电母的理论，只要这个理论它能够与现代的基于电荷、气压、温度、湿度等等物理量的气象理论给出大致相当的关于气象的预测、控制和理解就可以。但是，很有可能满足这些要求的雷公电母的理论必须是有户口的，也就是说，他们的脾气（例如生气和高兴的条件）是依赖于他们住在哪里。而反过来，我们却看见作为现代理论基础的电荷、气压、温度等概念，不仅仅可以用来预测、理解和控制风雨雷电，还可以被用在内燃机的汽缸、电视、电话以及各种问题上，而且它们的行为不管什么情况都遵循同一个模型（同一组方程）。因此，在引言部分，我们也设计了来自于物理学的例子来让大家体会到科学对普适性的要求。

第二章 一些具有系统科学特色的研究实例

2.1 热寂说、传统物理系统的相变与开放系统结构的产生

部分知识需要用到后面会学习到的内容，将来会理解的更好。Bénard 流，现象与计算，远离平衡产生结构，Prigogine, Haken, Bertalanffy

2.2 整体运动的激发模式

谐振子链的整体运动，相互作用系统的激发模式（孤子），Anderson

2.3 相变与临界性的例子

2.3.1 传统相互作用与相变、临界性：Ising 模型

相变的概念，引起相变的外界驱动，序参量，基本临界现象：临界慢化，长程关联，分布率，现实意义，Ising，杨振宁，李政道，Kadanoff

2.3.2 涌现、自组织临界：Bak-Tang-Wiesenfeld 沙堆

沙堆模型演示，什么是自组织，统计什么东西的规律，分布率的特殊地位，现实意义，Per Bak，汤超

2.4 相互作用，周期分叉与混沌的例子

二阶以下的位置函数势能代表没有相互作用。相互作用系统展现丰富的行为：初值敏感性，有限相体积，奇异吸引子，不可预测的决定性系统

2.5 相互作用与网络科学的例子：汉字研究

强调元素之间的关系，整体与个体的视角，现实意义

2.6 相互作用与投入产出分析：经济学、科学学、 数学

投入产出关系，一般的投入产出分析，Leontief 投入产出分析，非常简单的数学，非常广泛的应用，科学领域之间的关系，科学与技术和经济的关系

2.7 博弈与相互作用的例子

从囚徒困境到最后通牒，少体的博弈到多体的博弈，合作问题，如何用它来描述现实世界

2.8 演化与定态的例子

鱼群模型，现实意义

2.9 约束下定态与动态过程的优化的例子

2.9.1 约束下定态优化的例子：项目管理

项目管理的问题来源，问题的数学描述，现实意义，从一个项目的各个部分的依赖关系来考察整体系统的优化方法

2.9.2 动态过程优化的例子：银行存款

最优控制理论，把来源于物理系统的数学结构（最小作用量原理，Hamilton 方程）用于动态过程的优化，现实意义

第三章 概念地图与系统图示法

各种系统图示法，反馈图，一般关系图，系统的内部元素和元素之间的相互关系，举例，微分方程模型（Stella），在后面的内容中应用

3.1 如何描述一个系统：图示举例

3.2 几种特殊的图示法

3.3 概念地图图示法

第四章 为什么要学习数学物理基础

4.1 概念、数学模型的来源：前面的例子再阐述

4.2 处理方法的来源：例子

4.3 研究对象的来源：例子

第一部份中的每一节都会在下面展开一章或几章来论述。

第二部分

系统科学的数学物理基础

1. 数学是关于结构的科学，是思维的语言，物理学是在世界与数学结构之间建立起联系
2. 物理学的思想：理想化（抓住主要特征、主要问题）、模型化、数学化，追寻产生变化的原因。例如统计物理学应用于博弈论（思想与数学形式方面）。
3. 物理学关于具体学科或者具体系统的知识应用于系统科学：量子力学应用于博弈论（知识与数学形式方面），量子力学与复杂网络（数学形式方面）。
4. 物理学的计算方法，分析手段应用于系统科学：混沌与社会经济系统的可预测性（思想与知识方面），稳定性与演化博弈论（思想、知识与方法），统计物理学与经济物理学（思想方面，普适统计规律意味着深层的机制上的相似性，也有具体模型，具体分析方法方面的借鉴，临界性与自组织临界性，举例。相变在系统科学中的意义，举例）
5. 物理学问题中发展出来的数值计算和数值模拟方法应用于系统科学：最简单直接的应用，Monte Carlo 方法，Metropolis 方法，应用于多个体的系统，讨论个体之间的相互作用

第五章 线性代数

向量空间，线性变换和线性算符，线性方程与本征值问题，Dirac 抽象矢量记号，是我们这本书中所有的内容的基础。从算术计算到代数表达式，从代数表达式到具体的三维空间的变量，从具体的三维空间的变量到抽象的矢量，不一定在物理空间中的矢量，我们对于数学对象的描述是一个不断地深入和不断地抽象化、一般化的过程。同时，所有的数学计算到最后确实还需要具体的数字来代表。沟通这个从具体到抽象，再从抽象到具体的桥梁就是向量空间和线性代数。此外，数值线性代数也是其他所有的计算技术的基础之一（另外三个是 Monte Carlo、神经网络和 FFT——快速 Fourier 变换）。如果你了解过高性能计算，你就会知道所有的高性能计算的软件的基础都是 BLAS 和 Lapack 这两个线性代数包。

这部分推荐阅读材料：居余马《线性代数》，许以超《线性代数与矩阵论》，M. Artin《Algebra》，吴兆颜《高等量子力学》，喀兴林《高等量子力学》

5.1 集合与映射的语言

数学是关于思维和逻辑的学科，是思维和逻辑的语言。我们思考问题的时候，自言自语的时候，需要以我们语言（通常是母语）作为外在的形式。但是，不同国家的个人都可以用自己的语言来思考（当然，有可能有更合适的以及不太合适的问题。这里我们不讨论），说明各种不同的语言实际上都不是最终的形式，而且可以实现更深层次的统一。但是，当我们做逻辑和数学思考的时候就不一样了，我们基本上用同样的语言：数学。在这里，我们没有再一次统一与翻译的过程。这个统一的符号形式、统一的研究的形式，使得我们科学上的交流变得简单。在一个没有数学化的学科里面，很多概念的理解取决于阅读者的再解释。在艺术的层面固然是好的，但是在科学的层

面，却不是好事。

因此，学习数学，最最基本的目标是学到如何运用这个语言，让她成为所有学习者思考的基本媒介。其次，我们需要学会一些实现这个大目标（把数学作为语言）的一些基石——一些基本的概念和这些概念的表达式，例子，这些概念之间的关系。然后，我们需要学会一些计算技术。我们希望经过这部分数学的学习，除了提供学习以后其它科学的基础之外，还能够使得你对数学的观念有所改变：数学是描述这个世界的结构的语言。

我们先来看看我们最熟悉的自然数：随着描述对象的不同自然数扮演的角色至少分一下几种——记号，排序，距离（数量）。当我们描述的对象之间没有任何关系，仅仅表示它们是一个个体的时候，我们可以说某某东西是 1 号，某某东西是 3 号等等。这个就是给这些东西一个名字。这个功能称为自然数的记号功能。这个时候 3 号的物品和 1 号的物品之间没有其它更深层的关系，例如是否 1 号的重要性大于 3 号，或者 3 号的价格等于 1 号 3 倍。当物品之间确实有额外的关系的时候，他们对应着的自然数的含义也就不一样了，有了可能超过记号的其它功能。例如我们教材上的定理 1 和定理 3，第一当然表示记号功能，第二通常还表示定理 1 需要先讲，定理 3 在逻辑上有放在定理 1 后面讲的需要。这个时候自然数还起到排序的作用。不过，这个定理 1 和定理 3 之间通常不意味着它们之间的逻辑距离是定理 1 和定理 2 之间的逻辑距离的两倍。因此，记号 1, 2, 3 之间不是等距离的。如果我们的物品之间还可以度量距离而不仅仅是先后关系，那么这个时候，自然数就有了更加丰富的含义，例如当我们描述两地之间的空间距离而且在每一个 10 公里的地方树立一块牌子来供行驶在两地之间的司机估计距离的参考的时候，我们的每一个牌子不仅表示记号，序号，还表示 3 号牌子和 1 号牌子之间确实是 20 公里。

在上面的例子中，我们的集合是相同的：自然数 \mathbb{N} 。但是，当用来描述物品的时候，在每一个对象集合上的含义是不一样的。那么，这个不一样是如何造成的呢？数学的语言告诉我们，这个不同之处在于集合上定义了不同的映射。第一种情况，这个集合的元素之间没有任何映射，也就是通常语言说的没有任何额外的关系。在第二个例子中，集合上定义了一种叫做序的映射（实际上是从 $\mathbb{N} \otimes \mathbb{N}$ 到 $\{1, -1\}$ 的映射，从 $\mathbb{N} \otimes \mathbb{N}$ 中取一个元素 (m, n) ，如果 $m > n$ 大于关系成立则取 1，否则取 -1）。这个时候，集合加上序的映射，就成了有序集，能够用来描述存在更深刻关系的物品了。在第三个例子中，集合上还定义了一种更复杂的映射元素之间的减法。减法这个运算是建

立在 $\mathbb{N} \otimes \mathbb{N}$ 到 \mathbb{N} 之间的。以后我们会介绍更多的细节，这里只要体会到集合与映射是数学的基本语言就够了。

集合本身在数学的逻辑框架之中是不能通过其它数学结构来定义的，就是表示一个群体的对象。有了集合，我们来定义映射。

定义 5.1.1 映射：集合 A 与集合 B 之间的关系 f ，如果对于任意一个集合 A 的元素 α 存在一个集合 B 的元素 β ，使得 $f(\alpha) = \beta$ ，就称为集合 A 与 B 之间的映射。

这个关系利用通用的数学符号体系可以表达成 $f|_{A \rightarrow B} : \forall \alpha \in A, \exists \beta \in B \text{ s.t. } f(\alpha) = \beta$ ，其中 $f|_{A \rightarrow B}$ 表示集合 A 与集合 B 之间的关系 f ， \forall 表示任意， \exists 表示存在， \in 表示属于，*s.t.* 表述于是（英文的 *such that*）。有了集合与映射，我们就可以建立更丰富的数学结构了。

集合的运算，交、并、补，集合的卡氏积，幂集，都是非常重要的概念。我们假定读者已经学过，以后我们用得到的时候不再介绍。读者可以利用 Wikipedia 的资源（google “set” 或者访问 Wikipedia 的页面 [http://en.wikipedia.org/wiki/Set_\(mathematics\)](http://en.wikipedia.org/wiki/Set_(mathematics))）或者其它数学教科书 [?] 回忆一下这些概念。

第一个我们要定义的映射是示性函数或者称为指示函数（indicator function），记为 $I_A(x)$ 或者直接 I_A 。定义这个映射需要三个集合大集合 X ， X 的某个子集 A ，以及集合 $\{0, 1\}$ 。 I_A 要完成的事情是：给定任何一个 X 中的元素 x ，我们想知道这个 x 是否属于 A 。 I_A 是一个从集合 X 到集合 $\{0, 1\}$ 的由子集 A 所诱导出来的如下映射，

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} . \quad (5.1)$$

将来很多时候我们要用到这个映射。

5.2 从群到线性空间

单纯地一个集合 A 到 A 上的映射是平庸的，最多只能够建立集合元素之间的对应关系。一个稍微不平凡一点的映射，能够描述更丰富的现实的映射，是集合自身的卡氏积到集合的映射，或者集合的幂集到自身的映射。这里我们讨论前者。我们将从这个映射开始构建现代数学的大厦的一部分：

线性空间和线性映射。从计算的角度来说,线性空间上的运算是所有的其它数学计算的基础。从数学在其他学科的应用来说,线性代数和微积分是数学成为其他很多科学的基础和基本语言的原因之一。从数学理论的角度来说,线性空间的结构和分析方法是其它现代数学的基础。在这个意义上,微积分,或者说分析,都赶不上线性空间对于数学的重要性。

5.2.1 群、半群

集合 A 与 B 的卡氏积 $A \otimes B$ 是指这样的集合 $A \otimes B \triangleq \{(\alpha, \beta), \forall \alpha \in A, \beta \in B\}$ 。从一个集合 G 开始,考虑一个从 $G \otimes G$ 到 G 的映射 \oplus 。这个映射,记为 $\oplus|_{G \otimes G \rightarrow G}$ 在一些最基本的要求下成了整个数学的基础:群。第一,我们要求封闭性,也就是说 $f \oplus g \in G, \forall f, g \in G$ 。如果封闭性不满足,就连映射都构不成了。当然,实际上在描述现实的时候(以后我们会看到这样的例子),如果某些元素经过 \oplus 运算以后不再属于这个集合,我们经常扩充这个集合 G 使得运算封闭。在这个意义上运算,也就是关系,比集合的元素重要。关系,也就是运算,才是对现实世界的描述,而不是集合。集合在没有定义运算的时候,是平庸的。第二,我们要求集合与运算满足如下三条。

定义 5.2.1 群: 集合 G 配上映射 $\oplus|_{G \otimes G \rightarrow G}$ 称为群,如果满足:(1) 结合律, $\forall f, g, h \in G, (f \oplus g) \oplus h = f \oplus (g \oplus h)$; (2) 集合存在单位元 $\underline{0}$, *s.t.* $\underline{0} \oplus g = g, \forall g \in G$; (3) 任何元素 $g (\forall g \in G)$ 存在逆元 $f (\exists f \in G)$, *s.t.* $f \oplus g = \underline{0} = g \oplus f$ 。

这个时候,映射 $\oplus|_{G \otimes G \rightarrow G}$ 称为群乘。如果只有条件 (1) 得到满足,则称 G (配上 \oplus) 为半群。群这个概念是对很多数学运算进行了抽象,寻找这样的运算的最基本的要求得到的。以后我们会了解更多这个定义的含义。这里,我们先看见我们所熟悉的整数配上加法运算构成了群,有理数去掉 0 以后配上乘法构成了群。除了非常特殊的情况,对现实世界的物体的操作,基本上我们会保留某一个层次的不变性。例如翻转硬币的时候硬币还得是硬币,改变的仅仅是硬币哪一面向上。再例如考察一个天体的运动,改变的是天体的位置而不改变天体的构成。再如考察一个化学反应,我们改变的是电子的构型,而不改变原子核内的结构。再如考察核裂变反应的时候,我们改变的是核子的构型,而不改变核子自身。如果我们把所有的改变这个物体状态又保持某个层次的不变性的操作看成一个集合,那么这个集合显然对于连续的两个操作的结果是有定义的,也就是说连续两个操作的结果必须是这个

集合中的某一个操作。这个就是群乘映射的原型。这也是为什么封闭性是最重要的要求的原因。有了这个封闭性之后，我们就可以把对物体状态的研究转变成为对操作的研究：所谓物体任意时候的状态，就成了从某一个固定的初始状态出发，每次作用一次这些操作算符中的一个。也就是，算符的序列完全定义和替代了物体状态的序列。然后，我们将来总是要作进一步的计算的，例如 g, g^2, g^3, g^4 等等，如果我们没有结合律，那么 g^3 就可以解释成 $g^2 \oplus g$ ，也可以解释成 $g \oplus g^2$ ，两者就有可能不一样。这个时候，这些基本的运算就会变得非常复杂。于是我们看见，封闭性和结合律，是很多运算非常重要的性质。实际上，我们会看见，正是一个对象上的这个操作算符的集合的结构，大多数时候决定了这个物体的行为。在这个意义上说，群论是所有物理理论的基础：所有的物理对象都有一个状态，对状态的操作的集合，都构成群，而这个群的结构决定了这个物理对象的基本行为。

例 5.2.1 (\mathbb{R} 实数加法群) \mathbb{R} 实数加法群：集合 \mathbb{R} ，群乘定义为实数加法。封闭性，结合律，零元 (0) 和逆元 (x 与 $-x$) 都存在。

例 5.2.2 ($\mathbb{R}/\{0\}$ 实数乘法群) $\mathbb{R}/\{0\}$ 实数乘法群：集合 $\mathbb{R}/\{0\}$ ，群乘定义为实数乘法。封闭性，结合律，零元 (1) 和逆元 (x 与 $\frac{1}{x}$) 都存在。

实数集合上的这两个运算是非常重要的。实数集合上由于同时有这两个运算的存在，其结构要比群丰富的多，因此被称为域。这里给出域的定义，但是我们不详细讨论。

定义 5.2.2 域：集合 \mathbb{F} 上定义了加法运算 $+$ $|_{\mathbb{F} \otimes \mathbb{F} \rightarrow \mathbb{F}}$ 和乘法运算 \times $|_{\mathbb{F} \otimes \mathbb{F} \rightarrow \mathbb{F}}$ ，其中 $\mathbb{F}, +$ 和 $\mathbb{F}/\{0\}, \times$ 构成交换群，其中 0 是加法运算的单位元，而且加法和乘法之间满足分配律： $(a + b) \times c = a \times c + b \times c$ 。

可以验证有理数、实数、复数都满足这个条件。以后在乘法运算中，我们经常省略乘号 \times 。如果我们把加法运算放在有限大小的整数集合上，例如从 -100 到 100 ，群乘的封闭性就得不到满足。但是，非常简单地我们就可以扩大这个集合，使得其满足封闭性。例如，通过把上限加上 1 ，并且不停地加下去，把得到的所有的数以及它们的逆元合并上原来的集合构成一个新的集合。如果我们把域关于乘法的要求放在整数集合上，我们就会发现整数乘法运算的逆元不在整数集合之中，于是我们才需要把整数扩大成有理数。

例 5.2.3 Z_2 群：考虑一个理想硬币的状态， $\{\uparrow, \downarrow\}$ ，分别代表正面与反面。从现实经验来说，在不破坏这个硬币的结构的要求下，这个系统的可能

操作包含 $\{I, X\}$ 分别代表不翻转和翻转。例如翻转算符满足 $X \uparrow = \downarrow$, 而且 $X \downarrow = \uparrow$ 。不翻转算符保持硬币的状态。于是, 我们可以发现集合 $\{I, X\}$ 存在群乘 $I \oplus I = I, I \oplus X = X, X \oplus I = X, X \oplus X = I$ 。可以验证这个 \oplus 确实满足群乘的要求。这个具有两个元素的群称为 Z_2 群。这个两两元素之间乘法的结果称为乘法表。

在上面的例子中, 我们的群的元素是一个一般的操作, I, X , 而不是整数、实数等可计算的对象。目前为止, 我们通过查询乘法表可以知道这两个元素之间的所有的运算的结果。但是, 我们的每一次运算都必须依赖于查询乘法表, 而不能利用我们已经熟悉的整数、实数、复数等等计算规则简单明确的数学对象。由于群的元素的一般性和抽象性, 有一个关于群的理论就非常重要, 那就是群的表示理论。群 G 的一个表示, 就是为 G 中的一个群元 g 找到一个矩阵的形式 \hat{g} , 使得 $f \oplus g$ 的矩阵形式正好是 $\hat{f}\hat{g}$, 其中 $\hat{f}\hat{g}$ 就是矩阵乘法。如果有了这个表示, 我们抽象的计算就都可以用具体的数学计算 (在这里是矩阵运算, 注意实数也可以看作是矩阵) 来完成了。这个会使我们的研究大大简化。对于已经学习过线性代数的读者, 我们布置一道作业题。没有学习过线性代数的读者, 请在本章结束以后回过头再来完成这个作业题。

作业 5.2.1 一个正三角形 $\triangle ABC$, 经过 BC 边上的中位线翻转之后成了三角形 $\triangle ACB$ 。同样的, 这个三角形还可以围绕另外两条中位线做翻转。然后, 可以考虑连续的这样两个翻转组合起来的效果。请找出来这些翻转与这些翻转的所有的组合的集合, 并计算出乘法表 (任意两个翻转的组合相当于哪一个翻转), 然后用矩阵来表示这些翻转。

也许你需要注意, 两个翻转合起来的效果和其顺序有关, 反过来的效果和原来的效果可能不一样, 也就是 $f \oplus g$ 不一定等于 $g \oplus f$ 。如果群乘可交换, 这样的群称为交换群。

为了理解我们下一步要学习的线性空间, 我们关于群论的知识只需要这么多, 尽管更多一点的群论的知识对于学习数学、物理都是有意义的。

5.2.2 线性空间的矢量

也许矢量对于你来说意味着一个从原点出发到空间某一个位置的位置矢量, 或者在力学里面在质点 (力学里面把物体去掉形状大小质地颜色等等各种属性之后留下的代表物体位置的质量为 M 的点) 上放上一个箭头表

示其速度或者受力的矢量，或者是线性代数课程告诉你的一个矢量就是在某一套坐标系下写成各个分量形式的 $v = (v_1, v_1, \dots, v_N)^T$. 确实，这些都是矢量的表现形式。但是矢量的概念要远远比这些直观的矢量具有一般性。例如，实数域上的函数 $\{f(x)\}$ 配上合适的运算，也可以成为矢量。物体的颜色的集合，配合上合适的运算也可以成为矢量。上面例子中硬币的状态，配上合适的运算，再给运算找到现实物理系统的对应（就是运算出来的结果，例如 $\uparrow \oplus \downarrow$ ，也是某一个物理系统的状态），也可以成为矢量。

那么一个矢量成为矢量的最重要的特征是什么呢？矢量能够表示方向，能够被放大缩小，而且放大缩小就意味着在这个矢量代表的方向上的长度发生了变化。两个矢量放在一起能够做叠加运算。如果要制作一些适量的模型的话，只要我们做出来的满足上面两点要求，基本上就能够看成矢量的模型。例如，矢量自身我们用无限可伸缩的电视天线（实际上只能一节一节伸缩，但是我们可以要求节足够小足够密。实际上总长度也有限），矢量的加法可以用头尾相连的方式（也就是平行四边形法则）。当然，这个模型实际上是基于空间矢量的概念的，不够一般性。这里，我们通过这个模型，仅仅想告诉大家，矢量的放大缩小运算和两个矢量的加法，是矢量空间最重要的性质，而如何定义这两个运算使得矢量具有线性性（以后我们会看见，这个线性性就是指矢量的这两个运算之间满足线性分配律）就是整个矢量的概念的核心。

定义 5.2.3 数域 F 上的矢量：集合 V 上定义了加法运算 $+|_{V \otimes V \rightarrow V}$ 和数乘运算 $\cdot|_{F \otimes V \rightarrow V}$ 满足下列属性：

- (1) 加法运算“+”构成交换群；
- (2) 加法运算“+”与数乘运算“ \cdot ”满足双向分配律： $a \cdot (\mu + \nu) = a \cdot \mu + a \cdot \nu$,
 $(a + b) \cdot \mu = a \cdot \mu + b \cdot \mu$;
- (3) F 上的乘法单位元与 V 上的运算的统一性 $1 \cdot \mu = \mu, \forall \mu \in V$;
- (4) 混合结合律， $ab \cdot \mu = a \cdot (b \cdot \mu)$ ，这实际上是保证矢量与数的数乘和数域上的乘法的一致性。

注意，在这里“数域”保证了数之间的加法和乘法要满足作为一个“域”的基本要求。这里，我们用拉丁字母表示数域 F 中的元素，用希腊字母表示矢量集合 V 中的元素。当我们利用这个约定之后，数乘运算的符号“ \cdot ”就可以被省略，数乘的数既可以被放在左边，也可以被放在右边。在这个定义

中, 双向分配律是很重要的, 是线性性的保证。所谓一个运算的线性性就是当被操做的对象发生了线性的变化, 那么操作结果也相应会发生线性的变化: $\Gamma(a\mu + b\nu) = a\Gamma(\mu) + b\Gamma(\nu)$ 。这里这个 Γ 运算可以是数乘自己。例如 $c \cdot$ 。为了保持数域上的运算和向量集合上的运算的一致性, 我们还需要保证 F 上的加法零元与 V 上的加法零元的统一性 $0 \cdot \mu = O, \forall \mu \in V$ 。下面我们会证明, 这一条实际上是可以有其它要求推出来的。

例 5.2.4 证明 $0 \cdot \mu = O, \forall \mu \in V$: 我们先证明 V 上的加法零元是唯一的。假设有多个零元, 取其中的任意两个 O_1, O_2 , 我们按照加法构成交换群得到, $O_1 = O_1 + O_2 = O_2$ 。因此, 这些零元都是想同的。接着我们证明 0μ 就是其中一个零元。 $\forall \mu \in V, 0\mu = 0\mu + O = 0\mu + (\mu - \mu) = (0 + 1)\mu - \mu = \mu - \mu = O$ 。

最后一行, 我们首先利用了 V 中加法零元的性质, 然后利用了任何元素和它的逆元的和等于零 (逆元的存在性), 然后利用了 V 中的加法结合律, 接着利用了 F 中加法零元的性质, 最后又利用了 V 中加法逆元的性质。可以看到这个简单的证明实际上利用了很多前面给出的向量空间的定义的性质。实际上, $1 \cdot \mu = \mu, \forall \mu \in V$ 就没法用这些性质证明。鼓励你试试。直观的原因是如果我们强行地将 V 中的所有元素放大 2 倍以后再看成新的集合 V' , 这个集合还是向量集合, 但是在这个新的向量集合上, 就没有 $1 \cdot \mu = \mu, \forall \mu \in V'$ 而是 $0.5 \cdot \mu = \mu, \forall \mu \in V'$ 。也就是说数乘这个运算实际上把矢量放大大多少倍是无所谓的, 必须强行规定一个。当然, 这个时候一个好的规定的选择就是使得 F 上的乘法单位元的性质得到保持。

作业 5.2.2 证明 $aO = O$ 以及 $(-1)\mu = -\mu$ 。

例 5.2.5 (N 维欧氏空间矢量) N 维欧氏空间中所有从原点出发的位置矢量构成线性空间。其中数乘和加法可以采用分量的形式定义。因为可以写成分量的形式, 验证这个空间是向量空间非常简单。

例 5.2.6 (2 维欧氏空间可微函数在某点的切线) 2 维欧氏空间中所有经过原点的可微函数在原点的切线所代表的矢量构成线性空间。这个矢量可以写成分量的形式, 例如曲线 $(x(t), y(t))$, 曲线在原点的切线为 $(\frac{dx(t)}{dt}, \frac{dy(t)}{dt})$ 在原点取值。验证这个空间是向量空间也不难。

例 5.2.7 (1 维欧氏空间的无穷可微函数) 1 维欧氏空间上所有无穷可微函数构成线性空间, 其中加法数乘的定义就是通常的加法与数乘。首先,

我们来看一看加法的封闭性。取任意两个无穷可微函数 $f(x), g(x)$ 显然 $af(x) + bg(x)$ 还是无穷可微的，所以封闭。然后我们可以简单验证零元和逆元的存在性，加法结合律和双向分配律也不是问题。混合结合律，单位元统一性也很简单。

例 5.2.8 (1 维欧氏空间的平方可积函数) 1 维欧氏空间上所有平方可积函数构成线性空间，其中加法数乘的定义就是通常的加法与数乘。零元和逆元的存在性，加法结合律和双向分配律、合结合律，单位元统一性都比较简单。我们来看看加法的封闭性。取任意两个平方可积函数 $f(x), g(x)$ ，我们需要证明 $af(x) + bg(x)$ 还是平方可积的，所以封闭。证明过程大致如下 $(af(x) + bg(x))^2 = a^2f^2(x) + b^2g^2(x) + 2abf(x)g(x) \leq 2a^2f^2(x) + 2b^2g^2(x)$ ，后者可积，因此前者也可积。

为了以后可以查阅，我们把域 F 上线性空间 V 的定义中的所有要求重新写在这里。而现在开始，把满足这些要求的集合以及集合上的运算就称为向量空间或者线性空间。

1. F 本身是一个域：元素的加法乘法有定义，且满足域的要求
2. 加法与数乘的封闭性： $\mu + \nu \in V, a\mu \in V, \forall \mu, \nu \in V, a \in F$
3. 加法交换律： $\mu + \nu = \nu + \mu, \forall \mu, \nu \in V$
4. 零元存在： $\mu + O = \mu, \forall \mu \in V$
5. 逆元存在： $\forall \mu \in V, \exists \nu \in V. \text{ s.t. } \mu + \nu = O$
6. 加法结合律： $(\mu + \nu) + \eta = \mu + (\nu + \eta), \forall \mu, \nu, \eta \in V$
7. 双向分配律 $a(\mu + \nu) = a\mu + a\nu, (a + b)\mu = a\mu + b\mu, \forall \mu, \nu \in V, a, b \in F$
8. 数乘单位元一致性： $1\mu = \mu, \forall \mu \in V, \forall \mu \in V$
9. 混合结合律一致性： $(ab)\mu = a(b\mu), \forall \mu \in V, a, b \in F$

我们已经看到，这个线性空间的定义就是一个交换群加上一个数域，然后定义好相容的一致两者结合的操作：数乘。

到现在为止，我们的矢量之间只能做加法与数乘两种运算，我们所熟悉的矢量之间的夹角和点积（也称作内积），适量的长度，还没有定义。下面，我们来看一下矢量长度和矢量之间的内积的抽象定义是什么。以后我们会

看到, 适量的内积非常重要, 有了内积我们就可以定义长度和夹角, 而长度和夹角往往是我们能够测量的量。一个抽象的矢量, 例如一个无穷可维的函数, 例如一个平方可积的函数, 经常是难以直接测量的。

5.2.3 矢量内积

一个内积的是指从 $V \otimes V$ 到 F 的映射, 通常记为 (\cdot, \cdot) , 其中两个 \cdot 就是留给两个矢量的位置, 满足以下要求:

1. 封闭性: $(\mu, \nu) \in F, \forall \mu, \nu \in V$
2. 线性性: $(\mu, a\nu + b\eta) = a(\mu, \nu) + b(\mu, \eta), \forall \mu, \nu, \eta \in V, a, b \in F$
3. 长度大于等于零: $(\mu, \mu) \geq 0, \forall \mu \in V$, 取等号当且仅当 $\mu = O$
4. 复共轭: $(\mu, \nu) = (\nu, \mu)^*, \forall \mu, \nu \in V$

前面三条的动机很明确, 最后一条的目的需要好的反例才能说明白。在我们找到合适的反例之前, 只能先记住了。前面三条之中, 线性性是非常重要的和独特的。将来我们会在 V 上的线性算符与内积之间建立非常密切的联系, 而这些联系的基础都是内积的线性性。当我们把线性性与复共轭结合之后, 我们看到, 内积对于左边的矢量的线性性与右边的矢量的线性性不太一样:

$$(a\nu + b\eta, \mu) = (\mu, a\nu + b\eta)^* = a^*(\nu, \mu) + b^*(\eta, \mu). \quad (5.2)$$

当然, 如果数域 F 中, $a^* = a, \forall a \in F$, 那么实际上内积对左边和右边的矢量的线性性就是相同的。当 F 中一个元素的复共轭不再回到这个元素, 这就不一样了。

有了内积这个映射之后, 我们来看 (\cdot, μ) 和 (μ, \cdot) 是什么。我们发现, 这两个东西都可以看成是一个从 V 到 F 的映射, 它需要一个 V 的元素 ξ 的输入, 产生一个 $(\mu, \xi) \in F$ 的输出。而且, 这个映射与 V 的元素是一一对应的: 有一个 $\mu \in V$ 就有一个 (\cdot, μ) 和一个 (μ, \cdot) 。

记所有的这样的 (\cdot, μ) 和 (μ, \cdot) 的集合为 V^{**} 和 V^* (这里符号上有不完全严格的地方)。我们发现 V 与 V^{**} 以及 V^* 之间存在着——到上的可逆映射, 也就是有一个 V^{**} (V^*) 中的元素就有一个 V 中的元素。对于 V^{**} (V^*) 中的元素, 我们来看一看保持与 V 中的元素的加法与数乘一致的它自己的加法与数乘是什么样子。

我们说 $a(\cdot, \mu) + b(\cdot, \nu) = (\cdot, a\mu + b\nu)$ 。这个定义第一满足所有的线性空间的要求，第二满足与 V 上的加法与数乘的一致性。首先，这个运算是封闭的： $a\mu + b\nu \in V$ 。其次，我们来检验一下分配律。对于任意 V 中的元素 η ，我们有 $a(\eta, \mu + \nu) = a(\eta, \mu) + a(\eta, \nu)$ ，因此， $a(\cdot, \mu + \nu) = a(\cdot, \mu) + a(\cdot, \nu)$ ，因此满足一侧的分配律。另外一侧也能类似证明。其它性质就不一一证明了。在这里我们看到，实际上 V^{**} 上的加法和数乘实际上就是 V 的加法和数乘。我们说 V^{**} 不仅与 V 一一对应，连所有的运算都一模一样。于是，我们发现，实际上 (\cdot, μ) 就是 μ 的另一个写法，也就是说 V^{**} 就是 V 。

我们再来看看 V^* 与 V 的关系。我们定义 V^* 中的加法与数乘： $a(\mu, \cdot) + b(\nu, \cdot) = (a^*\mu + b^*\nu, \cdot)$ 。这个定义第一满足所有的线性空间的要求，第二满足与 V 上的加法与数乘的一致性。首先，这个运算是封闭的： $a^*\mu + b^*\nu \in V$ 。其次，我们来检验一下分配律。对于任意 V 中的元素 η ，我们有 $a(\mu + \nu, \eta) = a(\mu, \eta) + a(\nu, \eta)$ ，因此， $a(\mu + \nu, \cdot) = a(\mu, \cdot) + a(\nu, \cdot)$ ，因此满足一侧的分配律。另外一侧也能类似证明。其它性质就不一一证明了。其次，我们来看，如果我们把这个 V^* 中的矢量映射回 V 中的矢量，我们看看加法和数乘是不是就是 V 的加法和数乘。为了解决这个问题，我们考虑 V^* 与 V^{**} 的关系，反正 V^{**} 就是 V 。我们发现 $(\mu, \cdot) \in V^*$ 与 $(\cdot, \mu) \in V^{**} = V$ 的关系就是把矢量从放在内积算符的左侧改成放在内积算符的右侧。同时由于这个变换，矢量空间 V^* 中的元素的数乘需要把数变成复共轭之后再与矢量乘在一起。这个意义上，我们把 $V^{**} = V$ 中的元素称为右矢量，而把 V^* 中的元素称为左矢量。为了简化记号，记右矢量，左矢量分别为

$$(\cdot, \mu) \triangleq |\mu\rangle, (\mu, \cdot) \triangleq \langle \mu|. \quad (5.3)$$

同时内积的记号也可以相应记为

$$(\nu, \mu) \triangleq \langle \nu | \mu \rangle. \quad (5.4)$$

有了这个记号，称为 Dirac bra-ket 记号，之后，我们发现，其实抽象矢量与数域 F 中的数的差别更明显了，数就是通常的记号 a, b, c ，矢量就是加了尖括号的 $|\mu\rangle, |\nu\rangle$ 等等，而不依赖于拉丁字母和希腊字母的约定。所以，从现在开始，我们所有的矢量都用 Dirac 符号来表示。

这种用加括号明确标出矢量符号的方式还有一个好处：明确区分矢量与矢量在某坐标系下的分量。例如，以前我们经常以分量的形式来写下面的

矢量,

$$\begin{bmatrix} a \\ b \end{bmatrix} = a \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (5.5)$$

这个时候, 我们默认 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 与 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 是两个坐标基矢, 然后整体矢量 $\begin{bmatrix} a \\ b \end{bmatrix}$ 是这个两个基矢叠加起来的结果。这当然是对的, 但是, 换一个理解会更好更明确: $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 与 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 是两个坐标基矢在某一套坐标系下的投影分量形式, 而 $\begin{bmatrix} a \\ b \end{bmatrix}$ 是这个合起来的矢量在同样的坐标系下的分量形式。例如, 记这个坐标系的两个正交归一 (任意两个基矢 $|i\rangle$ 和 $|j\rangle$ 之间满足 $\langle i|j\rangle = \delta_{ij}$) 基矢分别为 $|0\rangle$ 和 $|1\rangle$, 则

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow [|0\rangle \quad |1\rangle] \begin{bmatrix} 1 \\ 0 \end{bmatrix} = |0\rangle. \quad (5.6)$$

于是, 一般地,

$$\begin{bmatrix} a \\ b \end{bmatrix} \Rightarrow [|0\rangle \quad |1\rangle] \begin{bmatrix} a \\ b \end{bmatrix} = a|0\rangle + b|1\rangle. \quad (5.7)$$

在这个符号系统里面, 矢量就是抽象的矢量, 通常线性代数里面的矢量记号是矢量在某坐标系下的分量形式。例如内积不再是坐标分量之间的运算, 而是矢量之间的运算,

$$\begin{aligned} & (a^* \langle 0| + b^* \langle 1|) (a|0\rangle + b|1\rangle) \\ &= a^* a \langle 0|0\rangle + a^* a \langle 0|1\rangle + b^* a \langle 1|0\rangle + b^* b \langle 1|1\rangle \\ &= a^* a + b^* b. \end{aligned} \quad (5.8)$$

也就是说运算不在分量的层次, 而在抽象矢量的层次。这个符号系统还有一个好处: 没有抽象矢量记号之前, 矢量的定义是从坐标变换下的分量的变换形式定义的, 说坐标变换下分量不变叫做标量, 分量如何如何变化的叫作矢量, 如何如何如何变化的叫做张量。这当然, 也是正确的定义。但是, 就好像你识别出一个朋友的脸, 你肯定不是通过他/她的脸如何变化来识别的, 而是通过某些内在特征来识别的。如果通过坐标变换下分量的变换形式来定义矢量, 我们就不能写下一个矢量的等式, 而只能够写下分量的等式, 例

如对于新坐标的分量形式 v' 通过某一个相似变换与旧坐标下的形式联系起来,

$$v' = S v. \quad (5.9)$$

而我们现在, 抽象矢量的记号告诉你, 抽象矢量是有定义的, 我们能够写下来矢量的等式

$$\sum_j v'_j |e'_j\rangle = |v\rangle = \sum_i v_i |e_i\rangle, \quad (5.10)$$

于是, 自然的我们就有,

$$v'_k = \sum_j v_j \langle e'_k | e_j \rangle, \quad (5.11)$$

于是,

$$S_{kj} = \langle e'_k | e_j \rangle, \quad (5.12)$$

而且矢量分量的定义就是矢量与基矢时间的内积,

$$v_k = \langle e_j | v \rangle. \quad (5.13)$$

抽象矢量记号就是那个可以写下矢量等式的内在属性, 而分量形式的记号是做不到的。有了这个记号, 我们终于可以说: 矢量在坐标变换下是不变的, 标量和张量在坐标变换下也是不变的, 改变的仅仅是分量形式。这个抽象矢量记号对于我们后面的内容非常重要, 所有的矢量将来都会写成这个抽象形式。只有这样, 我们认为, 才能够形成足够深刻的认识: **矢量不是数**, 数仅仅是矢量在某坐标系下的分量形式。

有了这个抽象记号以后, 我们还可以讨论以下符号的意义: 给定 $\mu, \nu \in V$, 定义

$$\hat{A}_{\mu\nu} = |\mu\rangle \langle \nu|. \quad (5.14)$$

用通常的线性代数的语言, 这个东西是一个矩阵, $|\mu\rangle \langle \nu|$ 表示一个列向量 (右矢) 乘上一个行向量 (左矢), 得到一个矩阵。按照我们新的记号系统, 我们来看如果我们把这个东西作用在 V 的元素上, 我们的到什么。

$$\hat{A}_{\mu\nu} |\eta\rangle = |\mu\rangle \langle \nu | \eta \rangle \in V. \quad (5.15)$$

所以, $\hat{A}_{\mu\nu}$ 是 V 到 V 的一个映射。同样的, 如果我们把左矢放在这个东西的左边, 我们看到 $\hat{A}_{\mu\nu}$ 是 V^* 到 V^* 的一个映射。如果我们考虑把所有的这样的映射放在一起构成的集合 $\{|\mu\rangle\langle\nu|, \forall\mu, \nu \in V\}$, 我们甚至可以讨论这个集合上的与 V 相一致的加法与数乘。这个就是 V 上的线性算符构成的集合。我们来看一下这个集合的结构。

5.3 线性算符

这一节我们主要解决如何在一个线性空间中选择合适的基矢的问题。这个问题与线性空间上的线性算符是连在一起的。我们先讨论线性算符。

定义 5.3.1 线性算符: $\hat{A}|_{V \rightarrow V}$ 是一个从线性空间 V 到 V 的线性映射满足如下线性性要求:

$$\hat{A}(a|\mu\rangle + b|\nu\rangle) = a\hat{A}|\mu\rangle + b\hat{A}|\nu\rangle. \quad (5.16)$$

记所有的这样的线性算符的集合为 $\mathcal{O}(V)$ 。

可以看到上一节末尾提到的算符 $\hat{A}_{\mu\nu} \in \mathcal{O}(V)$ 。如果已经在 V 里面确定一组 N 个 (先假设 N 有限, 无限的情况另说) 正交归一基矢 $\{|j\rangle\}$, 那么所有的矢量都可以在这一套基矢下展开, 那么给定任何一个算符 \hat{A} , 只要知道 \hat{A} 如何做用在所有的基矢上, 就知道了 \hat{A} 如何作用在所有的矢量上:

$$\hat{A}|\mu\rangle = \hat{A} \sum_j \mu_j |j\rangle = \sum_j \mu_j \hat{A}|j\rangle. \quad (5.17)$$

也就是说, 我们只需要知道 $|A_j\rangle = \hat{A}|j\rangle$ 这个算符。如果我们把这个矢量也做同样基矢下的展开, 其分量为,

$$|A_j\rangle = \sum_k A_{kj} |k\rangle. \quad (5.18)$$

由于某些我们将来会明白的原因, 这里我们把分量记为 A_{kj} 而不是 A_{jk} 。总结一下, 如果知道了所有的 A_{kj} 我们就知道了 $|A_j\rangle$, 于是相当于知道了 \hat{A} 作用在任何一个矢量上的结果, 也就是知道了 \hat{A} 。这个时候, 我们把所有的这些 A_{kj} 记录成一个矩阵,

$$A = (A_{kj})_{N \times N}. \quad (5.19)$$

我们发现，如果我们得到了这个矩阵的所有元素，那么我们就实际上知道了算符 \hat{A} 。从这个意义上说，矩阵就是我们的算符，在某一套坐标下的表现形式。

这里，我们为了以后计算形式上方面，证明一个恒等式。有了这个恒等式，上面的语言和计算结合的说明可以用一行公式写出来。对于一套完整的正交归一基矢 $\{|j\rangle\}$ ，

$$I = \sum_j |j\rangle\langle j|. \quad (5.20)$$

这个利用原来线性代数的语言很好理解，在一套正交归一基矢下，只有对角的而且对角元素值为 1 的矩阵，必然是单位矩阵。现在证明如下。对任意 V 中的矢量 $|\mu\rangle$ ，我们证明这个 I 作用的结果是 $I|\mu\rangle = |\mu\rangle$ ：

$$I|\mu\rangle = \sum_j |j\rangle\langle j|\mu\rangle = \sum_j \mu_j |j\rangle = |\mu\rangle. \quad (5.21)$$

有了这个等式之后，我们来重新算一下算符 \hat{A} 的分量形式，

$$A = \sum_{kj} |k\rangle\langle k|A|j\rangle\langle j| = \sum_{kj} A_{kj} |k\rangle\langle j|. \quad (5.22)$$

从这个表达式，我们非常清楚地看到矩阵 A 与算符 \hat{A} 的联系：矩阵元素是抽象算符在这一套基矢下的分量形式。进一步，由于任何一个线性算符都可以用这堆 $N \times N$ 个算符 $\{|k\rangle\langle j|\}$ 线性叠加，如果我们定义算符集合 $\mathcal{O}(V)$ 上的加法运算就是这个加法，数乘就是这个数乘（公式 (5.22) 同时代表了加法和数乘的定义）的话，我们可以证明 $\mathcal{O}(V)$ 构成了 $N \times N$ 维的矢量空间。定义要求的各个条件这里就不检验了。我们甚至可以定义这个空间上的内积，例如

$$(\hat{A}, \hat{B}) = \text{tr}(\hat{A}^\dagger \hat{B}), \quad (5.23)$$

其中一个算符的共轭 A^\dagger 的定义我们下面会讨论。在定义之后，读者可以检验这个内积满足所有的内积的定义的要求。

5.3.1 共轭算符

\hat{A} 是 V 到 V 的映射，由于 V 到 V^* 存在一一到上的映射，记为 Φ ，于是，通过 \hat{A} ，我们可以构造一个 V^* 到 V^* 的映射，记为 \hat{A} 的共轭算符（对

于无限维空间，我们需要区分共轭算符与伴算符。这里不区分) \hat{A}^\dagger

$$\langle \mu | \hat{A}^\dagger = \Phi(\hat{A}|\mu\rangle) = \Phi(\hat{A}\Phi^{-1}(\langle \mu |)) = \langle \mu | \Phi^{-1} \circ \hat{A} \circ \Phi, \quad (5.24)$$

也就是

$$\hat{A}^\dagger = \Phi^{-1} \circ \hat{A} \circ \Phi. \quad (5.25)$$

注意，这里的计算顺序是从左边开始的。现在， \hat{A}^\dagger 算符是作用在左边的， V^* 到 V^* 的映射。我们看一看能不能让这个算符也做用在右边。按照线性代数的符号，矩阵可以左乘矢量，也可以右乘矢量。也就是说，我们希望得到 $\hat{A}^\dagger|\mu\rangle$ 的定义。

这个矢量的定义稍微有点复杂，我们希望定义完了之后，在回到线性代数的符号的时候，这个运算真的就是把矢量乘在 \hat{A}^\dagger 对应的矩阵的右边，而不用做其它变换。我们先说有一组基矢的话，这个问题可以怎么处理，然后我们再定义不依赖于基矢的表达式。利用基矢 $\{|j\rangle\}$ ，我们得到

$$\begin{aligned} \hat{A}^\dagger|\mu\rangle &= \sum_j |j\rangle \langle j | \hat{A}^\dagger |\mu\rangle = \sum_j |j\rangle (\langle j | \hat{A}^\dagger) |\mu\rangle \\ &= \sum_j |j\rangle (\langle j | \Phi^{-1} \circ \hat{A} \circ \Phi) |\mu\rangle = \sum_j |j\rangle (\langle j | \Phi^{-1} \circ \hat{A} \circ \Phi |\mu\rangle). \end{aligned} \quad (5.26)$$

最后的表达式是一个右矢，于是我们得到了 $\hat{A}^\dagger|\mu\rangle$ 的表达式。基本的精神是给 $\hat{A}^\dagger|\mu\rangle$ 配一个左矢，然后，利用 \hat{A}^\dagger 如何做用在左矢上，我们就得到了右矢的形式。现在，我们来看一般的定义。在这之前，我们需要转变一下对左右矢的理解：把左矢当作从右矢空间 V 到数域 F 的映射，把右矢当作从左矢空间 V^* 到数域 F 的映射。一个左矢 $\langle \xi |$ 结合一个右矢 $|\nu\rangle$ 的时候，自然就成了 F 中的数， $\langle \xi | \nu\rangle$ ；同理，一个右矢 $|\eta\rangle$ 结合一个左矢 $\langle \psi |$ 的时候，自然就成了 F 中的数， $\langle \psi | \eta\rangle$ 。所以，等我们需要定义一个右矢的时候，我们只需要给出来一个从左矢空间 V^* 到数域 F 的映射。

现在，我们来看抽象的，不依赖于基矢的 $\hat{A}^\dagger|\mu\rangle$ 表达式是什么。给定任意一个左矢 $\langle \xi |$ ，我们希望得到一个 F 中的数，

$$\langle \xi | \hat{A}^\dagger |\mu\rangle = \langle \hat{A}^\dagger \xi | \mu\rangle = \langle \mu | \hat{A}^\dagger \xi \rangle^* = \langle \mu | \hat{A} | \xi \rangle^*. \quad (5.27)$$

注意，在方程的最左边，我们的 \hat{A}^\dagger 实际上是做用在左边的，我们希望定义作用在右边的表达式；在方程的最右边， \hat{A} 是作用在右边的。 \hat{A} 作用在左边还没有定义。上面的表达式就给出了对于任何的左矢 $\langle \xi |$ ，我们通过右矢

$\hat{A}^\dagger |\mu\rangle$ 得到了一个 F 中的数。也就是说, $\hat{A}^\dagger |\mu\rangle$ 确实是一个右矢, 于是 \hat{A}^\dagger 就能够作用在右矢上了。同理, 我们可以定义 \hat{A} 如何作用在左矢上。

现在, \hat{A} 和 \hat{A}^\dagger 都能够既作用在左矢上又作用在右矢上, 这个时候, 我们就能够问它们之间是否相等的问题了。如果 $\hat{A}^\dagger = \hat{A}$, 则

$$\langle \xi | \hat{A} | \mu \rangle = \langle \xi | \hat{A}^\dagger | \mu \rangle = \langle \mu | \hat{A} | \xi \rangle^* . \quad (5.28)$$

取这些矢量等于某一套完全的正交归一的基矢, 我们有

$$A_{ij} = A_{ij}^* , \quad (5.29)$$

也就是要求 \hat{A} 的矩阵形式满足转置以后取复共轭得到相同的矩阵元。这个要求被称为 Hermitian (厄米共轭), 记为

$$A^\dagger = A , \quad (5.30)$$

在有限维空间上, 还可以记为

$$(A^*)^T = A . \quad (5.31)$$

作业 5.3.1 证明 $(A, B) = \text{tr}(A^\dagger B)$ 满足 $O(V)$ 上内积的要求。

5.3.2 本征值与本征向量

有了 Hermitian 算符, 我们来考察这样的算符的本征值和本征向量。这个定理是本小节最重要的结果。

定理 5.3.1 Hermitian 算符本征向量正交性: Hermitian 算符 \hat{A} 的所有的本征值为实数 $\alpha_j \in \mathcal{R}$, 且不同本征值对应的本征向量相互正交: $\langle \alpha_1 | \alpha_2 \rangle = 0$, 其中 $\alpha_1 \neq \alpha_2$ 且 $A|\alpha_j\rangle = \alpha_j|\alpha_j\rangle$ 。

大致证明如下: 先证明本征值为实数,

$$\begin{aligned} \alpha_j \langle \alpha_j | \alpha_j \rangle &= \langle \alpha_j | A | \alpha_j \rangle = \langle A \alpha_j | \alpha_j \rangle^* = \langle \alpha_j | A^\dagger | \alpha_j \rangle^* \\ &= \langle \alpha_j | A | \alpha_j \rangle^* = \alpha_j^* \langle \alpha_j | \alpha_j \rangle \\ &\implies (\alpha_j - \alpha_j^*) \langle \alpha_j | \alpha_j \rangle = 0 \\ &\implies \alpha_j = \alpha_j^* \end{aligned} \quad (5.32)$$

假设 $\alpha_1 \neq \alpha_2$,

$$\begin{aligned} \alpha_2 \langle \alpha_1 | \alpha_2 \rangle &= \langle \alpha_1 | A | \alpha_2 \rangle = \langle \alpha_2 | A^\dagger | \alpha_1 \rangle^* \\ &= \langle \alpha_2 | A | \alpha_1 \rangle^* = \alpha_1^* \langle \alpha_2 | \alpha_1 \rangle^* = \alpha_1 \langle \alpha_1 | \alpha_2 \rangle \\ &\implies (\alpha_1 - \alpha_2) \langle \alpha_1 | \alpha_2 \rangle = 0 \\ &\implies \langle \alpha_1 | \alpha_2 \rangle = 0 \end{aligned} \quad (5.33)$$

如果算符 \hat{A} 的本征值都不相同, 那么只要我们把每一个本征矢量各自归一化, 我们就从 Hermitian 算符 \hat{A} 的本征矢量得到了空间 V 的正交归一基矢, 于是我们有完全性展开关系,

$$I = \sum_j |\alpha_j\rangle \langle \alpha_j|. \quad (5.34)$$

存在相同本征值对应不同本征向量, 这个称为简并本征值, 的问题技术上更复杂 (需要在相同本征值对应的本征向量的子空间内正交化矢量), 结论是一样的, 在此我们不讨论,

如果我们有两个不同的 Hermitian 算符 \hat{A}, \hat{B} , 我们问这个时候它们对应的正交归一基矢是否相同?

定理 5.3.2 Hermitian 算符 \hat{A}, \hat{B} 存在共同本征向量如果 \hat{A}, \hat{B} 对易: $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} = 0$.

同样我们假设本征值没有简并的条件下来讨论. 取 \hat{A} 的本征向量 $|\alpha\rangle$,

$$\hat{A}\hat{B}|\alpha\rangle = \hat{B}\hat{A}|\alpha\rangle = \alpha\hat{B}|\alpha\rangle \implies \hat{B}|\alpha\rangle = b_\alpha|\alpha\rangle. \quad (5.35)$$

于是 $|\alpha\rangle$ 也是算符 \hat{B} 的本征向量. 证毕 ■

反过来, 很简单就是可说明, 如果两个算符 \hat{A}, \hat{B} 的所有本征向量都完全一样, 则这两个算符必然对易. 直观地说, 在这个共同本征态所形成的基矢下, 这两个算符都只有对角矩阵元, 例如

$$\hat{A} = \sum_\alpha \hat{A}|\alpha\rangle \langle \alpha| = \sum_\alpha \alpha|\alpha\rangle \langle \alpha|, \quad (5.36)$$

完全对角. 而对角的矩阵永远是相互对易的.

通过这个定理, 我们看见, 如果我们需要讨论的算符不是所有的都相互对易, 那么我们就可以选择不同的 Hermitian 算符所代表的本征矢量来作为基矢. 这样, 我们就有很多种基矢的选择. 于是, 一个算符或者一个向量, 它们在不同的基矢下就有不同的表现形式. 这些不同的表现形式之间有什么样的关系呢?

5.3.3 线性变换与表象理论

这个不同的表象之间的关系被称为表象变换。一个线性算符在给定表象下就成了一个矩阵（一堆按照某种顺序排列的数）。因此，借助表象，我们就可以把抽象的算符的讨论变成比较具体的数之间的运算。可是，不同的基矢下，同一个算符对应着的矩阵是不一样的。这些不一样的表现形式对我们的解算结果有影响吗？这些不一样的形式之间有什么关系？这个关系就称为算符在不同基矢，也就是不同表象下的线性变换。

我们来看一个算符 S 在两个表象下的形式

5.3.4 矩阵的奇异值分解

我们已经看到，而且将要看到更多，当我们把一个线性算符算符表示为如下本征向量展开的形式的时候，

$$A = \sum_j \alpha_j |\alpha_j\rangle \langle \alpha_j|, \quad (5.37)$$

大量的问题就会非常容易的解决：在这个表象下矩阵不存在了，仅仅是一堆排成对角的数。于是，如何把矩阵变形成为对角的形式就成了一个非常重要的问题。前面，我们已经了解一个一般的矩阵不一定能够对角化。那么，是不是一定要要是 Hermitian 的矩阵才能够被对角化呢？矩阵的奇异值分解就解决了一般的矩阵的对角化问题。

按照复数域上的多项式解的完全性定理，我们知道对于 n 维矩阵 A ，方程

$$\det \|A - \alpha I\| = 0 \quad (5.38)$$

必然有 n 个解（允许重根）。因此，一个 n 维矩阵必然存在 n 个本征值，只不过有可能是复数。无限维的算符我们在这里不讨论。那么，是否 n 个本征值就会对应 n 个本征向量呢？我们来看下面的例子，

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \quad (5.39)$$

容易求得本征值方程为 $\alpha^2 = 0$ ，于是 $\alpha = 0$ 。可是，本征向量的方程

$$A \begin{bmatrix} a \\ b \end{bmatrix} = 0 \begin{bmatrix} a \\ b \end{bmatrix} \quad (5.40)$$

只有一个非平庸解 $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ 。那个和这个本征矢量正交的矢量 $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ 却不满足本征向量方程。因此，不是每一个矩阵都是可以对角化的，计算允许本征值取复数。实际上数学上这个问题的有充分必要条件。这里，我们不再讨论。

既然如此，有没有一种方法能够把任意的矩阵变成对角的形式，尽管这样的“对角化”可能不一定对应着本征值和本征向量？答案是用奇异值分解代替本征值分解。

定理 5.3.3 任意矩阵 A 存在 unitary 矩阵 U, V ，对角矩阵 Σ ，满足

$$A = U\Sigma V^\dagger. \quad (5.41)$$

其中 Σ 的各个元素非负。

在这里我们给出一个直观的解释和构造，证明请参考 Wikipedia 或者其它书籍 []。为了讨论简单，我们还是仅仅关心 $n \otimes n$ 的矩阵 A 。取 Hermitian 矩阵 AA^\dagger ，必然存在矩阵 U 和 D 使得

$$AA^\dagger = UDU^\dagger. \quad (5.42)$$

同样对于矩阵 $A^\dagger A$ ，存在矩阵 V 和 D 使得

$$A^\dagger A = VDV^\dagger. \quad (5.43)$$

可以证明 D 只包含正整数，而且从 AA^\dagger 和 $A^\dagger A$ 得到的相同（次序不同的话总是可以通过交换本征值的顺序来重新编号，例如 D 取从大到小的顺序。如果出现简并的情形，需要做相应的处理）。于是，我们定义，

$$\Sigma = \sqrt{D}, \quad (5.44)$$

就是取把对角的相应元素的正根。很容易验证这样定义的 $U\Sigma V^\dagger$ 能够满足上面两式中的 AA^\dagger 和 $A^\dagger A$ 。如果要完善我们的证明，我们还需要进一步证明 $U\Sigma V^\dagger$ 就是 A 。这个证明实际上要比我们的上面的直觉的构造更复杂一些。其原因是，原则上 AA^\dagger 和 $A^\dagger A$ 的本征向量分解是独立完成的，很难恰好满足我们要求的对应关系。不过，假设没有简并本征值，那么，按照本征值的顺序就能够对应起来，从而得到 A 的如上分解。

这个分解就被称为一般矩阵的奇异值分解。与此相关，还有一个矩阵的 Jordan 标准型的概念，在理论上也非常重要。在这里，我们也不讨论了。

到现在为止，我们把抽象矢量和一般线性算符的运算通过表象变成了矢量和矩阵的运算。我们还把矩阵的运算总结为线性方程的解的运算、矩阵本征值分解，以及矩阵的奇异值分解。而奇异值分解和本征值分解，我们已经看到，密切相关。那么，下面，我们来看一下在实际的计算上，我们可以如何求线性方程的解以及做本征值分解。

5.4 数值线性代数

这里，我们需要具体的可计算的数，所以，我们总是把抽象矢量和算符在某套坐标系下展开，也就是取某个表象以后再来做计算。所以，下面的所有内容都采用一般的线性代数的符号，矢量就是列向量，算符就是矩阵。

数值线性代数能够求解的核心问题有两个本征值问题和线性系统的解的问题。我们先讨论理论上这两个问题的定义和一般形式，然后我们介绍一下数值求解这两个问题的最基本的想法。我们会发现这两个问题的定义形式是不能用来真正计算比较大的矩阵的本征值问题和求解的问题的。实际上，我们将来还会发现，数值线性代数的计算是整个科学计算的核心（另外一个 Monte Carlo 方法），基本上所有的其他的数值计算的问题，例如数值差分，非线性方程的解，有限元方法等等，都需要把问题转化成为数值线性代数的问题，因此，学会数值线性代数的基本的精神有非常大的意义，是将来利用和发展多科学计算算法的基础。

在下面所有的介绍中，我们强调对计算方法的基本的理解，在具体实现上，我们只需要学会使用合适的工具包就可以。

5.4.1 矩阵相乘的 Strassen 算法

5.4.2 本征值与本征向量

本征值方程，幂方法

5.4.3 线性系统的解

代数余子式，子空间方法 BLAS, Lapack, Petsc, Slep, Sage, Matlab

第六章 概率论

我们不讨论这个世界到底是决定性的还是随机的，我们把确定性的事件也看成随机事件的一个特例，服从 δ 分布的一个特例。所谓 δ 分布就是在给定时刻这个“随机”变量仅有一个可能取值。在这个视角下，所有的经典世界都由概率论描述。

在这一章里面，我们要解决的问题是如何描述随机事件。从古典概型，我们已经知道随机事件可以用随机变量来描述，离散随机变量的每一个可能取值对应着一个概率，连续随机变量可以用概率密度分布函数描述，同时离散和连续随机变量都可以用累积概率分布函数来描述。这些还不够，我们要讨论一下随机事件的定义，概率论最核心的性质是什么（将来我们会看到这个核心是概率叠加原理：互斥事件的加法），然后利用上一章学到的矩阵和 Dirac 符号来介绍一种新的描述概率的方法——密度矩阵。最后，我们讨论一下离散时间的随机过程。

最后我们会简短地讨论一下 Monte Carlo 方法——产生给定分布函数的随机数的样本的一个方法，以及随机变量的测量。后者是一个基本上所有的教材都不讨论的问题。我们认为其实关于这个问题的认识非常重要。例如，我们考虑这样一个事件：扔出色子并且色子的某一面——例如 6 这一面——向上之后，观测这个正面的值，然后我们说这个随机事件的概率分布就是 6 这一面 100%，其他取值的概率为零。那么是否意味着在观测之前的概率分布——每一面都是 $\frac{1}{6}$ ——是错误的呢？我们认为，如果认为是错误的，则是对概率论和随机事件非常错误的理解。在原来的这个硬币的状态到后来的这个硬币的状态之间，存在着一个操作：测量。测量的前后，硬币的状态确实原则上可以发生改变的。在测量之后，确实硬币的状态是 100% 取值为 6，但是并不意味着测量之前的状态的描述也发生了改变。大多数经典客体的状态在测量前后不发生改变。这一点，在原则上，没有必要对于随机客体还保持。原则上，我们允许这个世界上存在着真正随机的客体。然而，

如果真的存在，那么我非常怀疑，一个没有深刻理解概率论的学者，可以理解这个世界的任何部分。

这个部分的推荐阅读材料是：Kolmogorov 《概率论导引》，王梓坤《概率论》，何书元《概率论》，A course in Probability Theory by Kai Lai Chung

6.1 古典概型：离散概率与几何概率

古典概型对于随机事件有一个基本的假设：存在着某一个基本的简单事件，其中每一个基本简单事件的概率是明确的甚至是相等的，然后复杂事件的概率总是可以通过分解成基本简单事件的方式来计算出来。在这里，我们先通过两个古典概型的例子，来了解一下概率论。然后，我们会展示这种基于简单基本事件的概率论存在一些问题，于是我们需要构造更加一般的概率论。

例 6.1.1 对称色子：计算两个六面的对称的色子出现的正面向上的数字之和的几率。

例 6.1.2 约会问题：Alice 和 Bob 约会，约定三点钟见面，可以这两个人都不太守时，耐心也不太好，会在三点钟的前后 5 分钟出现，而且出现以后只等对方 5 分钟。问他们能够实现这个约会的几率是多少？

例 6.1.3 圆内弦长的问题：单位圆内任意取一条弦，其长度小于 $\sqrt{3}$ 的几率是多少？

在这个例子中，我们发现存在着多种等概率的基本简单事件的选择。于是，基于基本简单事件的概率或者概率密度的古典概型，就有问题了。现代概率论通过下面的概率公理化定义来解决这个问题。我们先看一下这个公理化定义是什么，然后在做进一步的关于这个样的定义的含义和动机的讨论。

6.2 现代概率三元体

定义 6.2.1 概率空间：概率三元体 (Ω, \mathcal{F}, P) ，其中集合 Ω ，集合元素对应简单事件记号 $\omega \in \Omega$ ， Ω 的子集 A 构成集合 \mathcal{F} 是 Ω 上的 σ 代数，即 \mathcal{F} 满足

1. \mathcal{F} 至少包含 Ω ： $\Omega \in \mathcal{F}$ ；

2. 对集合的补集操作封闭: $\bar{A} \in \mathcal{F}, \forall A \in \mathcal{F}$;
3. 对可数个集合的交集封闭: $\cup_{j=1}^{\infty} A_j \in \mathcal{F}, \forall A_j \in \mathcal{F}$ 。

从 \mathcal{F} 到 $[0, 1]$ 的映射 P , 满足下列条件

1. 完全性:

$$P(\Omega) = 1; \quad (6.1)$$

2. 可列可加性: 对于可数个不相交的集合 (互斥事件, $A_i \cap A_j = \phi, \forall A_i, A_j$)

$$P(\cup_j A_j) = \sum_j P(A_j). \quad (6.2)$$

\mathcal{F} 是 σ 代数的要求是为了保证对于任意两个概率有定义的集合, 它们的集合操作——交并补——都有定义。因为在概率的意义上, 我们从两个集合的概率开始, 我们经常需要讨论这两个集合对应着的事件合起来的事件的几率, 或者是都发生的几率, 或者是一个发生另一个不发生的几率。映射 P 保证在 $[0, 1]$ 之间也是很好理解的, 概率最好是大于零的, 而且最大最大就是 1。完全性也很好理解, 所有的事件的整体的概率显然应该是 1。也就是说, 所有的概率论的定义的核心的一条, 就是为了保证“可列可加性”: 对于互斥的事件, 其整体的概率等于各个互斥部分的概率之和。我们称这一条为“概率性叠加原理”。也就是说, 就是这个概率性叠加原理体现了概率论的所有的意义。以后, 我们会有更深刻的体会。

在这里, 我们先来展示一下, 通过这一条概率性叠加原理来定义一下任意两个集合 A, B 的并的几率, 如果这两个集合 A 和 B 已经确定是 \mathcal{F} 的元素的话。我们先来看 $A \cap B$, 如果它等于空集, 则 A 和 B 是互斥事件, 于是 $P(A \cup B) = P(A) + P(B)$, 解决。如果 $A \cap B \neq \phi$, 我们把 $A \cup B$ 分解成三个互斥的集合 $A \cap \bar{B}$, $\bar{A} \cap B$ 和 $A \cap B$ 。于是, $P(A \cup B) = P(A \cap \bar{B}) + P(\bar{A} \cap B) + P(A \cap B)$ 。也就是说, 实际上, 如果我们知道每一个互斥的 \mathcal{F} 中的元素的几率, 实际上, 我们用概率性叠加原理就可以求出所有的 \mathcal{F} 中的元素, 也就是任意事件, 的几率。

回到我们的圆内弦长的问题, 可以证明, 从我们前面的任何一种等概率假设都可以构造一个满足上述要求的概率空间。

实际上, 我们还应该完成这样一件事情: 从概率论的公理化定义, 找出适当的条件, 在这个条件下随机变量的分布函数存在。但是, 这个任务超过

本书的水平，在我们发现一个好的能够在本书的概念和方法的范畴内讨论这个问题之前，我们直接就用分布函数来描述概率论。

6.3 条件概率与 Bayesian 公式

在整个概率论的框架中，还有一条重要的原理，就是，独立事件的交集的概率等于各自的独立事件的概率相乘。为了讨论这一条，我们定义下面的条件概率。

定义 6.3.1 条件概率：事件 A 已经发生的情况下事件 B 发生的概率称作事件 B 关于事件 A 的条件概率，记为 $P(B|A)$ ，

$$P(B|A) = \frac{P(B \cap A)}{P(A)}. \quad (6.3)$$

根据这个定义，我们有

$$P(B \cap A) = P(B|A)P(A), \quad (6.4)$$

如果

$$P(B|A) = P(B), \quad (6.5)$$

则

$$P(B \cap A) = P(B)P(A). \quad (6.6)$$

满足公式 (6.5) 的事件 A 和 B 被称为独立事件，也就是 B 发生的概率与 A 是否已经发生没有关系。

从条件概率的公式，还可以推导出来一个非常重要的定理：Bayesian 公式。

定理 6.3.1 Bayesian 公式：

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}. \quad (6.7)$$

简单证明如下：从公式 (6.4) 我们交换 A 和 B 的位置得到

$$P(B \cap A) = P(A \cap B) = P(A|B)P(B). \quad (6.8)$$

于是

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}, \quad (6.9)$$

而这个分母可以分解成互斥的两个部分 $A = (A \cap B) \cup (A \cap \bar{B})$ ，然后利用互斥事件的概率性叠加原理就可以得到 Bayesian 公式。证毕 ■

至此，概率论的基本理论就完成了。它包含概率的公理化定义，其中最主要的是概率性叠加原理，还包含 Bayesian 公式以及独立事件的概率乘法。在我们进入下面的关于大数定律和中心极限定理的讨论之前，我们通过几个例子来熟悉一下概率性叠加原理和独立事件的概率乘法，以及 Bayesian 公式。然后，我们举例介绍几个典型分布函数和基本数字特征，也就是基本统计量。

例 6.3.1 贝叶斯公式和条件概率的运用：考虑一个灯泡厂商，在考虑雇佣质检人员。灯泡厂商知道自己的工艺水平也就是合格率 q 。质检人员主要技术指标是 $P(\text{pass}|\text{good}) = p^{pg}$ 与 $P(\text{fail}|\text{bad}) = p^{fb}$ ，就是好的产品的通过率，以及坏的产品的检出率。理想的质检人员是两者都是 100%，但是实际上有的人要求严 (harsh 型)，好的也能被看成坏的；还有的人要求松 (easy 型)，坏的也能看成好的。只有质检合格的产品才进入市场。问：消费者考虑的出发点主要是 $P(\text{good}|\text{pass})$ ，也就是检出状态为“好”的产品之中坏产品的可能性，这个与厂商的选择有什么关系？如果你是厂商，为了增强消费者的信心，你应该雇佣什么样的质检人员？如果你是厂商，为了获取最高额的利润，你应该雇佣什么样的质检人员？假设这个产品的卖出价格为 Pr ，成本为 Ct ，如果消费者发现上当受骗，则惩罚为 Fn ，假设进入市场的坏的产品必然能够被消费者检出。如果这个厂商改进工艺水平提高产品合格率，会带来什么变化？

6.4 典型概率分布函数和基本统计量

各阶矩，特征函数

例 6.4.1 随机变量之和的随机变量：随机变量 x_1, x_2 同取值于标准正态分布随机数，计算两者之和 $y = x_1 + x_2$ 的密度分布函数。更一般地证明：对于通过对两个取值于独立同分布的随机变量取和所得到的随机变量，其分布函数的特征函数，等于原始分布函数的特征函数的平方。

6.5 大数定律与中心极限定理

大数定律的主要内容非常简单直接：如果我们有一大堆独立同分布的随机数 x_1, x_2, \dots, x_n ，每一个都从一个给定的分布函数 $f(x)$ 中抽样得来，然后我们取它们的平均 $x^{(n)} = \frac{x_1 + x_2 + \dots + x_n}{n}$ ，则我们希望这个平均值就是预先给定的分布函数的平均值参数 $\langle x \rangle = \int dx f(x) x$ 。因为这些独立同分布的随机数可以看成是从 n 个不同的独立同分布中抽样得来的，也可以看成是从一个分布函数中抽了独立的 n 个样本，于是按照后者，当 n 很大的时候，必然就要求这个平均值等于预先给定的分布函数的平均值参数。如果这一条不成立的话，那么分布函数的平均值参数也就失去了平均值的含义了。所以用一句话概括大数定律就是：实际样本的具体的平均值等于分布函数的抽象的平均值。这个也可以说成是频率等于概率。为什么能这么说呢？如果我们的 x_j 取值就是 0 或者 1 的话，其中 1 代表某事件发生了 0 代表没有发生，则这个时候 $x^{(n)}$ 实际上就是事件发生的频率，而 $\langle x \rangle$ 就是事件发生的概率。因此，大数定律也被称为频率就是概率。下面是更严格的大数定律的表述和证明。这里我们仅仅讨论弱大数定律。强大数定律可以参考其它书籍或者 wikipedia。

定理 6.5.1 大数定律：

在我们目前的逻辑框架之内大数定律其实是大数定理，是可以证明的。但是，概率论也存在着其它同样合理的逻辑框架，例如概率论的 Bayesian 定义，这某些框架内，这个定理是其他定义和定理的基础，而不能被其它定义和定理证明。缘此以及由于历史原因，通常我们还是把它称作大数定律。定律一般指从现实世界通过观察和实验得来的，不能或者至少在发现这个定律的当时还没有被证明的。定理是指从定义和公理通过逻辑推理能够证明的命题。公理是对定义或者说概念之间的关系的假设，也就是假定其正确的命题。

中心极限定理的含义是多个独立同分布随机变量的和的随机变量符合正态分布。这个定理的理解要比大数定律复杂的多：把很多很多个任意的给定（将来我们会看到还是有一些约束的）分布函数的样本的值合起来，我们就可以得到正态分布。也就是说，如果某个事件其数字特征可以看成是多个独立的原因造成的事件的数字特征之和，那么这个事件的数字特征就满足正态分布，不用管构成它的基础的事件的数字特征满足什么分布。这个结论非常强。它对吗，怎么证明，证明了以后又如何理解为什么会这样？我们现

在来讨论这些问题。

例 6.5.1 正态分布的实际生活中的例子：

定理 6.5.2 中心极限定理：

6.6 概率论的 Dirac 符号形式

结束一般教材上都可以见到的概率论的主体内容（除了中心极限定理的稍微另类一点的证明）之后，我们来呈现我们在这里介绍概率论的真正目的：概率论的 Dirac 符号形式。让我们从古典概型开始。一个古典概型的分布函数除了一个列表之外，还可以记作，

$$\rho = \sum_j p_j |j\rangle\langle j|, \quad (6.10)$$

其含义完全就是事件 $|j\rangle\langle j|$ 的几率是 p_j 。惟一的不同之处在于，现在，我们可以用加号（“+”，“ Σ ”）了，而以前我们只能够用列表。既然完全表达一样的信息，为什么要这么做？我们希望把概率性叠加原理强调出来。对于古典概型，我们知道各个基本简单事件本来就是互斥的，也就是 $\langle j|k\rangle = 0$ ，或者记为 $\text{tr}(|j\rangle\langle j|^\dagger |k\rangle\langle k|) = 0$ ，表示与基本简单事件对应的基矢或者矩阵相互正交。于是，这个密度矩阵反映的正好就是这个概率性叠加原理。例如，我们考虑一个事件 A 的概率，记事件 A 的每一个基本简单事件的元素为 α_j ，则

$$A = \sum_j |\alpha_j\rangle\langle\alpha_j|, \quad (6.11)$$

从而通过

$$P(A) = \text{tr}(A\rho), \quad (6.12)$$

我们得到

$$P(A) = \sum_{\alpha_j} p_{\alpha_j}. \quad (6.13)$$

如果有一个与 A 互斥的事件 $B = \sum_j |\beta_j\rangle\langle\beta_j|$ ，满足

$$\text{tr}(A^\dagger B) = 0, \quad (6.14)$$

也就是任意 α_j, β_k 满足

$$\langle \alpha_j | \beta_k \rangle = 0, \quad (6.15)$$

则

$$P(A \cup B) = P(A + B) = P(A) + P(B). \quad (6.16)$$

注意，这个通过把事件 A, B 写成矩阵 A, B 的形式，这个事件集合的加法真的可以表达成加法运算的形式。这个就是我们引入概率论的 Dirac 符号的目的。

这个记号还可以表达更进一步的含义。如果我们有一个 Ω 上的函数，例如色子的正面上的读数，或者是这个读数的函数，例如如果这个读数是奇数就输一块钱 ($m = -1$)，否则赢一块钱 ($m = 1$)，那么，我们可以把这个函数写成矩阵形式，例如前者为

$$S = \sum_{j=1}^6 j |j\rangle \langle j|, \quad (6.17)$$

后者为

$$M = \sum_{j=1}^6 (-1)^j |j\rangle \langle j|. \quad (6.18)$$

更一般的随机变量的函数可以记为

$$f_\Omega = \sum_{j \in \Omega} f_j |j\rangle \langle j|. \quad (6.19)$$

这个时候，我们仍然可以用公式 (6.12) 的推广形式，

$$\langle f_\Omega \rangle = \text{tr}(f_\Omega \rho) \quad (6.20)$$

来计算平均值，例如这两者分别为

$$\langle S \rangle = \sum_{j=1}^6 j p_j, \quad (6.21)$$

和

$$M = \sum_{j=1}^6 (-1)^j p_j. \quad (6.22)$$

总结一下，引入概率论的 Dirac 记号之后，我们的概率论基本公式成了公式 (6.10)，公式 (6.19) 和公式 (6.20)。以上的符号是以离散古典概型为例引入的，实际上对于连续古典概型也适用，甚至对于一般的概率空间也能够使用。但是，为了记号简单，而且以后也基本上仅仅在离散变量上使用这套符号，更一般的情形我们就不讨论了。

为了熟悉这套符号，学会熟练使用公式 (6.10)，公式 (6.19) 和公式 (6.20)，我们来用这个符号完成之前的一个例题。

例 6.6.1 Dirac 记号用于对称色子问题：计算两个六面的对称的色子出现的正面向上的数字之和的几率。

6.7 随机过程初步

概率论的基本理论完了之后，一个自然的进阶内容就是随机过程。随机过程实际上就是高维随机变量的分布函数的问题。这个高维随机变量又可以看成是一个随机变量的时间序列。而这个随机变量的时间序列又可以看成是每一个时刻一个随机变量在其分布函数中抽样的结果。这里，我们的目的是讲清楚什么是随机过程，对随机过程的描述有哪一些角度。后面我们有专门的章节来讲解随机过程，介绍典型的随机过程，随机过程的数学描述和求解。那部分，我们放到学完物理之后再一次回头来讲。那个时候，由于力学方程的概念的建立，我们的随机过程的描述会更具有直观性。

例 6.7.1 翻硬币操作的算符形式：硬币的状态和使硬币翻转的操作分别用什么数学形式描述？

硬币的状态是有两个可能状态（标记为 $S = (1, -1)$ ）的一个随机变量，因此如果以密度矩阵的形式描述硬币的状态，我们有

$$\rho = p|1\rangle\langle 1| + (1-p)|-1\rangle\langle -1|. \quad (6.23)$$

硬币翻转的操作使得 p 与 $(1-p)$ 相互交换，下面的操作实现这个交换的过程。定义

$$X = |1\rangle\langle -1| + |-1\rangle\langle 1|. \quad (6.24)$$

算符的作用就是把 -1 的状态变成 1 ，把 1 的状态变成 -1 。可以验算，按照下面的矩阵相乘的方式计算，这个算符就能实现翻转，

$$\rho = X\rho_0X^\dagger = X\rho_0X. \quad (6.25)$$

最后的等式是因为 $X^\dagger = X$ ，一般情况下不成立。实际上，我们还可以选择 $X \rightarrow iX$ ，把 X 算符乘上一个虚数单位 i ，来实现同样的操作。这个时候，这个最后的等式就不对了。Hermitian 共轭的操作就必须留着。

有了这个关于硬币翻转的算符描述之后，我们就可以来讨论多次硬币翻转的随机过程了。

例 6.7.2 硬币状态演化的随机过程：考虑一个硬币，初始状态为正面向上，经过如下的过程，考察最终状态以及长时间以后多次独立观察出现状态的几率：扔另一个随机硬币，此硬币正面（反面）向上的几率为 p ($1-p$)，此时翻转（不翻转）之前的硬币。

标记硬币的状态集合为 $S = (1, -1)$ 分别表示正面与反面。记 t 时刻硬币的状态为 $\xi_t \in S$ ，则在 $t+1$ 时刻，硬币的状态以 p ($1-p$) 的几率为 $\xi_{t+1} = -\xi_t$ ($\xi_{t+1} = \xi_t$)。按照上一个例子里面的硬币翻转的算符描述，这个随机翻转的过程也就是

$$\rho_{t+1} = pX\rho_tX^\dagger + (1-p)I\rho_tI^\dagger = pX\rho_tX + (1-p)I\rho_tI \triangleq \mathcal{L}\rho_t, \quad (6.26)$$

或者平均值满足

$$\langle \xi_{t+1} \rangle = -p\langle \xi_t \rangle + (1-p)\langle \xi_t \rangle. \quad (6.27)$$

其中第二式完全由对第一式求平均 $\langle \xi \rangle = \text{tr}(\hat{\xi}\rho)$ 得来。其中 $\hat{\xi}$ 是随机变量 ξ 的算符形式，

$$\hat{\xi} = |1\rangle\langle 1| - |-1\rangle\langle -1|. \quad (6.28)$$

如果回到之前的矢量形式的语言而不是算符的形式，记 t 时刻硬币的分布函数为 $P_t = (P(1, t), P(-1, t))^T$ ，则

$$P_{t+1} = pXP_t + (1-p)IP_t \triangleq \Gamma P_t, \quad (6.29)$$

也就是

$$\begin{bmatrix} P(1, t+1) \\ P(-1, t+1) \end{bmatrix} = (pX + (1-p)I) \begin{bmatrix} P(1, t) \\ P(-1, t) \end{bmatrix} \quad (6.30)$$

$$= \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \begin{bmatrix} P(1, t) \\ P(-1, t) \end{bmatrix}, \quad (6.31)$$

或者把这个随机变量的演化过程记为

$$\xi_{t+1} = \eta \xi_t, \quad (6.32)$$

其中

$$\eta = \begin{cases} -1 & \text{with probability } p \\ 1 & \text{with probability } (1-p) \end{cases}. \quad (6.33)$$

方程公式 (6.26), 公式 (6.29), 公式 (6.32) 从不同的角度描述的了相同的信息相同的过程, 但是公式 (6.27) 仅仅描述了平均值的信息。公式 (6.26), 公式 (6.29) 从随机变量的分布函数的角度进行描述, 公式 (6.32) 从随机变量的取值的角度来描述。注意, 公式 (6.26), 公式 (6.29) 都是两个算符取和的形式, 但是公式 (6.32) 不是两个随机变量的值取和的形式。公式 (6.32) 看起来形式简单, 但是所有的奇妙之处都被藏在了公式 (6.33) 引入的随机变量里面。如果我们回到这一章介绍的新的符号体系——用密度矩阵来描述随机变量, 我们会发现公式 (6.32) 完全就是公式 (6.26) 的另外一种写法。

这时候, 我们讨论 ρ_t 的稳态分步是什么, P_t 的稳态分步是什么, 或者 ξ_t 的整个轨道 $\vec{\xi} = (\xi_0, \xi_1, \dots)^T$ 的分布函数是什么, $\langle \xi_{t+1} \rangle$ 的长时行为是什么等问题。实际上, 如果读者已经从这个例子中理解了这几个不同角度的描述方式, 不同层次的描述方式, 那么这些问题的答案对于我们要阐述的相互体内容已经不重要了。但是, 为了这个例子的完整性, 我们在下一个例子中继续求解这些问题。

从这个例子当中, 我们看到随机过程描述的对象实际上是一个随机变量的演化过程的所有的轨道 $\{\vec{\xi}\}$ 的几率分布 $P(\xi_1, \xi_2, \dots)$, 也可以看成是每一个时刻的随机变量 ξ 的密度矩阵 ρ_t 或者分布函数 $P(\xi, t)$ 的演化过程 (如何由之前的密度矩阵 $\rho_{\tau < t}$ 分布函数 $P(\xi, \tau < t)$ 演化得来), 而且这个分布函数可以用密度矩阵的形式, 也可以采用概率矢量的形式。在特殊情况下 (以后我们会知道, 这个例子中的特殊情况是这个过程是一个 Markov 过程), 这个演化机制可以是分布函数的线性方程。通常在这个问题中, 我们关心分布函数的稳态形式, 随机变量的数字特征的长时行为, 或者更完整的信息 $\{\vec{\xi}\}$ 的分布函数。

明白了这个随机过程的一般图景之后, 我们下一步要解决的问题就是对于一些常用的随机过程, 随机变量的演化方程, 随机变量的分布函数的演化方程, 有哪一些基本的形式, 如何解析与数值求解。这些任务我们将来在

有关随机过程的专门章节里来完成。在下一部分，我们仅仅用几种不同方法来解决这个翻硬币的例子。

例 6.7.3 硬币状态演化的随机过程：续，硬币的长时状态。

让我们从公式 (6.29) 出发，我们已知 P_0 ，那么

$$P_t = \Gamma^n P_0, \quad (6.34)$$

其中

$$\Gamma = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}. \quad (6.35)$$

硬币的稳态分布，或者说硬币的长时状态实际上由 $\lim_{t \rightarrow \infty} P_t$ 决定，而这个极限由算符 Γ 的本征向量决定： Γ 最大本征值对应的本征向量会活下来，其它都会消失。这个可以验证这个最大本征值是 1，其对应的本征向量为

$$P_\infty = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}. \quad (6.36)$$

因此，这个硬币如果做多次独立测量（多次独立测量的含义是从头开始制备多个这样的硬币，过很长一段时间以后测量，然后记录结果，每次测量完成以后扔掉）的结果是一半的几率看到硬币正面向上，另一半的几率看到硬币反面向上。

这里我们主要采用了把随机分布函数看成向量，然后把随机过程看成这个向量的线性变换的方法，然后求解这个变换的不变向量。以后我们会看到更多其他的方法，但是这个基本的方法还是重要的方法之一。这个线性变换的过程更加一般地可以写成，

$$P(n, t+1) = \sum_m [W_{m \rightarrow n} P(m, t) - W_{n \rightarrow m} P(n, t)], \quad (6.37)$$

其中 n, m 表示状态的标号， t 是时间， $W_{m \rightarrow n}$ 称为状态 m 到 n 的跃迁几率。以后我们会做关于这个方程的更进一步的研究。

最后，我们就这个例子作一点点进一步讨论。这个讨论与公式 (6.32)、公式 (6.26) 以及公式 (6.27) 的形式有关，与将来量子力学的数学形式有关。

例 6.7.4 硬币状态演化的随机过程：续，考虑随机变量的可加性，也就是一个随机变量如果可能取两个或者多个不同的值，能把这两个值加起来（直接加起来，或者按照概率加起来）然后，还是代表这个随机变量吗？

前面我们已经用密度矩阵的演化方程式 (6.26) 描述这个翻硬币的过程（为了方面对比，重新写在下面），

$$\rho_{t+1} = pX\rho_tX + (1-p)I\rho_tI \triangleq \mathcal{L}\rho_t, \quad (6.38)$$

其中初始条件，

$$\rho_0 = |1\rangle\langle 1|. \quad (6.39)$$

这个时候算符是不随时间演化的，

$$\hat{\xi} = 1|1\rangle\langle 1| - 1|-1\rangle\langle -1|. \quad (6.40)$$

这个描述下，算符是不变的，密度矩阵是变化的。

现在，我们还可以写下算符形式的方程来描述这个系统的演化，

$$\hat{\xi}_{t+1} = pX\hat{\xi}_tX^\dagger + (1-p)I\hat{\xi}_tI^\dagger \triangleq \mathcal{L}^\dagger\hat{\xi}_t, \quad (6.41)$$

其中初始条件，

$$\hat{\xi}_0 = 1|1\rangle\langle 1| - 1|-1\rangle\langle -1|. \quad (6.42)$$

与公式 (6.41) 相配套的密度矩阵是

$$\rho = |1\rangle\langle 1|, \quad (6.43)$$

这里算符是变化的，密度矩阵是不变的。这是两套不同的描述方法，前者把分布函数或者说密度矩阵作为演化过程描述的对象，后者把算符作为演化过程描述的对象。我们看一看两者是不是等价的。这里，我们简单地通过对比从两者得到的平均值的演化方程来检验。

结合公式 (6.38) 和公式 (6.40)，利用

$$\langle \xi \rangle = \text{tr}(\rho\hat{\xi}), \quad (6.44)$$

我们得到

$$\langle \xi_{t+1} \rangle = -p\langle \xi_t \rangle + (1-p)\langle \xi_t \rangle. \quad (6.45)$$

在推导上面的等式的过程中，我们用到了

$$X\hat{\xi}X = -\hat{\xi}. \quad (6.46)$$

结合公式 (6.41) 和公式 (6.43)，利用公式 (6.44) 我们也得到公式 (6.45)。这个推导过程稍微有点复杂，详细说明如下。

$$\begin{aligned} \langle \xi_{t+1} \rangle &= \text{tr}(\hat{\xi}_{t+1}\rho) \\ &= \text{tr}(pX\hat{\xi}_tX + (1-p)I\hat{\xi}_tI\rho) \\ &= p\text{tr}(\hat{\xi}_tX\rho X) + (1-p)\text{tr}(\hat{\xi}_t\rho) \end{aligned} \quad (6.47)$$

现在，我们只要证明 $\text{tr}(\hat{\xi}_tX\rho X) = -\langle \xi_t \rangle$ ，我们就完成了整个证明。首先，我们从公式 (6.42) 了解到 $\text{tr}(\hat{\xi}_0) = 0$ ，其次，我们还从公式 (6.41) 看到 $\text{tr}(\hat{\xi}_{t+1}) = 0$ ：

$$\text{tr}(\hat{\xi}_{t+1}) = p\text{tr}(X\hat{\xi}_tX) + (1-p)\text{tr}(I\hat{\xi}_tI) = p\text{tr}(\hat{\xi}_tXX) = 0. \quad (6.48)$$

这也就是说

$$\langle 1|\hat{\xi}_t|1 \rangle = -\langle -1|\hat{\xi}_t|-1 \rangle. \quad (6.49)$$

于是，代入公式 (6.43) 有，

$$\text{tr}(\hat{\xi}_tX\rho X) = \text{tr}(\hat{\xi}_tX|1\rangle\langle 1|X) = \langle -1|\hat{\xi}_t|-1 \rangle = -\langle 1|\hat{\xi}_t|1 \rangle = -\langle \xi_t \rangle. \quad (6.50)$$

实际上，对于更一般的 ρ ，我们也有上面的关系。证毕 ■

实际上可以检验更多的数字特征来验证他们的等价性，但是我们不再一一检验，就认为两个描述等价：状态在变算符不动或者算符在变状态不动的描述是等价的。

注意，算符和密度矩阵都是可以做有意义的加法运算的。但是，我们不能写出随机变量之间的类似的看起来象几个部分相加的等式

$$\xi_{t+1} \neq -p\xi_t + (1-p)\xi_t, \quad (6.51)$$

因为这个等式给出的 ξ_{t+1} 不等于 1, -1 即使 ξ_t 等于 1, -1，也就是说，1, -1 这个集合对于通常数的加法这个操作是不封闭的，因此在不改变集合的情况下，这样的加法不能定义。这种随机变量本身的不可加性，随机变量算符或者是随机变量的密度矩阵的可加性是概率论的重要特征。

实际上, 我们仔细观察还会发现, 在这里

$$\mathcal{L}^\dagger = \mathcal{L}. \quad (6.52)$$

但是, 这是由于我们前面提到的 $X^\dagger = X$ 的原因, 这个等式在更一般的情况下 ($X^\dagger \neq X$) 是不对的。可以证明更一般的情况下, 这两个算符之间的关系正好是 Hermitian 共轭。以后, 在量子力学中, 我们会回到这个问题。

这个翻硬币的随机过程还可以通过抽取随机数做数值模拟的方式来分析: 每一步, 选取一个合适的随机数 (以概率 p 取值为 1, 则我们翻转硬币; 以概率 $1-p$ 取值为 -1, 则我们不翻转硬币), 然后经过很长时间以后我们就得到一条轨道, 经过独立的很多次的这样的轨道的模拟, 我们就得到了轨道的随机样本, 分析这个样本我们就能够得到所有我们要回答的问题的答案。这个方法, 实际上叫做 Monte Carlo 方法, 是我们下一节的内容。

6.8 Monte Carlo 方法

作业 6.8.1 用 Monte Carlo 方法实现上一节定义的翻硬币过程, 然后画出多次独立测量的平均值随着测量时间点的变化曲线。

6.9 随机变量的测量

第七章 力学

经典力学（在这里不包含相对论），作为一个学科分支，基本上已经停止发展，除了个别的理论领域，例如湍流的形成以及多体系统的轨道 [?, ?, ?] 等，以及一些应用领域，例如流体力学、结构力学等。那么，我们为什么还要学习经典力学呢，而且所有的物理以及相关专业的学生都要从经典力学开始学起呢？力学，对于物理学来说，除了知识，更重要的是，它也是一门物理学导论课。从思想上，从经典力学我们了解什么是物理学——物理学就是描述物体状态（包含运动状态和结构状态等）的变化，以及寻找导致物体状态变化的原因的科学的科学，以及了解把事物抽象成理想模型的概念；从概念上，我们知道运动的描述，建立位形空间，坐标系，相对性原理，力，能量，保守系统等概念等等；从技术上，我们将学习到最小作用量原理，分析力学的技术——这些技术将来在整个物理学里面都非常重要。

以上是物理学的学生为什么学习力学。那么，系统科学的学生为什么学习力学呢？技术上，分析力学的技术和把一个问题转化为极值优化问题的方式值得我们学习。思想上，力学的思想——从状态的描述到状态发生变化的原因的研究思路以及把事物抽象成理想模型的概念，不仅仅是物理学的思想，而是整个科学的思想。当然，具体概念的层次，倒是有可能不太重要，但是，如果没有这些具体概念，我们的思想和技术就是空中楼阁，或者无皮之毛。

本章推荐参考书：Feynman 《Feynman 物理学讲义第一卷》 [?], 方励之 《力学》 [?], 漆安慎 《力学》 [?], Landau 《力学》 [?].

7.1 力学思想

力学的基本研究对象是日常生活所见到的桌子以及桌子在一个推动的力作用下运动起来这个现象，或者一个乒乓球以及这个乒乓球在拍子挥动

的条件下发生运动这个现象，或者一个汽车以及这个汽车在发动机驱动下和方向盘的控制下发生运动这个现象，或者天上运动的星星以及它们的运动。对于这个研究对象，宏观的物体以及它们的运动，第一个要解决的问题是如何描述他们的运动状态。在这里，物理学引入了一个理想模型——质点，一个有质量没有体积没有形状没有其他任何属性的点。为什么除了质量以外其他的属性都不重要，这是一个很重要的问题。以后我们会知道，现在暂时先留着这个问题。没有体积和形状就要求我们只关心这个物体作为一个整体的运动，我们暂时不关心这个物体的自转等运动形式。这个运动的形式是所有的运动形式里面最简单的一种。我们称为物体的平动，除了平动还有转动等等。后面这些在我们这个精简的课程里面不再涉及。

如果我们把物体看成质点，只考虑其平动，那么这个描述的问题就简单了。一个没有体积和形状的质点，在任何时刻的位置，我们就用空间里的一个点来描述。所谓运动这个现象，就是随着时间这个点的位置发生了变化。于是，我们可以考虑进一步的问题，这个位置什么时候会发生变化呢？我们的生活经验告诉我们，当我们推动桌子的时候桌子会动，我们不推的时候就不动了，我们推的力大一点则运动的速度会变快一点。于是，我们可以建立一个运动的理论：运动需要力来维持，同时不同大小的力可以改变运动的快慢。这个理论在很多时候是可以用来解释我们观察到的现实的。物理学在这里有一个很重要的思想或者说期望、追求：用最少的原理解释最多的现象。也就是说，我们希望这个关于运动的理论能够解释非常多的各种各样的运动。这个时候，我们就发现刚才基于直接生活经验的理论就不好用了。例如，一个已经踢出去的小球，基本上可以认为我们的推力已经不存在了，但是小球不会立即停下来。当然，如果你一定要维持刚才的理论，你可以说推力还存在，甚至说还有一个其他的力作用在小球上，一直到小球停下来，这个力才停止。

这个说法就好像是用雷公电母（或者更多的控制单位）来解释风雨雷电现象一样，我们依赖于雷公电母的情绪来理解什么时候会有风雨雷电，于是，我们需要给他们建立庙宇或者什么的，才能一定程度上影响他们的情绪，于是实现一定程度上对风雨雷电的干预。第一，这个理论个性化非常强，这些神灵或者说控制单位有不同的性格和爱好等等等等，基本上不能实现用最少的原理解释最多的现象的目标；第二，这个理论的可验证性非常差，所以他们需要非常复杂的程序和规范，这样如果祈雨不成功，主持的巫可以把责任推给某个细节上的失败。

在我们这个小球的例子之中，如果我们不能给出来我们的推力在我们的脚离开小球以后，或者额外的力，在什么时候停止，我们就不能对小球的运动有一个有前瞻性/预测性的认识——我们总是可以把这些力的停止的时间定在小球实际运动停止的时间，也就不能用这个理论来指导实际小球的运动，或者说反过来拿小球的实际运动来检验我们的理论。这样的，利用小球实际运动停止的时间来确定力的停止的时间，然后又用力的停止作为导致小球的运动停止的原因，的理论永远是正确的，不能被推翻的。于是，这种理论也是不能够被检验的，不能成为科学。科学理论的基本要求是可证伪性 [?]: 一方面，原则上存在着可以用来否定我们的理论的现实；另一方面，如果我们的理论不这样，那么结果会不一样，而这个不一样的结果在现实的世界中没有被观察到。在这个意义上，如果我们的理论具备以上两条，我们说我们的理论是一个科学的理论。实际上，如果有一个基于炼金术或者草裙舞，甚至巫毒娃娃的理论，无论什么时候我们都可以通过观察例如巫毒娃娃的状态来正确地预测风雨雷电，而且这个预测的规则，或者说配方，也就是什么什么状态对应着什么什么天气，是事先就确定了描述，那么，这个巫毒娃娃的理论也将是科学的。尽管这样的理论可能普适性不是很强，也就是说，没准每一个地区需要一个不同的巫毒娃娃或者不同的规则。

物理学的科学性的要求比这个可证伪性的一般的科学要求稍微高一点点：物理定律要求在时间和空间上都是普适的，还追求最少的原则解释最多的现象。此外，如果有两个理论都可以解释一个现象，这个时候物理学偏好简单的。例如地心说也能比较好的解释很多天体现象，但是理论非常复杂，但是就算只解释同样的现象日心说就简单很多。当然，可证伪性和普适性的要求要比简单性有更高的优先级。

回到我们踢出去以后的小球的运动的例子，让我们尝试如何走得比思辩稍微远一点点。我们已经发现，之前的理论，力是维持运动的原因不是好的理论，不能证伪，也不能像前面描述的天气的巫毒娃娃理论一样管用。那我们来看一看，如何建立一个管用的又能够证伪的理论。其实，基于生活经验，我们也能够知道，力的作用是使运动的状态发生改变。我们知道当我们推一张桌子时候，如果我们希望我们停下来之后桌子动得快一点，我们需要推的时间长一点或者推的力气大一点。甚至在我们推动桌子的过程中，我们也可以通过改变这个推力的大小来达到控制其速度的目的。也就是说，当我们把力作用在一个质点上的时候，我们会改变质点的运动速度。当然，我们还不知道力和速度的改变这两者之间具体的关系是什么，和质点的什

么属性相关，例如大小、形状、质地、生产厂家等等。不过，有了这个思想认识上的突破之后，剩下的突破就容易多了。

解决这个具体什么关系的问题是物理学的另一个，也是最重要的思想是：做实验，做测量。这个来自于 Galileo 的思想使得物理学真正从思辩的层次独立出来，成为科学。Galileo 的实验很简单，把一个物块放在一个斜面上，然后让这个物块滑下来达到一个平面上。接着，改变这个平面的光滑程度。物块从斜面下来以后，我们看看这个物块走得有多远。这个远近是可以测量的。我们发现，随着光滑程度的不同物块运动的距离是不一样的，而且越光滑的表面物块运动的越远。在物块接触到平面之前，所有的过程是一样的，因此（其实你也可以怀疑这个因此），物块刚刚接触到平面上的状态，不管哪一种光滑程度的平面，都是一样的。于是远近不同只能是平面的光滑程度导致的。Galileo 根据他的实验猜测，如果有一个无限光滑的平面，则这个物块会永远运动下去。也就是说，光滑程度的不同可以看作是某种力的不同，而这种力使得物块从运动变成静止。对于无限光滑的表面，这种力为零，于是物体的运动状态不发生改变，一直运动下去。现在，我们当然知道这个力就是摩擦力。但是，当时把这个运动状态发生变化的原因归结为摩擦力是革命性的。然后，作为一个好的理论，我们必须至少满足管用的要求。我们跳过历史发展的逻辑，直接给出理论模型。这个管用的理论在滑块例子之中是这样的：对于任何表面，经过一个测量（可以是让一个标准物块先运动一下，测一下距离，然后经过一番计算），可以得到一个这个表面的性质的参数，叫做摩擦系数。然后，利用这个测量出来的摩擦系数，我们可以预测滑块的运动距离，而且这个预测的运动距离与实际实验相符合。这个就是管用的要求。这个理论本身其实还是不完备的，因为我们跳过了经过怎样的一番计算。将来我们会或者我们已经知道，实际上这个一番计算就是求解 Newton 第二定律对应的微分方程。我们还发现这个定律不仅仅在滑块例子中管用，在各种各样的许许多多的运动中都管用。而且，这里定律原则上允许不遵循这个定律的现实世界的存在，所以是可证伪的。当然，在经典力学（不包含相对论）所研究的对象的范畴内，这样的反例已知没有被发现。通过实验，我们还可以研究运动状态与哪些因素有关，例如，如果我们能够给力的大小定下一个度量的标尺，我们就可以研究给定大小的力对不同的物体的运动状态的改变的不同，例如这些物体可以有不同的形状、颜色、质地等等。

当然，最后我们发现这个因素仅仅是物体所包含的物质的量的多少的

过程，是一个非常非平庸的过程。首先这个物质的量不一定是这个物体的重量，本来的含义是包含的原子（甚至更基本的粒子）的多少。当然，实际上，在我们这个世界中，这个物质的量与这个物体的重量有直接关系。这个关系其实有着非常深刻的含义。广义相对论就是从这个关系出发开始的研究。

物理学还有一个非常重要的思想是理想实验，或者说与实验现象紧密相连的思辩过程。这里举一个非常著名的例子：Galileo 关于重物下落快的理想实验。我们考虑两个重量不一样的物体 M （重）和 m （轻）。我们把它们捆起来组成重物 Mm 。假设重的物体下落的速度快，那么捆起来的物体 Mm 的速度应该比两者分开都快。分开来考虑 Mm 中的 M 和 m ，我们发现由于 M 更快 m 更慢，所以 M 会拉着 m 运动得快一点， m 会拖着 M 运动得慢一点。因此合起来的速度介于两者之间。现在出现了矛盾。因此，我们一开始的假设——重的物体下落得快——是错的。后面，我们对这个理想实验有进一步的讨论。

小结一下，理想化模型化，实验与测量，还有理想实验，是物理学几个核心的重要的思想。质点是理想模型之一。科学性最低的要求是可证伪性，普适性，简单性是物理学对物理理论的要求。

7.2 Newton 力学的基本概念

通过质点这个理想模型，我们已经有了用空间位置来描述运动的简单想法。下面，我们把这个想法实现。首先，为了准确地描述这个空间位置，我们需要有一个坐标系：一个原点，三个独立的方向，一个测量单位。通常这个坐标系的原点需要建立一个参考系上：一个假设不动的东西。确定这个所谓的不动的东西，不是简单的事情。地球在自转和公转，所以在地球上，原则上所有的点都不能作为这个参考系。同样的，太阳系、甚至整个宇宙都在运动，没有一个点是真正意义上不动的。那我们如何选取我们的参考系，然后建立坐标系呢？如果我们的对象是一个在地球表面运动的汽车，那么在汽车的时间尺度内自转和公转造成的影响与汽车的车程相比比较小，于是，地球就是一个好的参考系。以这个汽车出发的时候的地球上的一个点作为原点，是可以接受的。如果我们考虑的是导弹的运动，那么，自转的影响可能就需要考虑。这个时候，参考系可能就要选取地球在某个时刻的点，然后考虑经过这段时间以后地球的运动，或者考虑非惯性参考系的问题。所谓惯性参考系，就是一个“不动”的参考系，或者说一个做匀速运动的参考

系。也就是说，这个参考系在相对于另一个参考系运动速度不变或者运动速度为零。所谓非惯性参考系，就是以速度变化的物体为中心建立的坐标系。

在下面的讨论中，我们假设通过对所关心的过程的考察，讨论这个问题所需要的惯性参考系以及相应的坐标系已经建立起来。那么这个时候，我们的运动的轨迹的描述就是一个 $\vec{x}(t)$ 的时间过程。知道这个在 t 时刻质点的位置 $\vec{x}(t)$ 由哪些因素决定，对于给定的问题计算出 $\vec{x}(t)$ 来，就是我们的理论的目标。通过微积分的学习，我们知道这个 $\vec{x}(t)$ 实际上是空间的一条带参数的曲线，如果我们知道这个曲线在任何时候的切线，我们就可以把这些切线连（积分）起来得到这条曲线。也就是说 $\vec{x}(t)$ 和 $\dot{\vec{x}}(t)$ 加上 t_0 时刻的原点 $\vec{x}_0 = \vec{x}(t_0)$ 是完全等价的。这个 $\dot{\vec{x}}(t)$ 被称为这个质点的速度，记为 $\vec{v}(t)$ 。类似地，我们可以定义加速度 $\dot{\vec{v}}(t) = \vec{a}(t)$ ，以及更高阶的时间导数。完整地知道了任何一阶导数时间过程加上初始条件都是完全了解了这个运动。

Newton 力学告诉我们，给定一个力的位置和速度的依赖函数 $\vec{F}(\vec{x}, \vec{v})$ （由于某些我们这门课程还不能讨论的原因，这个力可以依赖于位置、速度和时间，但是不能依赖于更高阶的时间导数。为了讨论和记号简单，我们进一步假设力不依赖于时间），运动过程由以下方程决定，

$$\ddot{\vec{x}} = \vec{a} = \frac{\vec{F}(\vec{x}, \vec{v})}{m}. \quad (7.1)$$

这个方程被称为 Newton 第二定律。我们还可以把这个方程改写成一阶封闭的微分方程的形式，

$$\dot{\vec{x}} = \vec{v}, \quad (7.2a)$$

$$\dot{\vec{v}} = \frac{\vec{F}(\vec{x}, \vec{v})}{m}. \quad (7.2b)$$

在进一步讨论这个方程之前，我们先来通过几个例子熟悉一下这个方程。

例 7.2.1 自由落体运动：质量为 m 的小球，在地球表面的高度为 h 的地方落下来，初始速度为 0，问经过多长时间落地。

例 7.2.2 经典一维谐振子（连着弹簧的一个小球）：质量为 m 的小球，连在一根弹性系数为 k 原长为 l 的弹簧上。初始时刻，拉开小球距离为 A ，求松开以后小球运动的轨迹。

例 7.2.3 经典一维谐振子 (连着弹簧的两个小球): 质量为 m 的两个小球, 连在一根弹性系数为 k 原长为 l 的弹簧上, 弹簧的另一端固定在墙上。初始时刻, 拉开两个小球距离为 $l + A$, 求松开以后两个小球运动的轨迹。

实际上, 在一个例子中的地球也是这个问题的参与者, 地球与小球之间的引力相互作用才是导致这个运动的原因。但是, 由于小球的运动对地球的状态产生的影响基本可以忽略不计, 所以我们把地球作为参考系, 把随着位置变化的引力用方向和大小都不变的与地球状态无关的重力来代替。第二个例子中, 弹簧的另一端, 也就是墙, 实际上也有运动, 但是幅度小很多。于是, 我们把墙看作参考系, 然后把墙与小球之间的相互作用看成与墙的状态无关的弹簧的弹力。在第三个问题中, 我们把墙替换成了另一个小球, 这个时候, 其运动也成了我们需要考察的对象, 两个小球之间的力与两个小球的状态都有关系。从这个角度来看, 第一个例子考察的是地球与小球之间的相互作用, 第二个例子是小球与墙壁之间的相互作用, 第三个是两个小球之间的相互作用。实际上, 我们相信, 各种各样的力, 都是物体与物体之间的相互作用造成的。有的时候能够看见相互作用的各个主体有的时候只能看见一个主体或者一部分主体, 是我们在考虑问题的时候根据需要划分的系统的不同造成的。让一个系统之外的外界对系统的影响不是很大的时候, 或者说倒过来, 当这个系统对系统的外界影响不大的时候, 我们可以把外界的主体忽略, 当成一个不变的影响或者直接忽略。

这个由一个质点或者多个质点的位置的集合所构成的空间称为一个系统的位形空间。除了位形空间, 我们常常还在相空间里面讨论问题。相空间就是由一个质点或者多个质点的位置加上速度的集合构成的空间。我们可以把公式 (7.1) 看作是位形空间的方程, 把公式 (7.2) 看成是相空间的方程。我们知道两个方程是等价的, 那为什么要引入相空间呢? 因为位形空间的轨道可以相交, 而相空间的轨道不能相交。很容易理解为什么位形空间的轨道可以相交, 撞车就是这样发生的。那为什么相空间的轨道就不会相交呢? 一个相空间的点对应了确定的位置与速度。而速度决定了下一个时刻的位置, 因此, 这个时刻的相空间中的一个点完全决定了下一个时刻这个点在哪里。于是, 如果两个轨道在某一个点相交, 则将来这两个轨道完全一样。同样的道理, 过去也完全一样。于是他们就是同一条轨道。也可以用下面的例子来说明。考察一个一维系统的两条相空间轨道。一维系统的相空间是二维的, 如图所示, 我们来证明有两条不同的相空间轨道不可能在某点相交。

公式 (7.2) 是一个确定性微分方程, 在给定初始条件的情况下解是唯一

的。于是，给定相空间上的某一个点，例如就是这个交点，由于这个交点给出了两个初始条件，通过这个交点的轨道只能有一条。因此，不同轨道不能在一点相交。公式 (7.1) 也是确定性方程，但是它是二阶方程，需要两个初始条件来确定一条轨道。给定位形空间的一个点只能给定一个初始条件，因此存在多条轨道通过这个给定的点。由于相空间的这个很好的性质，在以后的讨论中，相空间是一个很重要的概念。很多时候，速度会替换成动量 $\vec{p} = m\vec{v}$ ，两者相差一个常数（对于给定的系统，质量是一个常数）。所以相空间也经常指由位置和动量坐标联合确定的空间。

作为力学本身来说，主要概念和主要结构就只有这些，空间、时间、参考系、坐标、位置、速度（动量）、加速度、力、位形空间、相空间，还有 Newton 第二定律。这些概念描述了运动，把运动与运动变化的原因，也就是力联系起来。但是，什么是力的问题，力学是回答不了的。力学回答的问题是，给定了力，也就是物体之间的相互作用之后，运动将会如何发生。给定的力到底怎样的问题，由研究具体的力的学科来回答，例如电磁学研究电磁力，引力科学研究引力，核物理学研究核力等等。

有一种类型的力称为保守力或者有势力在物理学里面特别重要。我们甚至相信所有的力的背后，如果我们把我们的系统做的足够大，把足够多的外界包含在我们的系统之中，所有的力都应该是保守力。

7.3 势函数、保守力与能量守恒

保守力是指满足以下关系的力 $F(\vec{x}, \vec{v})$ ，存在一个函数 $V(\vec{x})$ 满足，

$$\vec{F}(\vec{x}, \vec{v}) = -\vec{\nabla}V(\vec{x}). \quad (7.3)$$

其中 $\vec{\nabla}$ 是位置坐标的导数。所以，这样的力，不能是速度 \vec{v} 的函数，而仅是位置 \vec{x} 的函数。

前两个例子中的力，就是保守力，也被称为有势力，也就是可以写成一个势函数然后把力看成是这个势函数的梯度的力；后面的例子中的力，一般看成是来自于相互作用，而这个相互作用的也可以写成势函数的形式。

例 7.3.1 前面三个例子中的势函数： $V_1 = -mgx$ ， $V_2 = \frac{1}{2}k(x-l)^2$ ， $V_3 = \frac{1}{2}(x_1 - x_2 - l)^2$ 。验证通过这三个势函数得到与前面的例子中完全相同的方程。

当方程中所有的力都是保守力的时候，我们来看一下方程式 (7.1) 的一个变形，

$$\begin{aligned} m \frac{d}{dt} \vec{v} &= -\vec{\nabla} V(\vec{x}) \implies \vec{v} \cdot \frac{d}{dt} \vec{v} = -\vec{v} \cdot \vec{\nabla} V(\vec{x}) \\ \implies m v_x \frac{d}{dt} v_x + m v_y \frac{d}{dt} v_y + m v_z \frac{d}{dt} v_z &= -\left(v_x \frac{\partial}{\partial x} + v_y \frac{\partial}{\partial y} + v_z \frac{\partial}{\partial z} \right) V(\vec{x}) \\ \implies \frac{1}{2} m \frac{d}{dt} (v_x^2 + v_y^2 + v_z^2) &= -\left(\frac{dx}{dt} \frac{\partial}{\partial x} + \frac{dy}{dt} \frac{\partial}{\partial y} + \frac{dz}{dt} \frac{\partial}{\partial z} \right) V(\vec{x}) \\ \implies \frac{d}{dt} \left(\frac{1}{2} m v^2 + V(\vec{x}) \right) &= 0 \end{aligned}$$

也就是

$$\frac{1}{2} m v^2 + V(\vec{x}) = E. \quad (7.4)$$

这里 E 是一个不依赖于时间的常数。这个等式被称为能量守恒，其中第一项被称为动能 T ，第二项被称为势能 V 。

例 7.3.2 能量守恒的应用：利用能量守恒求解第一个例子中小球落地以后的速度，以及第二、第三个例子中弹簧被压缩的最大长度。

现在，我们已经完成了 Newton 力学的主要内容的学习，我们来试着求解一下下面这个问题。

例 7.3.3 单摆的周期：固定在天花板上的轻质绳子连着一个质量为 m 小球，拉开绳子一小段距离后松开，求小球再次回到松开位置的时间。

受力分析的困难：最终合力的方向与运动方向的相互关联。需要对运动形式有一定的前瞻性，洞察力。

如果我们再考虑更复杂一点的平面摆，多级摆，这个受力分析就会更困难。一个复杂的机器，有很多部分构成。这个时候，受力分析的方法基本上就不能用了。Lagrange 发明了一个不用做受力分析的求解力学问题的方法。这就是我们下面要介绍的分析力学。

7.4 分析力学的技术

分析力学解决的问题是，只要我们能够写下动能 T 和势能 V 作为自变量 \vec{x} 以及 $\dot{\vec{x}}$ 的函数，或者作为自变量 \vec{x} 以及 \vec{p} 的函数，那么，所有的力学问题都可以直接变成求解微分方程的技术问题。我们先来看看逻辑上为什么可行，然后举几个例子学会如何运用，最后再讨论这个方法的意义。

7.4.1 Lagrangian 力学

对单个质点的守保守力作用的系统定义 Lagrangian,

$$L = T(\vec{x}, \dot{\vec{x}}) - V(\vec{x}, \dot{\vec{x}}) = \frac{1}{2}m(\dot{\vec{x}})^2 - V(\vec{x}), \quad (7.5)$$

Newton 方程与下列方程等价,

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_j} \right) = \frac{\partial L}{\partial x_j}, \quad (7.6)$$

其中 x_j 是 \vec{x} 的 j 分量的坐标。这个方程被称为 Lagrange 方程。

对于单个质点的系统, 我们可以简单验证这个方程确实给出 Newton 方程。

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}_j} \right) = \frac{\partial L}{\partial x_j} \implies \frac{d}{dt} (m\dot{x}_j) = -\frac{\partial}{\partial x_j} V \implies ma_j = F_j \quad (7.7)$$

例 7.4.1 单摆: 固定在天花板上的轻质绳子连着一个质量为 m 小球, 写出这个小球的运动方程。

7.4.2 Hamiltonian 力学

对单个质点的守保守力作用的系统定义 Hamiltonian,

$$H = T(\vec{x}, \vec{p}) + V(\vec{x}, \vec{p}) = \frac{1}{2m}(\vec{p})^2 + V(\vec{q}), \quad (7.8)$$

同时记位置坐标 \vec{x} 为 \vec{q} , 把自变量 $\dot{\vec{x}}$ 替换成 \vec{p} . Newton 方程与下列方程等价,

$$\frac{d}{dt} q_j = \frac{\partial H}{\partial p_j}, \quad (7.9a)$$

$$\frac{d}{dt} p_j = -\frac{\partial H}{\partial q_j}. \quad (7.9b)$$

其中 x_j 是 \vec{x} 的 j 分量的坐标。这个方程被称为 Hamilton 方程。

对于单个质点的系统, 我们可以简单验证这个方程确实给出 Newton 方程。

$$\frac{d}{dt} x_j = \frac{\partial H}{\partial p_j} \implies \dot{x}_j = \frac{p_j}{m}, \quad (7.10a)$$

$$\frac{d}{dt} p_j = -\frac{\partial H}{\partial q_j} \implies \dot{p}_j = -\frac{\partial}{\partial x_j} V = F_j. \quad (7.10b)$$

于是 $m\ddot{x}_j = ma_j = F_j$ 。

例 7.4.2 单摆：固定在天花板上的轻质绳子连着一个质量为 m 小球，写出这个小球的运动方程。

7.4.3 Langragian 力学与 Hamiltonian 力学等价以及最小作用量原理

如果我们从 Langrangian 力学出发推导 Hamiltonian 力学，则我们要做一个变量替换。替换方法是定义

$$p_j = \frac{\partial L}{\partial \dot{q}_j}(\vec{q}, \dot{\vec{q}}), \quad (7.11)$$

然后反解出 $\dot{\vec{q}}$ 作为自变量 (\vec{q}, \vec{p}) 的函数，代入 T 和 V ，得到 T 和 V 把 (\vec{q}, \vec{p}) 作为自变量的函数，然后定义

$$H = \sum_j p_j \dot{q}_j(\vec{q}, \vec{p}) - L(\vec{q}, \vec{p}), \quad (7.12)$$

接着我们就可以运用 Hamilton 方程了。现在，我们证明这个变量替换确实把 Langrange 方程变成 Hamilton 方程。这样的变量替换在统计力学中经常使用有一个专门的名字叫做 Legendre 变换。

这是一个非线性变换，我们不仅变化自变量还变化函数的值，同时也改变方程的形式，但是维持全导数关系不变。通常的变量替换维持函数形式不变或者维持函数值不变。我们这里的这个变换比通常的变量替换复杂。既然维持全导数关系不变，那么标准的做法就是利用全导数关系。为了记号简单，下面的推导假设一维的情况，省略下标 j ，整个过程可以直接推广到多维。

$$\begin{aligned} dH &= d(p\dot{q} - L) = pd\dot{q} + \dot{q}dp - \frac{\partial L}{\partial q}dq - \frac{\partial L}{\partial \dot{q}}d\dot{q} \\ &= pd\dot{q} + \dot{q}dp - \frac{\partial L}{\partial q}dq - p d\dot{q} \\ &= \dot{q}dp - p dq \end{aligned} \quad (7.13)$$

正好留下来的自变量（出现在微分算符后面的变量）是 (q, p) 。于是，

$$\dot{q} = \frac{\partial H}{\partial p} \quad (7.14a)$$

$$\dot{p} = -\frac{\partial H}{\partial q} \quad (7.14b)$$

正好就是 Hamilton 方程。同样的方法可以从 Hamilton 方程推导出 Lagrange 方程。所以两个方程完全等价。这个 Legendre 变换的含义是如果我们维持方程的内涵不变，但是希望改变自变量，新的方程是什么样子。这是一个非常有用的技巧。

到这里我们发现，一个力学系统的全部知识在以位置和速度为自变量的 Lagrangian 中，或者是在以位置和动量为自变量的 Hamiltonian 中。注意，自变量必须是配套的。力学的形式主义的这个发展还可以再深入下去。例如，最小作用量原理把力学纳入到一个极值问题的框架之内。给定一个力学过程的初始位置和末位置（也可以是给定一个力学过程的初始位置和初始速度，推导类似但稍微更复杂一点点 [?], 在这里不再讨论），实际运动遵循使以下函数取极小值的路径，

$$\mathcal{L} = \int_{t_0}^t L(\vec{x}(\tau), \dot{\vec{x}}(\tau)) d\tau, \quad (7.15)$$

其中 $x(t_0) = x_i, x(t) = x_f$ 固定。

现在我们来从最小作用量原理导出运动方程。 \mathcal{L} 是一个函数到数的映射，这个极值问题依赖于一个函数，而不是通常的一个函数的极值问题（这个时候这个极值依赖于一个自变量）。正规的做法是用变分法和泛函微分。这里我们采用物理学家的不严格的做法：把函数 $\vec{x}(t)$ 和 $\dot{\vec{x}}(t)$ 看成是很多很多个从 $t \in [t_0, t]$ 的点，于是，这个函数的函数就成了很多很多个独立自变量的函数，泛函的极值问题就成了多变量函数 $L(\vec{x}(t_0), \vec{x}(t_1), \vec{x}(t_2), \vec{x}(t_3), \dots; \dot{\vec{x}}(t_0), \dot{\vec{x}}(t_1), \dot{\vec{x}}(t_2), \dot{\vec{x}}(t_3), \dots)$ 的极值问题。所以当我们写下类似 $\frac{\partial \mathcal{L}}{\partial x_j(\tau)}$ 的时候，实际上我们是把 τ 时刻的变量 $x_j(\tau)$ 当作独立变量来做的偏导数。这个时候所谓的实际轨道（ \mathcal{L} 取极小值的轨道）和围绕这个轨道的无穷小变动可以用下面这个图来表示。于是，这个极值问题就是这样一个含义，在实际轨道附近的任意的任何一个时刻的无穷小微扰 $\delta \vec{x}(\tau)$ 以及相应的 $\delta \dot{\vec{x}}(\tau)$ ，只能增加 \mathcal{L} 的值。注意， $\delta \dot{\vec{x}}(\tau)$ 不是独立的，就像给定 $\vec{x}(\tau)$ 之后 $\dot{\vec{x}}(\tau)$ 就给定了，给定 $\delta \vec{x}(\tau)$ 就给定了 $\delta \dot{\vec{x}}(\tau)$ 。也就是说，我们的独立自变量其实仅仅是 $\vec{x}(t_0), \vec{x}(t_1), \vec{x}(t_2), \vec{x}(t_3), \dots$ 而不包含 $\dot{\vec{x}}(t_0), \dot{\vec{x}}(t_1), \dot{\vec{x}}(t_2), \dot{\vec{x}}(t_3), \dots$ 。

有了这个理解之后，我们再来从这个极值问题推导运动方程。为了记号简单，我们还是仅仅考虑一维系统，整个推导很容易就能够推广到多维。从

公式 (7.15) 我们得到,

$$\delta\mathcal{L} = \int_{t_0}^t \left(\frac{\partial L}{\partial x(\tau)} \delta x(\tau) + \frac{\partial L}{\partial \dot{x}(\tau)} \delta \dot{x}(\tau) \right) d\tau, \quad (7.16)$$

注意这里时间 τ 的积分可以看成是很多很多个不同的时间点的取和, 每一个时间点的取和对象就是这个普通的多元函数的导数乘上一个独立的任意的 $\delta x(\tau)$ 。同时, 我们需要注意到 $\delta \dot{x}(\tau)$ 不是独立的, 而 $\delta x(\tau)$ 是。于是,

$$\begin{aligned} \delta\mathcal{L} &= \int_{t_0}^t \frac{\partial L}{\partial x(\tau)} \delta x(\tau) d\tau + \int_{t_0}^t \frac{\partial L}{\partial \dot{x}(\tau)} d\delta x(\tau) \\ &= \int_{t_0}^t \frac{\partial L}{\partial x(\tau)} \delta x(\tau) d\tau + \frac{\partial L}{\partial \dot{x}(\tau)} \delta x(\tau) \Big|_{t_0}^t - \int_{t_0}^t \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}(\tau)} \right) \delta x(\tau) d\tau \\ &= \int_{t_0}^t \left(\frac{\partial L}{\partial x(\tau)} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}(\tau)} \right) \right) \delta x(\tau) d\tau. \end{aligned} \quad (7.17)$$

在最后一步我们用了初始条件的约束, 因此 $\delta x(t_0) = \delta x(t) = 0$ 。由于 $\delta x(\tau)$ 可以独立取值, 所以这个极值条件就相当于

$$\frac{\partial L}{\partial x(\tau)} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{x}(\tau)} \right) = 0 \quad (7.18)$$

以上的方程在泛函极值问题中经常出现, 而且专门有一个名字, 称作 \mathcal{L} 所对应的 Euler 方程。这里, 这个方程正好给出 Lagrange 方程的形式。于是, 我们从最小作用量原理推导出了 Lagrange 方程, 而 Lagrange 方程与 Hamilton 方程等价, 进而两者与 Newton 方程等价。于是整个力学就成了最小作用量原理。将来, 也许你还会看到不仅仅是力学, 光学、量子力学、场论、相对论等等所有的物理学都可以表达成作用量的形式。物理学的梦想之一就是要把所有的物理学分支全部统一到一个单独的作用量形式里面来, 当然现在这个还仍然没有实现。但是, 这个形式化的威力和普适性是毫无疑问的。

7.4.4 从单体到多体的分析力学形式

我们花了大量的篇幅来介绍 Lagrange 方程与 Hamilton 方程, 不仅是因为它们可以从受力分析中解脱出来, 还因为, 它们的形式非常容易推广到多体相互作用的系统。一个质点的动力学完全取决于 $V(\vec{x})$, 而这个 $V(\vec{x})$, 我们说过, 也是由于和隐藏起来的不在系统之内的物体相互作用产生的。那么, 当有多个质点, 而且这个多个质点之间存在相互作用的时候, 我们的理

论的这个形式还能不能用呢？这里，所谓相互作用就是从受力分析的角度，一个质点受到的力不仅仅和这个质点的位置相关，也跟其他的质点的位置相关。从势能的角度来说，就是系统的势能同时依赖于多个质点的位置，而不仅仅是一个依赖于单个质点的位置的势能函数的和。用数学表达式来说明就是，

$$V(\vec{x}_1, \vec{x}_2) \neq V_1(\vec{x}_1) + V_2(\vec{x}_2) \quad (7.19)$$

从 Newton 力学来说，单个质点的系统和多个质点的系统没有本质区别，但是受力分析会复杂很多很多，任意两个质点之间的受力都需要考虑，然后最后叠加起来。幸好，分析力学，无论是 Lagrange 形式还是 Hamilton 形式，我们都可以直接从单个质点推广到多个质点，仅仅需要包含，以两个质点的系统为例

$$L(\vec{x}_1, \vec{x}_2; \dot{\vec{x}}_1, \dot{\vec{x}}_2) = \frac{1}{2}m_1\dot{\vec{x}}_1^2 + \frac{1}{2}m_2\dot{\vec{x}}_2^2 - V_1(\vec{x}_1) - V_2(\vec{x}_2) - V_{12}(\vec{x}_1, \vec{x}_2), \quad (7.20)$$

或者

$$H(\vec{q}_1, \vec{q}_2; \vec{p}_1, \vec{p}_2) = \frac{\vec{p}_1^2}{2m_1} + \frac{\vec{p}_2^2}{2m_2} + V_1(\vec{q}_1) + V_2(\vec{q}_2) + V_{12}(\vec{q}_1, \vec{q}_2). \quad (7.21)$$

这个方便的形式在处理多体相互作用的系统的时候非常有用。

例 7.4.3 二级摆：固定在天花板上的轻质硬杆连着另一个轻质硬杆，第二个轻质硬杆连着一个质量为 m 小球，写出这个小球的运动方程。

例 7.4.4 三根弹簧连着的三个小球：三个小球通过三根弹簧首尾相连，放在一个光滑平面上，写下这个系统的运动方程。

分析力学的形式还能够帮助我们简化约束的处理，例如，考虑一个在曲面或者曲线上运动的物体。这里我们就不再讨论，有兴趣的读者可以去看一下 Dirac 的约束运动 [?]

7.5 关于什么是科学的进一步讨论

首先，我们想说明一下任何脱离实验和数学的思辩过程都是不可靠的，不管这个思辩看起来多么合理和漂亮。在本章一开始力学思想这一节，我们介绍了 Galileo 对于重物下落快的理想实验的反驳。这个反驳看起来很有道

理。这个理想实验的推理过程非常巧妙，结论也是正确的。但是，其实，它悄悄地用了一个假设，就是两个物体合起来的速度介于两者分开之间。这个假设实际上包含了对速度（其实是动量）的一个很重要的性质的把握：一定意义上的守恒性。如果我企图用同样的思路来证明越重的金子（质地一样）越值钱是错的，你就会发现，我在这里——两块金子合起来的价值介于两者分开之间——过不去，也就是说金子的价值的守恒性质和速度的守恒性质是不一样的。如果我们试着用数学公式来表达就是对于速度我们有，合起来的速度 $v_{mn} = \frac{Mv_M + mv_m}{M+m}$ ，于是 $v_m \leq v_{mn} \leq v_M$ 。合起来的金子的价值 $V_{mn} = MV_M + mV_m$ ，没有必要介于两者之间。但是，我们看见，整个理想实验的推理过程之中，没有明显地用到任何关于前一个或者后一个守恒律的地方。因此，整个逻辑链条实际上是不完整的。为什么说数学是物理学的语言呢，不仅仅因为数学帮助物理学做计算，还因为数学是逻辑思维的语言。当然，你应该接着问，我怎么知道速度的合成是这样的而金子的合成是那样的呢？我不得不告诉你，这个是实验定律。所以，动量守恒，在整个力学的框架内，是非常重要的实验定律。在漆安慎老师的《力学》的逻辑线路里面，动量守恒，而不是 Newton 第二定律，是整个力学的基础。这是一个很好的逻辑线路，因为力学本身没法定义力：各种各样的力是通过其它具体对象的学科定义的。例如电磁学可以定义电磁力，引力理论和一定义引力等等。当然，在我们分析力学的框架内，我们是不需要力的概念的。所以，我们的力学没有力。分析力学的框架需要的是如何写下各个系统的 T 和 V ，尤其是这个 V ，也是由研究各种相互作用的具体学科决定的，而不是力学的任务。力学处理的问题是，给定 T 和 V 以后的事情，以及为什么给定 T 和 V 就够了的这个框架。

第二，我们想进一步反思一下，科学到底是什么？有一个这样的两个人玩的游戏：一个人 A 在脑子里想一个物体，然后另一个人 B 问前一个人 A 总共 20 个问题， A 只能回答是或者否。通常 20 个问题结束之后， B 都会知道 A 脑子里想的是什么。有一天， A 作弊，他什么也不想，然后回答 B 的所有的问题，但是 A 有能力做到不自相矛盾。于是，20 个问题（甚至可以允许更多的问题）之后， B 还是得到了一个答案。问：这个答案是否正确？因为这个情境下， A 没有事先想好的物体，我们不能说 B 的答案是正确的。完全有可能， B 的答案依赖于 B 的问题是什么。有的人觉得科学就是这个 B ，世界就是这个 A ：世界本没有真理，科学的真理仅仅是科学研究中问的问题导致的。

这个思辩是有一定道理的。但是，我强烈鼓励你取玩一下这个游戏，然后你来做这个作弊的 A 。因为实际上允许无穷多次问问题，你需要获得一个不自相矛盾的答案，又没有一个明确的对象在脑子里面是非常非常困难的。如果有一个明确的对象，对于每一个问题， A 只需要按照这个对象的属性来回答，没有对象的话，就必须把所有的之前的问题和答案都记住，然后才能自洽。而且，物理学要求普适性——物理现象和定律不随着操作者的不同而不同（尽管科学原则上不要求可重复性），也就是说，如果另外一个 C 问你相同的顺序的问题甚至不同顺序但是一样的问题，你必须保证你给 B 和 C 的答案是一样的。而且，这样的 C 可以有无穷多个。这是一件非常困难的事情，就算 A 是世界或者上帝。

换一个角度，如果 B 得到了一个关于这个物体的最终答案，这个答案与之前的所有问题都不矛盾，而且还能利用这个答案来回答 B 想问的其他问题，而且这个答案与 A 将要给出的回答一样，那么，这个最终答案是不是 A 事先想好的，有什么关系呢？我要是 A 的话，就自己也按照 B 的问题推出一个答案，然后我按照这个答案来回答 B 的后续问题。这样，我就简单多了，不用一直维持一个避免自相矛盾的长长的列表。所以，科学到底是不是真理没有意义，只要科学给出的答案与之前的所有科学探索不矛盾，然后还能够用来回答进一步探索中会遇到的问题。

第三，我们举例来看一下佯谬对于科学发展的促进意义。这里，我们讨论 Zeno 佯谬：Achilles 追不上乌龟。Achilles 是一个走得很快的神，假设是乌龟的速度 v 的两倍。乌龟先走距离 L ，然后 Achilles 开始追。我们知道 Achilles 是肯定能够追上乌龟的，时间是 $t = \frac{L}{v}$ 。Zeno 是这样来推理的：我们考虑 Achilles 走到 L 的时候，乌龟在哪里，在 L 前面 $\frac{L}{2}$ 的地方。接着当 Achilles 把这个新的 $\frac{L}{2}$ 走完的时候，乌龟还在前面的 $\frac{L}{4}$ 处。这个过程会继续下去，但是，永远乌龟会在 Achilles 的前面的某个 $\frac{L}{2^n}$ 处，尽管这个距离越来越小。于是，Achilles 永远也追不上乌龟。这个推理没有任何逻辑错误。也许你学过极限，你知道这个距离趋于零，于是在 $n \rightarrow \infty$ 的极限下，Achilles 追上乌龟。但是在任何有限步，Achilles 还是追不上乌龟。而我们不可能等着这个无穷极限完成。那怎么办？其实，根本问题在于这个推理过程中，我们混淆了逻辑时间和实际时间。如果我们计算一下这个切分过程的每一步的实际时间，然后取和，我们就会发现，实际上总的时间是有限的， $t = \frac{L}{2v} + \frac{L}{2^2v} + \frac{L}{2^3v} + \cdots = \frac{L}{v}$ 。从这个佯谬，我们了解到极限的概念与计算，还了解到区分逻辑时间和实际时间。将来，我们还会讨论 Gibbs 佯谬，那

里，微观粒子的可区分和不可区分会导致不同的计算结果，于是我们需要讨论实际微观粒子的可区分性。因此，这些设计巧妙的佯谬其实都意味着当时的一个理论体系尚不完备的地方。这个 Zeno 佯谬其实有更深刻的含义：如果我们所谓的实际时间实际上也是某个逻辑时间呢？我们凭什么认为我们的时间就不可能是某个过程的逻辑时间呢？我们的时间都是通过某种周期现象来度量的，这个周期现象的周期就不一定是均匀的。实际上这个现象在接近黑洞视界的时候（或者说考虑任何高速运动的加速运动的物体的时候）就是需要考虑的。

7.6 本章小结

通过力学的学习，我们已经了解到物理学的基本思考方式是描述事物的状态，以及探索事物状态发生变化的原因。Newton 力学关心的是物体的时空状态——在任何时间点上物体在什么位置、速度是多少，以及状态发生变化的原因。利用力的概念，Newton 力学把原因与状态联系起来，也就是 Newton 第二定律。我们还了解到这个问题完全由物体的 Hamiltonian 或者 Lagrangian 描述，从其中我们可以得到运动方程，然后在一定的初始条件下求解运动方程我们可以得到物体任意时刻的状态，不管过去还是未来。最后，作为一个导出所有理论的原理（通常原理不能通过逻辑推导得来，也很难从现实世界中提炼出来），最小作用量原理提供了理解力学的形式的另一个角度。我们还了解到运用受力分析和 Newton 第二定律来处理相互作用的多体系统以及受约束的系统是很困难的，但是分析力学的技术使得这个问题更容易处理。

在纯粹的力学的图景内，我们的理论是如何得到检验的呢？假设我们做一个特定系统的测量，我们需要知道这个系统的某一个时候的完整的状态——力学告诉我们可以选择位置和速度，加上这个系统所处的环境——最好这个环境是不变的，如果变化我们有可能需要把这个环境也放到系统之中来，然后过一段时间之后如果我们再进行测量，则观测到的状态应该与我们理论预测的一样。在这个图景里面，我们需要知道某时刻系统的完整的状态。

从纯粹的力学的图景稍微扩大一点点，假设我们做一个系统的测量，但是我们对于这个系统的初始状态了解不够或者并不关心，我们也不太关心某一次测量的结果而是更加关心如果同时有很多个人独立地在完全相同的

系统上做测量得到的结果的所有可能性以及主要特征是什么，这个时候，我们怎么办？扔在空中的硬币就是这样一个系统。对它的测量我们关心的是出现人头和字的几率。首先，我们没必要或者不可能完整记录系统的某一时刻或者整个过程的完整状态，其次对于多次完全相同的实验来说初始条件可能也不是完全一样。这个时候，如果经过一段时间系统的状态和初始状态没有太大的关联，也就是说不同的初始状态可能一段时间以后导致类似的状态，那么我们就可以把时间因素和初始条件的因素去掉，只关心在达成那个类似的状态之后，对这个系统做测量得到的结果。这个时候，力学的问题，从 $\vec{x}(t_0) = \vec{x}_0$ 到 $\vec{x}(t)$ 的问题，就没有了 \vec{x}_0 也没有了 t ，成了 $\vec{x}(t \rightarrow \infty)$ 符合什么分布（这里假设连续变量，采用密度分布函数） $\rho(\vec{x})$ 的问题。这个就是我们下一章要解决的问题，称为统计力学的图景。

第八章 统计力学

在上一章的小结里面，我们提到在某些情景之下，力学问题 $\vec{x}(t_0) \rightarrow \vec{x}(t)$ 可以转化成为长时状态的分布函数（这里假设连续变量，采用密度分布函数）的问题 $\rho(\vec{x}(t \rightarrow \infty))$ 。这一章的基本任务就是讨论在什么样的条件下，物理系统的相应的长时状态的分布函数是什么样子。首先，我们会换一个角度（从分布函数的演化的角度）来重新写一遍力学，然后我们期望这个新的力学形式能够给我们一些长时状态的分布函数的启发。接着，我们会发现这样的启发相当有限。于是，我们从新的假设建立长时状态的分布函数的理论。最后，我们举几个例子学会如何做计算。

为什么系统科学的学生要学会统计力学。

推荐参考书：杨展如《高等统计力学》，Rachel 《Morden Course of Statistical Physics》, Kadanoff 《Statistical Physics》

8.1 从力学到统计力学

由于某些原因，例如初始条件不确定，例如所观测的事物本身就是随机客体，我们不能明确地写出来任意时刻的状态，于是我们利用状态的分布函数来描述这个系统： $\rho(\vec{x}, t)$ 表示在 t 时刻系统的状态符合这个分布。在这里，我们已经从对具体的单一系统的描述转到了对抽象的系综的描述：我们想象在 t 时刻有无穷多个完全相同的系统，它们合起来看（如果测量其状态的话）符合分布函数 $\rho(\vec{x}, t)$ 。

让我们通过一个类比来理解这个描述方式的转换。考虑一个两面（称为“正反”）完全对称的硬币。如果我们能够记录初始速度的大小方向，考虑重力和空气阻力，转面的弹性和几何条件等等，我们是完全就可以知道这个硬币在任何时候的状态的。这个就是力学描述。现在我们面对这样一个问题，我站在你面前，你和我都不能用任何辅助的仪器（意味着不仅是你，

我自己也基本上不能真正跟踪硬币的力学状态), 你来猜测硬币落在桌面上以后的状态。对于完成这个任务, 我们发现, 如果能够找到一个分布函数(这个分布函数可能是各种原因造成的, 例如初始速度、桌面条件等等的的不确定性等等, 或者这个硬币本身就是随机的), 而不是整个过程的力学状态, 问题就解决了。这个时候对于这个系统的测量就好像是从一个给定的随机分布中抽样, 力学过程完全藏了起来。理论与测量的对比也不再关于某次测量, 而是关于多次测量的典型值与偏差。这也是为什么这门物理学被称为统计物理学的原因。

现在, 我们用力学的眼光来看一看这个系统, 或者说这个由理论上无穷多个完全相同的系统构成的系综, 的演化, 也就是对于由 $\vec{x}(t) = T(t, t_0) \vec{x}(t_0)$ 描述的系统, 已知 $\rho(\vec{x}, t_0)$, 我们会得到什么样子的 $\rho(\vec{x}, t)$ 。为了描述简单, 我们已经形式上引入了演化算符 $T(t, t_0)$ 。由于环境不变的力学系统的演化由不显含时间的 $H(\vec{x})$ 决定, $T(t, t_0)$ 也肯定是 H 的函数, 具体什么样的函数关系我们暂时不需要。我们来看分布函数的演化。在整个 \vec{x} 的空间上(称为相空间, 每一个点决定了所有的将来的轨迹, 不同的轨迹不相交), 想象一下在 t_0 时刻我们按照 $\rho(\vec{x}, t_0)$ 给出了一个系统密度的分布, 我们想看看在 t 时刻这个密度会变成什么样。这个很简单, 我们来跟踪每一个系统的运动, 也就是

$$\rho(\vec{x}, t) = \rho(T(t, t_0) \vec{x}(t_0), t_0). \quad (8.1)$$

直观地来看, 就是在 t_0 时刻处于 $\vec{x}(t_0)$ 的代表点, 代表了密度 $\rho(\vec{x}(t_0), t_0)$, 随着时间演化, 这个代表点到了别的地方, 同时它代表的密度也跟着跑到了这个新的地方。而 $\rho(\vec{x}, t)$ 的视角是不一样的, 它描述的是每一个时刻分布函数 $\rho(\vec{x})$, t 是不同时刻的分布函数的一个标记。这两者描述的是同一件事情, 但是在图景上的不同是很重要的。前者是跟踪每一个代表点的运动, 它到了哪里就把相应的几率密度带到那里; 后者是忘了每一个代表点从哪里来, 盯住相空间中的某一个点 \vec{x} , 看某个时刻系统落在它这个点上的几率。我们想从 $\vec{x}(t) = T(t, t_0) \vec{x}(t_0)$ 得到 $\rho(\vec{x}, t)$ 的演化方程。前者是力学, 后者是随机分布函数描述的力学, 有可能可以得到统计力学, 也就是 $\rho(\vec{x}, t \rightarrow \infty)$ 的一些特征。

为了得到 $\rho(\vec{x}, t)$ 的演化方程, 我们需要一点点经典场的知识, 具体来说说是场的各种导数算符 $\vec{\nabla}$ 以及连续性方程,

$$\frac{\partial}{\partial t} \rho + \vec{\nabla} \cdot (\rho \vec{v}) = 0. \quad (8.2)$$

连续方程的含义是指在单位时间内在任意体积元内物质的量的增加（方程式 (8.2) 左端）等于进入这个体积元的物质的流的量（方程式 (8.2) 右端， $\rho\vec{v}$ 表示进入表面的流， $\vec{\nabla}$ 牵涉到体积分与面积分之间的转换）。具体连续性方程的推导可以参考电动力学的书籍 [?] 或者其他统计力学的书籍 [?], 或者本章末尾的附录。以一维系统为例，多维实际上是一样的，利用连续性定理，注意到这个时候 \vec{x} 实际上代表了 q, p 动量和坐标两个部分，我们得到，

$$\begin{aligned}\frac{\partial}{\partial t}\rho &= \frac{\partial}{\partial p}(\rho\dot{p}) + \frac{\partial}{\partial q}(\rho\dot{q}) \\ &= -\frac{\partial}{\partial p}\left(\rho\frac{\partial H}{\partial q}\right) + \frac{\partial}{\partial q}\left(\rho\frac{\partial H}{\partial p}\right) \\ &= -\frac{\partial H}{\partial q}\frac{\partial\rho}{\partial p} + \frac{\partial H}{\partial p}\frac{\partial\rho}{\partial q} \triangleq -\{H,\rho\}\end{aligned}\quad (8.3)$$

最后写成 $\{H,\rho\}$ 的形式仅仅是一个简单的记号。这个记号称为 Poisson 括号。这个方程就是 $\rho(\vec{x},t)$ 的运动方程，称为 Liouville 方程。其中我们用到了力学系统的 Hamilton 方程和场的连续性方程。后者是一个数学结果。

有了这个 Liouville 方程之后，我们顺便来证明一下 Liouville 定理。考察 $\rho(q(t), p(t))$ 的全微分，

$$\frac{d}{dt}\rho = \frac{\partial\rho}{\partial t} + \frac{\partial\rho}{\partial q}\frac{dq}{dt} + \frac{\partial\rho}{\partial p}\frac{dp}{dt} = \frac{\partial\rho}{\partial t} + \frac{\partial\rho}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial\rho}{\partial p}\frac{\partial H}{\partial q} = \frac{\partial\rho}{\partial t} + \{H,\rho\}$$

代入 Liouville 方程到上面的表达式，我们得到，

$$\frac{d}{dt}\rho = 0. \quad (8.4)$$

这个称为 Liouville 定理。我们看到 Liouville 定理没有用到 Liouville 方程之外的任何东西（实际上 Liouville 方程加上 Hamilton 方程，但是 Liouville 方程已经包含 Hamilton 方程），因此不包含任何新的信息。这也就是说 Liouville 定理包含的内容还是力学系统的运动方程加上场的连续性方程。从前面提到的盯住代表点的角度来看，Liouville 定理告诉我们代表点确实带着所代表的概率密度运动，而运动过程本身满足力学方程。

有了 Liouville 方程之后，我们就可以研究 $\rho(\vec{x},t)$ 的演化了。那么，是否存在一个长时极限 $\rho(\vec{x},t \rightarrow \infty)$ 呢，或者说存在定态分布函数满足下面的定义呢？

$$\frac{\partial}{\partial t}\rho(\vec{x},t) = 0. \quad (8.5)$$

Liouville 方程告诉我们如果 ρ 完全是 H 的函数, 则上述条件得到满足,

$$\frac{\partial}{\partial t}\rho(H) = -\frac{\partial H}{\partial q}\frac{\partial\rho(H)}{\partial p} + \frac{\partial H}{\partial p}\frac{\partial\rho(H)}{\partial q} \propto -\frac{\partial H}{\partial q}\frac{\partial H}{\partial p} + \frac{\partial H}{\partial p}\frac{\partial H}{\partial q} = 0. \quad (8.6)$$

这是我们能够得到的关于定态分布的重要信息。其实, 基本上这已经是以力学为出发点能够得到所有的关于定态分布的信息, 力学已经不能给我们更多的信息。这信息还不足以解决定态分布的问题。为了解决统计力学的定态分布的问题, 人们提出了新的假设。其中一个思路就是假设微正则系综(具体来说就是如果系统的能量固定, 则系统的状态在等能面上均匀分布), 然后希望从微正则假设推导出来其他更一般的定态分布。在推导的过程中, 再加上态密度(能量等于某个特定值的状态的数量)正比于能量的指数的假设, 就可以从微正则分布得到正则分布。在这里, 因为态密度正比于能量的指数函数的额外假设经常被大家忽略, 我们不采用这样一条看起来好像更有道理更容易接受的道路, 而是直接假设正则分布。这样, 我们强调这条看起来可行的道路实际上理论上是有问题的(需要态密度的额外假设)。如果有一天, 统计物理学基本理论的发展能够用其它的方式证明或者从其他的假设更自然地推导出来正则分布, 我们再来修改这一部分。

8.2 统计力学的基本理论: 系综理论

上面我们提到我们想得到 $\rho(\vec{x}, t)$ 在给定条件下的稳定分布。现在, 我们想象一下我们所面对的系统总能量和总粒子数都固定, 那么我们必然在相空间(所有的 \vec{x} 的集合)的等能面上运动。如果我们的系统可以和外界交换能量, 称为弛豫过程或者弛豫现象, 实际系统很难避免与外界交换能量, 那么总能量不是固定的。这里我们假设总粒子数还是固定的, 系统与外界不交换粒子。我们想知道这个时候系统分布函数是什么。注意这个分布函数给出的平均值与方差等等统计量将来必须测量结果一致。正则分布假设告诉我们,

$$\rho_{Eq}(\vec{x}, T) = \frac{1}{Z}e^{-\beta H(\vec{x})}, \quad (8.7)$$

其中 $\beta = \frac{1}{k_B T}$, k_B 称为 Boltzmann 常数, Z 是归一化常数,

$$Z = \int d\vec{x}e^{-\beta H(\vec{x})}. \quad (8.8)$$

有很多的理由和假说可以使得这个分布函数看起来更顺眼，更容易接受。这里我们一概不去管它们。我们下面的任务很简单，学会从这里开始如何做计算，展示几个例子这个分布函数给出来的计算结果确实与实验符合。

为了完成这个目的，我们来考虑理想气体。我们假定读者已经学过高中的理想气体定律：封闭在一个空间中的气体满足 $PV = Nk_B T$ （对于封闭气体 Nk_B 就是两个常数），而且做过实验进行了验证。这里我们的目的是从理想气体的 Hamiltonian 出发，利用正则分布证明理想气体定律。

例 8.2.1 三维理想气体 $H = \sum_{i=1}^N \frac{p_i^2}{2m}$ ，证明气体宏观量满足 $PV = Nk_B T$ 。

首先，我们计算一个气体分子（分子内部没有其他自由度，例如旋转等等，或者我们看成单原子分子）处在体积为 V 温度为 T 的盒子里的配分函数 Z

$$Z_1 = \int d\vec{x} d\vec{p} e^{-\beta \frac{p^2}{2m}} = V \left(\frac{2m}{\beta} \right)^{\frac{3}{2}} \int d\vec{p} e^{-p^2} \propto V \left(\frac{2m}{\beta} \right)^{\frac{3}{2}}. \quad (8.9)$$

省略的部分完全是一个常数。于是，如果我们有 N 个气体分子，则

$$Z_N = Z_1^N. \quad (8.10)$$

其次，我们来计算物理量，例如平均动能（或者平均速度大小）。为了简单，我们也从一个分子的开始，

$$\left\langle \frac{p^2}{2m} \right\rangle = \frac{1}{Z_1} \int d\vec{x} d\vec{p} \frac{p^2}{2m} e^{-\beta \frac{p^2}{2m}} = -\frac{1}{Z_1} \frac{\partial}{\partial \beta} \int d\vec{x} d\vec{p} e^{-\beta \frac{p^2}{2m}} = -\frac{\partial}{\partial \beta} \ln Z_1 = \frac{3}{2} k_B T. \quad (8.11)$$

这里我们利用偏导数简化了计算。实际上我们发现绝大多数的物理量都可以通过配分函数间接地算出来。这个背后有更深刻的原因（热力学势函数），我们不再深究。这里我们看到单个分子的平均动能满足

$$U = \langle H \rangle = \frac{3}{2} k_B T. \quad (8.12)$$

对于 N 个分子的系统，利用 $U = -\frac{\partial}{\partial \beta} \ln Z_N$ ，我们得到

$$U = -N \frac{\partial}{\partial \beta} \ln Z_1 = \frac{3}{2} N k_B T. \quad (8.13)$$

在我们计算气体的压强之前，我们需要知道压强的微观物理对应，这里我们直接给出答案 $P = \left\langle \frac{1}{3} n m v^2 \right\rangle$ ，其中 n 为单位体积内的粒子数。我们会在附录

中??补充一个说明。有了这个公式，我们看到

$$P = \left\langle \frac{1}{3}nmv^2 \right\rangle = \frac{2}{3}n \left\langle \frac{1}{2}mv^2 \right\rangle = \frac{2}{3}n \frac{3}{2}k_B T = \frac{N}{V}k_B T. \quad (8.14)$$

这个是实验可以验证的理想气体定律。

例 8.2.2 一维 Ising 模型 $H = -J \sum_{j=1}^N S_j S_{j+1}$ 的精确解。其中 $S_j = \pm 1$, $S_{N+1} = S_1$ 求出配分函数 Z 。

首先，对于 $N = 2$ 的系统，我们的计算很简单：总共有四个状态， $|S_1 S_2\rangle = |11\rangle, |1-1\rangle, |-11\rangle, |-1-1\rangle$ ，其能量分别为 $E = -J, J, J, -J$ ，因此

$$Z = e^{\beta J} + e^{-\beta J} + e^{-\beta J} + e^{\beta J}. \quad (8.15)$$

对于 N 比较小的情况，我们类似可以通过计算 2^N 个状态及其能量得到。当 N 很大的时候， 2^N 是一个很大的数字，直接求和计算就不太好用了。

这里我们用两个例子展示了配分函数的计算。下面一个例子，我们将不做计算，但是利用其他人已经得到的配分函数，转而讨论有了配分函数以后我们可以对系统的性质了解到什么程度。

例 8.2.3 已知二维 Ising 模型的配分函数 Z ，求出相变点，画出相图。

8.3 相变与临界现象

从二维 Ising 模型的配分函数开始，我们计算了相变点，画出了相图，下面我们来计算一些典型的临界现象。

8.4 Metropolis 方法

我们已经看到一切的信息都包含在配分函数之中，对于自由度很大的系统，配分函数非常难以计算。我们从 Monte Carlo 方法了解到给定一个概率分布函数，我们有办法产生一个符合这个概率分布的样本集合。现在我们问能不能针对给定的 Boltzmann 分布，给出一个符合这个分布的样本。这个问题的答案就是 Metropolis 算法。

按照通常的 Monte Carlo 算法，分布函数自身是通过显式函数定义好的，然后我们可以考虑反函数方法、离散化、或者变化的反函数方法，或者中心极限定理等等办法来获得抽样。但是，在 Boltzmann 分布中，我们的

分布函数本身的定义不是显式的，其中包含归一化常数，也就是配分函数。如果我们能够显式地计算出来配分函数，也就用不着抽样了。看起来这是一个循环的问题，无解。但是 Metropolis 等物理学家提出了一个很巧妙的办法。

以两个能级 ϵ_1, ϵ_2 状态的系统为例。我们知道每一个能级的几率是 $p_1 = \frac{e^{-\beta\epsilon_1}}{e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2}}, p_2 = \frac{e^{-\beta\epsilon_2}}{e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2}}$ 。因此，重复多次一个简单的抽样：每次产生均匀分布随机数 $\xi \in [0, 1]$ ，如果 $\xi < p_1$ 则取样状态 ϵ_1 否则取样状态 ϵ_2 ，就能够得到我们期望的样本集合。

Metropolis 方法说，我们这样来避过求 $e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2}$ ：假设系统目前处于状态 ϵ_1 (ϵ_2)，随机选取一个不同的状态，这里只能是状态 ϵ_2 (ϵ_1)，然后计算这个可能状态到原始状态能量差 $\Delta E = \epsilon_2 - \epsilon_1$ ($\Delta E = \epsilon_1 - \epsilon_2$)，如果 $\Delta E < 0$ ，则系统转变到新的状态，否则系统以概率 $e^{-\beta\Delta E}$ 转变到新的状态。我们按照这个过程计算一下稳态分布函数。假设 $\epsilon_1 > \epsilon_2$ 。

例 8.4.1 已知转移矩阵 W 的矩阵元， $W_{11} = 0, W_{12} = e^{-\beta(\epsilon_1 - \epsilon_2)}, W_{21} = 1, W_{22} = 1 - e^{-\beta(\epsilon_1 - \epsilon_2)}$ ，求稳态分布。

我们知道这个状态转移的动力学过程可以由 $P_{n+1} = WP_n$ 来描述，其中 W 就是上面的转移矩阵。因此，稳态分布由方程

$$P_{eq} = WP_{eq} \quad (8.16)$$

决定，也就是 W 的本征值为 1 的本征向量。把这个方程整理一下更详细地写下来，

$$\frac{P_{eq}(\epsilon_1)}{P_{eq}(\epsilon_2)} = \frac{e^{-\beta\epsilon_1}}{e^{-\beta\epsilon_2}}. \quad (8.17)$$

于是我们发现

$$P_{eq} = \left(\frac{e^{-\beta\epsilon_1}}{e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2}}, \frac{e^{-\beta\epsilon_2}}{e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2}} \right)^T. \quad (8.18)$$

这个与我们期望得到的 Boltzmann 分布完全一致。于是我们看到 Metropolis 方法确实不用计算 $e^{-\beta\epsilon_1} + e^{-\beta\epsilon_2}$ ，而且也得到了 Boltzmann 分布。把这个两个状态的过程一般化，从任意一个状态出发，随机选择一个不同的状态，然后计算能量差，如果大于零则转移到新的状态，否则按照一定的概率 $e^{-\beta\Delta E}$ 来转移。这个过程的稳定分布就是 Boltzmann 分布。在整个状态

转移过程的设计中，我们遵循了两条原则：遍历性——从任何一个状态能够到达系统中的任何状态；细致平衡条件，

$$\frac{P(S_1)}{P(S_2)} = \frac{e^{-\beta E(S_1)}}{e^{-\beta E(S_2)}}. \quad (8.19)$$

可以说正是细致平衡条件的使用使得我们能够避开求配分函数而直接抽样。保证了这两条，就相当于我们对于状态空间中的任意两个状态都有细致平衡的成立，于是，所有状态按照这个比例归一化一下自然就成了 Boltzmann 分布了。实际上这个方法，对于任何定义明确又不能直接显式写出来的分布函数都可以使用。

例 8.4.2 二维 Ising 模型的 Metropolis 方法模拟，画出相图。

按照一般的 Metropolis 方法的思想，我们需要制定两个步骤：第一、如何选择尝试的状态；第二、选定状态以后转移到这个状态上的几率。第一个不着要注意遍历性，第二个步骤要注意细致平衡。按照这两个原则，我们制定如下的模拟方案。Ising 模型的 Monte Carlo 模拟：

1. 从某一状态 S_0 出发，随机或者顺序选取一个自旋 S_l 。
2. 对自旋 S_l 尝试翻转，计算反转的能量 ΔE 。
3. 如果 $\Delta E < 0$ ，接受翻转以后的新状态；不然，对 $[0, 1]$ 之间均匀随机数抽样 ξ ，如果 $\xi < e^{-\beta \Delta E}$ 接受新状态，否则拒绝。
4. 选取另一个自选，重复以上过程，直到系统的宏观状态（力学量平均值，热力学量）不发生变化。

其中的第三步可以替换成：按照 $\frac{1}{e^{\beta \Delta E} + 1}$ 来抽样，不管 ΔE 大于还是小于 0。或者更一般地，假设系统的末态有 k 种选择，计算各个可能末态的能量差 $\Delta E_k = E_k - E_0$ ，其中 E_0 为初态能量，然后按照分布函数 $\frac{e^{-\beta \Delta E_k}}{\sum_{k'} e^{-\beta \Delta E_{k'}}$ 来抽样。对于只有两个状态的自旋，这相当于取分布函数

$$\begin{cases} \frac{e^{-\beta \Delta E_{\uparrow}}}{e^{-\beta \Delta E_{\uparrow}} + e^{-\beta \Delta E_{\downarrow}}} = \frac{e^{\beta(S_{l-1} + S_{l+1})}}{e^{\beta(S_{l-1} + S_{l+1})} + e^{-\beta(S_{l-1} + S_{l+1})}} & \text{取 } \uparrow \\ \frac{e^{-\beta \Delta E_{\downarrow}}}{e^{-\beta \Delta E_{\uparrow}} + e^{-\beta \Delta E_{\downarrow}}} = \frac{e^{-\beta(S_{l-1} + S_{l+1})}}{e^{\beta(S_{l-1} + S_{l+1})} + e^{-\beta(S_{l-1} + S_{l+1})}} & \text{取 } \downarrow \end{cases} \quad (8.20)$$

可以验证这些选择都满足遍历性与细致平衡。其中最后一个方案是比较特殊的，其转移概率不依赖于 S_l 本身的状态，选择的时候包含 S_l 的所有的可能状态，这里包含 $S_l = \pm 1$ 。

从 Metropolis 方法提出之后, 物理学家们, 尤其是计算物理学家们, 统计物理学的研究者们, 提出了很多更有效率的变形。例如集团翻转算法 [?], 蠕虫算法 [?], 甚至具有类似思想但是方法完全不同的 Wang-Landau 算法 [?]. 尤其是 Wang-Landau 算法的思考方式, 非常有启发性, 值得了解。Metropolis 算法的基本思想是避开配分函数的计算、避开跑遍所有空间, 通过按照 Boltzmann 分布的重要程度来抽样。Wang-Landau 算法也是避开配分函数的计算、避开跑遍所有空间, 但是按照某一个能量下系统的状态个数 (称为态密度或者简并度) 来抽样。也就是说, 态密度越大 (小) 的能量附近样本点更多 (少), 这样不用跑遍整个状态空间, 就可以获得相空间的代表点。然后, 在每一个代表点上计算物理量还有相应的代表点的几率 (这时候原则上需要利用所有代表点来计算配分函数)。

8.5 在物理学以及其他学科中的熵

给定分布函数 ρ , 一般的熵的定义如下

$$S = -k_B \langle \ln \rho \rangle. \quad (8.21)$$

对于确定性事件, 随机变量只取确定值, 于是 $S = 0$ 。对于均匀分布, 例如微正则系综, 每一个状态上的几率都一样, 因此 $\rho = \frac{1}{W}$, 其中 W 为总的状态数, 于是

$$S = -k_B \sum_{j=1}^W \frac{1}{W} \ln \frac{1}{W} = k_B \ln W, \quad (8.22)$$

S 取最大值 (给定 W 个状态, 均匀分布熵最大。这一点可以证明, 见附录??)。确定性事件表示最有序的情况, 均匀分布表示最无序的情况。前者 $S = 0$, 后者 S 取极大值。所以, S 是无序程度的度量。另外, 微正则系综的熵的公式 $S = k_B \ln W$ 被认为是很重要的公式, 被刻在 Boltzmann 的墓碑上。

由于熵是针对一般的分布函数定义的, 是一个数学的定义, 尽管来源于对物理系统的描述, 其有物理学之外的更一般的意义。Shannon 把熵做为信息论的基础 [?].

8.6 从熵最大推导正则分布

之前, 我们提到 Boltzmann 分布的来源是一个尚未解决的问题。现在我们利用微正则系统的熵最大原理来推导正则分布。我们考虑这样一个问题: 考虑 L 个盒子, 每一个盒子能够容纳 0 到 ∞ 个粒子, 每一个在第 l 个盒子中的粒子具有能量 ϵ_l , 假定总能量固定 $\sum_l n_l \epsilon_l = E$, 总粒子数固定 $\sum_l n_l = N$, 问使得熵最大的粒子在每一个盒子中的个数的分布 n_l 是什么形式?

这里隐含的逻辑是这样的。第一, 熵最大的分布形式表示这个宏观状态的微观状态数最多。给定一系列的 n_l 系统的微观状态仍然没有确定, 可以通过交换两个不同盒子里的粒子来实现相同的分布形式。我们想得到这个宏观状态, 其所对应的微观状态最多, 这样在微观状态等几率的假设下, 这个宏观状态出现的几率最大。第二, 所有的能级上的粒子合起来是一个孤立系统, 但是粒子可以在能级之间转移, 因此把单个的能级看成系统, 则这样的系统构成巨正则系统 (可交换能量和粒子), 把所有的能级合起来看成一个系统, 这样的系统构成一个微正则系统。

首先, 给定 n_l 我们的熵可以表达成 $S = k_B \ln \frac{N!}{\prod_{l=1}^L n_l!}$ 。利用 Lagrangian 乘法, 我们需要最大化的目标函数 (常数 k_B 已经被提出去了) 是

$$F = \ln \frac{N!}{\prod_{l=1}^L n_l!} + \beta \left(\sum_l n_l \epsilon_l - E \right) + \alpha \left(\sum_l n_l - N \right). \quad (8.23)$$

这个函数在 $\frac{\partial F}{\partial n_l} = 0, \frac{\partial F}{\partial \beta} = 0, \frac{\partial F}{\partial \alpha} = 0$ 的条件下取极值, 得到以下等式

$$n_l = \frac{N}{Z} e^{-\beta \epsilon_l}, \quad (8.24)$$

$$\sum_l n_l \epsilon_l = E, \quad (8.25)$$

$$\sum_l n_l = N. \quad (8.26)$$

其中 Z 与之前引入的待定常数 α 做过变量代换。现在只剩下两个待定常数 Z 和 β , 他们满足上面的方程。通过计算可以进一步简化, 例如我们可以得到 $Z = \sum_l e^{-\beta \epsilon_l}$, 待定常数 β 满足 $Z(\beta) = e^{-\beta \frac{E}{N}}$ 。

完成这个推导之后, 我们要指出一些要注意的地方。第一, 熵最大原理能够推出 Boltzmann 分布不表示我们了解了 Boltzmann 分布的来源。第二, 这个推导过程中所有的能级构成的系统被看做一个孤立系统, 所以温

度的倒数 β 不是一个外界给定的参数，而是由约束 E 和 N 给定的等价有效温度。这个与真正的正则分布的概念并不完全一样。当然，在平均值的意义下，忽略涨落，他们的含义是相同的：平均能量的一种度量。

8.7 本章小结

8.8 本章技术性附录

第九章 量子力学

学习量子力学的动机，系统科学的研究者学习量子力学的动机，这一章的思路：量子力学的实验经典理论不能解释，能够解释的理论的数学结构，这个数学结构的特殊之处。推荐参考书：Feynman《物理学讲义第三卷》，喀兴林《高等量子力学》，吴兆颜《高等量子力学》，L. Ballentine《Modern Quantum Mechanics》。

9.1 量子系统的实验

在介绍量子系统的实验之前，我们来讨论一个经典对象——一把不太可靠的枪射出来的子弹——的理想实验。

实验 9.1.1 子弹的双缝实验：一把准确度有问题的固定好的机枪在点射同一个地方，但是由于准确度有问题，射出的子弹以一定的概率 p ($1-p$) 到达目标上（下）方 $\frac{d}{2}$ 的地方。然后，在 $\pm\frac{d}{2}$ 的地方各有一个奇怪的小缝，这个小缝会把这个子弹在比较小的仰角 θ_0 的范围内以均匀的概率弹射到距离 L 之外的屏幕上。每一个打在屏幕上的子弹会在屏幕上留下一个荧光的斑点。实验中，我们观察屏幕上子弹留下的痕迹。

我们通过分析单个小缝的结果然后组合两个小缝的行为来分析这个实验。

我们不知道这个实验是不是真的做过，如果做过小缝的行为是不是接近实验中的实际情况。但是，通过经典概率论我们可以得到如下的大致结果，而且这个结果有一个重要的特征：在任意一个小缝的子弹辐射范围内，在两个小缝的情况下都有子弹的痕迹。这个特征是在实际的实验中验证的。如果小缝的行为比我们假设的简单情况复杂，实际的轨迹就会不完全一样，但是这个特征是不变的。一个随机变量的平均值只能够取值于其所有的可能取值之间。

现在，我们来看一下量子系统的真实实验的结果。

实验 9.1.2 电子的双缝干涉实验：一个电子射线枪，射出与上面的情况完全相同的子弹——电子（其中，小缝的行为可能不是子弹的理想实验中的情况），我们观察电子在屏幕上留下的痕迹。我们发现，在某些地方——某些在一个小缝辐射范围内的地方，当两个小缝都打开的时候——电子几乎没有留下痕迹。

我们发现了一个新的特征：两个小缝都打开以后，电子打到某些地方的几率比任何一个小缝独自打开的时候都小。看起来，过两个小缝的电子之间存在着某种相互作用。但是，我们可以让电子枪也点射，每次只能够射出一个电子。这样的话，存在相互作用的解释就非常不可能了。在整个我们关心的空间，每时每刻只有一个电子存在，它跟谁相互作用？还有一种可能：一个电子也能够以某种方式劈开两半，然后同时过双缝，从而实现自己跟自己相互作用。我们不能否认这个可能性。一个简单的检验方法就是在两个小缝的地方各自放置一个探测器，然后看一下是否在某一个时刻可以在两个探测器上同时都探测到半个电子。实际上，这个实验是做过的，我们从来没有探测到过半个电子。将来我们会知道，这个额外的探测器有可能带来一个额外的问题：经过先经过小缝然后再探测器以后的电子的状态有可能就不是仅仅经过小缝不经过探测器的状态了。因此真的靠这个额外的探测器来回答电子如何过两个小缝是有问题的。不过，在这里，我们关心的事情不是电子如何过双缝，而是是否电子在有的时候会分开两半。或者说，在整个空间，我们有很多很多个小缝，难道说电子都需要分成这么多个部分？电子又如何知道需要分成几个部分呢？所以分成几个部分一起过去的假说是不太合理的。当然，我们面对的对象是有独特行为的量子客体，我们不清楚哪些不合理是可以接受的。如果我们没有任何其它方式可以建立电子行为的数学模型和物理图景，那么不合理指出也是必须接受的。

幸好，我们还有别的选择。我们沿着这个想法继续走下去：电子在这个过程中没有与任何其他的东西发生未知的相互作用，除了电子从电子枪里面发射出来，然后在小缝处发生有可能是复杂的但是行为是完全已知的相互作用之外。也就是说，在单个的上（下）方小缝打开的时候，电子的痕迹如图 a（b）所示——这个图样可以不一样，跟小缝有关，但是在两个小缝都打开的时候，成了图 c 的样子。这个图 c 绝对不是 a 和 b 的概率叠加 $pa + (1 - p)b$ 能够得到的。如何理解这个现象？

下面，我们再用两个类似的实验，来展示量子客体的行为的独特性。

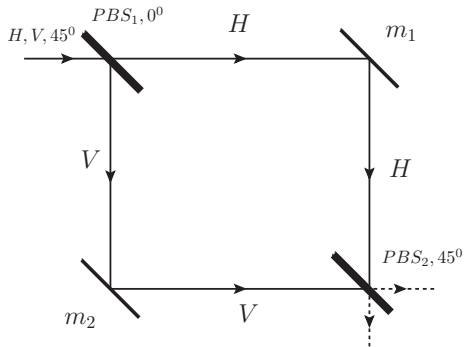


图 9.1: 光子 which-way 实验装置示意图

实验 9.1.3 光子 which-way 实验

实验 9.1.4 电子自旋 which-way 实验

电子的双缝干涉实验，光子 which-way 实验，电子自旋 which-way 实验，都告诉我们假定粒子选择了哪一条路径来走，加上概率论的互斥事件概率可加性，我们预期看到的现象与实验不符合。第一个可能的出路，我们可以修改概率论的互斥事件概率可加性。这个可能性我们学习过，是概率论的核心。我们需要尽可能的保留它：它在经典的时候确实是正确的，而且很容易理解——一件事情有不同的发生方式，那么其结果可以看成是所有方式的结果的某种叠加。第二个可能是我们需要放弃粒子选择了某一条路径这个概念。那么，是否可以替换成光子同时走了所有的路径呢？进一步的实验表明，任何时刻做测量，粒子总是一个一个地被测到，没有测到过半个光子或者半个电子。如果我们假设粒子在测量之前分成许多份，那么测量的时候，在探测到的位置就要求发生某种奇怪的事情，导致其他部分的粒子瞬间到被测量的位置集合。这一点也是很难理解的：被测量到这件事情瞬间传给粒子的已经分离的所有部分。

也就是说，第一、粒子确实选择了某一条路径，尽管我们在观测之前不知道走了那一条路径，或者换一个表达方式：在任何时刻任何地方测量粒子，得到的结果都是这个地方测到了或者没有测到一个完整的粒子；第二、粒子在所有路径——而不是某一个路径——上的某种平均或者某种取和，而且这种取和不能是概率性叠加。这个时时刻刻走一条路径和所有路径的非概率性叠加的矛盾并不在“一条”和“所有”上面，而是在“一条”和“所有的非概率性叠加”上面。我们已经看到，对于经典的随机子弹，注意是随

机的子弹，其在任何时刻走的是“一条”路径，并且观测结果与“所有”的路径的结果的概率性叠加不矛盾。现在，我们需要提供一种数学模型，能够解决上面的这个“一条”和“所有的非概率性叠加”之间的矛盾，从而解决对量子粒子被测量的时候的行为的理解。这个答案就是下面我们要学习和讨论的目前主流量子力学的主要假设。

9.2 量子系统的基本理论

9.2.1 量子系统的静态描述

公理 9.2.1 量子系统的状态由 Hilbert 空间 V 的归一化的矢量 $|\mu\rangle$ 表示，或者，等价地，由 Hilbert 空间上的密度矩阵 $\rho = |\mu\rangle\langle\mu|$ 表示。

实际上密度矩阵的形式更一般，可以表示混合态。我们稍后回到这一点。Hilbert 空间矢量存在加法、数乘、内积的操作。我们需要定义各个操作对于量子系统的意义。

公理 9.2.2 量子系统的可观测量由 Hilbert 空间的 Hermitian 算符矢量 O 表示。

是不是所有的 Hermitian 算符矢量都可以成为量子系统的可观测量是另一个问题。理论上应该可以，但是实际测量过程中，满足要求的仪器不一定能够设计的出来。Hilbert 空间的线性 Hermitian 算符存在算符的加法、数乘、线性变换等操作，我们也需要定义各个操作对于量子系统的含义。

公理 9.2.3 对于给定状态 $|\mu\rangle$ 或者 ρ ，测量可观测量 O 的所有可能结果是 O 的本征值 o 的集合，每一个状态 o 出现的几率满足

$$P_o = |\langle o|\mu\rangle|^2 = \langle o|\rho|o\rangle. \quad (9.1)$$

因此，测量的平均值满足

$$\langle O \rangle = \sum_o o P_o = \text{tr}(O\rho) = \langle \mu|O|\mu \rangle. \quad (9.2)$$

其中我们用到了 $\sum_o o P_o = \sum_o o \langle o|\rho|o\rangle = \sum_o \langle o|O\rho|o\rangle$ ， $\langle o|$ 以及 $|o\rangle$ 是 Hermitian 算符 O 的本征向量。

公理 9.2.4 对于给定状态 $|\mu\rangle$ 或者 ρ ，测量可观测量 O 并观测到状态 o ，则测量以后，系统的状态成为特定的观测到的状态 $|o\rangle$ 或者 $\rho = |o\rangle\langle o|$ 。

测量以后的状态就是所测量到的状态。这一点，在经典力学的世界里不用特别指出：这本来就是测量的含义——测量就是测得系统当时的状态，因此系统在测量之后的当时，必然就是测得的结果反映的状态。在量子的世界里面，这一点不是那么平庸。我们把这个公理单独拿出来。通常的教科书把它与上一个公理合在一起。

下面，我们用这三条来解释上面提到的量子力学的实验。所谓解释就是我们给出一张配方（菜谱），在实验的每一个阶段，你会知道如何对状态进行操作，当时的状态是什么，而且这样的配方对所有的量子系统都适用。至于理解这个配方，是以后的事情。我们先来看一下光子 wick-way 实验。入射光子的状态为（以后我们会了解为什么是这样的状态）

$$\rho_1 = \frac{1}{2} (|H+V\rangle\langle H+V|). \quad (9.3)$$

如果我们在第一个偏振分束镜后面就放一个屏幕，那么我们测量的就是竖直 H 和水平 V 方向的偏振。我们来看一下，按照我们的理论，我们会得到什么结果。我们的可观测量这时候是

$$O_1 = |H\rangle\langle H| - |V\rangle\langle V|, \quad (9.4)$$

其中，为了记号方便，我们把我们的屏幕的荧光点标记为 \pm ：如果水平方向（这个时候是光路方向，不是偏振方向，记为 h ）过来的光达到屏幕上记为 $+$ ，如果竖直方向（记为 v ）过来的光达到屏幕上记为 $-$ 。按照公理9.2.3，我们有测量结果的概率分布，

$$\rho_1^c = \frac{1}{2} |h\rangle\langle h| + \frac{1}{2} |v\rangle\langle v| = \frac{1}{2} |+\rangle\langle +| + \frac{1}{2} |-\rangle\langle -|, \quad (9.5)$$

也就是说水平方向、竖直方向收到光子的概率都是 $\frac{1}{2}$ ，或者说屏幕上记录一个 $+$ 或者 $-$ 的可能性都是 $\frac{1}{2}$ 。

这个预测与真实的实验结果完全吻合。现在，我们用同样的配方，来看一下在第二个偏振分束镜做的测量。由于经过第一个分束镜以后，两个光路被反射镜又完全合到了一起，我们知道（更多的细节以后会补充）这个时候，

$$\rho_2 = \frac{1}{2} (|H+V\rangle\langle H+V|). \quad (9.6)$$

光子的偏振状态完全与入射第一个分束镜的时候相同。第二个分束镜是按照 45° 来分开光束，因此（更多细节以后会补充，粗略地理解就是 $|H+V\rangle$ 代表 45° 的方向， $|H-V\rangle$ 代表 135° 的方向），

$$O_2 = \frac{1}{2} |H+V\rangle\langle H+V| - \frac{1}{2} |H-V\rangle\langle H-V|. \quad (9.7)$$

于是，我们得到可能观测结果的几率分布，

$$\rho_2^c = 1 \cdot |h\rangle\langle h| + 0 \cdot |v\rangle\langle v| = 1 \cdot |+\rangle\langle +| + 0 \cdot |-\rangle\langle -|, \quad (9.8)$$

也就是说水平方向、竖直方向收到光子的概率分别是 1,0，或者说屏幕上记录一个 + 或者 - 的可能性分别是 1,0。这个结果也与实验结果完全吻合。

下面，我们来解释为什么合回来的光子状态是 ρ_2 ，以及为什么 $|H+V\rangle$ 代表 45° 的方向， $|H-V\rangle$ 代表 135° 的方向。同样的，这个解释不是为了理解，而是配方层次的解释。我们先规定 $|H\rangle$ 代表 0° 的方向， $|V\rangle$ 代表 90° 的方向的偏振光。那么 45° 的偏振光如何表示呢？按照我们的公理9.2.1，所有的各个方向的偏振状态的集合是一个 Hilbert 空间，而且这个空间还是两维的：原因是无论哪里一个方向的偏振光过来，经过偏振分束镜以后只有两种可能——要么走水平方向要么走竖直方向。因此， $|45^\circ\rangle$ 必然能够表达成 $|H\rangle, |V\rangle$ 的线性叠加，也就是说，

$$|45^\circ\rangle = \alpha |H\rangle + \beta |V\rangle = \cos\theta |H\rangle + \sin\theta e^{i\phi} |V\rangle, \quad (9.9)$$

其中我们用归一化条件找出了独立变量 θ, ϕ 。对于高维的系统，这样的独立变量会更难找。对于这个状态，我们重复前面 O_1 的测量，可以看到观测的所有可能值的分布函数满足，

$$\rho_{1,45^\circ}^c = \cos^2\theta |h\rangle\langle h| + \sin^2\theta |v\rangle\langle v|. \quad (9.10)$$

这个可以做实验来比较。通过进一步的实验，我们可以发现， 45° 的偏振光束给出的出射光是等几率的，于是

$$\cos^2\theta = \sin^2\theta, \quad (9.11)$$

我们得到 $\theta = 45^\circ$ 或者 $\theta = 135^\circ$ 。至于到底是 45° 还是 135° 的问题，人为规定一个就可以。通常，我们取逆时针方向作为角度增加的方向。 ϕ 的确定更加复杂一些，我们暂时采取最简单的 $\phi = 0$ ，以后我们会回到这个问题。

进入下一部分量子系统的演化之前，最后的问题：为什么 ρ_1 经过分束镜，然后经过反射镜，合回来的光子状态是 ρ_2 。同样地，这里给出的是配

方层次的解释。考虑光子的偏振和光子走的光路构成的整个系统，我们称先走水平方向然后反射回到垂直方向的光路为光路 1，先走垂直方向然后反射回到水平方向的光路为光路 2。在光子进入偏振分束镜之前，整个系统的状态为

$$\rho_0^{P,T} = \frac{1}{2} (|H + V\rangle\langle H + V| \otimes |0\rangle\langle 0|), \quad (9.12)$$

其中 P, T 分别表示偏振 (polarization) 和路径 (Trojectory), 0 表示在光子经过分束镜之前的光路。此时它不区分光路 1, 2。经过分束镜进入反射镜之前,

$$\rho_1^{P,T} = \frac{1}{2} (|H \otimes 1 + V \otimes 2\rangle\langle H \otimes 1 + V \otimes 2|). \quad (9.13)$$

这表示偏振为 H (V) 的光子走的是路径 1 (2)。经过反射镜之后到达最后的分束镜之前,

$$\rho_2^{P,T} = \frac{1}{2} (|H \otimes 0 + V \otimes 0\rangle\langle H \otimes 0 + V \otimes 0|) = \frac{1}{2} (|H + V\rangle\langle H + V| \otimes |0\rangle\langle 0|). \quad (9.14)$$

这表示偏振为 H 和 V 的光子走的是路径重新不可区分。对于这个状态，如果我们只关心偏振状态而不是光子的路径的话，那么这个系统的状态就跟如下状态等价

$$\text{tr}^T (\rho_2^{P,T}) = \frac{1}{2} (|H + V\rangle\langle H + V|). \quad (9.15)$$

我们得到的状态正好是前面的 ρ_2 。这里，我们用了部分迹的运算 $\text{tr}^T (\cdot)$ ，它表示对变量 T 的自由度求和。当我们需要计算的可观测量只与 P 的自由度有关的时候，按照公理 9.2.3,

$$\langle O^P \rangle = \text{tr} (O^P \rho^{P,T}) = \text{tr}^P (O^P \text{tr}^T (\rho^{P,T})). \quad (9.16)$$

因此， $\text{tr}^T (\rho^{P,T})$ 就好像是这种情况下系统在 P 自由度下的状态，称为约化密度矩阵。在经典概率论中，对应的数学概念就是约化密度分布函数。

用以上配方，我们可以类似地解释双缝干涉实验，以及电子自旋 which-way 实验。再一次强调，这里所谓的解释，其实是一个预测量子系统状态的机器，一个数学模型：从一个最开始的状态出发（这个状态本身在给定实验条件的情况下有固定的规则写下来），测量的结果是什么（可观测量原则上由给定的实验仪器与步骤确定），测量以后的状态是什么；然后测量结果与

测量以后的状态能够通过进一步的实验检验。物理学的理论，可以认为是对世界如何运行的描述，甚至有的时候可以回答为什么的问题。但是，其基本任务是回答是什么，怎么样的问题。也就是说，如果有一个理论，对于给定的实验条件、实验仪器、实验步骤的系统的实验结果能够给出正确的计算结果，那么这个理论就是这个现象的好的理论。当然，一个更好的理论还要求能够解释更多的同类的现象。但是，只要能够达到以上的目标，一个理论就能够建立起来。在量子力学的理论中，路径积分量子化的物理图像就与我们这里的正则量子化的图像有所不同。一个好的物理学家，对于核心的现象，头脑中常常是多于一个理论的，而且还要能够在这些不同形式的理论之间自由的转换。

在上面的例子中，我们有一个小小的未完成的步骤，为什么状态是 ρ_0^{PT} 光子经过偏振分束镜就能够变成状态 ρ_1^{PT} 了呢？这个问题与量子系统的演化有关。在解释它之前，让我们来学习一下量子系统的演化。

9.2.2 量子系统的演化

公理 9.2.5 对于给定初始状态 $|\psi(0)\rangle$ 或者 $\rho(0)$ 的系统，系统自身动力学完全由系统的 Hamiltonian H 通过如下方程决定，

$$i\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle, \quad (9.17)$$

或者等价地

$$i\frac{\partial}{\partial t}\rho(t) = [H, \rho(t)]. \quad (9.18)$$

这个方程被称为 Schrödinger 方程。有时候，后者也被称为 Liouville-von Neumann 方程。如果 H 不显含时间，形式上以上的两个方程可以解开如下，

$$|\psi(t)\rangle = e^{-iHt}|\psi(0)\rangle, \quad (9.19)$$

或者等价地

$$\rho(t) = e^{-iHt}\rho(0)e^{iHt}. \quad (9.20)$$

利用 H 的本征值和本征向量（记为 $E_n, |n\rangle$ ），我们还可以得到

$$U(t) = e^{-iHt} = e^{-iHt} \sum_n |n\rangle\langle n| = \sum_n e^{-iE_n t} |n\rangle\langle n|. \quad (9.21)$$

也就是说，如果我们能够得到 H 的本征值和本征向量，那么，从给定的初始状态出发，任何一个时刻的状态我们都可以得到。得到新的时间点的状态之后，如果我们需要在这个状态上对某一个可观测量做测量，我们在利用前面的公理9.2.3来解决测量结果的问题。

例 9.2.1 经过 x 方向磁场以后，在其正方向出射（反方向完全被挡主）的 $\frac{1}{2}$ -自旋系统，经过 z 方向磁场 B_z 中 $t = \frac{\pi}{4B_z}$ 时间以后出射。求正方向和反方向接收到自旋的几率。

作业 9.2.1 初始状态是 ρ_0^{PT} 光子经过偏振分束镜就能够变成状态 ρ_1^{PT} ： $H =$

到此为止，量子力学的基本框架的内容就有了。剩下的，多粒子系统的 Hilbert 空间是所有的粒子的 Hilbert 空间的直积，全同多粒子系统的状态需要考虑对称性与反称性等等，以及更多的坐标与动量空间的 H 如何求本征值的问题，我们都留给有兴趣的读者去阅读给物理学专业学生的量子力学教材。我们的基本任务是帮助读者理清思路，建立最基本的框架，掌握最核心的概念，学会运用最基本的计算分析技术。在厘清概念方面，我们认为除了以上的基本框架，量子力学的数学结构与经典力学、经典概率论的数学结构之间区别也是非常重要的。量子力学部分的最后，我们讨论一下这个区别。

9.3 量子与经典密度矩阵的区别

这一部分，我们来尝试用经典密度矩阵来解释双缝干涉现象、光子 which-way 实验和电子 which-way 实验。在那之前，我们尝试更简单的：来解释一下 Stern-Garlach 实验。我们说满足如下分布的纯随机客体，给出的计算结果能够解释 Stern-Garlach 实验，

$$\rho_c = \frac{1}{2} (|H\rangle\langle H| + |V\rangle\langle V|). \quad (9.22)$$

我们同样测量前面的可观测量 O_1 ，按照9.2.3（同时也是经典系统遵循的），我们得到测量结果的分布函数

$$\rho_c^c = \frac{1}{2} |+\rangle\langle +| + \frac{1}{2} |-\rangle\langle -|. \quad (9.23)$$

这表示测量的结果有两种可能（重复多次的话，屏幕上会有两个斑点），这两种可能的几率都是 $\frac{1}{2}$ 。那么这样的经典密度矩阵是不是也能够解释双缝干涉现象、光子 which-way 实验和电子 which-way 实验呢？我们试试。

还是按照前面的配方，第一步，到达第一面偏振分束镜之前的状态是

$$\rho_0^{PT;c} = \frac{1}{2} (|H\rangle\langle H| + |V\rangle\langle V|) \otimes |0\rangle\langle 0|. \quad (9.24)$$

经过第一面分束镜之后，到达反射镜之前，

$$\rho_1^{PT;c} = \frac{1}{2} (|H \otimes 1\rangle\langle H \otimes 1| + |V \otimes 2\rangle\langle V \otimes 2|). \quad (9.25)$$

经过二面分束镜之前，

$$\rho_2^{PT;c} = \frac{1}{2} (|H\rangle\langle H| + |V\rangle\langle V|) \otimes |0\rangle\langle 0|, \quad (9.26)$$

回到入射光的状态。然后我们需要计算按照 45° 和 135° 方向偏振的几率。原则上，为了计算这个，我们需要代表这两个方向偏振光的算符 \hat{P}_θ ，然后利用 $P_\theta = \text{tr}(\hat{P}_\theta \rho)$ 算出这两个偏振方向的几率。实际上，经典力学框架之内，我们找不到这样的算符的数学形式。我们之能够通过实验来给一个说法。

首先，我们注意到，对于我们的具体情况，

$$P_\theta = \text{tr}(\hat{P}_\theta \rho_2^{PT;c}) = \frac{1}{2} \text{tr}(\hat{P}_\theta \rho_{2,H}^{PT;c}) + \text{tr}(\hat{P}_\theta \rho_{2,V}^{PT;c}) = \frac{1}{2} P_\theta^H + \frac{1}{2} P_\theta^V, \quad (9.27)$$

其中， $\rho_{2,H}^{PT;c}$ 代表 $\rho_2^{PT;c}$ 的第一部分（关于 $|H\rangle\langle H|$ 的部分）， $\rho_{2,V}^{PT;c}$ 代表 $\rho_2^{PT;c}$ 的第二部分（关于 $|V\rangle\langle V|$ 的部分）。这个公式表明，不管 \hat{P}_θ 是什么，概率性叠加公式是成立的：某事件有两种发生的可能，则其概率是这两种事件发生概率按照全概率公式取和。现在，我们需要知道 $P_{45^\circ}^H, P_{45^\circ}^V$ 和 $P_{135^\circ}^H, P_{135^\circ}^V$ 。因为没有算符的具体形式，我们通过如下的实验来推算这几个数值。制备光子的状态 $|H\rangle\langle H|$ ，然后做 45° 偏振分束镜的测量。也就是让上面的实验中的光路 1（标注了 H ）的光子直接通过第二面偏振分束镜，同时把光路 2（标注了 V ）的光子挡住。实验告诉我们，分束镜之后有两个可能的输出光斑，多次实验告诉我们这两个光斑的强度在误差范围内完全相同。于是我们知道，

$$P_{45^\circ}^H = \frac{1}{2} = P_{135^\circ}^H. \quad (9.28)$$

类似的，我们可以通过让光路 2 的光子单独打到分束镜上得到

$$P_{45^\circ}^V = \frac{1}{2} = P_{135^\circ}^V. \quad (9.29)$$

于是，

$$P_{45^\circ} = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2} = P_{135^\circ}. \quad (9.30)$$

最终我们得到：按照经典力学的几率表达式公式 (9.24)，我们会得到两个可能的输出光斑，而且其多次实验得到的强度相同。这个计算结果与实验完全不相符。

对比公式 (9.3) 和公式 (9.22)，我们看到其区别就是

$$\Delta\rho = \frac{1}{2} (|H\rangle\langle V| + |V\rangle\langle H|). \quad (9.31)$$

这个部分只有在量子的密度矩阵中存在，经典的没有。然而这部分对于解释量子系统的实验行为是非常重要的。由于量子系统的状态由 Hilbert 空间的矢量描述，而矢量可以用基矢的叠加来表示（例如 $|\mu\rangle = \mu_1 |v_1\rangle + \mu_2 |v_2\rangle$ ），当我们把这样的矢量转变成成为密度矩阵的时候（ $\rho = |\mu\rangle\langle\mu|$ ），必然会带来以上形式的交叉项。我们粗略地称之为密度矩阵的非对角元。之所以“粗略”是因为这些非对角元在其它的表象下，可以成为对角元。而且因为密度矩阵是 Hermitian 的这样的表象总是存在的，尽管这个表象和具体实验中的可观测量的表象经常不一样。如果两者一样，那么在这个一样的表象下，密度矩阵和可观测量都是对角的。这个时候，整个系统的数学形式就完全回到了经典概率论的形式。至少在这个可观测量的测量的角度来说，量子性（相干相消或者相干相长）消失了。当然，同样的系统，如果我们考虑另外一个可观测量，只要这个可观测量对应的算符与前一个可观测量不能对易（可对易算符有共同本征矢量集合），那么这个时候量子性又会体现出来。因此，也可以说量子力学的量子性体现在算符的非对易性上：物理量不再是数，而是矩阵或者说算符。

尽管我们一再指出，通过量子实验表现出来的量子系统的行为使得我们有必要引入事件之间的加法操作，才是量子力学的数学采用矩阵的形式的原因，其实，形式上，我们完全可以从对易关系开始构建整个量子力学。这个就是下面的正则量子化。

9.4 正则量子化与表象理论初步

正则量子化的基本思想是，找到一个经典系统的正则坐标 q 和正则动量 p ，然后把 q 和 p 都升级成算符 \hat{q} 和 \hat{p} ，并假设其对易关系为

$$[\hat{q}, \hat{p}] = i. \quad (9.32)$$

这两个物理量之间的对易关系有时候被称为正则对易关系。然后其它物理量，它们是 q, p 的函数，就成了 \hat{q}, \hat{p} 的算符。其中，对应关系的细节（例

如 $q^2 p$ 应该成为 $\hat{q}^2 \hat{p}$ 还是 $\hat{q} \hat{p} \hat{q}$ 还需要进一步讨论, 但是有了正则对易关系之后, 其他的物理量都成了一般情况下不相互对易的算符。于是, 共同本征矢量不存在了, 密度矩阵的表述需要非对角元了, 量子干涉现象出现了。我们的下一个例子就是一维谐振子的量子化。

例 9.4.1 一维谐振子的量子化与求解: 经典一维谐振子的 Lagrangian 是 $L = \frac{1}{2} m \dot{x}^2 - \frac{1}{2} m \omega_0^2 x^2$, 求解对应的量子系统的 Hamiltonian, 以及最小能量值的本征态对应的本征函数, 以及其时间演化。

首先, 我们需要求出正则坐标和动量, 其次要把它们看成算符, 最好能够在某一套基矢下把这两个算符的具体分量形式写下来, 然后我们把这些分量形式代入 Schrödinger 方程, 求解本征函数以及时间演化。

取 $q = x$, 则 $p = \frac{\partial}{\partial \dot{q}} = m \dot{x} \implies \dot{q} = \frac{p}{m}$, 于是 $H = p \dot{q} - L = \frac{p^2}{2m} + \frac{1}{2} m \omega_0^2 x^2$ 。然后, 我们需要把 q, p 看作算符 \hat{q}, \hat{p} , 并且让其对易关系 $[\hat{q}, \hat{p}] = i$ 。这里我们取 x 的表象, 也就是我们的基矢满足

$$\hat{x}|x\rangle = x|x\rangle. \quad (9.33)$$

也就是说

$$\hat{x} = \sum_x \hat{x}|x\rangle\langle x| = \sum_x x|x\rangle\langle x|. \quad (9.34)$$

我们来看看满足以上对易关系的 \hat{p} 会是什么? 我们先给出答案, 再给出一个直观的说明。答案是在 $|x\rangle$ 的表象下, 任意一个函数 $\psi(x)$ 成了一个如下的矢量,

$$|\psi\rangle = \sum_x \psi(x)|x\rangle, \quad (9.35)$$

算符 \hat{x} 作用在 $|\psi\rangle$ 可以表达成其如何做用在分量 $\psi(x)$ 上:

$$\hat{x}|\psi\rangle = \sum_x \hat{x}\psi(x)|x\rangle = \sum_x x\psi(x)|x\rangle, \quad (9.36)$$

于是, 从分量形式来看,

$$\hat{x}\psi(x) = x\psi(x). \quad (9.37)$$

那么, 在这个意义上, 从分量形式来看,

$$\hat{p}\psi(x) = -i\frac{\partial}{\partial x}\psi(x). \quad (9.38)$$

可以验证如下:

$$(\hat{x}\hat{p} - \hat{p}\hat{x})\psi(x) = -ix\frac{\partial}{\partial x}\psi(x) + i\frac{\partial}{\partial x}x\psi(x) = i\psi(x). \quad (9.39)$$

于是,

$$\left[x, -i\frac{\partial}{\partial x} \right] = i. \quad (9.40)$$

现在我们来给一个更数学的说明(更严格的证明见喀兴林的《高等量子力学》[?])。从正则对易关系出发,

$$\langle x' | \hat{p} | x \rangle = \langle x' | (\hat{x}\hat{p} - \hat{p}\hat{x}) | x \rangle = \langle x' | \hat{p} (x' - x) | x \rangle = (x' - x) \langle x' | \hat{p} | x \rangle \quad (9.41)$$

因此

$$\langle x' | \hat{p} | x \rangle = i \frac{\langle x' | x \rangle}{(x' - x)} = i \frac{\delta(x' - x)}{(x' - x)} = -i \frac{\partial}{\partial x'} \delta(x' - x). \quad (9.42)$$

最后一步,我们利用了 Dirac δ 函数的性质, $x\delta'(x) = -\delta(x)$ 或者说 $-\delta'(x) = \frac{\delta(x)}{x}$ (这两个表达式可以用分步积分证明。更严格的数学证明可以避免这两个等式[?])。这个就是算符 \hat{p} 在 \hat{x} 的本征矢量的表象下的分量形式。这个表象称为位置表象。如果我们愿意,我们可以验证

$$\langle x | \hat{p} | \psi \rangle = \sum_{x'} \langle x | \hat{p} | x' \rangle \langle x' | \psi \rangle = \sum_{x'} -i \frac{\partial}{\partial x} \delta(x - x') \psi(x') = -i \frac{\partial}{\partial x} \psi(x). \quad (9.43)$$

以后,在位置表象下,我们就简单把动量算符 \hat{p} 记作

$$\hat{p} = -i \frac{\partial}{\partial x}. \quad (9.44)$$

现在,我们准备好了来求解 Schrödinger 方程了。对于一维谐振子系统,其在位置表象下的形式是,

$$i \frac{\partial}{\partial t} \psi(x, t) = \left(-\frac{1}{2m} \frac{\partial^2}{\partial x^2} + \frac{1}{2} m \omega_0^2 x^2 \right) \psi(x, t). \quad (9.45)$$

在位置表象下,我们求解的程序还是和一般的抽象的算符,或者任何其他表象,一样的。首先,求得 H 的本征值 E_n 和本征向量 $\phi_n(x)$, 这里称为本征函数;然后,利用这些个本征函数加上时间项组合起来,得到,

$$\psi(x, t) = \sum_n c_n \phi_n(x) e^{-iE_n t}, \quad (9.46)$$

其中

$$c_n = \int_{-\infty}^{\infty} dx \phi_n^*(x) \psi(x, 0), \quad (9.47)$$

$\psi(x, 0)$ 为初始状态对应着的波函数。

现在, 我们来看看谐振子的本征波函数, 其满足

$$\left(-\frac{1}{2m} \frac{d^2}{dx^2} + \frac{1}{2} m \omega_0^2 x^2\right) \phi_n(x) = E_n \phi_n(x). \quad (9.48)$$

这个方程当然可以想办法直接求解。我们这里, 作为一个例子, 我们仅仅考虑这个方程最简单的解 (除了 $\phi_n(x) = 0$ 以外)。我们看到, 如果 $\phi_n(x)$ 的二阶导数刚好给出 x^2 项, 那么方程左边的系数就会不依赖于 x , 这样整个方程就简单了。也就是说大概 $\phi(x) = e^{\alpha x^2}$ 就行: $\frac{d^2}{dx^2} \phi(x) = 2\alpha e^{\alpha x^2} + 4\alpha^2 x^2 e^{\alpha x^2}$ 。于是 $\alpha^2 = \frac{1}{4} m^2 \omega_0^2 \implies \alpha = -\frac{1}{2} m \omega_0^2$ (实际上正的 α 是没有物理意义的, 在 x 很大的时候发散。物理系统一般要求 $\phi(x)$ 有限), 而这个时候的本征能量 $E_0 = \frac{1}{2} \omega_0$ (能量下标 0 是有含义的。这里, 我们不讨论这个问题)。因此,

$$\phi_0(x) = e^{-\frac{1}{2} m \omega_0 x^2}, E_0 = \frac{1}{2} \omega_0. \quad (9.49)$$

系统如果处于这个状态, 那么其时间演化为

$$\psi(x, t) = e^{-\frac{1}{2} m \omega_0 x^2} e^{-i \frac{1}{2} \omega_0 t}. \quad (9.50)$$

在这个例子中, 我们讨论了位置表象下的 \hat{x} 和 \hat{p} , 然后利用他们的形式, 求解了一维谐振子的基态波函数 (我们没有证明这个是基态, 在下一个例子中, 我们会顺便补充这个证明)。我们还以谐振子为例讨论了正则量子化的一般手续。下面, 我们再举一个例子, 还是求解一维谐振子的 Hamiltonian 的本征值和本征态, 但是不在任何表象中求解, 而是直接利用对易关系。这个解法也因此被称为代数解法。这个解法是科学发展历史上的一件艺术品, 是很多量子力学的进一步理论的发展的基础。

例 9.4.2 一维谐振子的代数解法: 经典一维谐振子的 Hamiltonian 是 $H = \frac{\hat{p}^2}{2m} + \frac{1}{2} m \omega_0^2 \hat{x}^2$, 求解其本征能量。

作业 9.4.1 一维谐振子的湮灭算符的本征向量 $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$ 在粒子数表象下的形式。

除了正则量子化，一个经典系统的对应的量子理论还可以通过路径积分量子化等其它方法来构建。这里就不再讨论关于路径积分量子化的方法，尽管正像 Feynman 说的，任何一个严肃的物理学家都应该在脑子里对一种理论存在多种等价的表述方式。我们这里提到的表象理论，是指一个抽象定义的算符，通常仅仅通过对易关系定义的算符，可以在某一个算符的本征矢量做基矢的情况下表示出来（把各个分量求出来）。这个时候，有可能一个算符在一套基矢下的形式和另外一套基矢下的形式完全不一样，但是它们表示同一个算符。更一般地来说，一个理论的不同形式，也可以看成是这个理论的不同表象。

9.5 Schrödinger 绘景与 Heisenberg 绘景

在《概率论》一章中，我们利用翻硬币这个简单随机过程举例说明了随机过程的演化过程可以通过两种方式来表达：分布函数（密度矩阵）的演化，以及随机变量的演化，且两种表述方式等价。其中，随机变量的演化的描述中，我们假设分布函数不变，改变的是随机变量的算符形式。这里，我们用类似的两种角度来看量子系统的演化。第一种分布函数（密度矩阵）的演化的描述，就是我们在这一章的前面一直在使用的方法，称为 Schrödinger 绘景。这里，我们介绍另一种算符的演化，分布函数保持不变的描述方法，称为 Heisenberg 绘景。

我们已经知道，对于 H 不显含时间 t 的量子系统，其的演化可以表达为

$$\rho(t) = e^{-iHt} \rho(0) e^{iHt}, \quad (9.51)$$

物理量 A^S （为了表示 Schrödinger 绘景的算符与将要定义的 Heisenberg 绘景的算法的区别，我们把之前的算符的记号稍作修改，加上一个上标 S ）的观测值 α 和相应的几率 p_α ，由如下方程决定

$$p_\alpha = \text{tr}(|\alpha^S\rangle\langle\alpha^S| \rho(t)) = \langle\alpha^S | \rho(t) | \alpha^S\rangle = \langle\alpha^S | e^{-iHt} \rho(0) e^{iHt} | \alpha^S\rangle. \quad (9.52)$$

现在，我们定义

$$A^H = e^{iHt} A^S e^{-iHt}, \quad (9.53)$$

$$\rho^H = e^{iHt} \rho^S(t) e^{-iHt} = \rho^S(0). \quad (9.54)$$

然后记相应的 A^H 的本征向量为 $|\alpha^H(t)\rangle$, 则

$$|\alpha^H(t)\rangle = e^{iHt} |\alpha^H(0)\rangle \quad (9.55)$$

因此,

$$\text{tr}(A^H \rho^H) = \text{tr}(|\alpha^H(t)\rangle \langle \alpha^H(t)| \rho(0)) = \langle \alpha^S | e^{-iHt} \rho(0) e^{iHt} | \alpha^S \rangle = p_\alpha. \quad (9.56)$$

也就是说, 计算各种观测量的方式和得到的结果都不变。这种保持变换前后基本公式的形式和计算结果都不变的变换, 在物理学里, 有非常重要的地位。于是, 我们得到量子力学的另一种形式,

$$A^H = e^{iHt} A^S e^{-iHt}, \quad (9.57)$$

$$\rho^H = \rho^S(0), \quad (9.58)$$

$$\langle A \rangle = \text{tr}(A^H \rho^H). \quad (9.59)$$

这里, 最后一个公式中的算符 A 可以是任何一个可观测量的算符。因此, 如果我们用这个绘景来讨论物理问题, 则我们需要求解所有的物理量对应的算符的演化方程。这个, 有可能比求解密度矩阵的演化 (只有一个方程) 来的复杂, 除非我们能够找到一个最小的固定的算符的集合, 所有的其它算符都可以用这个集合里面的算符来表达。对于任意算符, 我们还可以得到一般的算符的演化方程。对 A^H 求时间 t 的导数, 我们有

$$i \frac{\partial}{\partial t} A^H = [A^H, H]. \quad (9.60)$$

这个方程被称为 Heisenberg 方程。实际上, 在量子力学的进一步发展, 我们会看到, 很多时候 Heisenberg 绘景反而使问题变得更容易处理, 因为这个最小而固定的算符 (其实, 一个例子就是我们在谐振子问题中用到的产生湮灭算符) 集合经常能够被找到。这个, 将来有机会学习量子场论的过程中, 我们会看见。

9.6 再论表象理论: 相干态表象

我们看到量子力学可以在位置空间描述, 也可以在动量空间描述, 也可以在任意的 Hermitian 算符的本征态下描述, 但是不能在位置加上动量的相空间描述, 因为 \hat{x}, \hat{p} 算符不对易, 没有一般的共同本征态。当然, 这些描述空间的不同, 仅仅是空间基矢不一样, 空间本身都是一样的: 所有的平方

可积函数构成的 Hilbert 空间。这些不同的表象之间通过一个基矢变换或者称为矢量的相似变换相联系。其实，Schrödinger 绘景与 Heisenberg 绘景也可以看成是一个通过一个矢量的相似变换相联系的两个表象，只不过，这个变换是时间依赖的。

对照经典力学，我们看到，经典力学需要在相空间描述而相空间是一个位置和动量的联合空间。于是，一个自然的问题就是，是否可以把量子力学也在相空间里面描述。简单的答案是不能。复杂一点的答案是也可以，只是这个描述看起来比较奇怪，对这个描述下的变量和分布函数的理解也与经典力学和我们已经学习到的量子力学不同。但是，这个描述的威力非常大。所以在这里，我们稍微说明一下这个描述的思想，更具体的内容推荐大家通过量子光学的教材，例如 D.F. Walls 的《Quantum Optics》[?], C.W. Gardiner 和 P. Zoller 的《Quantum Noise》[?]

这个奇怪的描述的基本思想是引入一个特殊的非 Hermitian 的算符，然后利用这个算符的本征态来作为整个 Hilbert 空间的基矢。这个算符对于一般的位置空间算符 \hat{x} 和动量空间算符 \hat{p} 来说，定义成湮灭算符（见谐振子的例子）

$$\hat{a} = \frac{\sqrt{2}}{2} (\hat{x} + i\hat{p}). \quad (9.61)$$

可以发现这个算符的本征值是复数，如果我们记这个本征向量为

$$\hat{a}|\xi\rangle = \xi|\xi\rangle, \quad (9.62)$$

那么，我们就有可能可以把位置空间上的函数 $\phi(x)$ 变换成为复数域上的函数 $\rho(\xi)$ 。实际上这个变换是如何做的，我们就不讨论了。在这个变换下，所有的算符也可以变换成复数域上的函数。在这里，我们指出这个变换的一个好处：存在着这样一种表象，在这个表象下，原来位置空间的形如 $\phi(x, x')$ 的非对角的密度矩阵，可以成为对角的复数域上的分布函数 $\rho(\xi)$ ，只不过这个“分布函数”可以取值不仅仅是大于零的实数。由于形式上这个表象中的密度矩阵成了“密度分布函数”，很多的经典概率论的方法和概念都可以尝试迁移过来。因此，这个表象是一个非常有用的表象。

9.7 写在量子力学后面的话

把量子系统的演化方程写成

$$\frac{\partial}{\partial t} \rho^S = L \rho^S \quad (9.63)$$

的线性方程的形式，其解为

$$\rho^S(t) = \mathcal{L}_t \rho^S(0) = e^{Lt} \rho^S(0). \quad (9.64)$$

这里，我们注意到，算符 \mathcal{L} 与 L 之间实际上是 Lie 代数到 Lie 群的关系，也就是指数映射。我们已经知道经典力学的演化方程也可以写成相同形式的线性方程，有相同形式的解。我们也通过翻硬币的例子了解到，随机过程也可以表达成类似的形式。在下一章关于随机过程的讨论中，我们会更清楚地看到这一点。也就是说，确定性的经典力学、随机性的经典力学、量子力学，都可以表达成线性方程的形式，而且演化算符就是某一个代数的算符的指数映射。当然，离散时间的过程就不一定是 Lie 群，但是肯定也是某一个群，以及与这个群的生成元的关系。

于是，我们发现，确定性的经典力学、随机性的经典力学、量子力学的数学结构实际上都是群论。实际上，如果了解多一点群论，我们还可以看到，生成元在群的研究中是非常重要的：生成元之间的对易关系决定了所生成的群的结构。如果了解一点点微分几何，我们还会发现，我们的群元素之间除了有群乘，通常还算有加法运算，而且由于有生成元之间的非对易关系，通常这个加法运算确实可以构成某个矢量空间（其实是切矢量空间，如果现在不明白，跳过这一部分，将来再回来），但是这个空间没有全局的基矢。一个矢量空间，如果有全局基矢，那么在很大程度上它就可以利用坐标来讨论；如果没有全局基矢，则需要研究矢量平移、坐标变换与一致性等等问题。这个就是微分几何要研究的问题。于是，在这里，我们说，这些演化过程的背后的数学结构都是群论与微分几何，尤其是群论。关于这部分，推荐吴兆颜的《高等量子力学》[?]

了解了这一点之后，我们就会发现，为什么线性代数在这个理论框架中的重要地位：群的表示理论就是线性代数。在实际计算的层次，我们永远需要通过矩阵、矢量等等计算来表示群，而群是所有这些演化理论的基础。

此外，这个统一的形式，还有一个额外的好处，演化方程都是线性的。当然，这个会产生一个自然的问题，那么非线性动力学研究的是什么？以后我们会看到，如果我们把一个自治（关于一个封闭系统的方程是自治的）的

非线性动力学方程所描述的过程转化为密度矩阵形式的描述，它也必然是线性的。只有把密度矩阵的方程变成随机变量的数字特征的演化方程的时候，方程会包含非线性。而且，实际上，当密度矩阵形式的方程中的算符项 L 包含相互作用的时候，这个非线性项才产生。也就是说，非线性实际上代表了相互作用。

关于这个写在后面的话，如果你暂时看不懂，请你不要担心，这个只有理解上的意义，没有任何本书的计算过程实际上需要这些理解。迄今为止，在我自己的研究工作中，也仅仅起到理解上的作用，尽管我一直在寻找合适的问题，让这些思想以及这些思想相关的技术在实际研究中发挥作用。我只能说，我有一些方向和想法，有待于探索和实现。

下一章，我们来看一下如何把随机过程的演化方程也写成这个线性方程的形式，而且代数算符 L 决定了演化算符 \mathcal{L} 。

第十章 随机过程

通过概率论一章（第六章）的“随机过程初步”一节（6.7节），我们已经了解到离散时间的随机过程可以看成是多变量随机数的联合分布，可以看成是每一个时刻的随机变量的分布函数的演化过程，也可以看成是一个随机变量的演化的历史样本的集合。同时，我们还学会了用密度矩阵、随机变量算符、概率分布“矢量”以及相应的概率转移矩阵，来讨论离散时间的随机过程。这里，我们第一要讨论更多的随机过程的例子，这样我们会对于这几个不同的描述方法更熟悉；第二，我们还要介绍一点点连续时间的随机过程的概念和几个典型的例子，以及处理方法。

从离散时间过渡到连续时间，在数学上存在着一个非常大的跃迁：离散时间随机过程的分布函数是很好定义的——多维随机变量而已，但是在连续时间上，我们实际上要讨论某一个连续时间随机轨道——也就是一个时间过程一个函数——的概率分布问题。一个一般的对于连续时间的随机过程的讨论实际上需要一个轨道的分布函数 $p(x(t))$ ，如果我们要做极限和微积分的运算，我们还需要类似于 $d[x(t)]$ 的一个量——轨道的微分。一个实数作为自变量的微分我们有清楚的概念——本质上是实数域上的测度的问题但是实际上当作一个小小线段的极限也无伤大雅，但是函数作为自变量的微分，我们就不好定义了（对于学有余力的读者，这个问题相关的概念是“路径积分”——路径做自变量的微积分，独特的物理学家 Feynman 的重要贡献）。在这里，我们不在抽象的层次来讨论这个问题，而是继续我们物理学家的不太严格的直觉图景，把连续时间的过程的分布函数看成是很多很多个离散变量的联合分布 $p(x(t_1), x(t_2), \dots)$ ，然后把两者认同起来。我们也暂时不讨论用随机变量的分布函数不能描述的问题。

现在，在学习了力学、统计力学和量子力学之后，我们再来看随机过程。物理系统的力学方程不管经典力学还是量子力学都是确定性的。但是，经典力学 Liouville 方程描述的是密度分布函数的演化，密度分布函数可以

做几率解释；量子力学 Schrödinger 方程描述的是密度矩阵的演化，密度矩阵可以做几率解释。那么有没有系统在本质上就需要包含随机性的方程呢？如果有，怎么处理呢？

根本上是否存在随机系统是一个哲学问题，我们不讨论。我们讨论在描述的层次，什么时候需要随机描述。考虑一个完全确定的力学系统：一群小分子连着一个大热浴，就像我们在统计物理学中经常见到的正则系综。整体必然满足确定性的力学方程。但是，如果我们只关心这群小分子，而不去测量大热浴的状态，那么这群小分子就好像是处于一个满足正则系综分布的随机个体。我们抛硬币，确实每一次的力学过程都是完全确定的，但是初始速度的大小和方向，空气阻力的细节，我们不能完全控制，于是这个硬币离随机个体差不了多远。所以，环境和初始条件不可完全把握，可以造成描述上的随机性。随机性的描述的层次有可能还有一个来源，演化方程对应的轨道存在不稳定性：两条某段时间离的很近的轨道，在其他时间可以离得很远，或者反过来。这样的系统我们也会在下一部分讨论非线性动力学的时候看到。随机性还可能存在别的来源，例如量子力学的测量，每次总是会选择一个被测量算符的本征态，但是选择哪一个是不确定的。这个过程是否存在更深刻的确定性描述我们还不知道，但是至少在描述的层次，这里是有随机性的。所有的这些随机性，都会导致我们需要随机性的描述。

换一个思考的角度，我们看到，经典力学可以采用随机描述的框架 (Liouville 方程)，量子力学可以采用随机的描述 (Liouville-von Neumann 方程)，所以随机描述是力学系统的正常的描述。当然这些方程本身都是确定性的，仅仅是方程描述的变量可以做随机解释。在这个意义上，就算在一个非常明确的力场中的确定性的经典力学系统，我们也完全可以用随机描述的讨论它。结合这两个方面的因素，我们说，随机描述才是描述的主要手段，确定性的描述仅仅是一个特例。

我们发现，确定性的力学过程完全可以等价地用相空间分布函数的演化方式来描述，而统计力学的根本问题就是这个相空间分布函数的定态分布的问题，而且统计物理学的 Metropolis 方法就是构造某个随机过程然后得到这个随机过程的稳态，进而量子力学也是一个概率意义上的描述（只不过，这个概率的简单事件满足态的直接叠加原理而不是概率论性叠加原理）。因此，我们说力学、统计力学和量子力学就是概率论（以及量子概率论，或者说满足态叠加原理的概率论）。在统计物理学中还有一个很重要的分支我们没有谈到：扩散和输运等非平衡过程的统计力学研究。这样的现

象的例子有液体中花粉粒子等悬浮小颗粒的扩散，温度不同的两个热浴连着的管道上流动着的热流等等。这些过程也可以通过随机过程来描述。花粉粒子的运动的数学模型就是 Brown 运动。这个将是我们这一章的主题之一。郝伯林在有一次课上谈到统计物理学的学习的时候大致说过，第一、具体例子在学习统计物理学的时候很重要；第二、在统计物理学的范畴内，这个例子包含 Ising 模型和 Brown 运动。这个我完全同意，以问题为中心的学习方法，在了解一定程度的基本理论和基本框架的基础上，是非常好的学习方法。我也同意，在统计物理学之内，好像其它的问题就用不着了。

回到我们的主题，以上的讨论主要是为了说明，很多物理学实际上就是随机过程，而随机过程也从物理学的研究中获取进一步发展的动力和方向。所以，这一章我们来学习一下随机过程。我们主要通过以下几个具体例子：Wiener 过程，Brown 运动，Ornstein-Uhlenbeck 过程，以及这些过程的物理背景，来学习以下几个概念和技术：转移概率，转移速率，稳态分步，Fokker-Planck 方程，随机微分方程，Langevin 方程，Master 方程，Monte Carlo 方法。

本章推荐参考书：H. Risken 的《Fokker-Planck Equation》

10.1 Langevin 方程：随机力

我们从一个连在弹簧上的小球的 Newton 方程开始，这里弹簧的另一端连着一个做独立自主的随机运动的墙壁。

例 10.1.1 与振动墙壁相连的小球：

接下来，我们考察水中的花粉粒子的运动。

例 10.1.2 水中的花粉粒子：

10.2 Master 方程与 Fokker-Planck 方程

如果一个系统包含了所有的世界，物理学家相信，这个系统必然是由 Hamiltonian 描述的，于是这个系统的演化动力学必然是仅仅取决于这个系统的当前的状态的。所有的相互作用不依赖于历史而仅仅依赖于当前的位置。于是，我们把这样的仅仅依赖系统当前状态，而不依赖这个系统的历史状态的过程称为 Markov 过程。实际上，我们任何一个系统的选取都是在整

个世界中选取了小小的一个部分，然后我们把这个小小的部分看成几乎孤立的，不太受外界影响的。如果，我们的系统真的满足这个要求，那么这个系统的动力学将是 Markov 的，而实际上这一点不一定满足。所以大量的随机过程，不是 Markov 的。在这里，我们仅仅研究 Markov 的随机过程。由于前面提到的物理学的原因或者说物理学家的信仰的原因，这样的过程实际上比从数量上看起来——在所有的随机过程中 Markov 的应该占有的比例为零——要有代表性的多。

10.2.1 Markov 过程的 Master 方程

10.2.2 从 Master 方程到 Fokker-Planck 方程

10.2.3 几个典型的随机过程

10.3 Langevin 方程与 Fokker-Planck 方程的关系

10.4 平稳性

10.5 随机过程的模拟与计算

10.6 本章小结

在 6.7 节中我们学习了离散时间随机过程的处理方法，包含概率分布“矢量”以及相应的概率转移矩阵，密度矩阵、随机变量算符等。通过本章的学习，我们学会了用连续时间的随机过程的数学方程来描述物理现象。这样的数学方程包含 Langevin 方程、Fokker-Planck 方程和 Master 方程，这样的典型物理现象包含 Brown 运动，Ornstein-Uhlenbeck 过程。我们还了解了 Langevin 方程、Fokker-Planck 方程和 Master 方程之间的关系，还有这些过程的模拟计算。

研究随机过程还有一个重要方法：生成函数方法或者叫做母函数方法。在我们目前的逻辑框架之内，我没有找到合适的位置。如果以后想好了，会把它加进去。在此之前，请阅读其它随机过程的教材来学习生成函数方法。随机过程的研究还可以通过把随机方程转化为各阶矩的演化方程来讨论。这个转化为各阶矩的方法在物理学里面叫做 BBGKY (Bogoliubov-Born-Green-Kirkwood-Yvon hierarchy) 方法，或者称为 Green 函数方法，关联

函数方法。这个部分，如果以后有可能会在更加专门的部分，例如《高等统计物理学》的教材中介绍。在这之前，也只有请读者从专门的物理学教材中去学习这个方法了。

第三部分

系统科学的基本理论

第十一章 线性与非线性动力学简论

动力学系统的基本问题是从一个给定的动力学方程开始，这个方程可能是某个物理过程的数学描述，给定的初始状态，我们如何了解将来（或者过去）任意时刻的状态。这个问题，看起来是一个平庸的技术问题，如果方程能够被精确求解，那我们就用精确解好了。如果不能精确求解，我们就用数值解。微分方程的典型结构如下：

$$\frac{d}{dt}x = f(x, t), \quad (11.1)$$

$f(x, t)$ 可能包含空间导数算符或者其它更复杂的形式，例如空间积分算符。我们原则上只需要把初始条件代进去，用足够小的 Δt ，就可以得到

$$x(t_0 + \Delta t) = x(t_0) + f(x_0, t_0) \Delta t. \quad (11.2)$$

然后，一直继续迭代下去。这个基本上就来自于 $\frac{d}{dt}x$ 的定义。也有其它更复杂，但是误差更小，收敛更快的迭代方法，例如 Runge-Kutta 方法，但是基本思想就这么简单。如果这个思路能够解决问题，那么我们关于动力学系统的进一步研究就完全没有必要了。动力学系统的一般理论，或者说微分方程定性理论，告诉我们说，真实微分方程的行为，跟这个简单的思路相差很远。其中，非常重要的一条就是：某些系统从非常相近的初始条件出发，可以得到相差非常大的轨道。失之毫厘，谬以千里。对于这样的系统，上面提到的计算思路本质上使不能使用的：每一小步的计算都有一些小小的误差，这些误差，还有不同的 Δt 的值，会带来轨道很大的不同。那么，这个时候，我们怎么办呢？我们也许掌握不了每一个时刻的状态，也许就不用掌握每一个时刻的状态，那么我们有没有办法对这样的系统取得一些有意义的认识呢？动力学系统的一般理论告诉我们说，这个时候，我们看看动力学方程

的定性行为。

在这部分，我们就来介绍一下，有哪一些定性行为，如何做这样的定性分析。首先，我们从最简单的线性系统开始，然后进入到非线性动力学系统。通过这个部分的学习，我们要学会使用线性稳定性分析的技术来对动力系统的定性行为做初步的分析，然后需要了解典型动力系统的行为，包含不动点、周期轨道、混沌现象、分支行为，有一个了解。其他的更加专门的技术应该在专门的《非线性动力学》的书籍中去学习。我们这一章的目的就是领读者进门，然后铺垫一定的基础，这样读者可以自行进一步学习。

这个部分推荐的阅读材料是 Edward R. Scheinerman 的《Invitation to Dynamical Systems》(简称 IDS)，仅推荐这一本，而且这一本就足够了。IDS 这本书的写作风格是带领读者一起思考，然后主要通过例子来学习最基本和最核心的内容。完全满足我们推荐的可以采用概念地图学习方法的教材的要求。其实，你仔细看的话，你应该会发现，本书的写作风格，正是受了 IDS 的深刻的影响。只是后来我们注意到其实这样做还有学习方法上的理论依据。另外，推荐科普书 James Gleick 的《Chaos: Making a New Science》(中文译名《混沌开创新科学》)。

11.1 线性系统的演化

一个一般自治的线性系统定义如下，

$$\frac{d}{dt}x = Ax. \quad (11.3)$$

如果 A 含有 t ，也就是非自治的系统，那么我们可以把 A 本身的动力学也包含进来，例如引入变量 y ，使得新的 $\hat{x} = (x, y)^T$ 的方程满足上面的定义。当然，如果 A 本身的动力学(也就是有关 y 的方程)是非线性的，那么我们最后的方程也就成了非线性的，成了我们下一节讨论的对象。这里我们只讨论能够表达成自治方程的线性系统。

这个方程的解非常简单：

$$x(t) = e^{A(t-t_0)}x_0 = \sum_n e^{\alpha_n(t-t_0)}|n\rangle\langle n|x_0. \quad (11.4)$$

其中，我们用到了 A 的本征值 α_n 和本征向量 $|n\rangle$ 。我们看到，如果 x_0 有其中一个 $|n\rangle$ 的分量，而且这个分量对应着的本征值 $\alpha_n > 0$ ，那么 $x(t)$ 趋向无穷大。如果所有的 $\alpha_n < 0$ ，则系统收敛到 $x(t) = 0$ 。对于某一个 $\alpha_n = 0$ ，其它小于零的情况，系统收敛到对应的状态 $|n\rangle$ 上。

对于差分方程的情况，我们类似得到，

$$x_{n+1} = Ax_n, \quad (11.5)$$

则

$$x_n = A^n x_0 = \sum_m \alpha_m^n |m\rangle \langle m| x_0. \quad (11.6)$$

于是，如果 x_0 有其中一个 $|m\rangle$ 的分量，而且这个分量对应着的本征值 $\alpha_m > 1$ ，那么 x_n 趋向无穷大。如果所有的 $\alpha_m < 1$ ，则系统收敛到 $x_m = 0$ 。对于某一个 $\alpha_m = 1$ ，其它小于 1 的情况，系统收敛到对应的状态 $|m\rangle$ 上。

总结一下，对于线性方程，我们只需要知道 A 的本征值和本征向量就够了。而且，对于一般的 x_0 ，我们假设其在各个本征向量上都有分量，于是 A 的本征向量也不需要了。我们只需要 A 的本征值。有了这个线性系统的认识之后，我们来看一下非线性系统。

11.2 非线性系统的线性化：定态与线性稳定性

对于非线性自治系统（同样，如果非自治，原则上可以通过引入更多的动力学变量使得系统自治）， $f(x)$ 必须是 x 的高于一阶的多项式。对于这个系统，第一个我们可以把握的特征是定态，也就是满足

$$f(x^*) = 0 \quad (11.7)$$

的状态，或者对于差分方程 $x_{n+1} = g(x_n)$ 满足 $g(x^*) = x^*$ 的状态 x^* 。在这个状态下，

$$x(t) = x^*, \quad (11.8)$$

系统的状态不再发生变化。也就是说在一个系统的所有的变量构成的相空间中（这里相的意思是状态），存在一些特殊的点，系统只要处于这些点上就不再动了。那么自然的一个问题是，如果系统在这些点的附近呢？假设系统在这些点附近就会跑到这些点上去，那么，如果我们把所有的这样的定态点（也称为不动点）的影响力区域画出来，我们就基本上知道了这个系统的行为：在各自的影响力区域之内，系统被吸引到不动点上去。所以，不动点是否是吸引性的，各自的影响力多大，这样的问题就非常重要了，甚至超过具体的轨道的重要程度。于是，非线性动力学的研究就关注不动点以及不

动点的稳定性。所谓不动点稳定性，就是如果系统的状态在这个不动点附近，系统将来的状态是不是会收敛到这个不动点上去。更一般地，我们可以讨论不变轨道的稳定性。例如，某一条轨道很特殊，只要系统跑到这条轨道的任何点上，系统就会沿着这条轨道一直运动下去。那么，我们就问如果系统处在这个轨道的附近，是不是系统将来的运动就会在这条轨道上运动呢？当然，在我们下面的例子当中，为了计算简单，我们只局限在讨论不动点以及不动点的稳定性上，轨道的稳定性不做讨论。

不动点的稳定性的一种讨论方法是线性稳定性分析。其基本思想是如果系统离不动点足够近， $x_0 = x^* + \delta x$ ，那么，系统的演化方程就可以做线性近似：方程的两端只保留 δx 的一次项，也就是

$$\frac{d}{dt}\delta x = f(x^* + \delta x) = f'(x^*)\delta x. \quad (11.9)$$

而这个是线性方程。其行为前面我们已经了解，非常简单，完全取决于 $f'(x^*)$ 的本征值。注意，有可能 $f'(x^*)$ 的左导数和右导数是不一样的。对于一维的情形，我们只存在三种稳定性的情况，如下图：完全稳定，完全不稳定，临界点（一侧稳定另一侧不稳定）。我们还可以把这样的情况结合起来，讨论如果有两个相邻的不动点会怎样。例如，我们组合成下图的流动的情形。很多时候，这个确实被称为相空间的流图。对于一维情况，总共我们也仅仅有这些基本的结果以及它们的组合：要么跑到无穷大，要么收敛到某个点上，要么从某个点离开跑到无穷大或者收敛到其他的点上。

那么，更高维的系统有哪一些比较典型的定性行为呢？这里，我们举例子来介绍几种定性行为。

例 11.2.1 Logistic 映射，差分形式，

$$x_{n+1} = \gamma x_n (1 - x_n). \quad (11.10)$$

也可以表达成微分方程的形式，

$$\dot{x} = \gamma x \left(1 - \frac{1}{K}x\right). \quad (11.11)$$

这个方程描述了存在种内竞争时候种群大小的变化。例如连续方程的第一项表示出生的新人口数量和现有的人口成正比，可是这个比例系数（出生率）依赖于人口总量：当人口总量远远小于 K 的时候，出生率高；当人口总量接近 K 的时候，出生率小；超过 K 的时候，出生率为负值。让我们来分析这个动力学过程的定性行为。以差分形式为例。

计算不动点, 我们得到 $x_1^* = 0$ 以及 $x^* = \frac{r-1}{r}$, 计算得到 $g'(x^*) = \gamma - 2\gamma x^*$ 。于是 $g'(x_1^*) = \gamma$, $g'(x_2^*) = \gamma - 2\gamma x^* = 2 - \gamma$ 。于是我们发现, 当 $0 < \gamma < 1$ (出生率, 总是假设大于零, 小于零是平庸的) 的时候, x_1^* 是稳定的, 于是种群消失; 当 $1 < \gamma < 2$ (出生率, 总是假设大于零, 小于零是平庸的) 的时候, x_2^* 是稳定的, 于是种群维持一个非零的水平。当 γ 继续增大, 这个不动点也就不稳定了。也就是说, 系统大小非零, 也不是常数。有没有更多的信息能够获取呢?

有, 讨论轨道的稳定性。但是, 我们说了这个有难度。我们来用一个取巧的办法: 我们考虑方程 $x_{n+1} = g(y_n), y_n = g(x_{n-1})$, 或者说等价地我们讨论 $x_{n+1} = g(g(x_n))$ 的不动点及其稳定性。这样做的目的, 其实是考察原来的动力学系统的周期为二的轨道的稳定性。迭代两次的方程的不动点是原始方程的周期为二的轨道。为了获得一个更直观的感觉, 在这个计算之后, 我们通过做图来分析原始的方程。从那里, 我们会看见周期二轨道的出现。推荐用 google[?] 搜索一下 Logistic mapping, 可以找到 wikipedia[?] 以及 WolframAlpha[?] 等平台上的很多关于这个做图过程的讨论。计算 $g(g(x))$ 的不动点, 我们得到四个不动点 ($g(g(x))$ 是四阶方程)。其中的两个不动点就是 $g(x)$ 的不动点 (它们在 $\gamma > 3$ 的时候不稳定), 多出来的两个不动点是我们关心的对象。再计算 $g(g(x))$ 的导数, 讨论其在两个新的不动点的值, 我们就可以知道这个周期二的轨道的稳定性。

类似的, 我们可以讨论周期三的轨道的稳定性, 周期四的轨道的稳定性, 等等。我们得到了下面的分支图: 某些参数值 (横轴) 的时候, 系统只有一个稳定不动点, 有的时候有两个, 三个等等, 图中显示了稳定不动点的值 (纵轴)。有几个稳定的不动点, 在相应横轴参数的位置就有几个纵轴的值, 于是就得到了我们看见的比较复杂的分支曲线。有一个问题, 我们并没有说明, 实际上在图中的某些区域, 系统有比周期分叉更复杂的行为, 成为混沌的系统。下面, 我们会更详细地举例说明。

其实, 到目前为止除了我们尚未说明的混沌区域的行为, 我们发现, 所有的上面的讨论过的系统的行为都可以通过 Runge-Kutta 数值计算来得到, 我们仅仅通过跟踪演化轨道, 就可以把各个稳定的不动点找到, 然后, 把所有的稳定不动点画到刚才的坐标上, 也能够得到这张图。当然, 完全从数值计算的角度来说, 也许我们对于某些轨道失去稳定性的图景 (在什么地方失稳, 为什么失稳) 就没有好的了解, 但是基本的分支图还是能够画出来的。那我们就又回到了一开始的问题: 如果这样, 我们研究微分方程定性理论有

什么用, 直接数值计算好了。

前面, 我们已经演示了, 从技术上来说, 如果能够找到不动点, 做稳定性分析, 那么我们就可以不做数值计算而了解系统的行为, 尽管这个不是本质进展和必要性的。下面, 我们来看看更复杂的动力系统的现象, 看看这个时候, 数值计算轨道还够不够用, 我们需要在数值计算轨道的基础上观察哪一些新的量, 才能对系统的行为有好的了解。

11.3 混沌与分支

我们首先来观察一个展现混沌行为的简单系统, 然后我们来体会一下混沌系统的一些特征。其实, 例 11.3.2 中的 Logistic 映射在 γ 取值为 3.57 的时候就可以用来作为我们的例子。我们把它放在后面, 先讨论著名的 Lorentz 吸引子。

例 11.3.1 Lorentz 吸引子的轨迹: 用 Runge-Kutta 方法以及数学计算软件 Mathematica 或者 SageMath, 来计算以下方程通过初始点 $p_1 = (x_0 =, y_0 =, z_0 =)$ 以及 $p_2 = (x_0 =, y_0 =, z_0 =)$ 的轨迹,

$$\frac{dx}{dt} = \sigma(y - x), \quad (11.12a)$$

$$\frac{dy}{dt} = x(\rho - z) - y, \quad (11.12b)$$

$$\frac{dz}{dt} = xy - \beta z. \quad (11.12c)$$

其中 $\rho = 28, \sigma = 10, \beta = \frac{8}{3}$ 。

通过前面的例子我们已经了解到, 不动点最简单, 其实任意阶的周期解 (也就是 $f^k(x)$ 的不动点) 也是一个容易想象的情况。现在, 我们来看一个复杂的多情况。我们来跟踪一条这个方程决定的轨迹, 我们从 $x_0 =, y_0 =, z_0 =$ 出发, 由于整个方程是决定性的, 我们很容易就能够用 Runge-Kutta 画出这条轨迹。首先, 从 x 的取值来看, 我们发现这条轨迹没有跑到无穷去, 看起来也不像是周期的, 不过粗略看起来确实有一定的周期的成分在里面。我们对它做一个 Fourier 变换, 看起来周期性不强, 典型的周期不存在。现在, 我们来考虑 $x_0 =, y_0 =, z_0 =$ 的轨迹。这个点和前面的初始值非常非常接近。然后, 我们观察到, 在 $t >$ 之后, 两个轨迹的差别非常大。这个很大的差别表示, 两个非常非常接近的初始条件, 可以导致相差很大的轨迹。这个现象

让我们甚至怀疑我们的计算：我们的数值计算是在有限精度下计算的，存在舍入误差，而这个舍入误差可以造成将来的轨迹上的非常大的区别。也就是说，我们计算出来的轨迹以及不大，不知道什么地方的舍入误差就让我们结果远远地离开真实的结果。那么，能不能依靠提高计算保留的精度来提高计算结果的准确度呢？对于这个系统，我们发现只要演化时间足够长，无论多小的舍入误差，都可以造成结果上巨大的区别。这一点，我们会在下一个例子中给一个数学上比较严格的说明。在这个例子中，我们可以通过控制计算的精度——这个可以在 Mathematica 之类的软件中实现——然后观察更多的轨迹来验证这一点。这个本质上的很小的初始条件上的差别可以造成将来的轨迹的很大的区别的现象被称为“初值敏感性”，也被称为“蝴蝶效应”，是混沌系统的重要特征。另外一个这个例子展现的特征是整体轨迹在有限区域内（不趋向无穷），又不是简单的不动点或者周期运动。

我们通过下一个例子来更加细致地看一下一个系统进入混沌的方式。

例 11.3.2 差分形式的 Logistic 映射的混沌行为

$$x_{n+1} = \gamma x_n (1 - x_n). \quad (11.13)$$

其中 $\gamma = 3.57$ 。

11.4 更多的具体系统上的研究实例

第十二章 相变、临界性与自组织

临界性

相变是指一个系统的状态从一种宏观性质转变成了另一种宏观性质。这个宏观性质通常是定性的。例如，考虑一个盒子的气体，我们把气体的某一对分子的位置发生了交换这样的两个状态之间的区别看作是微观上的差别，细小的差别。如果我们通过降低温度的方法把分子之间的平均距离缩小，分子之间的联系加强，直到这个盒子的气体转变为液体，那么这个时候，我们就说这个系统发生了相（就是宏观状态）的变化——相变。更加准确的定义在下文会讨论。在此之前，我们注意到，一个相变通常对应着构成这个系统的微观组成部分之间的联系突然加强或者减弱了。而这样的变化在自然界中非常普遍，而且非常值得关注。在讨论很多物理系统的问题，甚至社会系统的问题的时候，我们不太关心一个系统的状态的细小的变化，但是我们特别关注一个系统的定性的变化。例如，一个人群交往很正常，不太多也不太少的社区，突然之间变成了紧密联系在一起经常聚集在一起的社区，或者是相互忽视的社区，这个必然意味着某种内在的结构发生了变化，而且这种结构变化通常有一个内在或者外在驱动力。在这个例子中，我们自然关注的表面特征的变化，内在结构的变化，以及变化的内在或者外在驱动，就是我们物理学中研究一个相变最关心的问题。他们分别是，以后会定义，序参量、关联函数和相互作用、相变的控制参数。我们还关心在这个控制参数取值为多少的时候，这个系统的相变会发生，也就是相变点的问题。

如果相变问题的研究没有共性，也就是说每一个相变的发生，上面的序参量、关联函数、控制参数和相变点之间的关系，完全依赖于具体的系统，那么我们也不需要一个关于相变的一般的研究的分支学科。我们完全依赖于研究那个具体系统的学科好了。实际上，我们发现，一旦发生相变的时候，很多共性的现象会发生，很多有共通之处的研究方法可以使用。

我们这一章的主要目的就是讨论几个相变的例子，通过这些例子来了解有哪一些有共性的现象以及有哪一些有共通之处的研究方法。这些例子包含 Ising 模型、沙堆模型；这些现象包括：临界慢化、临界点附近的关联长度发散、标度关系、幂律等；这些研究方法包含 Landau 的相变的一般理论与平均场、Monte Carlo 模拟（尤其是 Metropolis 以及类似的方法）、重整化群方法等。

本章推荐阅读材料：Kadanoff 的《Statistical Physics》，于禄、郝伯林和陈晓松的小册子《相变与临界现象》。说明，对于熟悉统计物理学的读者，我们这里仅仅讨论二级相变或者与二级相变类似的连续相变。在这一章里，在保证科学性的前提下，我们尽量呈现粗糙的描述而不是定量的计算。这部分内容的更加严格的呈现应该从关于“高等统计力学”的教材中去寻找。例如 Michael Plischke 和 Berger Bergersen 的《Equilibrium Statistical Physics》。

12.1 相变的研究方法

我们在统计物理学一章已经讨论过 Ising 模型的相变以及通过 Monte Carlo 模拟来寻找相变点。我们发现，Ising 模型在无外场的情况下有两个相：一个整体磁矩为零的状态，一个整体磁矩不为零的状态。后者又有两个可能的状态，整体磁矩大于零或者整体磁矩小于零。对于磁性材料，我们称后者具有磁性，前者没有磁性。有的磁铁在 140°C 左右——这个从一个状态变成另一个状态的温度就称为临界温度，在这个例子中被称为居里温度——就可以失去磁性，所以这个相变的现象是实际生活中也能够观察到的。如果你想寻找这样的材料来做一个实验，google “Curie temperature”（居里温度），然后选择一个居里温度比较低的材料，例如 100°C 左右的，来购买。然后你可以通过把这个材料放在炉子上加热来感觉其磁性的改变。在做这个实验的时候最好是在两个人在一起。这个实验比较安全，但是通常有一个不参与操作的观察和保护实验者是有益的。当然，如果是仅仅为了体验相变的话，完全没有必要做这个实验，我们日常生活中就能够发现，温度降低，水会结成冰，温度升高，水会化成汽。

在这一章，我们来观察这个相变的一些更仔细的性质，以及学习研究这些性质的方法。

例 12.1.1 通过 Monte Carlo 模拟分析 Ising 模型的临界点、关联函数、集团大小分布、比热：

通过这个例子，我们复习了如何使用 Monte Carlo 方法来分析统计力学系统，还看到在临界点附近关联函数和集团大小分布的行为。一般地说，相变就是一个系统中的某个性质——通常用一个值来描述——从零变成非零或者倒过来的过程。这个时候，这个描述这种性质的值就称为序参量。对相变的研究最基本的方法就是按照这个系统的 Hamiltonian H 来计算出这个系统的配分函数 Z ，然后从这个配分函数计算出序参量随着某个或者某几个控制参数的变化函数，然后识别出来在什么地方这个函数会发生一个从零到非零（或者反之）的变化。然后，原则上我们也能够通过这个配分函数计算出来关联函数等等其它物理量。但是，这样一个方法，我们已经看到，在简单的 Ising 模型上都是非常困难的，能够用这种方法求解出来的都是艺术品。科学原则上是希望任何人都能够做这个计算的。于是，我们需要寻找别的方法。Monte Carlo 方法就是其中一种。当然，实际上的 Monte Carlo 方法总是在一个比较小的系统上做的计算。而实际上的物理系统，其大小通常都远远大于我们模拟计算的系统。我们只能通过模拟计算给出一个系统大小从小到比较大的时候的渐进行为的推测。在物理学中，有对于相变的更加严格的定义，以及发生的条件的研究。在这里，我们仅仅给出定义。相变是热力学函数以及热力学函数的导数不连续。热力学函数自身不连续称为一级相变。热力学导数的不连续称为二级相变。更高级的相变对应着更高级的导数不连续。其实，我们从 Ising 模型的讨论已经看到，导数 $\frac{dm}{dT}$ 或者 $\frac{dm}{dh}$ 呈现一个 λ 的形状，而且随着系统大小变大，这个尖峰变尖，趋于无穷大。

然而，我们所作的任何有限系统上的计算，实际上都不会给出这个不连续，只不过越来越陡。在有限系统上，配分函数的每一项都是可微的，很难想象把它们都加起来就不可微了。实际上这个可微性在无限系统上在加起来之后，都是很难想象其发生改变的。Lee 和 Yang 的工作给出了一个非常巧妙的回答：考查配分函数在复平面上的性质，而不仅仅是实轴。他们发现，如果配分函数的零点在复平面上随着系统大小的增加（大小增加，密度不变，称为热力学极限，也就是很多很多同样的系统连接在一起）越来越接近实轴，那么这个时候配分函数存在不连续性。这个结果被称为 Lee-Yang 定理。这个定理从原理上回答了相变在理论上是可以计算和分析的。当然，在实际计算过程中，人们不一定通过计 Lee-Yang 零点来分析相变。

目前为止，我们分析了一个具体的模型，尽管我们交代过这个模型的典型行为：关联函数临界点附近有大的变化，集团大小的分布也发生定性变换，比热出现了 λ 形的转变，是比较普遍的，我们并没有交代任何关于相变的一般理论。我们还是停留在具体系统具体分析的水平。有没有什么更加一般的分析方法和关于相变的一般理解呢？有了 Ising 模型的现象作为基础，我们现在来讨论这个问题。

除了计算配分函数和做 Monte Carlo 模拟，我们要介绍的使得临界现象的分析一定程度上从统计物理学独立出来的技术是重整化群。这个技术的思想是在临界点附近我们改变系统的大小相当于改变系统的温度或者其它驱动参数，或者说系统的状态离临界点的距离。粗略地说，当我们把原始系统的几个单位看成一个单位——称作标度变换或者尺度变换（这里其实我们需要同时重新定义这个新的单位的状态以及这个新的状态与旧的状态的联系。暂时，在这个讨论中，我们粗略地使用多数原则），也就是缩小系统的尺寸的时候，每一个这样的变换，都会使得系统离这个相变的临界点。如果系统处于无序态，多数原则只能使得系统更无序。如果我们的系统已经进入到新的有序相当中去，那么这个变换会使得我们的系统越来越走向这个新的相：因为这个新的系统的每一个单元的状态越来越趋同了。在 Ising 模型中，前者看起来就好像系统的温度越来越高了，后者看起来好像系统的温度越来越低了。这里，我们利用 Monte Carlo 模拟的到的结果做一个演示：如图，随着这个尺度变换——利用多数原则把原来的 $2 \times 2 = 4$ 个看成 1 个，一样多就随便取一个——确实看起来系统在温度高于临界点的一侧看起来温度在升高，系统在温度低于临界点的一侧温度在升高。也就是说，临界点是不动点——关联长度无穷大，但是是一个不稳定不动点。当然，实际上我们要求我们的尺度变换满足两个要求：第一、保持配分函数的值不变；第二、保持相互作用，也就是 Hamiltonian，的形式不变。我们的这里的这个 4 变成 1 的多数原则的变换实际上不满足这一条。在实际分析中，我们需要更好的变换。不过，对于我们理解为什么临界点是一个不稳定不动点来说，这个例子足以说明问题。

为什么我们要维持配分函数的值不变，还要维持相互作用的形式不变呢？原始的相变的临界点的确定需要讨论配分函数的对数以及配分函数的对数的导数的连续性。现在，我们把这个问题转化成为讨论配分函数的对数或者对数的导数随着标度变换如何变化的问题，然后我们企图通过寻找这些配分函数以及它的导数之中出现的关于这个系统的参数，例如 Ising 模型

中的 J 和 T ，来寻找一个关参数变化的不稳定不动点的方式来确定临界点。也就是说，我们想知道随着标度变换，这个原来的系统在它的参数 (J, T 等) 取值的情况下的性质，相当于一个新的系统在什么样的新的参数取值的情况下的状态。于是，我们当然必须保证这个两个系统的状态是一样的，而且是可比的。所谓一样的，就是如果我们取计算物理量，这些物理量的取值应该一样，所谓可比的，就是这个配分函数与各个参量的函数关系也是不变的。也就是说，我们想找到一个标度变换同时保证配分函数这个东西关于系统的参数的函数形式和函数值都不变。如果我们找得到这样的变换，那么，就算我们不能计算出这个函数形式，我们也能够得到这个系统什么时候发生相变的信息。其实，我们还可以得到更多：其实所有的临界行为都可以得到。

这是一个非常聪明的办法。利用变换的不稳定不动点代替求解配分函数。当然，这个变换的逻辑有一点点问题：我们在假设相变点存在的前提下发现，这个相变点就是不稳定不动点。如果我们的变换找不到这样的一个不稳定不动点，不表示相变点不存在，有可能是所用的变换不合适。不过，这个逻辑上的细节在使用重整化群技术的时候常常被忘记。下面，我们来看几个例子。首先，最简单的，一维 Ising 模型。这个模型是没有相变点的。我们来看看如果我们一定要做一个变换，这个变换告诉我们什么。

例 12.1.2 一维 Ising 模型的重整化群求解相变点： $H = \sum_{j=1}^N s_j s_{j+1}$ 。

这里，我们用最简单的格点消去法：定义相邻的两个格点我们只保留一个，如何保留呢，假设我们只保留奇数位的那个自旋。其实这个变换和多数原则是基本上一样的，如果两个一样，则留哪一个无所谓；如果不一样反正需要随便选一个，就一直选奇数位的好了。有了这个变换之后，我们希望保证下面的等式

$$\sum_{s_2} e^{-\beta J(s_1 s_2 + s_2 s_3)} = A e^{-\beta J'(s_1 s_3)}. \quad (12.1)$$

同样的，对于其他的格点，例如 s_3, s_4, s_5 之间，我们也仅仅需要保留 s_3, s_5 形如上式右侧的表达式。只好保证公式 (12.1)，我们就能够保证

$$\sum_{\vec{s}} e^{-\beta H(\vec{s})} = A \sum_{\vec{s}'} e^{-\beta H'(\vec{s}')}. \quad (12.2)$$

其中 \vec{s}' 就是保留下来的奇数位的自旋的状态。这个推理成立是因为 $e^{\beta \sum_{j=1}^N s_j s_{j+1}}$ 可以展开成两两相乘的形式。其中的常数 A 不依赖于系统的参

数，所以将来也不进入物理量的表达式。另外，由于 β 总是和其它参数合在一起的，在讨论重整化群的时候，我们总是把 βJ 看成新的 J ，或者说记 $J = 1$ ，然后我们出现在以后的等式中的 J 其实就是 β 。公式 (12.1) 实际上就是，

$$e^{-J'} = Ae^{2J} + Ae^{-2J} \quad (12.3a)$$

$$e^{J'} = 2Ae^0 \quad (12.3b)$$

也就是

$$J' = \frac{1}{2} \ln \frac{e^{2J} + e^{-2J}}{2}. \quad (12.4)$$

利用非线性动力学中学习到的线性稳定性分析的方法我们可以来确定这个变换的不动点。我们发现，这个方程除了 $J = 0$ ($J = \infty$ 的稳定性更复杂，这里不讨论) 是不动点而且是稳定不动点之外，没有其它不动点。这个结果与我们的预期：一维 Ising 模型没有相变点也是符合的（但是我们要注意我们提到的逻辑上的问题）。

这个例子展示了如何利用某个满足要求的标度变换来寻找不稳定不动点的方法。这是非常神奇的事情。在这个视角下，这个迭代变换是研究相变的核心。从这个迭代变换，我们不仅可以得到不动点，还可以得到临界指数。这是我们下一节的内容。在那之前，我们来看一个稍微复杂一点的重整化群的例子。

例 12.1.3 多层次菱形晶格上的 Ising 模型： $H = \sum_{\langle ij \rangle} (s_i s_j)$ ，其中菱形晶格由一个菱形通过理论上的无穷多次迭代生成。每一次迭代把菱形的一条边扩展成一个菱形。

我们还是尝试用格点消去法：去掉一个菱形的中间的两个点，让这个菱形回复到上一级的晶格。由于理论上我们这个晶格通过无穷多次迭代生成，所以这个迭代不改变整个晶格。

通过上面的两个例子，我们看到，从重整化群的角度来看，相变就是一个系统的特征关联长度从有限的变成无限的于是标度变换不变，而不在这个点上的系统在标度变换下不稳定。所以，这个重整化群的方法是一个可以研究出现这种临界现象的相变的一般方法。对于关联长度不出现这个变化的相变系统，这个方法当然也就不适用了。不过，我们在这里考察的都是出现这样的关联长度的变化的相变。

了解了直接求解配分函数、Monte Carlo 模拟和重整化群方法之后，我们再用 Ising 模型来讨论平均场的方法。

例 12.1.4 一维 Ising 模型的平均场： $H = J \sum_{j=1}^N s_j s_{j+1}$ 。

假设 $\langle s_j \rangle = m$ 而且 $s_j = m + \delta_j$ ，于是，

$$H = (N-1) J m^2 + 2Jm \sum_{j=1}^N \delta_j + o(\delta). \quad (12.5)$$

其中常数系数 $N-1$ 是考虑到边界条件。如果我们用周期边条件，则应该是 N 。这个表达式的第一项是常数，忽略不计，最后一项高阶小量，忽略不计。因此，我们唯一要考虑的是中间一项。然后，由于一会儿我们会看到的原因，我们再用原始的 s_j 变量换回来，我们得到

$$H = 2Jm \sum_{j=1}^N s_j. \quad (12.6)$$

其中，我们再一次扔掉了常数项。这是一个非常简单的单体问题，于是

$$m = \langle s_j \rangle = \frac{e^{-\beta 2Jm} - e^{\beta 2Jm}}{e^{-\beta 2Jm} + e^{\beta 2Jm}} = \tanh(\beta 2Jm). \quad (12.7)$$

这个方程除了 $m = 0$ 之外，根据 $\beta 2J$ 的情况还可能存在着另外两个解或者零个解。这个条件是，

$$\beta_c 2J = 1 \rightarrow T_c = 2k_B J. \quad (12.8)$$

其实，我们可以把这个例子的自由能算出来，把忽略掉的常数都加回来，我们得到

$$F = -k_b \ln Z = Jm^2 - \frac{1}{\beta} \ln(2 \cosh(2\beta Jm)). \quad (12.9)$$

这个表达式将来在讨论 Landau 的相变的一般理论的时候有用。

如果我们考虑一个稍微复杂一点的例子：二维 Ising 模型，我们可以类似地得到平均场近似下的解。

例 12.1.5 二维 Ising 模型的平均场： $H = \sum_{i,j} (s_{i,j} s_{i+1,j} + s_{i,j} s_{i,j+1})$ 。

假设 $\langle s_{i,j} \rangle = m$ 而且 $s_{i,j} = m + \delta_{i,j}$ 。

按照前面的例子，我们忽略二阶小项采用自由边条件而不是周期边条件，则

$$H = (L-1)^2 Jm^2 + 2Jm \sum_{\langle ij \rangle} \delta_{ij} + o(\delta). \quad (12.10)$$

这个 Hamiltonian 跟一维的情形完全一致，于是所有的下面的讨论也完全一致。相变会发生，而且发生在相同的临界点上。

其实，任意维数的 Ising 模型，这个平均场技术给出来的都一样。然而，我们根据 Monte Carlo 的计算以及配分函数的计算，已经知道，一维 Ising 模型没有相变现象而二维才有。这一点，是平均场理论描述不了的。那么平均场理论到底是什么层次的描述呢？这个问题，我们会留在专门的课程里面去回答，不过，下面的进一步讨论可以给出一些提示。

最后，我们给出来一个二级相变的唯象理论，这个唯象理论其实可以看作是平均场方法的理论基础：Landau 的二级相变理论。

假设一个物理系统的状态形式上完全由其序参量描述。实际上一个序参量的值可能对应着很多个微观状态，而且不同的序参量的值这个微观状态的个数可能不一样。这个我们一概不管。假设这个序参量仅仅是位置的函数 $\phi(\vec{x})$ 。我们从统计物理学知道一个系统的宏观状态在合适的条件下就是某一个热力学势函数取极值的状态。这里，我们考虑自由能，假设系统的状态就是自由能取极值的状态。于是，我们的任务，给定一个系统就是寻找使得自由能函数取极小值的序参量。对于 Ising 模型，我们猜测这个自由能差不多有如下的形式：

$$F = \int dV \{a|\phi(\vec{x})|^2 + b|\phi(\vec{x})|^4\}. \quad (12.11)$$

考虑到我们感兴趣的状态实际上处处均匀的状态，我们让 $\phi(\vec{x}) = \Phi$ ，或者说我们先考虑均匀状态做为零级近似（关于为什么均匀状态是 F 函数的可能极值，其实是有其他的道理的： F 中原则上包含 $|\nabla\phi(\vec{x})|^2$ 而这一项对于非均匀系统总是取大于零的值）。于是，

$$f = \frac{F}{V} = a|\Phi|^2 + b|\Phi|^4. \quad (12.12)$$

没有一次项和三次项的原因是我们的 F 需要关于 $\Phi \rightarrow -\Phi$ 对称（来源于 H 的相同性质，关于所有的自旋翻转不改变 H 的值）。现在，我们来探讨这个函数的最小值。

$$f' = 2a\Phi + 4b\Phi^3. \quad (12.13)$$

于是，这个最小值可能出现在 $\Phi_0 = 0$ ， $\Phi_0 = \sqrt{\frac{-a}{2b}}$ ， $\Phi_0 = -\sqrt{\frac{-a}{2b}}$ 处，对应的 $f'' = 2a, -4a, -4a$ 。后面两个点是最小值当且仅当 $a \leq 0$ 。所以，如果我们要

求出现相变这样的现象：在某个温度以上，系统只有一个最小值而且其值为零；在这个温度一下，系统有两个最小值而且其值为非零，则需要，

$$a = a_0 (T - T_c), \quad (12.14)$$

其中 a_0 是大于零的常数。Landau 就是这样对相变的唯象理论提出的要求。在这个要求下，相变的物理图景非常清楚，见下图：在 $T > T_c$ 的时候，系统只有一个极小值 $\Phi = 0$ ；在 $T < T_c$ 的时候，系统只有两个极小值 $\Phi \neq 0$ 。

这个关于 F 的函数形式的假设看起来很突然，关于系统 a 的要求也好像硬凑起来的。但是，这个理论在很多的具体的系统的情况下，可以在一定的近似下被推导出来。超导的研究就是一个这样的例子 [?]。Landau 理论还可以用来讨论一级相变。这个关于 $\phi(\vec{x}) = \Phi$ 的理论，假设系统处处一致，也是平均场方法的基本精神。

其实，如果把公式 (12.9) 做一个 $m = 0$ 附近的展开，我们就会发现，其形式正好满足上面假设的情形，

$$f = 2J(1 - 2\beta J)m^2 + bm^4. \quad (12.15)$$

12.2 临界行为

重整化群与临界指数，普适性，有限大小标度

12.3 自组织临界性

沙堆模型，幂律统计，自组织驱动与外界驱动的对比如

12.4 更多的具体系统上的研究实例

第十三章 复杂网络概论

我们已经强调过相互作用在复杂性研究中的重要性。任何无相互作用的系统，或者是可以通过某些变换变成无相互作用的系统，我们都可以把这个系统的可观测量的演化方程写成一个线性方程，于是这个系统的动力学行为总是简单的。对于相互作用系统，例如经典力学、量子力学、随机过程，尽管用概率层次的描述的话形式上这个系统的状态演化的方程还是线性的，但是如果我们关心这个系统的可观测量的演化方程，这个方程必然是非线性的。于是，如果沿着力学的思想来研究所有的系统，对于包含相互作用的系统的研究就非常复杂。

复杂网络，为研究相互作用系统提供了一条新的思路。举个例子。我们在一个人群中考虑接触传染的传染病现象。如果我们关心整个系统任何时刻的状态，那么，我们需要了解这个人群中每一个人在什么时刻与什么人发生接触，这个接触是什么形式的，因为不同的形式可能传染的概率不一样，这个概率甚至会依赖于这个两个接触的人的身体的进一步情况，甚至精神状况而不仅仅是身体的物理状态。没准，这个还与这个接触发生在哪里有关。如是，我们发现，一旦这个人群比较大，这个研究的思路，基本上就是没有希望的。我们首先需要建立描述这个人群的每一个人的合适的变量，然后考虑环境的描述，然后这个疾病与这些变量的关系，然后讨论这个系统的动力学。传统的传染病模型对这个问题做了一个非常简化的研究：个体同质化，忽略环境，于是个体均匀混合。相当于人群中的每一个人都被其他人包围着，人群的密度是一致的，然后与病毒携带者或者病人接触就会以一定的几率染病。在物理学，这个处理方法被称为平均场理论。这个方法获得结果往往能够反映大致的情况，但是细节上的描述能力有所欠缺。复杂网络的研究思路告诉我们，这里存在着一条介于完全的动力学描述和同质化均匀化的平均场理论的中间道路。这条中间道路说在这个问题中，如果我们能够加上这个一个信息，考虑每一个人的经常接触的人有可能是不可

一样的，然后忽略什么地方什么时间形成接触，有必要的話，也可以考虑每一个人的体质等健康情况导致传染和被传染的几率不一样，那么这个时候，我们是不是就可以讨论一些比较细节的问题了呢：例如对于这个传染病来说，谁最危险，谁对于整个人群来说最危险（一旦其得病其他人得病的可能就大大升高），谁是控制核心（一旦这个人与其他人的接触切断，整个人群得病的几率就会大大降低）。这些问题，在平均场描述下都是解决不了的问题，而全信息的力学研究又难度太大。而且，实际上，在传染病的这个问题中，精确地掌握每一个个体的信息的重要程度远远比不上之前我们问的几个谁是各种各样的核心的这些问题。

于是，我们发现，在这个角度下面做研究，第一，我们需要一个每一个个体与其他人如何相连——这个被抽象为邻接矩阵，将来会定义；第二，我们需要每一个个体的一个体质的或者是其他的关于转染和被传染几率的一个综合描素——这个被抽象为顶点的权重；第三，我们也可以考虑每一条边可能是异质的——接触频率或者方式不一样导致传染的几率不一样，这个被抽象为变得方向和权重。合起来，我们把这三个数学结构的整体，称为网络。

当然，在有了这个网络之后，我们仍然可以想办法加上其他信息构成完整的动力学描述。但是，网络科学的魅力就是，很多时候，这个结构以及相应的结构信息的分析计算，就能够给我们要解决的问题很好的提示甚至答案。下面，我们就通过一点点基本理论的介绍——也就是分析及算哪些结构和怎么算的问题，然后通过具体的例子来体会如何把网络科学用在描述复杂的一般的相互作用的系统之上。

本章推荐阅读材料：Barabasi 的《Network Science》，吴金闪和狄增如的《从统计物理学看复杂网络》，王小帆和陈关荣的《网络科学导论》，MEJ Newman 的《Introduction to Networks》

13.1 复杂网络的描述方法

在上面的复杂网络的描述基本思想之下，复杂网络的研究有几个主题。第一个是网络作为现象发生的舞台在多大程度上如何决定了现象的什么性质，我们称为网络的静态量（有时候也称为集合量或者结构量）的问题。例如，当我们考虑上面疾病传播问题的时候，显然网络的平均最短距离，也就是说平均来看网络上任意两点之间需要经过几次接触才能连到一起。然后，

平均来看一个人有多少个有效的一定意义上独立的接触也是一个重要决定疾病传播过程的参数。这个参数实际上可以用两个量来描述：第一，顶点的度；第二、顶点的集聚系数，也就是这个顶点的邻居有多少的比例是直接相连的。当然，实际上，还存在着一些实际上不太能从直觉中体会到的关于这个舞台的描述也对这个过程有很大的影响，例如我们以后会介绍的度分布的指数，也就是说不仅仅是度的均值，这个分布的方差以及其它数字特征也有影响。关于这个第一个主题，我们回顾一下当时复杂网络兴起的初期的一个工作是有意义的。这个工作就是 Strogatz 和 Watts 的小世界网络模型以及小世界网络上振子同步的工作。我们在 13.2 节会把这个工作和欣赏这个工作所需要的基本概念放在一起介绍。

第二个是网络上顶点重要程度的刻画，也称为中心性问题。在一个异质的多体系统中，我们总是希望能够把这些个体的不同意义上的重要程度给一个描述，这样我们可以为了特定的目的把这些顶点区别对待。例如要保护连通性缩短距离我们需要有限保护或者提高哪一些顶点，要破坏连通性增加距离需要有限破坏哪一些，或者在上面的传染病模型中，哪一些对于控制传染病的传播是最关键的顶点等等。这个重要程度的刻画与其他的非网络科学的方式有一个不同之处，我们希望通过一个顶点在网络上的地位的不同来描述其特殊的地位，而不是通过这个顶点的其它不能放在网络上的性质。一定程度上说，我们希望描述一个位高权重者，不是通过这个位高权重者实际在完成的事情或者特有的人格魅力或者独特的学识和魄力，而是通过这个位高权重者如何与其他的顶点所代表的个体发生相互作用的角度来衡量。而且，在这里我们也不是预先就知道其它顶点所代表的个体的地位。整个分析都是基于网络结构以及网络描述本身来讨论的。

静态量和中心性问题这两个方面有的时候是交织在一起的：网络的某一个描述量可以用来解释网络上发生的现象，而这个量有经常能够用来给网络上的顶点做一个排序。当然，其实这两个问题的角度是完全不一样的。

第三个方面，是问具有某种属性的网络是如何发展演变而来的，也就是网络的演化问题。当然，某些网络是可以完全通过人工设计而来的，但是绝大多数的网络现象没有一个最终设计部门，是自发演化得来的。那么这个时候，如果我们发现，这些演化得来的网络存在着一些共性和个性，我们就会问这些共性的特征还有这些个性的特征是如何形成的。这个问题的研究主要是满足好奇心的问题，但是如果需要对这样的一个自发演化得来的网络进行干预，那么这个问题的研究的结果也是有益的。

第四个方面，是问有哪一些动力学现象在网络上的表现和其它背景舞台上的表现不一样的，也称作网络上的动力学。其实，这个问题和第一个问题在一定程度上是相关的，只不过这个时候我们的关注的重点，也就是研究的对象是动力学过程。之前的问题中关注的主要是描述背景舞台，也就是网络，的量。

第五个方面，是问有哪一些问题是必须通过网络科学的角度和技术来回答的或者说通过网络科学的角度和技术能够回答的更好，是其它的研究方法比不上的，也就是做网络科学的独特的应用的问题。在一个技术、一个思想或者说一门学科成熟的过程中，这个独特的应用问题总是一个非常核心的问题。我们认为，网络科学在这一点上是很有潜力的。

理解了网络研究的这五个主题之后，我们来分别阐述网络研究的这几个方面。

13.2 基本静态量与中心性

举例：小世界模型举例：K-core 和 K-shell 的原始文章度、集聚系数、介数、K-core、K-shell、平均最短路径、集团结构等

13.3 网络演化

BA 模型、局域机制问题和局域机制模型

13.4 网络上的动力学

13.5 网络科学应用

13.5.1 用网络研究方法增进人类交往，提高创造性和生产力

Alex Pentland

13.5.2 用网络研究方法发现容易发生脑退化的人

13.6 本章小结

在这一章里，我们主要讨论了网络研究的思想——把相互作用抽象成存在与否——和研究的主题——网络研究的静态量、中心性、演化、动力学模型和独特应用，捎带着用一些例子讲解了一部分网络研究的概念和技术：度、集聚系数、介数、K-core、K-shell、平均最短路径、集团结构、BA 模型等等。网络科学还是一个处在发展阶段的学科，新的概念会继续产生，新的独特的应用也不会不断地被创造出来。这些基本的概念和技术仅仅给读者提供一个在这个领域里面进一步思考和工作的基础。例如，网络的控制[?]：在网络上放上一个线性的动力学系统，通常每一个顶点一个力学变量，然后讨论使得这些力学变量到达期望的状态的路径的问题，是近几年兴起的一个研究方向。但是，实际应用的背景以及在实际问题中如何控制的问题还有待于进一步的研究。

第十四章 博弈论与演化博弈概论

博弈论，也叫做对策论，的基本问题是：当有利益冲突的博弈者在做策略选择的时候，有没有一个理论可以预测，以及如何预测，这些博弈者的选择。一个没有利益冲突的策略选择问题是决策问题：在那里每一个决策者的可以选择的策略集合是一样的，利益是共同具有的或者是完全一样的。决策问题是一个约束下的极值问题，因此尽管约束下的极值问题的求解还是有值得进一步研究的地方，理论框架本身是没有问题的。但是，博弈论的问题不是这样：决策者 A 的收益不仅仅取决于 A 的策略选择还依赖于 B 的策略选择，也就是我们所说的相互作用收益或者说能量是两个变量的函数而且不是简单叠加 $V(s^A, s^B) \neq V(s^A) + V(s^B)$ 。如果我们假设这个对策论的问题可以通过下面的方法转化成一个决策问题：固定其它决策者的选择，让某一个决策者按照他自身的目标优化其策略，然后把这个决策者的策略固定考虑另外一个决策者的决策问题，那么看起来，博弈论也就不复存在了。这个思路叫做博弈者的交替的短视的最优反应动力学 (Myopic Best Response Dynamics)。但是，我们发现，这个理论预测得到的博弈问题的决策经常不能与实际的决策相符合。所以，博弈论需要构建自己的基本理论。

在这里，我们仅仅讨论非合作完全信息静态战略博弈。这个博弈的理论是其他的博弈的理论的基础。在本章中，我们简称“非合作完全信息静态战略博弈”为博弈。Nash 所构造的博弈论基本理论试图寻找一种特殊的策略状态，在这个策略状态下，所有的决策者不会产生独自的偏离。也就是说，如果自己独自选择其他的策略，那么这个偏离的博弈者的收益不会增加。这个特殊的策略状态后来就被叫做 Nash 均衡。如果一切顺利，那么我们剩下的问题也就是对于给定的博弈如何求解 Nash 均衡的问题。但是，这个世界不想让我们这么轻松这么无趣，某些问题的最终策略确实能够被 Nash 均衡解释，但是有很多博弈，其博弈者采取的策略不能由 Nash 均衡解释，还有的问题包含了太多个的 Nash 均衡于是解释和预测能力不够。于是，现在我

们面对的问题是：有没有更好的均衡的定义，它对于给定的博弈给出很少的预测，而且是很准确的与实际博弈过程的结果很相符的预测。目前这个问题还是一个开放的正在得到研究的问题。

同时，博弈论中还有一个基本问题就是如果大家一开始不是处在某个不变的均衡态上，例如对于能够用 Nash 均衡解释的问题来说，假设一开始很多博弈者没有经过精确的计算所以没有采取 Nash 均衡的策略，随着博弈的进行，大家的状态是如何演化到这个均衡上去的。这个实际上提出了比均衡更高的要求：稳定性。

在这一章中，我们还是主要介绍 Nash 的理论体系，然后介绍演化博弈理论——一个尝试从许许多多的均衡解中把某些特殊的少量的均衡挑选出来以及回答解的演化和稳定性这两个问题的理论。其他的更多的均衡的定义，以及从一大堆均衡中选择均衡——称作均衡的提炼——的方法就留待读者从专门的博弈论的教材中去学习，或者自己取提出新的定义或者新的提炼方法。

实际上，这个学科是对有主观能动性的决策主体的策略选择过程和选择结果的描述。可以想象，如果我们讨论的是人类或者是人类的集合（例如一个有一群人构成的组织、机构），那么实际上这个决策过程和人的思考、决策的神经过程是相关的。于是，本质上这门学科成了神经科学的一个应用性学科。那么，是不是在我们解决人类的思维的规律之前，这本学科都不能取得发展了呢？不是的。人类的思维科学如果建立的话，当然对博弈论是一个非常大的促进，但是当这样一个理论没有的时候，我们也是有可能可以解决某些人类决策的行为的。某些决策行为是可以存在详细思维过程的基础和细节之上的共性的。我们的博弈论研究的正是这部分共性。例如，我从来不期望我们的博弈论的研究可以解决绘画的运笔，作曲的曲调的选择等艺术创作的问题。但是，原则上，这些现象也应该是思维科学的研究对象。一个更基本的层次的学科可以促进上层学科的发展，上层学科依赖于更基本的学科，但是又不是完全由更基本的学科决定，上层学科可以有自身的规律，这个正是系统科学的核心基本思想之一：涌现性——更高层次的系统可能涌现出来与底层的系统看起来完全不相关的行为。

我们为什么认为博弈问题是系统科学的重要研究方向呢？系统科学的任务是运用数学概念、自然科学尤其是物理学的概念和方法，来构建自然的或者社会系统的数学模型。这个研究的角度就是各门科学研究的共性，而研究的结果，也就是所构建的数学模型，体现的，则是各门学科自身的特性。

我们也提到过，在这样的数学模型需要考虑的因素中，最主要的，就是对于相互作用的处理。我们认为，社会科学的问题，其根本在于各个主体之间相互作用的描述，以及从如何计算分析这些相互作用产生出来的现象。在这一点上，我们认为，博弈这个问题，博弈论这个学问，是所有其他的牵涉到多个个体的整理行为的社会科学研究的基础。系统科学就是需要研究从少量个体到多个个体的行为，而博弈这个问题正好就是这样的研究或者至少是这样的研究的基础。

本章推荐阅读材料：《Strategic Thinking》

14.1 博弈、博弈论与 Nash 均衡

Nash 均衡理论的基本内容

14.2 理论与实际的对比：Nash 均衡的描述能力

理论与实验的对比

14.3 演化博弈：均衡实现的问题

演化博弈的基本理论，Replicator Dynamics，演化稳定性分析

14.4 新的均衡的问题

14.5 更多的具体系统上的研究实例

第十五章 优化理论

15.1 运筹学概论

- 15.1.1 基本问题及其数学描述
- 15.1.2 问题的解决方法举例
- 15.1.3 更多的具体系统上的研究实例

15.2 控制论概论

- 15.2.1 基本问题及其数学描述
- 15.2.2 问题的解决方法举例
- 15.2.3 更多的具体系统上的研究实例

第四部分

附录：计算技术与学习方法

第十六章 计算机语言与工具的学习

16.1 Linux 系统中安装、编辑、编译、执行程序

Linux 系统是一款由爱好者们一起借鉴 Unix 系统开发出来的一般来说免费提供给用户使用的开源的（能够获得原始代码并做修订）操作系统。其在网络功能（例如远程登录服务器、Web 服务等等）、稳定性、编程开发、作为服务器使用方面具有超过其它大多数其他系统的表现。更多的关于 Linux 的历史信息可以参考 Wikipedia 的 Linux 条目。在这里，我们仅仅需要知道以下简单的信息：Linux 的核心是一个完成最基本的操作系统的功能的内核；爱好者们为 Linux 系统开发了各种软件工具，而且大多数开源免费通过应用程序库提供；Linux 系统的软件通常遵循“我只做一件事，并且做好它”（Do one thing, and do it well）的来自于 Unix 的原则；基本上所有能够用图形界面通过鼠标点击做的事情都可以通过命令行来实现，因此需要重复多次的事情最好通过命令行批处理文件来执行；Linux 系统的每一个文件和和目录（实际上也是文件）都通过权限来控制——例如每一位用户的文件和目录原则上只有这个用户自己能够访问。

很久以前的 Linux 基本上是爱好者为了爱好者群体开发的，图形界面以及系统配置的易用性不是其主要考虑的因素。因此，对于非爱好者来说，入门的门槛比较高。现在，大多数发行版本由专门提供整合版本的机构或者公司来提供。这些机构已经把普通用户作为目标，因此，基本上所有的整合版本，称作 Linux 发布版，的 Linux，例如 Ubuntu、Debian、Fedora 等等，已经完全不需要用户来配置，也直接提供直观易用的图形界面。如果你是手机或者平板 Apps 的使用者，你还会发现，从软件库中搜索和安装软件的过程和 Apps 非常接近。实际上，在 Apps 的这种管理方式推出之前，大多

数 Linux 系统已经开始使用这样的通过网络来提供的由专门机构进行管理的软件库，通常称作软件中心，的软件管理方式。如果没有这样一个软件中心——一个网上的软件存储、分类、更新的中心，加上用户电脑上的软件中心客户端软件——的话，浩如烟海的 Linux 上的软件会让每一个用户迷失。

由于 Linux 开源免费（因此也没有病毒，所有的开源程序的代码是可以获得的，因此可读的。实际上在 Linux 里面，除了程序的源代码，所有的配置控制文件也都是文本形式的、人类可以读的。这一点对于深入了解 Linux 非常重要）、运行稳定、网络应用基础架构丰富，在 web 服务和高性能计算服务方面，已经具有非常重要的地位。大量的 web 和计算服务器，尤其是 cluster 结构的服务器——有多个电脑通过光纤、网线网络相连——上的占有率更高。另外，高性能计算服务器如果不是 Linux，那么也很有可能是 Unix 服务器。两者在普通用户层次的使用上，基本不可区分。因此，在这里我们学习如何在 Linux 环境下做高性能计算。

Linux 发布版的安装非常简单，搜索并下载一个发布版，例如 Ubuntu。本书写作的时候，Ubuntu 的版本是 14.04。下载的文件通常是一个 ISO 光盘镜像。下载完了之后，按照提示制作一个启动 USB 或者启动光盘，或者下载的时候选择一个在 Windows 下可以直接安装的版本。下载了 ISO 文件之后，你也可以选择在虚拟机中安装。如果你的电脑内存比较高（> 2G）而且你首次安装和使用 Linux，我推荐你在虚拟机安装。

除了下载完整的 Linux 发布版，还有 Windows 下的软件，提供 Linux 模拟器——不用安装新系统，安装一个软件就行，在这个软件下，编程、配置、操作和 Linux 一样。Cygwin 就是这样的软件。如果你能够拿到 Linux 服务器的帐号，你也可以不在自己的电脑安装 Linux，而选择直接使用 Linux 服务器。通常，我不推荐这个方案，因为在自己的电脑上安装非常简单，而且出错了只影响自己的电脑，而在服务器上出错了可能影响其他人的工作。以下假设你在自己的电脑上安装了 Ubuntu 的 14.04 版本。

如果你是 Mac OS 用户，那么你实际上在运行一个 Unix 的变体，就不用再安装 Linux 系统了。以下介绍的软件编译、执行的过程在 Mac OS 下同样适用。当然，在 Mac 机器上，你也可以独立或者在虚拟机上安装一个 Ubuntu。

对于大多数功能来说，你都能在软件中心找到合适的软件。找到以后，只要简单点击几下鼠标就能够安装好。安装的过程中，有可能让你输入一个能够在本机上安装软件的用户密码。注意，由于 Linux 系统的管理方式

的基础就是权限，所以并不是每一个用户都能够在一个系统上安装软件的。如果是你自己安装的 Ubuntu，在安装过程中你建立的自己的账号和密码，那么，Ubuntu 默认这个用户具有安装软件以及调整其它设置的权利。我们一会儿会更加深入地来介绍这个安装的过程实际上系统做了什么事情。

现在，我们来看一个从源代码开始的安装程序的过程。有一些软件是不用编译可以直接用解释器执行的，例如 Python 或者 Java 等语言写的软件。这个时候只需要把源程序复制到一个地方，然后设置好路径就行了。有的时候解释语言的程序也会做一个安装包，那么这个安装包实际上就是把文件复制到合适的地方然后设置好路径。

下面，我们以需要编译执行的软件为例。通常情况下，源代码包含一些文本文件，例如 C 语言格式的文件（后缀.c），或者 C++ 格式的文件（后缀.cpp），还有一个非常重要的 makefile 文件（文件名 Makefile 或者 makefile）。有的时候，还会有函数说明文件（例如后缀.h）。把函数说明，或者说函数的形式化定义，把函数的真实实现程序分离开来的好处是，对于大多数软件的用户而言，实现的细节是可以不了解的，只需要知道基本的功能和输入输出就可以了。因此，单独的说明文件在存在多个子程序的代码中使用的非常普遍。

注意，如果你有在 Windows 下通过 Visual Studio 集成环境编写程序的经验，你是可以直接在一个界面下面实现编辑源程序、编译程序和编译所得到的结果的，甚至可以把这几个步骤组合成一个按钮。但是，我们要了解，实际上这个过程是可以分解完成的。编译的步骤，在任何文本编辑器里面都可以完成。很多人喜欢用 Emacs。我自己最近喜欢用 Geany。

下面就是一个简单的 C 语言 HelloWorld 程序。

有了源程序，下一步就是编译。编译的过程的原理和细节就不介绍了。简单来说，就是把源程序给这个语言的编译器来产生这个机器能够运行的机器码，也就是把程序员能够读懂的语句，变成机器能够执行的语句。这个过程通常需要如下形式的语法，例如对于 C 语言，我们可以执行如下的一系列命令：

```
$ gfortran -c HWF.o HWF.F （把Fortran源程序编译成为可以供其他程序调用的O文件）
$ gcc -c HWC.o HWC.c （把C源程序编译成为可以供其他程序调用的O文件）
$ gcc -c HelloWorld.o HelloWorld.c example.h （把C源程序编译成为可以供其他程序调用的O文件）
$ gcc -o HelloWorld HelloWorld.o HWF.o HWC.o （把各个O文件编译成为可以供其他程序调用的可执行文件）
```

注意，在这里，我特意实现了三种方式来输出 HelloWorld（以后你还会见

到并行的版本): C 语言主函数、C 语言子函数、C 语言调用 Fortran 子函数。这个命令的更多细节稍候再介绍。

通过编译, 我们得到了文件名为 HelloWorld 的可执行文件。这个时候, 我们就可以通过一个包含完整的路径的文件名的方式来执行这个文件。例如当我们在这个文件所在的目录 (/home/jinshanw/Example, 在任何目录下通过 `cd /home/jinshanw/Example` 到达这个目录) 下的时候, 我们来运行 `./HelloWorld` 就可以看到程序执行的结果。

这个就是一个典型的在 Linux 上编辑、编译、运行程序的过程。如果有安装过程, 那么实际上就是把最后得到的编译好的文件复制到合适的地方, 然后让系统记住这个程序已经被复制到新的位置。例如 HelloWorld 文件可以被一定到 Installed 子目录下面, 然后运行的时候输入命令 `/home-jinshanw/Example/Installed/HelloWorld`, 或者修改环境变量 Path 让系统知道在这个目录下寻找 HelloWorld 可执行文件。至于如何修改路径, 请 google 之。

如果我们的源程序包含多个子程序, 那么每一次的编译执行我们都需要把所有的以上的命令。为了减少重复劳动, 人们引入和 makefile 文件。我们只需要把以上四行命令按照顺序排列好 (如下图), 然后给这一组命令一个名字, 下次我们只要简单运行 `make HelloWorld` 就可以执行所有的这个模块的命令了。同时, 我们还可以把其他的命令也放到这个 makefile 中去。例如, 这里我们把 Makefile 的执行部分当作一个模块, 叫做 runHW。你可以试试 `make runHW` 的结果。使用 makefile 的另外一个好处是程序之间的依赖关系是显性的, 有利于后期的修改和维护。

实际上, 安装其他人编写的软件的过程也就是跳过了编辑这一步, 需要完成编译、复制、设置路径这几步。在编译过程中, 有一个依赖性的概念很重要。例如, 这里, 我们 HelloWorld 程序实际上需要 HWF 来实现一部分功能 (C 语言程序调用 Fortran 来完成一部分功能), 这个就是 HelloWorld 对 HWF 的依赖性。绝大多数时候, 程序开发不是从头开始的, 需要使用很多其他人已经开发好的模块, 这个时候都需要处理好程序之间的依赖关系。有的时候这样的模块, 如果被使用的比较普遍, 会专门编译成为库文件 (Windows 下的 .DLL 文件、Linux 下的 .a 文件)。关于如何使用库文件见下一个例子。在上面的例子中, 假设我们能够拿得到所依赖的程序的源程序, 例如 HWF.F, HWC.c。实际上, 在编译 HelloWorld 过程中, 我们直接需要的是 .o 文件。所以, 将来我们会看见, 只要能够拿到 .a 文件, 也就是一大

堆.o 文件的集合，我们就可以把这些子程序模块用到我们自己的程序中来。

总结一下，自己编写程序的过程包含编辑、编译、运行等几个步骤；安装他人的程序也需要通过下载源程序（解压缩）、编译、复制、设置路径等几个步骤；以上步骤可以利用 makefile 文件整合，从而可以用一个命令来完成所有的操作。在安装他人程序过程中，需要注意程序之间的依赖性问题。因此，通常在安装他人程序的时候，需要执行一个依赖关系探测的步骤——这个通常通过 configure 命令来完成。因此，一个典型的安装程序的过程是

```
$ ./configure (或者某个路径XX下的: XX/configure)
$ make
$ make check (或者make test)
$ make install
```

中间加入的一步通常是为了验证编译好的可执行文件是否产生期望的结果。

实际上，软件中心完成的事情就是，在服务器端已经有人完成了针对特定版本（也就是你的机器上的版本）的 Linux 的已经编译好的文件（也就是 make 已经被执行），并且在服务器端已经对软件质量和代码的可获得性做了一定的控制，然后检验依赖关系（相当于运行 configure）并补充必要的依赖软件，接着复制已经编译好的软件到合适的目录下面，修改必要的配置文件和环境变量（运行 make install）。

在实际编写程序的过程中，还有一个重要的步骤：调试。很少有人写一个程序能够一次就写好，不出错的。因此，在程序编写的阶段编译运行程序，测试其功能是一个非常重要的步骤，调试和编辑、编译、运行实际上是一个需要反反复复来执行的过程。关于调试，以及编程风格——为了方便调试和减少错误的总结和发展出来的如何写程序的一些惯例，例如变量名的大小写和长短、子函数的拆分和长度、缩进的长度、标注的方式和详细程度等等，请通过 google 自己找到更多的参考资料来学习。

除了这个编辑、编译、执行、调试、安装的过程，我们还需要学会简单的操作系统的功能，如何建立、复制、移动、删除、修改文件。这个都可以通过图形界面来完成，但是了解一下这些最基本的功能的命令行实现也是有好处的。Linux 系统的简介以及 Linux 下安装软件就介绍到这里。

16.2 C 语言、Python 与编程风格

任何语言的学习，或者说，任何工具的学习，都是通过使用来实现的。语法规则的总结、例子例句的使用是为了帮助学习者更好地通过是用来学习，或者是作为一种补充提高做中学的系统性。因此，计算机语言的学习首先需要学习一般流程：如何编辑、编译、执行程序，如何调试。然后，学习基本的关键词和语法。最后，通过做一些小例子，以及比较大的程序来真正实现有意义的学习。千万不要追求新、追求全。除了基本的语法，最基本的算法，其他的都可以在工作中用到的时候再学习。而且，最好带着一个问题来学习计算机语言，而不是泛泛地学习。例如，我们可以为了实现网络分析来学 C 语言，也可以为了求解线性方程的解来学习 C 语言，也可以为了图像处理来学习 C 语言。

第一，我们需要大概了解一下 C 语言提供的最基本的数的类型：逻辑型（名义上没有这个类型，实际上有，以及相应的逻辑运算）、整数（int）、长整型（long）、实数（float）、双精度实数（double）、单精度复数（complex）、双精度复数（double complex）与相应的运算，以及最基本的 C 语言自带的数组——以及与数组相关的指针和内存。当然，C 语言也允许你通过结构体之类的东西来定义由这些基本数据类型构成的自己定义的数据类型。第二，我们还需要了解 C 语言的基本的语法结构：条件（if-else）与分支（case-switch）、循环（for, while, do-while）、跳转（goto, break, continue, return）。第三，少量的自带库函数，例如三角函数对数等数学运算，输入输出函数等。有了这些就够了。就可以开始用中学了。而且这些基本的内容也是需要了解一下而已，更多的细节也需要在用中学。这样的话，所谓 C 语言的学习，也就是几个小时的事情。

当然，如果这是你第一次接触编程，那么，你需要从其他人的漂亮的程序中学习一些基本的算法。例如，找一个排序算法（例如冒泡或者二分）体会一下，然后自己动手模仿一下。你还可以尝试把数组这个基本的结构拓展成矩阵（物理上还是一维的，逻辑上当作二维来使用），然后对这个矩阵做基本的矩阵加法乘法运算。你还可以尝试利用数组和链接表来表示一个网络，然后实现最简单的网络分析功能，例如计算度、集聚系数等。

有了这些准备，你就可以开始你的研究工作了。注意，所谓数学，就是对心智中的世界的一种抽象；所谓物理学就是对现实中的关于物体状态和物体运动的那部分世界的抽象化；所谓计算机科学就是对这个世界能够通过步骤和计算这个方式描述的那部分世界的抽象化。因此，实际上，你的研究

工作可以是完全和计算机或者计算机语言没有关系的问题。你所要做的就是将所研究的问题转化成为一个可以用明确的步骤来计算的一个问题。如果你还没有合适的问题，在这里，我建议你可以考虑一下 google PageRank 算法的实现。WWW 网络的数据可以从很多地方下载到。其实其他网络的数据也可以拿来用。有了问题，有了基本的语言单位，有了算法，那么剩下的事情就是简单地通过使用增长经验的问题了。

C 语言尽管具有灵活、高效率的特点，但是其基本的功能模块比较少。例如，在最基本的 C 语言里面，也就是不包含各个 C 语言的商业编辑器提供的额外的功能，链接表——可以看作可变长度的数组——都是需要程序员自己定义的结构。处理字符串的功能也比较少。当然，这些额外的功能其实都可以通过在基本的 C 语言包上增加额外的库函数来实现，而不需要 C 程序员重复发明轮子——” C 程序员一直在重复发明轮子 “是一种流行的说法。另外，C 语言本身如果要实现面向对象的编程，尽管时可能的，但是需要程序员在思想上比较重要的转变。因此，实际上更高级（而不是更好地）的语言，提供更多的功能模块的和更多的对象抽象和封装的功能的语言，也是有相当的市场的。对于某些目的来说，可能直接采用更高级的面向对象的语言，甚至有的时候解释语言而不是编译语言是有好处的。由于在网络功能、做图功能、字符串处理功能、扩展程序包、可移植性等方面的优越性，我推荐学习 Python 语言。

有过比较大——几十个子程序几百个小时就行——的程序的开发经验的程序员，就会从最基本的编辑、编译、调试、执行的循环注意到变成风格的问题：当你需要从一个程序跟踪回去检查另一个程序的可能的的问题的时候，你就需要两个程序从形式上看起来比较一致，这样才容易阅读。你还需要两个程序都有比较完整的说明，最好每个程序关键的几行也有比较好的说明。你还会希望你的变量名、子程序名称都比较直观，也容易阅读，也就是不太长，名称又具有提示意义的作用。你希望缩进和大括号的运用是一致的，这样可以用你的眼睛就看出结构化的结构，而不是必须依赖于集成开发环境的提示。你希望你的子程序的功能比较单一又相对完整，还要不太长，能够一口气看出来在做什么。而且，你会希望你在所有的这些看起来无关程序内容的小问题上的选择是一以贯之的，也就是不会今天这样选择，明天那样。这些无关内容的小问题的选择就称为“编程风格”。

对一个程序员来说，编程风格是一个非常重要的问题，尽管没有固定的风格也不影响你写出算法上、效率上优秀的程序。更多的关于编程风格

的内容可以在“Linux kernel coding style”看到。Brian W. Kernighan 和 Dennis M. Ritchie 的《The C Programming Language》[?] 是一本具有明确、一致和良好的编程风格的 C 语言教材，俗称 K&R。

16.3 BLAS、Lapack 及其实现

16.4 并行计算初步与 BLAS、Lapack、Petsc、Slepc

一个计算机程序抽象地来说，就是一个接受两个输入一个输出的机器。两个输入分别是：指令和数据，一个输出是：数据。在这个过程中完成的事情就是把接收到的指令作用到接收到的数据上，然后给出新的作用完了之后的数据。所以，一个计算单元的核心就是存储和执行指令的地方，以及一个存储数据的地方。这两个地方分别叫做中央处理器（CPU）和内存。现在 GPU 也可以成为存储以及执行指令的地方。CPU 通常也带着一定量的内存。这些细节，在我们的抽象的计算单元的模型中都忽略不计。

当我们需要的计算任务耗时非常长，或者需要大量的内存的时候，我们可以考虑让多个机算计或者计算单元一起工作来解决这个问题。这个就是并行计算。基本的并行计算可以是把一个需要重复非常多此的任务放在不同的计算单元上计算，然后把计算的结果合起来。Monte Carlo 计算天生适合这个并行计算的思路。大量的独立轨道需要模拟，这样的独立轨道都可以分开有不同的计算单元来完成。不过这个时候，要注意一个问题，各个计算单元的随机数不能是一样的，最好能够通过并行随机数发生器来产生随机数。稍微复杂一点的并行计算是把一个任务分成多个子任务，然后这些子任务之间需要交流数据来完成这个大任务。利用多个计算单元来处理矩阵和矢量相乘就属于后者。例如一个 CPU 对应着的内存存储着这个矩阵以及这个矢量的一部分的元素，另外一个 CPU 对应着的内存存储着这个矩阵以及这个矢量的另一部分元素，假设计算的结果——乘出来的矢量也将会分散地存储在两个 CPU 相应的内存中，那么，计算过程中第一个 CPU 可能需要用到第二个 CPU 对应着的内存中的数据。这个时候，我们就要有办法能够把数据交流过去。这个就是并行计算的基本的抽象模型。

现在，我们来看一下实现这个任务需要哪一些稍微具体一点的结构。第一，我们需要一个相互连起来的计算单元，这样数据才能够在这些计算单元中交流。所以，一个并行计算的物理基础是一堆连在一起的计算单元。由

于计算单元之间的连接方式非常多种多样，例如通过主板相连，通过光纤、网线相连，甚至通过 Internet 网络相连，如果我们要直接面对这个物理的连接的话，我们的程序需要具体连接具体处理。这个会导致程序的硬件依赖，非常复杂。所以计算机科学家们提出了一个抽象模型：消息传递机制 (MPI)。不管底层的物理连接是什么，我们希望得到一个资源（包含 CPU，内存，可用来计算的时间）的池子，在这个池子里面，消息通讯的问题能够统一地解决。这个机制包含两个方面，第一，把物理硬件通过某种方式整合起来成为池子；第二，在这个池子里面，可以根据任务的需要通过计算机程序来建立一个某个计算任务所需要的资源的整体，然后在这个整体上来做计算。作为并行编程和并行计算的使用者，我们通常只关心第二个层次。第一个层次一般是硬件、软件工程师和系统管理员的事情。但是，我们要了解这样一个概念：首先需要有一个资源池子，然后我们计算的时候根据需要在这个大池子之中建立一个抽象的超级计算机。然后，我们实际上面对的对象是如何在这个抽象的超级计算机上做计算。

现在，假设我们已经有了一个能够自动根据所获得的资源（CPU、内存的数量，可计算的时间）的多少来自动决定如何做计算的程序。例如下面的 MPIHelloWorld.c 就是一个简单的这样的 MPI 程序。它根据获得的 CPU 的数量来决定打印几个“Hello World!”，而且还显示少量关于这个抽象的超级计算机的信息（第几个 CPU 总共多少个 CPU 构成这个超级计算机）。实际上在这个程序中，各个计算单元之间没有真正的数据交流，只有两个变量的数据（rank, size）以及一个真正的指令（print）。第一行建立一个抽象的超级计算机。接着获取这个超级计算机的 CPU 数量，以及每一个 CPU 的序号，加上这个 CPU 所在的主机的名字。然后，打印这些信息到屏幕上。

编译并行程序完全和串行程序一样，唯一的不同是编译器需要改成相应的 mpicc 或者 mpif77。mpicc 等实际上还是系统默认的编译器，只不过加上了一些额外的编译参以用来跟下面的执行程序 mpiexec 配合。在我们的例子中，我们只需要运行：`mpicc MPIHelloWorld.c -o MPIHelloWorld`。那么，我们如何执行这个程序呢？我们需要根据这个程序的需要以及这个池子的空闲资源来指定这个程序用到的的计算资源的数量。假设我们发现这个大池子当中有 4 个计算单元空闲，而且我们的程序无所谓需要多少个计算单元，那么，我们就可以利用如下的命令来执行我们的程序：`mpiexec -n 4 ./MPIHelloWorld`。典型的运行结果见下图。

这个命令的第一部分表示，后面的程序需要通过一个 mpiexec 的程序

来简介运行。这个 `mpiexec` 做的事情就是从资源池子之中划分出 $n = 4$ 个计算单元来建立一个超级计算机。然后，它在这个超级计算机中运行后面的程序 `MPIHelloWorld`。所以，其实，在这个 `mpiexec` 的后面由这很多其他的程序保证它能够完成这个任务。这个就是我们前面说的硬件、软件工程师和系统管理员完成的，我们不需要了解太多的事情。

以上这个情景就是一个并行计算的典型例子：管理员搭建好池子，程序员编号利用了 MPI 的完成一定任务的程序，程序员选择合适大小的资源来运行程序，程序运行之后给出结果。这个典型情形一般被称作分布式并行计算，或者 MPI 并行计算。在实际并行计算过程中，有可能程序员不能也不许要直接运行 `mpiexec` 来执行其所编辑与编译以后的程序，而是通过一个任务队列管理器来提交的。例如，程序员写一个关于所需要的资源的说明文件，然后这个说明文件会自动被任务队列管理器来阅读和安排什么时候执行。一个典型的这样的说明文件包含所要运行的程序的名字（必要的话还包含路径和所需要的参数），资源的数量（包含 CPU、内存数量，估计运行时间），如下图。那么这个任务队列管理器实际上做了什么事情呢？它寻找合适的机会，按照一定的优先级，把这个任务所需要的资源从池子里面留出来，然后发布一个 `mpiexec` 命令加上合适的资源数量和所执行的程序来组成一个完成这个任务的超级计算机。

这个是一般的情况，如果我们稍微了解一下这个池子的物理连接方式的话，有可能我们可以更好地利用这个抽象的超级计算机。一般而言，这个池子的连接方式有几种，单个计算单元单个存储单元（见示意图），通过主板相连的多个计算单元单个存储单元（见示意图），通过主板相连的多个计算单元多个存储单元（见示意图），通过网络（包含局域网和 internet）相连的多个计算单元多个存储单元（见示意图），其中最后的情况可以更复杂，例如通过网络相连的处理单元，每一个处理单元包含用主板相连的多个计算单元和多个存储单元。现在的多核 CPU 实际上使得第一种情况也变成了第二种情况。每一种情况的物理连接，我们都可以把它当作一个池子来使用。所以上面的一般情形在大多数情况下都是可以使用的。

但是，如果我们考虑物理连接的实际情况，有可能可以进一步提高计算的效率。例如，我们考虑一个通过网络相连的多个计算单元多个存储单元：8 台电脑，每一台配一个 8 核 CPU16G 内存。如果我们直接把这个系统当作拥有 64 个计算单元 126G 内存的池子，那么有的时候我们需要在本来同属于一台电脑的核心之间交流数据，而且这个数据的交流在我们的程序看

来，和不同的电脑之间交流数据是没有差别的。我们知道实际上这个是有差别的：在我们这个情况下，同一个电脑之内的不同核心之间内存是共享的，不用交换；而就算需要交换的话，不同电脑之间的交换速度要远远低于同一个电脑内部。因此，如果我们能够利用好这个体系的实际物理连接的情况，减少电脑之间的数据交流，增加电脑内部的数据交流，我们就可以提高计算的效率。这种在同一个电脑之内的并行与跨越电脑的并行不一样，称作多线程计算。对于我们的这个物理连接的情形，我们需要混合多线程和 MPI 并行：在同一台电脑上做多线程并行，在电脑之间做 MPI 并行。实际上，只要是共享内存的计算机，我们都可以用多线程来做并行。但是，大规模计算机大部分都是通过网络相连的，分布式内存的计算机。所以，MPI 并行是最普遍的情况。

小结一下，并行计算有多线程并行和 MPI 并行计算，多线程用于共享内存计算机，MPI 用于分布式内存计算机，有的时候，我们需要混合两种并行方式来提高效率，但是 MPI 并行总是可用的。在进行 MPI 并行之前，我们需要一个已经建立起来的资源的池子，然后我们从池子中获取一定量的资源组成一个抽象的超级计算机，然后在这个超级计算机中完成计算。这个获取资源的方式有的时候需要通过向任务队列管理器提需求来实现。有了对并行计算的这个一般认知，我们再来讨论几个实现并行线性代数计算的软件 `Petsc` 和 `Slepc`。

编写并行程序最主要的步骤是把算法当中的可以并行的部分并行，然后通过消息传递机制来实现数据交流。例如，在我们之前的矩阵和向量相乘的问题中，我们的两个 CPU 之间就要通过 MPI 来交流数据。这些并行计算线性代数软件实际上就帮我们完成了这两件事情，同时它们的算法本身也是基于数值线性代数的已经被证明的优秀算法的例如 `Lapack` 和 `BLAS` 当中包含的一些算法。在我们之用这些软件的时候，我们甚至不需要直接来面对 MPI 数据交流的问题，仅仅需要一个 MPI 数据交流的概念。也就是说，了解在哪些时候，这样的交流会发生，就可以了，而不需要我们自己来写出 MPI 语句。目前，也有研究人员用多线程和 MPI 并行混合的模式改造数值线性代数软件包 `Petsc`，而且已经取得一定的效率上的提高。易用性和稳定性上还有待于进一步提高。

16.5 xmds2

16.6 sage

16.7 NetLogo

16.8 Swarm

16.9 C 与 R 中的 igraph

第十七章 概念地图与理解型学习方法

学习的目的是为了创造知识和创造性地使用知识。重复性的平庸的使用知识——指的是其他人已经创造性地展示了某知识如何用于解决某问题，而学习者的任务就是学会这个运用——实际上不需要经过历经大学、研究生院、导师这样一个研究性学习的过程来实现。其最好的培养方式是培训，技能型的学习，例如通过职业教育，或者工作以后就接受岗位培训，甚至直接通过在工作实践中的摸索得来。我这里的学习方法，尽管原则上还是可以在技能型学习中使用的，不包括这样的重复性的使用知识的学习。尽管人类财富的创造需要这样的重复性的运用，但是人类文明的进步是由创造知识和创造性地使用知识来推动的。

我们知道，为了创造知识和创造性地实用知识，最关键的是需要理解知识，内化知识。一个学习者、研究者，只有把知识变成自己内在的能够自发运用的工具，才能以此为基础真正创造新的知识，或者把这个知识运用到其他人没有用过的地方，来解决现实世界的问题。

名词索引

质点 只包含质量一个属性的运动物体的一种抽象. 37

人名与常用翻译

Newton 牛顿. 70

插图

9.1 光子 which-way 实验装置示意图 109

举例目录

5.2.1 例 (\mathbb{R} 实数加法群)	43
5.2.2 例 ($\mathbb{R}/\{0\}$ 实数乘法群)	43
5.2.3 例	43
5.2.4 例	46
5.2.5 例 (N 维欧氏空间矢量)	46
5.2.6 例 (2 维欧氏空间可微函数在某点的切线)	46
5.2.7 例 (1 维欧氏空间的无穷可微函数)	46
5.2.8 例 (1 维欧氏空间的平方可积函数)	47
6.1.1 例	62
6.1.2 例	62
6.1.3 例	62
6.3.1 例	65
6.4.1 例	65
6.5.1 例	67
6.6.1 例	69
6.7.1 例	69
6.7.2 例	70
6.7.3 例	72
6.7.4 例	73
7.2.1 例	82
7.2.2 例	82
7.2.3 例	83
7.3.1 例	84

7.3.2 例	85
7.3.3 例	85
7.4.1 例	86
7.4.2 例	87
7.4.3 例	90
7.4.4 例	90
8.2.1 例	99
8.2.2 例	100
8.2.3 例	100
8.4.1 例	101
8.4.2 例	102
9.2.1 例	115
9.4.1 例	118
9.4.2 例	120
10.1.1 例	129
10.1.2 例	129
11.2.1 例	138
11.3.1 例	140
11.3.2 例	141
12.1.1 例	144
12.1.2 例	147
12.1.3 例	148
12.1.4 例	149
12.1.5 例	149