第一章 量子力学的理论框架

在前言部分，笔者提到自己受朱守华老师在《电动力学》教学团队采访过程中“课堂教学实际上是为了让学生与老师之间达成共识”这句话的影响，在2026年5月重新整理了本章。其中，主要的改变就是避免了大量之前计划的关于量子力学背后的哲学基础的讨论。修改后，我们的出发点就是纯纯的物理内容。这样做的目的，是用一种让“我们物理学专业的学生感觉踏实的方式”给出量子力学的理论框架，为后续讨论提供一个语言上我们能够达成共识的基础。

既然是纯纯的物理，那大量的公式推导就无法避免了。按笔者对我们的学生的了解（当然也包括笔者自己），大家是喜欢这种方式的。但同时，笔者也会强调我们这里并不会一上来就*“通过数学公式来讲量子力学的理论架构”*。在笔者看来，这样做的问题很明显，主要有两个。第一，是会让读者不知道为什么要建立这个数学上的架构？第二，是为了把理论说的尽量严格，在说这些理论架构的时候，总是要做各种补充讨论[1]。这就使得我们的理论不是那么的简洁、漂亮。作为一个结果，在课程结束后，第一点会让*我们的学生不理解我们为什么要把理论框架设置成这样*？第二点会让我们的学生*不容易记住这个框架*[2]。教学效果，自然就不会太好。笔者当年也是带着这些困惑从学生时代一路走过来的。

在这两点中，关于第二点，我们在这里要做的改进比较简单，就是在写作的过程中以启发性而非严格性为主要考虑。不是太强调严格的话，理论本身很容易

[1]说话严格是让人痛苦的。比起严格，笔者确实更有启发性与有趣。当然，我们不能离谱！这也是一个“度”的问题。

[2]实际上，不理解和记不住也是伴随的。记不住的原因很大程度上还是不理解，这个实际上是第一点。在第二点，我们强调的记不住的另一部分的原因，即“理论太复杂，不简洁，因此记不住”。

简洁，学完也好记住。针对第一点问题（关键点在理解），我们需要进行的操作就比较多了。具体而言，就是我们会花比较多的精力来讲述在量子力学发展过程中人们是怎么意识到要基于线性代数这个数学工具来建立量子力学的理论框架的？这里，我们需要抓住两个关键词：*稳定性问题*、*运动学*。我们要强调的数学工具是《线性代数》，这是背后的数学。在第四章中，我们还会强调《群论》。《线性代数》和《群论》合在一起，构成了量子力学理论框架中所有的数学。本书中，我们选择逐一突破！在第一章，先讲《线性代数》，从*稳定性问题*和*运动学*两个关键词出发来讲为什么要用线性代数？第四章，我们会从《群论》的角度把这个理论的数学结构进行一个重建。

先说稳定性问题。我们首先要明确的是在历史上，也就是上世纪初，物理学关心的核心问题就是“*原子中的电子运动为什么稳定？*”[3]。在那个时期，我们熟悉的那些物理学家们也正是通过对类似这样的稳定性问题的研究建立量子力学理论框架的。在这个理论中，线性齐次方程很关键，它被用以描述稳定性问题。这也明确指出了线性代数这样一个量子力学背后的数学工具。当然，在量子力学的理论架构建立之后，人们意识到它不光可以描述稳定性问题，也可以描述非稳定性问题。同时，实际计算中，我们还可以针对各种具体情况进行方法的进一步发展，比如二次量子化、多粒子张量积空间、开放系统。这些，都是我们《高等量子力学》课程要学习的重要内容。但是，在课程开始的时候，在理解量子力学理论架构的时候，我们还是要抓住“稳定性问题”这个最根本的关键词，而不是

[3]黑格尔讲过哲学就是哲学史，实际上物理学也是物理学史。我们在学习物理学的过程中，一定要充分地理由这部分内容来理解我们的学科。关于这一点，笔者在文献【李新征，2026】中强调过很多次。

做过多展开。这样，我们会比较容易理解我们的前辈在上世纪二十年代中后期为什么要进行那样的量子力学理论架构设置？

此外，回到“物理学理论而非仅仅是量子力学理论这个层面”略作扩展，如果我们回顾更远一些的历史（比如到科学革命初期），我们也会发现不光量子力学是这样的（从稳定性问题出发发现理论形式），经典力学同样是这样的。在经典力学的发展过程中，牛顿就是在针对行星轨道的理论描述这一经典力学中的稳定性问题的研究中建立了牛顿力学的理论框架的。这个时间节点是 1687 年。之后，整个 18、19 世纪，以法国为中心的数学家、物理学家对其进行的发展就是把牛顿力学应用至更加复杂的力学体系，包括考虑耗散。直到二十世纪初，朗之万这样的物理学家还在关注类似稳定性问题之后的拓展。因此，就像我们在学习经典力学的时候可以先从稳定性问题出发来接受这个力学理论，然后再去想耗散的事情[4]。在学习量子力学的时候，我们也需要先从稳定性问题出发来理解这个理论，之后再去理解更复杂的情况。这里和之前不同的地方，是在讲《经典力学》的时候我们不需要强调稳定性就可以很容易的理解牛顿三大定律了，但在《量子力学》的学习过程中，如果不强调稳定性，人们就很难理解为什么要在五条公设中做那些设置？仅此而已。

“稳定性问题”之后，第二个关键词是“运动学（Kinematics）”[5]。Kinematics

[4]行星运动是个摩擦可忽略的力学问题，会帮助我们发现基本规律（牛顿三大定律以及万有引力定律）。

[5]在我们的教育中，一般一说运动就是力学。因为我们的物理学习运动就是从初中的力学开始的。但实际上，在描述力与运动的时候，可以有很多词，比如静力学（statics，重点是静）、动理学（kinetics，也翻译为动力学，重点是动）、动力学（dynamics，重点是动因，dyn 强调动因）、运动学（kinematics，这个就是强调运动的描述本身，不强调力）。当然，还有 mechanics（这个词根是 machine，和机械有关），它是我们在讲到力学时更愿意用到的总称。

代表的实际上就是我们脑子里对运动的最直接、最本真的描述，比如托勒密的地心说模型、哥白尼的日心说模型。它们的共同点，都是人们在脑子里要通过几何的方法构建出的星体的运动轨迹。那个时候（直到 17 世纪初之前），成熟的数学只有简单的几何、简单的代数(初中和初中之前的数学就可以搞定,会画图即可)。换句话讲，科学革命前的物理学使用的都是*传统的几何学中的轨道*的。

在牛顿力学中，力的定量描述是需要以解析几何（基于代数描述几何）为基础的微积分（在解析几何的基础上引入微积分）的。相对于牛顿之前的行星“运动学”（只有轨迹，没有力）[6]，在牛顿力学中，物理学理论中的“运动学内容”实际上已经有了一个形式上的根本的改变（增加了基于微积分的 mechanics 的内容)。这个改变促成了近代物理学的诞生。但同时，我们也需要意识到这里所使用的依然是*依赖于直观的几何*。和牛顿力学密切相关的哲学思想就是机械论，也是比较直观的哲学理论[7]。这些概念和我们的生活经验极其符合。因此，我们学习这些经典力学的概念的时候都不会有什么问题。

量子力学最反常识的地方就在于它对粒子运动的描述中没有轨道的概念,这也是我们在这里需要强调“运动学”这样一个关键词的原因。我们量子力学的很多特色，实际上都来源于此。说到底，这和我们物理学关心的核心问题，即物质

[6]这里，大家请把自己定位到那个时间节点，企图通过自然哲学来理解自然界的人。针对笔者说的当时没有力的这句话，可能有读者会说亚里士多德的哲学里会通过“自然位置”引入“力”。这个笔者当然认同，笔者在这里强调的是我们所熟悉的那种“力”，牛顿之后人们心中的“力”。这个，在早期的宇宙学的运动学模型中是不存在的。

[7]在热力学、统计物理、量子力学这些物理学理论提出后，人们逐渐摒弃了机械论。

的存在形式与运动规律相关[8]。这里，考虑到任何事物都是以运动的形式存在的[9]，这个存在形式里面已经包含了物质是怎么运动的内容。而运动规律，即包含物质存在时的运动规律，也包含存在形式改变（与之相应会有运动状态的改变）时，我们的物理系统需要尊重的规律。

在量子力学诞生之前，人们不需要用量子态、物质波这样的概念去描述微观世界的物质存在形式。这实际上也是海森堡、德布罗意这样的人厉害的地方。他们对物理学的贡献是在“怎么描述运动”这个层面的。这都直接带出了《线性代数》这个基本的理论工具。在笔者作为学生学习《量子力学》、《高等量子力学》的时候，虽然能理解量子态、物质波这些概念并可以完成一些解题练习与应用，但从来没有注意到 kinematics 这个词，也没有理解前面那几段讨论的内容。因此，对这量子态、物质波这些概念的理解是受限的、不系统的。2021 年春，笔者开始讲授《今日物理》课程，并需要把量子力学是怎么产生的这个问题说清楚。在这个过程中，笔者认识到这个关键词对于我们理解量子力学的重要性[10]。这几年，在笔者课题组针对量子多体系统的科研实践中，笔者也有不少体会[11]。因此，我

[8]在经典物理学中，这个就是物理学的全部。在现代物理学（基于量子力学、相对论）中，还有包含“观察者如何感知”的内容。后面我们讲五条公设的时候还会详细说明。更多哲学上的解释，请参考【李新征，2026】。

[9]在很小的时候学哲学，应该都学过“这个世界是建立在运动上的”这样的观点。这个观点是被很多经验所证实的，是牢靠的。

[10]这里笔者首先看的是矩阵力学的原始文章【Heisenberg，1925】，它的题目就是“On the quantum-theoretical reinterpretation of kinematical and mechanical relationships”。之后，笔者大量参考了陈童老师的《初始量子力学》与《量子力学新讲》、Van der Waerden 在 1968 年整理出版的文集【Van der Waerden，1968】，进而整理出了文献【李新征，2026】的第八章。本章的内容是在文献【李新征，2026】的第八章的内容的基础上进一步整理的。

[11]比如，近期笔者和学生朱禹丞等人一起完成的关于超导体系中与软声子相关的物理【朱禹丞，2026】，我们的理解的关键点就是晶格的“量子多体”的“运动学”。

们在这本讲义中也对其进行强调。在笔者看来，这个或许也是我们现有《量子力学》教材普遍漏掉的一颗遗珠。

基于如上考虑，我们从“稳定性问题”、“量子力学的新的运动学内容”两个角度出发来介绍量子力学的理论框架。其中，理解“量子世界的新的运动学”是人们在历史上认识到要构建完全不同于经典理论的“量子力学”的关键。这里，大家可以把关注点放在“如何描述一个微观粒子的运动”上。而对“稳定性问题”的理解，则是我们构建这个量子力学理论体系的数学语言的抓手，它是上世纪20年代的物理学关心的问题。因为我们考虑的是稳定性问题，我们需要把力学量（严格意义上应该叫可观测量，后面会解释）处理为厄米算符[12]。因为我们考虑的是稳定性问题，我们的运动方程是线性齐次方程。这些要求带来的后果是量子力学描述的微观运动的空间是希尔伯特空间，描述运动的态是其中的矢量，而这些矢量之间是需要满足线性叠加原理的。通过这些基础性的介绍，我们希望先能够建立这样一个概念，就是微观粒子的运动学描述与经典粒子完全不同。之后，借助线性齐次方程，我们可以提出希尔伯特空间这样的概念来描述微观粒子的运动。这是我们这个理论的数学结构。这个数学结构明确之后，很多计算和理论推导方面的操作就很容易执行了。

在思路明确后，下面的重点就是本章的各节应该怎样设置了。

[12]厄米算符大家可以简单理解为“实数”的“矩阵对应”。在量子力学的语言中，由于量子态是希尔伯特空间中的矢量，可观测量必须是矩阵。但是为了稳定，这个矩阵必须是厄米矩阵。这里，大家可以用一个谐振子来理解。在经典力学中，它的位置、速度，都可以用$e^{-i\omega t}$来描述。ω是实数保证了稳定振荡。对于一个量子力学系统，这个ω会被哈密顿量$\hat{H}$取代。我们对$\hat{H}$的要求既是它是一个厄米矩阵。换句话说，本征值都为实数。这样做，也是为了保证稳定振动。只是在这个振动中，位置、速度相对于经典系统，不能同时确定了。

首先是*第一节*。本着从“强调量子力学中的新的运动学内容”出发来引入量子力学理论框架的思路，我们在本章的第一节会借助*海森堡提出并由玻恩和约当完善的矩阵力学*来解释在上世纪二十年代中期人们是如何意识到要建立目前这个量子力学理论框架的。这样做的好处，是有了这些公式推导的积累，按照我们中国学生学习的习惯[13]，我们会觉得心里更有底一些。同时，通过彻底重复海森堡当初的创新性的想法以及玻恩和约当的整理，我们也可以更容易理解人们在这里说的“运动学新内容”是怎么先通过矩阵建立最早的量子力学的[14]？由于矩阵和波函数在线性代数的语言中的等价性，在理解这些后，我们就很容易接受量子力学的数学工具是线性代数这个关键点了。

之后，在*第二节*，笔者想强调“虽然量子力学的 *kinematics* 和经典力学非常不同，它的 *mechanics* 和经典力学还是很相似的”。因此，我们在理解量子力学的过程中也一定不要把量子力学和经典力学撕裂。这样做的好处，是能够让学生理解物理学理论的时候保持一个整体性的考虑。实际上，对应性原理的主旨就是这一点[15]。我们会通过薛定谔方程的推导来针对这一点进行强调。在这个推导过程中，除了电子运动的波函数表述（来自德布罗意），貌似仅仅依赖于经典力学，

[13]这个真的就是我们的习惯，我也有这个习惯。这个习惯本身没有好或者不好，这是个客观存在。

[14]矩阵力学是最早的量子力学的形式，这一点大家应该都会承认。

[15]对应性原理也被称为对应性关系。它最初的思想是由玻尔在 1913 年在他的玻尔模型中表达的，即在量子数很大的时候，他的原子论回归经典力学。1918 年，在他的文章 On the Quantum Theory of Line-Spectra 中，他把这一点明确提出。这个原理在量子论发展期间（1923 年之前），是很多新工作的指导思想。在量子力学的理论架构中，它并没有作为一个基础被强调。但是这个架构形成的量子力学符合经典极限回归经典力学的特征。这时，可观测量（注意，在量子力学中，我们最好不要叫力量学，叫可观测量，因为量子力学的所有思考都是建立在观测上的）变为经典力学中的传统力学量。

我们也可以推出薛定谔方程。从这个角度来想，大家应该也会理解为什么后期在“关于波函数的概率性”的辩论的时候，薛定谔是属于相对保守的那一派了。一定程度上，我们也可以说虽然他推出了薛定谔方程，但在波动力学发展的过程中，他的贡献更多地体现在物理学理论的数学形式的[16]。革命性的思想，更多地是来自于德布罗意的。

有了“基于矩阵力学与物质波理论对量子力学新的运动学内容的强调”，有了“薛定谔方程的推导对量子力学与经典力学内在联系的强调”，*在第三节*，我们就可以基于“稳定性问题”、“线性齐次方程”、“希尔伯特空间”来讲解量子力学的理论体系。在这个理想的体系中，大家可以很容易理解*为什么我们会有我们常讲的五条公设？*[17]同时，我们也要说明就像牛顿基于行星运动这个稳定性问题建立了经典力学但是经典力学的应用场景可以扩展至所有的经典系统，这里，虽然我们微扰稳定性问题建立了量子力学的理论框架，它的应用场景也适合于所有的情况下的微观粒子的力学描述的。当然，该补充的地方，在本章后半部分以及后面几章我们还是要补充。

基于如上考虑，本章按三节展开：1）量子力学的新的“运动学”内容（这是从区别于经典力学的角度理解量子力学）；2）量子力学与经典力学之间的联系（这是从与经典力学的相似性的角度理解量子力学）；3）基于稳定性问题理解量子力学的理论框架（这是借助稳定性问题，构建最简单的量子力学理论框架，方

[16]思想上比较活跃的，比较符合现代我们对量子力学的认识的，是玻恩、玻尔、海森堡、德布罗意这些人。泡利实际上也是一个保守的代表，这也是他一次次错过重要发现或者在重要发现的过程中起了负面作用的一个原因。

[17]虽然五条公设的内容在不同的地方会有不同的版本，也没有谁说量子力学一定要按五条公设来讲，但从五条公设出发，确实比较方便建立理论架构。

法理解与记忆)。目的,是给大家建立一个比较容易理解且能够记住的理论框架,作为后续讨论的基础。

1.1 量子力学的新的“运动学”内容

经典力学背后的数学是微积分，量子力学背后的数学是线性代数[18]。此外，经典力学是建立在实数域上的，量子力学是建立在复数域上的。相信这些话，多数同学听说过,并且也会认可。这是量子力学与经典力学的数学工具的根本差别。其背后的原因,在物理上就是前面提到的:对一个微观粒子和一个宏观粒子而言，我们关于其运动的描述必须是不一样的。换句话说，粒子的“运动学”本身存在根本区别。

为了让读者更好的理解这一点，我们在本节就以海森堡提出后经玻恩、约旦总结的“矩阵力学”这个历史上的具体工作为例，来带大家看看当时的海森堡是做了什么具体的处理，进而让物理学完成了这个至关重要的进步的。在这个进步中,“矩阵”扮演了最重要的角色，它是一个线性空间在确定一组基后力学量（在量子力学应该叫可观测量）的表现形式，也是海森堡描述一个量子系统的新的运动学内容时所使用的数学工具。这个,我们可以从海森堡的一人文章的题目看出。这个文章的题目叫“Uber quantentheoretische Umdeutung kinematischer und mechaniseher Beziehungen”， 英文对应“On the quantum reinterpretation of kinematical and mechanical relationships”。

这篇文章的核心意思，就是从“quantum（量子)”的理论的角度，对微观粒

[18]至少在初步理解的层面，需要线性代数，深入理解还有对称性的内容。

子的“kinematical and mechanical relationships（运动学与力学关系）”进行一个“reinterpretation（重新解读）”。相对于之前的 quantum theory（量子论），它意识到了 kinematics 的关键不同，进而实现了玻恩在 1924 年提出的将 quantum theory 转变为 quantum machanics 的历史任务【Born，1924】[19]。这里，海森堡提倡应该抛弃玻尔模型中电子*经典轨道*的概念。由于两个定态之间的能量差、跃迁几率是物理学上的“可观测量（observable）”，他把坐标、动量这些经典力学理论中的物理量基于玻尔提出的“定态”与“定态跃迁”的概念进行了重新描述[20]。基于“可观测量”来构建量子力学，是量子力学相对于经典的物理学理论的最大特征。后面，我们会反复强调可观测量这个概念在量子力学中的重要性。

具体而言，海森堡提出的重新表达就是*用厄米矩阵取代实数，来表达一个物理学量*。在经典物理学里面，任何物理学量都是一个实数，它天生就是可以被观测到的。但是在量子力学中，任何物理学量都是一个矩阵，它们被统称为可观测量（observable）。可观测的意思，是它们的本征值可以被测量，相应的本征态也有物理学的信息。但这个矩阵本身，在不同的基组下可以有不同的表达。不管是哪个基组，只改变矩阵形式，不改变本征值。这些本征值相应的本征态的信息不

[19]海森堡当时是玻恩的博士后。1924 年，玻恩在 Zeitschrift für Physik 上以 Über Quantenmechanik（关于量子力学）为题提交了一篇讨论量子理论的论文【Born，1924】，首次提出了“量子力学”而非“量子理论”的概念，吹响了人们将量子论转化为量子力学的号角。这也是人类历史上第一次出现“quantum mechanics”这个关键词。考虑到海森堡的一人文章后，是玻恩第一个意识到他使用的工具是线性代数，并和约当、海森堡一起完成了两人文章与三人文章，我们可以说海森堡的这个突破绝非偶然，而是物理学发展到那个阶段，人们对问题定位到一个特定高度后的必然结果。

[20]这段话里面其实也有康德哲学中“物自体”和“现象界”这两个概念的影子。但在我们的课上，不适合做过多展开。感兴趣的同学可以参考文献【李新征，2026】来进行理解。我们在那里，想强调的就是量子力学、相对论这种现代物理学相比于物理学革命之前的物理学体现的就是康德哲学的内容。他们都强调观测者在物理学理论中的地位。这在经典物理学中，是可以被忽略的。

变，但其列向量（在用右矢空间表达的话）是会随基组的选择变化的。只是它们组合后的结果不变。这些内容，我们都会在后面的五条公设中详细解释。

同时，因为是矩阵，这些可观测量之间就可以不对易了。它们对易的情况是它们的测量不影响对方，但多数情况下，可观测量之间是不对易的。比如，位置和动量这两个可观测量的算符不确定，这就是著名的不对易关系（具体推导见1.3.3 小节）。在本章，这些内容我们都会涉及。现在，我们从矩阵力学这个多数教科书中不太介绍的语言出发，来说明这个问题。目的，是强调矩阵在量子力学中作为“数学语言”的重要性。我们重点参考的文献是《初始量子力学》与《量子力学新讲》，均来自东华理工大学的陈童老师。笔者很感激这些讲义帮自己解困扰多年的惑！[21]

我们考虑的问题是一个简谐势中的电子的运动（图 1.1）。按照海森堡的语言，在经过玻恩、约当整理后，位置在这个表述下就不是一个数，而是一个算符了。这个算符可以用矩阵表示，其矩阵元的具体形式是：

$$[\hat{x}]_{nm} = [\hat{x}_0]_{nm} e^{-i\omega_{nm}t}$$

公式 1.1

其中，$[\hat{x}_0]_{nm}$为每个矩阵元振动的振幅[22]。这里，ω_{nm}乘上$\hbar$对应的就是图 1.1b 中水平红线的能量差（$\hbar\omega_{nm} = E_n - E_m$）。在 1925 年海森堡的一人文章发表的时间节点，爱因斯坦关于定态跃迁与自发辐射、受激辐射之间的关系的那篇奠定激光原理的文章已经在 1917 年发表【Einstein，1917】。因此，海森堡知道公式 1.1 中的跃迁是既可以从低能级到高能级，也可以从高能级到低能级的。换句话说，

[21]笔者总结后，将这些内容在文献【李新征，2026】中已经进行了呈现。在我们这个讲义中，我们进行了一定程度的二次整理，但改的地方不多。

[22]但这个振幅大家不要认为是实数，量子力学至始至终都是用复数来表达的力学理论。

ω_{nm}的两个下标n、m对应的就是他说的定态（对应到稳定的原子中，就是玻尔说的那个电子轨道对应的定态，只是这里海森堡摒弃了轨道的概念）。并没有大小之分。

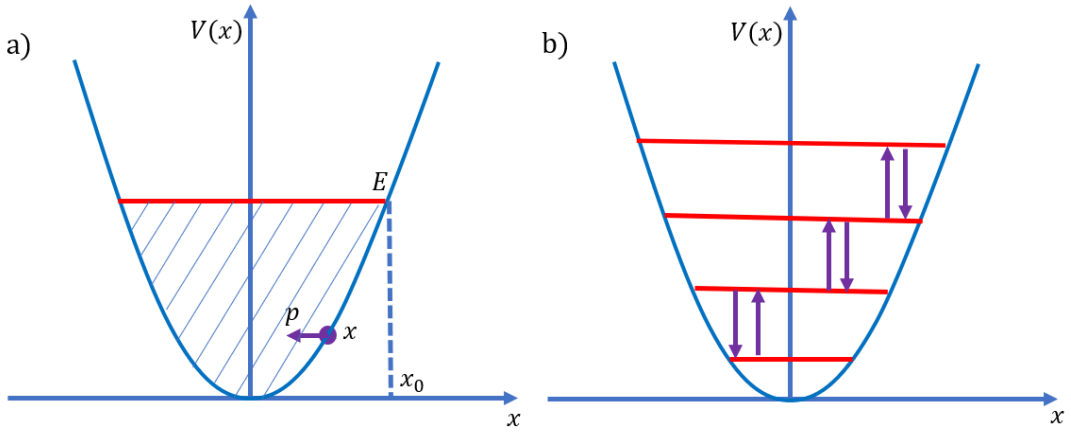


图 1.1 简谐势阱中带电例子运动示意图。这里，我们强调的是经典世界与量子世界的“运动学”的不同。简谐势的弹簧是同一个弹簧，因此势能函数$V(x)$相同。在经典的世界中（左边图 a），我可以任意的把弹簧拉到要给特定的位置x_0。这个时候动能为零，势能为E，系统总能量也是E。松手后，粒子开始做简谐振荡，动能势能不停转换，红线一下的阴影区域都是粒子可以触及的区域。粒子运动（the kinematics of the particle），是可以通过这个图像（左边图 a）来描述的。在量子的世界中（右边图 b），弹簧的弹性系数只决定能级与激发。唯一的可观测量是系统在不同定态（本征态，红色水平线）之间的跃迁，这又紫色箭头标记。海森堡正是认识到了量子世界与经典世界这个不同，才开始将任何可观测量（包括坐标、动量）基于跃迁来描述（紫色箭头），进而完成了 1925 年那篇最重要的工作。这里，跃迁发生在相邻能级之间，这是简谐振子的特征。在学完产生、湮灭算符之后，这个很好理解。在非谐势中，这个限制就不再存在了。

同时，海森堡还注意到由于$e^{-i\omega_{nm}t} = \left[e^{-i\omega_{mn}t}\right]^*$。这样的话，他可以进一步规定：

$$[\hat{x}_0]_{nm} = [\hat{x}_0]^*_{mn}$$

公式 1.2

进而，$[\hat{x}]$这个矩阵的矩阵元就有如下性质：

$$[\hat{x}]_{nm} = [\hat{x}]^*_{mn}$$

公式 1.3

这个性质，定义的就是厄米矩阵（矩阵的转置共轭等于其本身）。*我们可以把它当做实数的矩阵对应来看待即可*。换句话讲，海森堡通过矩阵（他后来才知道这是矩阵），对简谐势中的电子运动中的电子位置进行的运动学表达。表达完之后，这个位置不再是传统意义上的“位置”本身，而是一个“可观测量”。只有在其本征态（位置算符的本征态），电子才具有确定的位置。在其它可观测量算符的本征态，位置就是一个分布。

除了位置，动量也可以表现为一个算符，其矩阵元形式为：

$$[\hat{p}]_{nm} = [\hat{p}_0]_{nm} e^{-i\omega_{nm}t}$$

公式 1.4

这里，类比公式 1.1，$[\hat{p}_0]_{nm}$是动量算符的每个矩阵元的振幅。而且，$[\hat{p}_0]_{nm}$、$[\hat{p}]_{nm}$也具有类似公式 1.2、公式 1.3 的性质。在量子力学中，所有的可观测量都是类似的厄米算符[23]。下面我们要做的，就是基于公式 1.1-1.4 对应的新的量子的“运动学描述”，再根据对应性原理套用一个经典力学中的运动方程进行量子对应，这套*将传统的“力学量”基于“可观测量”进行新的运动学表达*（Kinematical

[23]前面我们强调过，在上世纪二十年代，量子力学诞生的时候，人们关注的核心问题是原子中电子轨道的稳定性问题（加速运动不会辐射能量进而让电子掉到原子核上）。为了保持稳定，公式 1.1 到 1.4 中指数函数的指数需要是实数（振荡不衰减），可观测量（对应经典力学中的力学量）对应的矩阵也需要是厄米矩阵（本征值为实数）。但如果我们在量子力学中关心的问题从稳定的量子态推广至有产生与湮灭的情况，就像后面要讲的量子多体系统。那这个系统中描述局域的准粒子激发的自能（self-energy）就是非厄米的。它的实部对应准粒子的激发能，虚部的倒数对应这个准粒子的寿命。这个是上世纪中叶，量子场论提出的时候，人们已经知道的东西。这几年，非厄米的物理学忽然变成了一个热门课题，对应的也是这里说的情况。因此，笔者的建议是“力学量（严格来讲，应该说是可观测量）在量子力学里面要表达为厄米算符”这句话大家可以在学习《量子力学》的初期先简单粗暴地接受，方便理解很多习题。但同时，我们一定也要知道这个仅仅对应某个稳定的量子态。非厄米的可观测量作用到量子态上之后，该量子态会衰减或者增强。换句话讲，它会“不稳定”。

Description）的力学理论（Mechanical Theory）就成功了[24]。

应该说，在量子论中无法解决但可以被矩阵力学（其三篇文章）解决的问题有不少。也正是因为这些成功，才使得人们可以说量子力学在这一刻诞生。从阅读友善的角度，我们基于图 1.1 中的谐振子，选择比较有代表性的两个进行逐一攻破：1）关于定态为什么会稳定（在玻尔那里，这是一个假设）；2）一个量子的系统的能量为什么会分立（在普朗克那里，这也是一个假设）。我们可以说：*当量子论中的假设可以作为这个基于矩阵的力学理论的结论被推出来的时候，量子力学就建立了。*

我们可以先看“定态稳定问题”。这个问题只需利用对应性原理，将公式 1.1 代入经典电动力学的电磁辐射方程即可解决[25]。在经典的电动力学理论中，我们知道一个加速运动的电子会辐射电磁波。这样，我们无法在玻尔的原子模型中将电子轨道理解为电子绕原子核的圆周运动。玻尔提出了“定态假设”在一定程度上解决了这个问题，但那就是一个“假设”。在矩阵力学中，定态会成为“结果”而非“假设”。进而，量子论得以完成向量子力学的进化。

---

[24]这里，我们需要强调两点。第一，就是*量子力学中没有力学量，只后可观测量*。很多教科书会按照经典力学的习惯用力学量这个词（包括笔者刚出的一本书【李新征，2026】，很抱歉笔者在那个时候还没有意识到这一点），这是有误导性的！在量子力学的理论体系中，可观测量有明确的物理意义。这个物理意义在经典力学没有对易，我们必须按量子力学的方式来理解。关于这一点，我们在学习量子力学过程中需要不停地体会。第二，就是虽然可观测量与经典物理学中的力学量物理意义完全不同，但关于*可观测量之间的关系的形式*，往往符合经典物理学中力学量之间的那个形式。这也是我们常说的对易性原理的内容。比如，在 1.1 节后面的推导中，我们会强调只要把经典力学中用实数表达的“力学量”换成量子力学中用矩阵表达的“可观测量（有本征值，本质上可以被观测）”，矩阵力学就建立了。这个大家很快就可以看到。

[25]在量子力学的理论架构建立后，对应性原理可以被当作结论推出（量子世界在经典极限下回到经典世界）。但是目前这个阶段，量子力学的理论架构还没有被建立。这时，对应性原理的作用是帮我们来理解量子力学的理论架构。这个在逻辑上，有点像先有鸡还是先有蛋的问题。在这个地方，既然我们还没有量子力学的理论架构，我们就从对应性原理出发看量子力学的理论架构为什么应该这样？

具体而言，这是因为当一个电子加速运动时，借鉴经典电磁学理论中的公式，把电磁场辐射的功率P和电子位置x都处理成算符，我们有：

$$\hat{P} = \frac{e^2}{3\pi\epsilon_0 c^3}|\ddot{\hat{x}}|^2$$

公式 1.5

这里，我们出于简单的考虑使用一维系统，$|\ddot{x}|$维电子的加速度，前面的系数都是数学常数或物理学常数。这样的话，在海森堡所建立的运动学表述中，将公式 1.1 代入公式 1.5，我们就可以得到：

$$[\hat{P}]_{nm} = \frac{e^2\omega_{nm}^4}{3\pi\epsilon_0 c^3}|[\hat{x}\hat{x}]_{nm}|$$

公式 1.6

从公式 1.6 我们可以看出，对于$n \to n$这样的跃迁，其辐射功率由于$\omega_{nn} = (E_n - E_n)/\hbar = 0$，自然为零。这样，定*态稳定性的问题就解决了*。可以说，这个问题在这里被解决的方式简单得甚至有点让人无法置信。

下一步我们需要做的，是要理解电子是如何从一个定态到另一个定态进行跃迁进而产生原子光谱？历史上，矩阵力学在氢原子中的应用，是由泡利在 1926 年完成的【Pauli, 1926】，这是在矩阵力学的三篇文章之后【Heisenberg, 1925】【Born, 1925】【Born, 1926a】。这个工作让人们彻底接受了矩阵力学。在海森堡的文章中，他处理的是如图 1.1 那样的简单力学系统（简谐振子以及简单的非谐模型）。考虑到原子只是数学形式更复杂，其物理思想在我们前面讲的简谐振子体系基本都已经包括，在理解“电子是如何从一个定态往另一个定态进行跃迁进而产生原子光谱”，我们将注意力集中在图 1.1 中的能量为什么可以分立即可。这里，能级分裂是等间距的。*为了得到这个结论，在逻辑上，我们需要走三步。这三步是笔者从东华理工大学陈童老师在网上分享的讲义《初识量子力学》的第一章《矩阵革命》*

中看到的。在征得陈老师同意后，我们在下面分享一下。在已出版的文献【李新征，2026】中，笔者也已展示了这些步骤。

我们需要先得到可观测量矩阵的运动方程，再建立位置、动量的不对易关系。之后，利用这个不对易关系，我们可以得到谐振子的分立的能量。需要指出的是矩阵运动方程和不对易关系在海森堡的一人文章中并没有出现，它们都是在玻恩、约当的两人文章中被第一次提到的。下面，我们就按照这个逻辑，分 1）矩阵运动方程、2）不对易关系、3）谐振子能量量子化这三步来解释量子力学建立过程中的这个关键成就。

先看运动方程。海森堡在他的一人文章中引入了一个“用两个指标来描述位置、动量，进而进行乘法运算”的计算规则【Heisenberg，1925】。那两个指标，对应的都是定态。很快，玻恩就意识到海森堡所建立的乘法规则也是矩阵乘法规则，这在十九世纪中叶就已经出现了[26]。这样的话，参考公式 1.1 与公式 1.4，哈密顿量同样可以进行下面这样一个表达：

$$[\hat{H}]_{nm} = [\hat{H}_0]_{nm} e^{-i\omega_{nm}t}$$

公式 1.7

写成矩阵的形式的话，它就等于：

$$\hat{H} = \begin{bmatrix} [\hat{H}_0]_{11} & [\hat{H}_0]_{12} e^{-i\omega_{12}t} & \cdots \\ [\hat{H}_0]_{21} e^{-i\omega_{21}t} & [\hat{H}_0]_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

[26]在文献【Born，1978】的第 217－218 页，玻恩有这样一段描述：“After having sent Heisenberg’s paper to the Zeitschrift für Physik for publication , I began to ponder about his symbolic multiplication, and was soon so involved in it… For I felt there was something fundamental behind it… And one morning, about 10 July 1925, I suddenly saw the light: Heisenberg’s symbolic multiplication was nothing but the matrix calculus, well known to me since my student days from the lectures of Rosanes in Breslau”。这里的 Rosanes 是雅可比·罗萨内斯（Jakob Rosanes，1842－1922 年）。他是玻恩在 Breslau（布雷斯劳）读大学前两年的课程的时候遇到的数学教授，当时玻恩选了他关于线性代数的课。

公式 1.8

这里对角项里面没有e指数是因为$\omega_{nn}=(E_n-E_n)/\hbar=0$，进而$e$的指数项等于一。非对角项按照与公式 1.1（位置）、公式 1.4（动量）类比得到的公式 1.7（哈密顿量），是有e指数的形式和时间依赖的。

下面，我们要使用到一个从经典力学中的封闭力学系统直接拿过来的假设：*能量守恒*。这个假设在经典力学中是以一个定律的形式出现的，来自于经验的总结。通过对应性原理，我们可以理解为“在公式 1.8 中所有非对角项的$\left[\widehat{H}_0\right]_{nm}$必然为零”。这是因为只有这样，才可以保证系统的哈密顿量不随时变化[27]。按照诺特的语言，时间平移对称性对应能量守恒，说的也是同样的道理。

这样，公式 1.9 的哈密顿量就可以简化为对角矩阵的形式。这时，按照海森堡的规则，n、m对应的都是定态。定态能量是E_n，公式 1.8 简化后的样子就是：

$$\widehat{H}=\begin{bmatrix}E_1 & 0 & \cdots\\ 0 & E_2 & \cdots\\ \vdots & \vdots & \ddots\end{bmatrix}$$

公式 1.9

我们再次强调*这里是用定态的n、m作为指标来对行列式的行、列进行标识*。如果不用定态而用一般的一组基函数，公式 1.9 就不是对角矩阵了。为了尊重海森堡当时的习惯，同时也为了简单，我们这里还是基于定态来得到矩阵力学中的运动方程[28]。

从公式 1.9 出发，我们可以推导出任何一个可观测量算符需要满足的方程。

[27]严格意义上，是哈密顿量的本征值和本征态的模的平方（对应物理真实）不随时间改变。这里，由于海森堡用的基就是定态，哈密顿量随之变成了对角项为定态能量、非对角项为零的形式。

[28]运动方程得到后，我们可以看出它在基组变化下是不变的（基组可以不对应本征态）。这是因为基组的变换对应一个矩阵的相似变换，运动方程仅需进行相同的相似变换即可，不用改变形式。

同样，类比公式 1.1（位置）、公式 1.4（动量），这个可观测量的算符可写作：

$$\left[\hat{A}\right]_{nm} = \left[\hat{A}_0\right]_{nm} e^{-i\omega_{nm}t}$$

公式 1.10

基于这个公式，就有：

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\hat{A}\right]_{nm} = -i\omega_{nm}\left[\hat{A}\right]_{nm} = -i\frac{(E_m - E_n)}{\hbar}\left[\hat{A}\right]_{nm} = -\frac{i}{\hbar}\left(\left[\hat{A}\right]_{nm}E_m - E_n\left[\hat{A}\right]_{nm}\right)$$

公式 1.11

这个等式如果用矩阵写出，就是：

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\left[\hat{A}\right]_{11} & \left[\hat{A}\right]_{12} & \cdots \\ \left[\hat{A}\right]_{21} & \left[\hat{A}\right]_{22} & \cdots \\ \vdots & \vdots & \ddots\end{bmatrix} = \begin{bmatrix}\left[\hat{A}\right]_{11}E_1 & \left[\hat{A}\right]_{12}E_2 & \cdots \\ \left[\hat{A}\right]_{21}E_1 & \left[\hat{A}\right]_{22}E_2 & \cdots \\ \vdots & \vdots & \ddots\end{bmatrix} - \begin{bmatrix}E_1\left[\hat{A}\right]_{11} & E_1\left[\hat{A}\right]_{12} & \cdots \\ E_2\left[\hat{A}\right]_{21} & E_2\left[\hat{A}\right]_{22} & \cdots \\ \vdots & \vdots & \ddots\end{bmatrix}$$

$$= \begin{bmatrix}\left[\hat{A}\right]_{11} & \left[\hat{A}\right]_{12} & \cdots \\ \left[\hat{A}\right]_{21} & \left[\hat{A}\right]_{22} & \cdots \\ \vdots & \vdots & \ddots\end{bmatrix}\begin{bmatrix}E_1 & 0 & \cdots \\ 0 & E_2 & \cdots \\ \vdots & \vdots & \ddots\end{bmatrix} - \begin{bmatrix}E_1 & 0 & \cdots \\ 0 & E_2 & \cdots \\ \vdots & \vdots & \ddots\end{bmatrix}\begin{bmatrix}\left[\hat{A}\right]_{11} & \left[\hat{A}\right]_{12} & \cdots \\ \left[\hat{A}\right]_{21} & \left[\hat{A}\right]_{22} & \cdots \\ \vdots & \vdots & \ddots\end{bmatrix}$$

公式 1.12

即

$$i\hbar\frac{\mathrm{d}\hat{A}}{\mathrm{d}t} = \hat{A}\hat{H} - \hat{H}\hat{A}$$

公式 1.13

这时，如果我们记

$$\hat{A}\hat{H} - \hat{H}\hat{A} = \left[\hat{A}, \hat{H}\right]$$

公式 1.14

就会有：

$$i\hbar\frac{\mathrm{d}\hat{A}}{\mathrm{d}t} = \left[\hat{A}, \hat{H}\right]$$

公式 1.15

这个方程，就是人们经常说的海森堡运动方程。它的首次出现，也是在玻恩、约

当的两人文章中【Born，1925】【沈顺清，2025】。

现在，第一步完成。下面我们要做的就剩两步了：1）有了运动方程，我们来推不对易关系；2）有了不对易关系，来推谐振子能量的量子化。需要说明的是我们要推的不对易关系是位置与动量的不对易关系（所谓的“海森堡不对易关系”），它的表达式是：

$$[\hat{x},\hat{p}] = i\frac{h}{2\pi} = i\hbar$$

公式 1.16

这里，$\hbar$ 是约化普朗克常数，它和普朗克常数h差一个 2π，由狄拉克在后期引入。换句话说，我们要推公式 1.16。其它的力学算符之间当然也会不对易，我们先不管。我们的目的是要知道公式 1.16 的右边为什么是$i\hbar$?

不失一般性，我们可以先假设：

$$[\hat{x},\hat{p}] = C$$

公式 1.17

下面，我们要理解为什么C等于$i\hbar$？在经典力学中，我们知道位置与动量存在如下关系：

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{p}{m_e}$$

公式 1.18

同时，根据前面提到的矩阵力学运动方程，我们有：

$$\frac{\mathrm{d}\hat{x}}{\mathrm{d}t} = \frac{1}{i\hbar}[\hat{x},\hat{H}] = \frac{1}{i\hbar}\left[\hat{x},\frac{\hat{p}^2}{2m_e} + \hat{V}(\hat{x})\right]$$

公式 1.19

由于：

$$[\hat{x},\hat{x}] = \hat{x}\hat{x} - \hat{x}\hat{x} = 0$$

公式 1.20

而$\hat{V}(\hat{x})$是$\hat{x}$的函数，所以：

$$[\hat{x}, \hat{V}(\hat{x})] = 0$$

公式 1.21

进而，公式 1.19 可以演化为：

$$\frac{\mathrm{d}\hat{x}}{\mathrm{d}t} = \frac{1}{i\hbar m_e}\left[\hat{x}, \frac{\hat{p}^2}{2} + \hat{V}(\hat{x})\right] = \frac{1}{2i\hbar m_e}[\hat{x}, \hat{p}^2] = \frac{1}{2i\hbar m_e}([\hat{x}, \hat{p}]\hat{p} + \hat{p}[\hat{x}, \hat{p}])$$

$$= \frac{1}{2i\hbar m_e}\left(C\hat{p} + p\hat{C}\right)$$

公式 1.22

这样，结合公式 1.18（$\mathrm{d}x/\mathrm{d}t = p/m_e$，与经典力学方程对应），就有：

$$\frac{1}{2i\hbar m_e}(C\hat{p} + \hat{p}C) = \frac{\hat{p}}{m_e}$$

公式 1.23

为了达到这个效果，最简单的取法就是$C = i\hbar$。这个$C = i\hbar$再结合公式 1.23，就可以得到公式 1.16 那个所谓的“海森堡”不对易关系【Born，1925】[29]。

最后，我们利用位置与动量的不对易关系，来理解谐振子的能量量子化。对于一维谐振子，它的经典振动方程是：

$$\ddot{x} = -\omega^2 x$$

公式 1.24

按照矩阵力学版本的量子力学的习惯，可以把公式 1.1（位置算符）代入，将这个式子两边的坐标都表示为矩阵。因为两个矩阵相等则它们的每个矩阵元都相等，就有如下等式：

[29]这里，可能会有同学说这个推导不严谨。我们的解释是物理学发展过程中，很多重要的东西确实是靠猜的。成功的物理学的特征就是：猜完，建立起相应的理论的数学形式后，我们会发现现实的世界就应该具备这样的数学结构！

$$(\omega^2 - \omega_{nm}^2)[\hat{x}_0]_{nm} e^{-i\omega_{nm}t} = 0$$

公式 1.25

在这个公式中，$\omega_{nm} = (E_m - E_n)/\hbar$。它是由两个量子定态的能量差别决定的。$\omega$是由势能的曲率决定的，对于一个特定的简谐势$\hat{V}(\hat{x})$，它就是一个正实数。这也就意味着在由公式 1.24 通过对应性原理建立的量子力学中，*由坐标x通过量子跃迁联系起来的两个定态的能量差必须满足*：

$$(E_m - E_n)/\hbar = \pm\omega$$

公式 1.26

才能使公式 1.25 中的$[\hat{x}_0]_{nm}$*才可以不为零*。

这样的话，我们可以把相互之间有跃迁的定态安排为相邻的定态（这也即使了图 1.1b 中为何跃迁都是相邻能级之前的跃迁），它们的能量差就是$\hbar\omega$。图 1.1b 所示的那个定态能量等间距分布这一点就明确了。换句话说，仅仅通过公式 1.24 到 1.26 这几个简单的式子，普朗克的黑体辐射中的最关键的假设就有了力学理论的基础！

至此，前面我们费了半天力气推的不对易关系甚至还没有用上（第一步推运动方程是为了第二步推不对易关系）。我们花了半天的精力去推这个不对易关系，除了上面这条，还因为它与一个重要的量子现象相关。这个现象就是图 1.1b 蕴含的另一个重要信息：*即使在基态，对于量子体系而言能量也不为零，这个能量叫零点能*。下面，我们就集中精力说明零点能的存在，并给出简谐振子的量子化的能量的绝对值。

我们的出发点是公式 1.16，将$p = m_e\dot{x}$代入公式 1.16，有：

$$m_e\left([\hat{x}\hat{x}]_{mn} - [\hat{x}\hat{x}]_{mn}\right) = i\hbar\delta_{mn}$$

公式 1.27

取对角项$m=n$的情况，这个式子进一步等于：

$$m_e\sum_l\left(-\omega_{lm}[\hat{x}_0]_{ml}e^{-i\omega_{ml}t}[\hat{x}_0]_{lm}e^{-i\omega_{lm}t}-(-\omega_{ml})[\hat{x}_0]_{ml}e^{-i\omega_{ml}t}[\hat{x}_0]_{lm}e^{-i\omega_{lm}t}\right)$$
$$=i\hbar$$

公式 1.28

进而：

$$m_e\sum_l(-i\omega_{lm}[\hat{x}_0]_{ml}[\hat{x}_0]_{lm}+i\omega_{ml}[\hat{x}_0]_{ml}[\hat{x}_0]_{lm})=i\hbar$$

公式 1.29

这也意味着：

$$2m_e\sum_l(\omega_{ml}[\hat{x}_0]_{ml}[\hat{x}_0]_{lm})=\hbar$$

公式 1.30

在讲公式 1.2 和公式 1.3 的时候，我们说过$\hat{x}_0$、$\hat{x}$都是厄米矩阵。因此，公式 1.30 可进一步变为：

$$2m_e\sum_l(\omega_{ml}|[\hat{x}_0]_{ml}|^2)=\hbar$$

公式 1.31

从公式 1.26 以及与之相关的讨论中，我们又知道$[\hat{x}_0]_{ml}$只有在$l=m\pm1$的时候不为零，且这时$\omega_{m,m\pm1}=\omega$。这样，公式 1.31 的加和还可以进一步简化为：

$$2m_e\omega\left(\left|[\hat{x}_0]_{m,m+1}\right|^2-\left|[\hat{x}_0]_{m,m-1}\right|^2\right)=\hbar$$

公式 1.32

进而：

$$\left|[\hat{x}_0]_{m,m+1}\right|^2=\left|[\hat{x}_0]_{m,m-1}\right|^2+\frac{\hbar}{2m_e\omega}$$

公式 1.33

下面，我们就可以基于这个公式，把$\hat{x}_0$这个矩阵写出来。

假设把能量最低的能级编号为0，由于-1的能级并不存在，因此有$[\hat{x}_0]_{0,-1} = 0$。我们可以从它出发，通过公式 1.33 来写出$\hat{x}_0$这个矩阵。按上升的顺序，依次有：

$$\left|[\hat{x}_0]_{0,1}\right|^2 = \frac{\hbar}{2m_e\omega}$$

公式 1.34

x_0的厄米性会导致：

$$\left|[\hat{x}_0]_{1,0}\right|^2 = \left|[\hat{x}_0]_{0,1}\right|^2 = \frac{\hbar}{2m_e\omega}$$

公式 1.35

之后，有：

$$\left|[\hat{x}_0]_{1,2}\right|^2 = \left|[\hat{x}_0]_{1,0}\right|^2 + \frac{\hbar}{2m_e\omega} = \frac{\hbar}{m_e\omega}$$

公式 1.36

以及：

$$\left|[\hat{x}_0]_{2,1}\right|^2 = \left|[\hat{x}_0]_{1,2}\right|^2 = \frac{\hbar}{m_e\omega}$$

公式 1.37

总结起来就是：

$$\left|[\hat{x}_0]_{m+1,m}\right|^2 = \left|[\hat{x}_0]_{m,m+1}\right|^2 = \frac{(m+1)\hbar}{2m_e\omega}$$

公式 1.38

而$[\hat{x}_0]_{m+1,m}$、$[\hat{x}_0]_{m,m+1}$的本意是振幅。因此，矩阵$\hat{x}_0$可进一步写为：

$$[\hat{x}_0] = \sqrt{\frac{\hbar}{2m_e\omega}}\begin{bmatrix} 0 & \sqrt{1} & 0 & 0 & \dots \\ \sqrt{1} & 0 & \sqrt{2} & 0 & \dots \\ 0 & \sqrt{2} & 0 & \sqrt{3} & \dots \\ 0 & 0 & \sqrt{3} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

公式 1.39

基于这个表达式，我们就可以来求总能了。

我们要做的事情很简单，就是将公式 1.39 以及尤其确定的矩阵x代入下面这个简谐振子哈密顿量的表达式：

$$\hat{H} = \frac{\hat{p}^2}{2m_e} + \frac{1}{2}m_e\omega^2\hat{x}^2$$

公式 1.40

并且记住$\hat{p}$、$\hat{x}$都是矩阵，以及$\hat{p}^2$、$\hat{x}^2$的意思就是$\hat{p}$这个矩阵乘上$\hat{p}$这个矩阵、$\hat{x}$这个矩阵乘上$\hat{x}$这个矩阵，就可以。这样操作的话，得到的$\hat{H}$这个矩阵的对角项就是：

$$\begin{aligned}\left[\hat{H}\right]_{nn} &= \frac{1}{2}m_e\left(\left[\dot{\hat{x}}^2\right]_{nn} + \omega^2[\hat{x}^2]_{nn}\right) = \frac{1}{2}m_e\sum_l\left(\left[\dot{\hat{x}}\right]_{nl}\left[\dot{\hat{x}}\right]_{ln} + \omega^2[\hat{x}]_{nl}[\hat{x}]_{ln}\right) \\ &= \frac{1}{2}m_e\sum_l\Big((-i\omega_{nl})[\hat{x}_0]_{nl}e^{-i\omega_{nl}t}(-i\omega_{ln})[\hat{x}_0]_{ln}e^{-i\omega_{ln}t} \\ &+ \omega^2[\hat{x}_0]_{nl}e^{-i\omega_{nl}t}[\hat{x}_0]_{ln}e^{-i\omega_{nl}t}\Big) \\ &= \frac{1}{2}m_e\sum_l(\omega_{nl}^2[\hat{x}_0]_{nl}[\hat{x}_0]_{ln} + \omega^2[\hat{x}_0]_{nl}[\hat{x}_0]_{ln})\end{aligned}$$

公式 1.41

其中$\hat{x}_0$这个矩阵又满足公式 1.39，将起直接代入，就有：

$$\begin{aligned}\left[\hat{H}\right]_{nn} &= m_e\omega^2\left(\left|[\hat{x}_0]_{n,n+1}\right|^2 + \left|[\hat{x}_0]_{n,n-1}\right|^2\right) = m_e\omega^2\frac{\hbar}{2m_e\omega}(2n+2) \\ &= \left(n+\frac{1}{2}\right)\hbar\omega\end{aligned}$$

公式 1.42

这也就意味着对于第$n+1$个简谐振子的定态（这里是从$n=0$开始算的），它的能量是$\left(n+\frac{1}{2}\right)\hbar\omega$。这个能量首先是量子化的。其次，对于第一个定态，也就是最低能量态，它的能量是$\frac{1}{2}\hbar\omega$而非零。这个能就是量子系统的零点能。

到这里，矩阵力学的核心内容我们就介绍完了。简单来说，就是海森堡从*力学量应该基于可观测量*来进行描述这样一个思想出发【Heisenberg，1925】，建立了一套基于可观测量的运算规则来描述量子的粒子的运动。这个规则后来被玻恩发现就是矩阵运算规则【Born，1925】【Born，1926a】。基于这样一个对微观粒子的运动学（kinematics）的描述的改进，人们可以通过对应性原理直接将经典的力学方程（Mechanical Equation）用于量子系统的描述，唯一需要注意的就是运动学需要遵循量子的规则。这样，既可以推出量子力学的不对易关系，又可以在一些简单的系统（比如谐振子）中得到量子化的能量。1926 年，泡利将矩阵力学应用于氢原子电子能级的描述【Pauli，1926】。至此，量子力学的第一种表述形式（矩阵力学）彻底建立。

这里，海森堡天才的想法是他用矩阵替代了数来描述经典力学中的力学量（量子力学中的可观测量）。他对这个矩阵所施加的限制（比如它必须是厄米算符）恰恰帮助他解决了电子轨道的稳定性问题。在矩阵力学之后，薛定谔非常及时地提出了波动力学的量子力学表达。在波函数的展开过程中，线性变换就是那个发挥关键作用的东西。这两种量子力学的表述，自然也就是等价的。借助矩阵与线性变换，或者与之等价的波函数，量子力学的另一种表述也成为了可能。

而在与之等价的量子力学的波函数表述的构建过程中，虽说最后的结果也就是我们现在都在使用的薛定谔方程是由薛定谔给出的，但其革命性的思想，在笔者看来却是德布罗意的物质波理论。在德布罗意的工作中，他创新性地将之前爱

因斯坦对光的二象性理论推广至电子的“运动学”描述中。这个理论在 1923 年提出【de Broglie，1923a】【de Broglie，1923b】【de Broglie，1923c】。发表后，玻恩的学生埃尔萨塞（Walter Maurice Elsasser，1904－1991 年）注意到之前美国科学家戴维森（Clinton Joseph Davisson，1881－1958 年）与孔斯曼（Charles Henry Kunsman，1890－1970 年）的一个实验工作或许与之有关【Davisson，1921】，于是写了一个通讯来评论这个事情【Elsasser，1925】。

之后，更关键的是在 1926 年夏天，牛津大学举办了一个英国科学促进会的会议。在这个会议上，玻恩建议同样参会的戴维森重新设计他的实验，来验证电子的物质波特性。戴维森听从这个建议，与革末一起重新设计了实验，这就是我们课本上经常提到的戴维斯－革末实验（Davidson－Germer Experiment）。这些实验的历史作用就是验证了电子的物质波属性【Davisson，1927a】【Davisson，1927b】【Davisson，1928】。G. P. 汤姆孙也在这个会议上。显然，他也从这个讨论中获益。加上，他的父亲就是阴极射线方面的老专家与电子的发现人 J.J. 汤姆孙。他也很快做出了类似的实验【Thomson，1928】。因为这个原因，戴维森于与 G. P. 汤姆孙一起分享了 1938 年的诺贝尔物理学奖。他们的获奖理由是：因他们通过实验发现电子在晶体中的衍射（for their experimental discovery of the diffraction of electrons by crystals）。

此外，我们还可以把这个和后面我们要讲*波函数的几率性解释*联系起来。实际上，这个概念也是玻恩在利用薛定谔方程研究电子的散射问题的时候提出的。这里，笔者说的这些内容看似有点散，落脚点无非两个：1）量子力学的波函数表述中，新的“运动学”由电子的物质波这个概念来负责，后面，这个概念可以推广至所有微观粒子；2）处于尊重历史的考虑，笔者想强调一下玻恩在电子物

质波理论的实验验证中的助力，这种助力不会通过“奖项”的形式来体现，但我们不能忽视，物理学的内容是远远大于“奖项”的；3）后面要讲的波函数的统计性解释实际上也和这种思考相关。

1.2 量子力学与经典力学的联系

上一节，我们通过“kinematics”这个关键词强调了量子力学与经典力学的不同。本节，我们想讨论一下这两种貌似完全不同的物理学理论之间，实际上是存在很深的内在联系的。这种联系，在笔者看来，以“mechanics”为关键词，通过“薛定谔方程的推导”来进行理解或许最为直接。为了简单起见，我们讨论一个一维体系，从经典的哈密顿-雅可比方程出发。来推导薛定谔方程。我们的目的，就是想说只要接受了电子的物质波表述的运动学新概念，那么，用经典的哈密顿-雅可比方程来描述这个波，我们就可以*推出薛定谔方程*。这种从经典出发推量子力学方程的过程，体现的恰恰就是*量子力学与经典力学的内在联系*。

经典的哈密顿-雅可比方程的具体形式是：

$$\frac{\partial S(x,t)}{\partial t}+\frac{1}{2m}\left(\frac{\partial S(x,t)}{\partial x}\right)^2+V(x)=0$$

公式 1.43

这是一个经典力学的方程，其中$S(r,t)$被称为欧几里得作用量（Euclidean Action，简称作用量，Action），它是拉格朗日量（Lagragian，记为L，它是动能和势能的差）的积分，定义如下：

$$S(r,t)=\int_{t_0}^{t}Ldt=\int_{t_0}^{t}(T-V)dt$$

公式 1.44

这里，我们这个讨论的出发点是公式 1.43，是一个*经典方程*。

对于我们要研究的处在特定能量E状态的系统，时间平移对称性决定了它的作用量可以写成：

$$S(x,t) = W(x) - Et$$

公式 1.45

把公式 1.45 代入公式 1.43，就意味着在经典力学的框架下，我们可以有：

$$\left(\frac{\partial S(x,t)}{\partial x}\right)^2 = 2m[E - V(x)]$$

公式 1.46

在一个特定的时刻t_0，在经典力学的理论框架下，公式 1.45 中的作用量为定值S_0的就是一个位置x_0。它们之间的关系是：

$$S(x_0,t_0) = W(x_0) - Et_0$$

公式 1.47

在薛定谔的推导中，他的关注点是*这个定值点随着时间的移动*。对于高维的情况，这个点对应一个曲线或者曲面。*对应于经典波，大家可以把它理解为波前的位置的移动*。

这个关注点明确后，下面的事情就好办了。我们需要考虑的，是这个作用量为这个定值的点在下一个时刻的位置。我们把它记为：$(x_0 + \mathrm{d}x, t_0 + \mathrm{d}t)$。这里，我们知道的条件是：

$$S(x_0,t_0) = W(x_0 + \mathrm{d}x) - Et_0 - E\mathrm{d}t = W(x_0) + \left.\frac{\mathrm{d}W(x)}{\mathrm{d}x}\right|_{x=x_0} \mathrm{d}x - Et_0 - E\mathrm{d}t$$

公式 1.48

这样的话，用公式 1.48 减去公式 1.47，就有：

$$\left.\frac{\mathrm{d}W(x)}{\mathrm{d}x}\right|_{x=x_0}\mathrm{d}x-E\mathrm{d}t=0$$

公式 1.49

我们看公式 1.45，它左边的$S(x,t)$有两个变量，右边的$W(x)$有一个变量。因此，有：

$$\left.\frac{\partial S(x,t)}{\partial x}\right|_{x=x_0}=\left.\frac{\mathrm{d}W(x)}{\mathrm{d}x}\right|_{x=x_0}$$

公式 1.50

这样，公式 1.49 就又进一步化为：

$$\left.\frac{\partial S(x,t)}{\partial x}\right|_{x=x_0}\mathrm{d}x-E\mathrm{d}t=0$$

公式 1.51

这样的话，这个点在时间$\mathrm{d}t$内移动位置$\mathrm{d}x$，这两个小量之间的关系就是：

$$\frac{\mathrm{d}x}{\mathrm{d}t}=E/\left(\left.\frac{\partial S(x,t)}{\partial x}\right|_{x=x_0}\right)$$

公式 1.52

而公式 1.46 告诉我们：

$$\left.\frac{\partial S(x,t)}{\partial x}\right|_{x=x_0}=\sqrt{2m[E-V(x_0)]}$$

公式 1.53

因此，公式 1.52 进一步演化为：

$$\frac{\mathrm{d}x}{\mathrm{d}t}=\frac{E}{\sqrt{2m[E-V(x_0)]}}$$

公式 1.54

同样，这个运动方程是经典的。

这个公式的$\mathrm{d}x/\mathrm{d}t$，我们可以记作u。它代表的是作用量为S_0的点在x轴上移动的速度。我们可以把它理解为一个经典波的波前移动的速度。如果是高维情况，这个点对应曲线或曲面，这个速度就是这个曲线或曲面在其法向方向移动的速度。这样，由这个作用量定值点定义的相速度就是：

$$u = \frac{E}{\sqrt{2m[E - V(x_0)]}}$$

公式 1.55

至此，薛定谔的推导中的经典力学的部分就结束了。下面，我们会基于作用量再定义一个波函数。之后，就是基于这个波函数中的相位的运动，结合公式 1.55 来得到薛定谔方程了。换句话说，在这条路下，薛定谔方程与经典力学之间的天然的、紧密的纽带就应该清楚了。薛定谔方程中从概念上来讲的话最重要的创新点，依然是电子的运动学表述，也就是需要基于波函数而非轨道来描述电子运动。这一点，来自于德布罗意。

在薛定谔 1926 年的一系列文章中【Schrödinger，1926a−1926f】，他假设在一个能量守恒的系统中，电子具有如下形式的波函数：

$$\Psi(x,t) = A(x)e^{iS(x,t)/K} = A(x)e^{i[W(x)-Et]/K}$$

公式 1.56

如果电子是自由电子，在德布罗意那里，就已经被描述为了平面波，其波函数的含时项为$e^{-\frac{iEt}{\hbar}}$。而这里，波函数的含时项为$e^{-iEt/K}$。因此，薛定谔说：One cannot resist the temptation of supposing K to be a universal constant, independent of E and independent of the nature of the mechanical system, because if this be done and K be given the value $h/(2\pi)$, then the frequance will be given by $\nu = E/h$。

这里，薛定谔说的$h/(2\pi)$就是约化普朗克常数$\hbar$。换句话说，公式 1.56 中的

K就是$\hbar$了。相应的，公式 1.56 变为：

$$\Psi(x,t) = A(x)e^{iS(x,t)/\hbar} = A(x)e^{i[W(x)-Et]/\hbar}$$

公式 1.57

之后，薛定谔想到了*经典波动方程的普遍形式*：

$$\frac{\partial^2}{\partial t^2}\Psi(x,t) = u^2\frac{\partial^2}{\partial x^2}\Psi(x,t)$$

公式 1.58

这是一维体系中描述固定边界条件的振动的经典方程，早在 18 世纪就由达朗贝尔提出。

下面，我们要做的仅仅就是把公式 1.58 这个经典的力学方程拿过来，把 1.56 式那个电子波函数代入。这样，公式 1.58 的左边就等于：

$$\frac{\partial^2}{\partial t^2}\Psi(x,t) = -\frac{E^2}{\hbar^2}\Psi(x,t)$$

公式 1.59

而其右边，又因为从经典力学推出的相速度（公式 1.55 代入公式 1.58 的右边），等于：

$$u^2\frac{\partial^2}{\partial x^2}\Psi(x,t) = \frac{E^2}{2m[E-V(x)]}\frac{\partial^2}{\partial x^2}\Psi(x,t)$$

公式 1.60

这样，将公式 1.58 至公式 1.60 联立，我们就有：

$$\frac{E^2}{2m[E-V(x)]}\frac{\partial^2}{\partial x^2}\Psi(x,t) = -\frac{E^2}{\hbar^2}\Psi(x,t)$$

公式 1.61

它可进一步化为：

$$\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\Psi(x,t) = -[E-V(x)]\Psi(x,t)$$

公式 1.62

进而：

$$\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}+\hat{V}(\hat{r})\right]\Psi(x,t)=E\Psi(x,t)$$

公式 1.63

这时，再结合波函数的定义式（公式 1.57），我们也就可以得到薛定谔方程：

$$i\hbar\frac{\partial}{\partial t}\Psi(\boldsymbol{r},t)=\left[-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}+\hat{V}(\hat{r})\right]$$

公式 1.64

从上述推导，我们应该可以感受到薛定谔这个人对经典力学的掌握、对波动方程的理解对于他完成那个工作是发挥了极其关键的作用的。在笔者看来，思维的跳跃并不大，一切都是力学很合理的推导。突破的，就是对运动（kinematics）的描述。相对于经典力学中的问题，他将微观的粒子处理为一个波。

最后，我们强调一点：*这样的一个波动方程应用到束缚电子体系，对应的就是固定边界条件的偏微分方程*。它的解是分立的，*这是达朗贝尔这些人在 18 世纪中叶已经知道的经典力学中波动方程的性质*。在我们的生活经历中，鼓声的音阶也是不连续的，与这个也是一个道理（图 1.1）。薛定谔将类似的一个波动方程应用到了氢原子体系[30]、简谐振子体系，得到的量子化的结果都很完美。对于氢原子体系，它的解析解可以直接解释与自旋无关的所有实验结果（薛定谔方程中也并没有自旋自由度）。对于简谐振子，其解与之前海森堡利用矩阵力学的解也

[30]这里不像矩阵力学，海森堡、玻恩、约当都没有把它用到氢原子体系，直到泡利的工作的出现。所以我们前面也说泡利的工作对于多数人接受矩阵力学是非常重要的！波动力学用到氢原子体系并不需要非常繁琐的数学，薛定谔直接就做了。这在一定程度上也能反映出量子力学的这两种表达在应用层面的难易程度和实用性差别。玻恩在 1926 年研究散射问题的时候直接使用波动力学，也就很好理解了。

是一致的。考虑到波动方程在数学上更易于求解，同时这套理论对于当时的物理学家来说也更友善，波动力学很快就收到了广泛的欢迎。就连玻恩本人，在研究散射问题的时候，使用的也是薛定谔的方法。也是在那个工作中，他给出了波函数的物理意义的统计性几率诠释【Born，1926】。这个解释是其 1954 年获得诺贝尔物理奖的一个重要的理由[31]。



图 1.1 固定边界条件的波动方程是一些经典力学中的基本问题与薛定谔方程的共同的数学结构。

1.3 量子力学的理论框架

现在，我们在第一节通过矩阵力学的形式来对量子力学所蕴含的新的“kinematics 的内容”进行了强调。这一点定义了我们需要用线性代数这个工具来描述这种新的力学。量子力学的理论架构（详见五条公设）基本上基于这个数学工具来构建。在第二节，我们通过薛定谔方程对量子力学与经典力学在“mechanics”方面的联系的说明。这一点蕴含着量子力学在经典极限下必须回归经典力学这样一个限制。与这种限制对应，有个对应性原理。它本质上讲的就是如果我们把量子力学中的可观测量对应为经典力学中的力学量，那么在量子力学

[31]原话是：for his fundamental research in quantum mechanics, especially for his statistical interpretation of the wavefunction（他在量子力学中的基础研究，尤其是波函数的诠释）。这里，组委会既认可了他对量子力学的架构的基础性的贡献，也强调了波函数的统计性诠释的重要性。

中的*可观测量之间的关系*与经典力学中*力学量之间的关系*可以对应。基于这两节，我们在下面就可以用一个稳定的力学系统（封闭、无耗散）作为着手点，来构建量子力学的理论体系了。

在樱井和 Napolitano 的书中【Sakurai, Napolitano, 2017】，他们用 Stern-Gerlach 实验很好地用引入式的方式构建了量子力学的理论体系。对于读者而言，那种引入方式是令人印象深刻的，因为自旋态确实最能说明量子力学中量子态的本质。毕竟，这个态连经典对应都没有（虽然他在介绍中借助了偏振光的概念，但那实际上是完全不同的物理）。按同样的方式来总结本讲义的这部分内容的话，笔者感觉再写也不可能写过他们。因此，本着打不过就躲的原则，我们采取另一条路[32]。前面，我们已经对量子的运动学进行了足够多的强调。我们的目的就是让大家直接接受描述量子力学的数学语言是线性代数这个关键点。从它出发，直接构建量子力学的理论框架在逻辑上应该是没有跳跃的。

基于此，我们就可以直接用一种（笔者认为的）我们中国学生学习过程中比较习惯的思路来构建这个理论架构：*通过思考应该如何描述一个理想的物理学问题来构建一个公理化的理论体系（axiomatic theoretical system）*[33]。这个理想的物理问题是一个*封闭的、稳定的量子系统*。我们要把这一点进行强调。然后，结合前面将的量子力学的新的“运动学”与它和经典力学的联系，把“*为什么这个量*

[32]笔者在这里采取的“通过运动学引入线性代数、通过薛定谔方程强调与经典力学之间的联系，然后建立理论体系”或许对有些同学也会更友善一些。至少笔者在准备讲义的时候把自己放在学生的位置，反复琢磨，还是觉得这个方式更合适的。

[33]好的理论体系都是公理化的，比如欧氏几何（从公设出发）、牛顿力学（从定律出发，这里定律强调是可被验证的经验的总结，逻辑地位和公设一致，都在理论最底层）、热力学理论（四大定律）、电动力学理论（麦克斯韦方程组）、狭义相对论（狭义相对性原理与光速不变原理）、广义相对论（广义相对性原理和等效原理）。

*子力学理论体系应该这样构建”*讲清楚。这样做的好处是我们不会觉得这个理论体系就是天上掉下来的，我们会觉得它有道理！同时，因为它是公理化的，内容就会会非常的简洁、漂亮。之后，基于这个架构，我们再去做题、做拓展，就很容易了。这也回应了我们在本章引言部分讲到的笔者自己在当时学《量子力学》遇到的不理解理论为何这样构架以及学完就忘的问题。

但需要说明的是对《量子力学》这门课而言，目前其公理化形式并不是那么让人满意。比如，关于五条公设，不同的教科书是有不同的说法的。我们采取的，只能说是其中的一种，与量子力学的哥廷根/哥本哈根诠释密切相关，但不完全相同。其共同性，是把测量带来的量子态向相应的可观测量的某本征态的投影作为一个公设来对待[34]。不同点，是针对测量，我们用投影取代塌缩来进行表述[35]。我们需要强调的是量子力学的讲法目前没有标准答案。从这五条公设出发的好处，是《高等量子力学》课程中的很多标准内容都可以作为其扩展来进行详细讨论。

同时，我们也注意到在很多《高等量子力学》课程中的理论基础部分，人们往往会讲测不准关系、互补性原理与量子纠缠这些内容。*但我们想强调的是在公理化的理论体系中，它们**不是理论的出发点而是结果**。它们是可以通过前面提到的五条公设推出的*[36]。基于这个考虑，我们把它们放在本节的第二小节讲，中间

[34]这里，我们称之为哥廷根/哥本哈根，而不是像其它教材直接称之为哥本哈根诠释。原因，是想强调哥廷根学派在其中的贡献，毕竟我们去看海森堡的关键文章和后面的两人、三人文章的话，单位都是哥廷根。

[35]这与孙昌璞老师的讲法是一致的【孙昌璞，2024】，毕竟塌缩过程很难被实际看到。*把测量当作是一次向相应的可观测量的算符的本征态的投影*，这是量子力学理论架构中更本质的东西。它充分利用了第一条公设中讲的量子态是希尔伯特空间中的一条射线的特征，同时也保证了整个量子力学理论体系的自洽。

[36]就像在牛顿力学中，三大定律是逻辑基础。万有引力定律是基于三大定律定义的力学理论来描述行星运动，然后以开普勒三定律这样的实验观测结果为依据进一步发现的规律。它在牛顿力学

有比较详细的推导。

最后，我们也想强调一下“电子自旋”是一个量子力学中微观粒子对的全新的“运动学”内容。这个自由度在经典力学中没有对应。我们在后面会基于对称性的考虑来详细讲解这个概念。但是，考虑到课程前面的讨论会时不时地用到它，我们还是在第一章用一个小节来初步地引入自旋这个概念。这样，即便在“广义”的层面，多数教科书上讲的量子力学的理论框架在包含这三节内容后基本也就讲完了。当然，我们还可以讨论一些简单体系中的解的问题，但这些内容整体而言是《量子力学》课程的内容，不应该由我们负责。在《高等量子力学》课上，我们需要在第一章给出量子力学的理论框架，方便后续讨论。但我们不需要把这个框架进行过度展开。

基于如上考虑，我们的三个小节的题目分别是：1）五条公设；2）测不准关系、互补性原理；3）电子自旋。我们按这个顺序进行讨论。

1.3.1 五条公设

1.3.1.1 态矢量公设

在量子力学的公理化理论的构建过程中，第一条公设是用来描述一个量子系统中物质的存在形式的。这和我们物理学关心的核心问题，即*物质的存在形式与运动规律*，以及我们如何对其进行感知，是密切相关的[37]。在一个量子系统中，

中的逻辑地位和三大定律是存在本质区别的。物理学基本理论中各个定律、公设、原理、定理、引理、推论在其中的逻辑地位，我们在学习的过程中一定要分清！我们学习的目的并不是简单的掌握知识点，而是要理解这些知识点，进而建立自己的知识体系。

[37]我们后面在介绍其它公设的时候，也会按这句话展开。对于近代物理学（科学革命后、物理学革命前建立的经典物理学）和现代物理学（物理学革命建立的以量子力学、相对论为支柱的现代

我们用量子态来描述其物质存在形式：*一个物理系统在任意时刻的状态，由希尔伯特空间中的一个矢量——态矢量——完全描述*。

在 1.1 节，我们已经花了很多时间去跟大家解释一个核心信息，就是描述量子力学的数学工具是线性代数。既然是线性代数，就自然的有了一个线性空间，即希尔伯特空间。这就使得我们量子态有了一个数学对应：希尔伯特空间中的向量。这是上世纪初，经过二十多年的探索，人们在上世纪 20 年代中期这个时间节点忽然认识到的事情[38]。我们将这个向量记为：$|\psi\rangle$。这里$|\ \ \rangle$是一个符号，看起来像是一个括号(bracket)的右边。因此，按照狄拉克引入的习惯，人们取 bracket 的后面几个字母，称之为 ket。这个 ket 矢量形成的空间，称之为 ket space（右矢空间）。*$|\psi\rangle$里面这个ψ是这个量子态的标识。它可以对应任意我们可以观测到的物理学量（可观测量）或者一组标识这个可观测量的数字*。比如，取x的话，$|x\rangle$代表的是在x点出现的量子态；取p的话，$|p\rangle$代表的是动量为p的量子态。更多情况下，我们直接用$|\psi\rangle$，意思就是这个量子态用ψ来标识。

略作扩展的话，我们可以说这里的ψ有点康德哲学中现象界的意思【李新征，

物理学），物理学研究的是物质的存在形式与运动规律这一点都是显然的。*物理学革命中，物理学家认识到我们对物理世界的感知同样重要。*这也是笔者在讲授完《今日物理》的课程后，在我们这门课里面总是强调“现代物理学一定程度上讲是康德哲学的胜利”的根本原因（详见文献【李新征，2026】【王若熙，2026】）。在以相对论、量子力学为代表的现代物理学中，观测是非常重要的。这和经典物理学的理论是根本不同的。因此，我们可以讲经典物理学关注的是物质的存在形式与运动规律。现代物理学关注的是物质的存在形式与运动规律，*以及我们如何感知*。第一条公设先看存在形式与运动规律（考虑到运动是永恒的，这里的运动规律本身也有存在形式的内容），后面讲感知（算符公设、测量公设）。

[38]海森堡的大胆的数学处理打开了一扇门，玻恩和约当的整理将其系统呈现，目前我们用到的规范的语言来自狄拉克。当然，薛定谔也很重要。他基于德布罗意的思想给出的波动方程给出了量子力学到目前为止最实用的表述形式。在第八章，我们会专门讲费曼的路径积分。笔者对这个方法的感觉是在量子单体或少体问题的研究中，它并还没有太大的优势。但在量子多体系统中，它是很有潜力的。这个观点，或许还需要一定的时间去检验。

2026】【王若熙，2026】。它描述的是我们可以感知的物质存在形式的信息。我们用这种可以被感知的信息作为标识，来定义一组希尔伯特空间中的基，描述其中的向量（即量子态）。这组信息如果是完备的信息（比如三维空间中，在不考虑自旋自由度的时候，所有的x），那么基于这些信息所定义的向量合在一起构成的就是这个系统的希尔伯特空间中一组完备的基。基于这组基，我们可以描述希尔伯特空间中的任意向量。同样，所有的$|p\rangle$合在一起也构成一组完备的基，它们和所有的$|x\rangle$之间通过傅里叶变换联系。如果这种可以被感知的信息并不包含所有的量子态的信息，那么它所形成的基组就不能覆盖希尔伯特空间中所有的维度。后面，我们会将可观测量完备集（Complete Set of Commuting Observables，简称 CSCO）的概念，对应的就是这样的物理。

既然是希尔伯特空间，其向量就有内积。为了计算这个内积，就需要定义一个 ket 空间的*对偶空间*（图 1.2）。我们取 bracket 中的前三个字母 bra，称为 bra space（左矢空间）。对于它里面的向量$\langle\psi|$，如果标识ψ和右矢空间中的向量$|\psi\rangle$一致，标识的就是 ket space 中的向量$|\psi\rangle$在其对偶空间中的对偶向量。而$\langle\psi|\psi\rangle$，表示的就这个用ψ标识的向量和其自身的内积，对应这个向量的模的平方。右矢空间中的向量$|\psi\rangle$与左矢空间中的另一个向量$\langle\varphi|$之间，也是可以定义内积$\langle\varphi|\psi\rangle$。它是一个复数。往往用来定义向量$|\psi\rangle$与向量$|\varphi\rangle$之间的夹角。

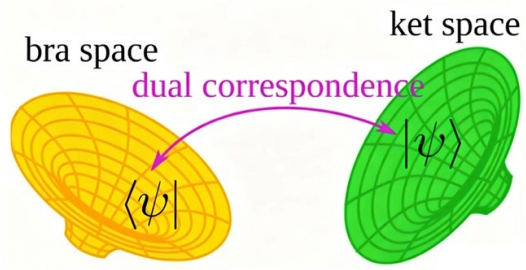


图 1.2 ket space 与 bra space 之间的向量对应关系。

这些，应该说都是线性代数中大家比较常见的内容。在量子力学中，唯一新颖的地方就是狄拉克引入了 bra 和 ket 这样的语言，来把左矢、右矢、对偶空间这些概念进行了一个更为简洁的描述。*希尔伯特空间是一个完备的复线性内积空间，它有三个结构上的要求*。

第一，该空间具有线性的性质。也就是说如果$|\alpha\rangle$与$|\beta\rangle$都是这个空间中的向量，那么$c_1|\alpha\rangle + c_2|\beta\rangle$必然也是这个空间中的向量。这里，$c_1$和$c_2$是任意复数。*这个看似普通的数学要求，实际上是量子力学最核心的物理原理（态叠加原理）得以成立的先决条件*。在经典物理中，系统的状态空间（相空间）不是线性的：两个允许的粒子运动轨迹的“和”通常不是一个允许的运动轨迹。例如，地球绕太阳的轨道和金星的轨道，其“叠加”毫无物理意义。而在微观的世界中，很多实验事实（如电子双缝干涉）迫使我们承认，如果一个系统可以处于通过狭缝 1 的态，也可以处于通过狭缝 2 的态，那么它必定可以处于一个“同时通过两条狭缝”的态，即两者的线性叠加。希尔伯特空间的线性结构，正是为容纳这种经典逻辑无法理解的、波的形式比较善于表达的存在方式而量身定做的数学框架。没有线性，就没有叠加，也就没有了整个量子力学。

第二，该空间中的向量在复数域上定义了内积。线性代数中向量内积我们都很熟悉。这里讲一下它的物理对应。在量子力学理论中，测量很重要。关于这一点，我们在第二条公设中马上就会进行说明。内积定义的物理意义是我需要通过内积来描述对于目前的这个量子态如何通过测量表达？比如，对其中的任意一个向量$|\beta\rangle$，我可以用可观测量的本征态形成的基对它进行表达。假设这组基是一组$|\alpha\rangle$向量（或者叫态），*那么，测量后它落在某个$|\alpha\rangle$态上的几率就是$\langle\alpha|\beta\rangle$的模*

的平方。这里的这个$\langle\alpha|\beta\rangle$，就是$|\beta\rangle$和$|\alpha\rangle$的内积。线性代数定义的数学结构，在它建立了几十年后，通过量子力学，在物理学的世界中得到了完美的对应[39]。

在很多地方，或许读者会看到这样一句话：*在标识量子态的时候，方向相同但大小不同的向量描述的是同一个量子态*。这句话，实际上是可能为读者带来一些困惑的（至少给笔者带来过困惑）。比如，我用$|\beta\rangle$标识一个量子态。那么，按这句话，$c|\beta\rangle$也标识这个量子态（这里c为复数）。但显然，$\langle\alpha|\beta\rangle$和$\langle\alpha|c\beta\rangle$大小是不一样的。这就影响了测量所对应的量子态的物理实在。针对这个我们，我们需要修改的地方是：*量子态确实是只看方向，但这个方向必须通过单位长度的向量来标准化表达，否则概率计算就失去了统一的基准*。这实际上也在一定程度上揭示了统计对于量子力学的重要性[40]。不光希尔伯特空间本身的数学结构会影响量子力学理论的架构，统计规律（比如正定、归一）也会影响量子力学理论的架构。

现在，有了*线性的要求（背后是态叠加）、有了内积的定义（背后是波函数的统计性诠释）*，下一个对希尔伯特空间的要求就是它一定要*完备*。这是我们对这个线性空间的*第三点*要求，它保证了我们在进行极限运算、定义无穷级数展开（如将一个态用一组完备基展开为）时，结果不会“跑出”我们定义的物理空间。因此，我们也可以说它是之前已经发展成熟的微积分和泛函分析工具能在量子力学中畅通无阻的通行证。

最后作为本小节的落脚点，我们说：*有了线性结构来保障叠加原理，有了内积结构来为量子力学的概率诠释提供代数基础，有了完备性来保证我们所熟悉的*

[39]线性代数的核心思想（线性空间、线性变化、矩阵）建立的时间节点是十九世纪下半叶。量子力学是二十世纪二十年代中间几年。

[40]因为统计规律的限制，在计算任何概率的时候，态矢量必须先归一化。

极限运算、级数展开这样的数学操作的合法性，一个完美的基于希尔伯特空间所建立的数学家园就存在了。它是一个超越经典力学但又在经典极限下可以包含经典力学的内容的，用于描述量子力学的数学工具。我们对其要求看似抽象，但结合其背后的物理实在的意义是非常具体的。我们可以利用这个数学工具，对一个量子系统的量子态进行严格、完备的理论描述。

1.3.1.2 算符公设与测量公设

现在，有了描述量子态的态矢量公设（量子态是希尔伯特空间中一条射线）。下一步，从“物理学关心的是物质的存在形式与运动规律，以及我们如何对其进行感知”来看，我们就应该关注如何去感知量子态的存在了。在量子力学这个描述微观世界的物理学基础理论中，算符公设（第二条）与测量公设（第三条）都是针对这个“感知”来进行设计的[41]。和经典世界存在和被感知之间平铺直叙的关系不同[42]，量子世界中，这里有很多奇妙的、我们从日常生活经验中很难直接抽象出来的特征。考虑到这两天公设的内容是相关的，我们在本小节把这两条公设放在一起讨论。

其中，算符公设针对的就是我们用什么来感知微观世界？我们的答案是“可

[41]后面的第四条公设，即演化公设，对应的是“运动规律”。这样，前四条合在一起，就把“物理学关心的是物质的存在形式与运动规律以及我们如何对其进行感知”包含全了。最后一条多粒子对称公设，即全同性原理，是我们在面对量子多体问题时进行的补充。五条合在一起，就构成了量子力学最基本的理论架构。

[42]从牛顿利用其力学描述了行星运动开始，人们对经典世界中运动的描述就是有个运动，它有轨道。我描述这个轨道，观测也不影响这个轨道。轨道的存在与轨道被感知之间的关系是 trivial 的。

观测量”。这个可观测量，在量子力学理论体系中，要用算符进行表达。因此，算符公设的具体内容可表述为：*可观测量应该由线性的、厄米的算符来表达*[43]。既然是可观测量，就必须有个测量过程，它符合某种规律。测量公设描述的就是这个测量需要满足的规律。它告诉我们：*测量可观测量的时候，所有可能的结果都是该算符的本征值。我们测得的结果是其本征值，某次测量所得到的量子态就是算符的这个本征值相应的本征态（这里自然就要有统计解释的问题）*。

这两条公设在量子力学形式体系起到的作用是承上启下。如果说第一公设搭建了“舞台”（希尔伯特空间）并定义了“演员”（态矢量），那么这两条公设就是定义了“演员身上可被观察的属性（可观测量）”以及“我们如何从演员身上读取信息”。下面，我们将从：1）使用算符的物理必然性；2）本征值与本征态的物理意义；3）厄米性的要求；4）测量过程的非幺正本质；5）一些算符的对易与不对易关系这五个维度，来统一地针对该这两条公设的内容进行一个相对详细的剖析。

先看在这里*“使用算符而不是数的必然性”*。这个在 1.1 节实际上我们已经讲过了。经过二十多年的探索，人们就是认识到了不能用传统的基于微积分描述的轨道的“运动学”把位置、动量当作数来理解微观世界的“运动”。与之相应，海森堡在 1925 年提出了基于光谱这样的可观测量来构造量子力学的关键的物理思想。他的理论经过玻恩、约当的整理，成为了我们在 1.1 节介绍的矩阵力学。在矩阵力学里，没有明确地讲态矢量是什么？这是因为我们需要求解的本征态在那里已经被当作基来进行处理了。这样的话，整个理论只需要围绕可观测量算符

[43]在经典物理中，我们常说力学量。但在量子力学中，实际上没有力学量，只有可观测量。可观测量在经典极限下回到了经典的力学量。

构造即可。这也为矩阵力学带来了实用性的困难，毕竟在复杂的真实体系中，本征态我们并不知道。矩阵力学之后，在波函数方法这条路上，人们实际上就是把这里的新的运动学内容分成两部分：*波函数描述态，算符描述可观测量*。这样，各司其职，量子力学理论在数学处理上就会变得更加友善[44]。现在，教科书中，我们就是按照这个划分来理解量子力学的。不管是那条路，只要我们是基于希尔伯特空间来描述量子世界的运动学[45]，算符（厄米算符）必将取代数（实数）来描述量子世界中的可观测量。

之后，我们来看“本征值与本征态的物理意义”这一点。本征值是测量可能结果的集合，本征态是测量结果确定的那些特殊状态，对应的是这个系统的希尔伯特空间中的一条射线。理解这句话，我们可以从一个可观测量算符$\hat{A}$的本征值方程出发来解释这个事情：

$$\hat{A}|a_n\rangle = a_n|a_n\rangle$$

公式 1.65

这里，$|a_n\rangle$的物理意义很明确，就是用a_n来标识对这个可观测量$\hat{A}$进行测量获得的值是a_n的量子态。这和我们前面讲的$|x\rangle$、$|p\rangle$是一致的。换句话讲，对$|a_n\rangle$这个态，如果用算符$\hat{A}$来测量的话，获得a_n这个值的几率就是一。由于是物理量，这

[44]在第二章，量子动力学的部分，我们会讲三种绘景。在那三种绘景中，薛定谔绘景遵守这个习惯。海森堡和相互作用绘景实际上不遵守这个习惯。当然，它们之间有等价性，可保障我们在实际计算过程中的严格。

[45]后面，我们会解释路径积分方法。它和这里我们介绍的内容形式上略有不同，但本质上是等价的。在路径积分的语言中，人们是基于构型空间（在单体系统就是位置空间）来描述量子世界的运动学。这样的话，用希尔伯特空间来描述量子世界的运动学可以被用路径积分来描述量子世界的运动学来绕过。更多解释请大家耐下心来，看后面的第八章。

个a_n必须为实数[46]。同时，因为一次测量的结果是确定的，另一次测量如果是不同本征值，那不同本征值对应的本征态必须是彼此正交的，因为它们描述的必须是互斥的物理情况。因此，对于$|a_n\rangle$标识的量子态，它不可能包含任何$|a_m\rangle$的成分（如果$a_m \neq a_n$）。当然，这里$|a_n\rangle$并不一定对应希尔伯特空间的一个维度，存在本征值简并的情况。这时，$|a_n\rangle$可以构成希尔伯特空间的某个高维的子空间。这是对公式 1.65 所定义的本征态与本征值的解释。

在这些解释的基础上，从公式 1.65 出发，我们是可以想象人们是能够通过可观测量算符$\hat{A}$所对应的操作进行测量，进而筛选出结果为a_n的一批$|a_n\rangle$态的样本的（若本征值简并，则制备到对应的子空间）。这也构成了现在大家经常听到的量子态工程和量子信息处理的基础。

同时，基于公式 1.65 所定义的算符$\hat{A}$的本征态，我们可以把该量子系统的任意一个量子态进行如下展开：

$$|\psi\rangle = \sum_n c_n |a_n\rangle$$

公式 1.66

这个公式的基础，是公设一里面的态叠加原理。这里，各个$|a_n\rangle$本身是量子态，它们的线性组合也是量子态。但是，测量公设告诉我们用$\hat{A}$作用到这个$|\psi\rangle$上的话，测量结果一定会让这个量子态在某个$|a_n\rangle$的投影。出现这个结果的几率是$|c_n|^2$。而多次测量的统计平均得到的这个可观察量在量子态$|\psi\rangle$上的期待值则是：

$$\langle\psi|\hat{A}|\psi\rangle = \sum_n |c_n|^2 a_n$$

[46]关于这一点，不光量子力学有这个要求，在著名的李-杨相变理论中，实数对应稳定的物理世界（在那里就是相）也是核心点【Yang，1952】【Lee，1952】【叶麒俊，2024】【王方成，2025】。

公式 1.67

这个期待值的解释是基于统计的。

在“本征值与本征态的物理意义”这一点的最后，我们还需要说明单个可观测量的本征态通常不能唯一地标记希尔伯特空间中的每一个独立状态。我们往往需要一组互相兼容（对易）的可观测量来共同标记一个量子态。它们合在一起，对应的才是描述这个量子系统的完全的信息，才是希尔伯特空间的一组完备基。这组算符称为完备对易可观测量集（Complete Set of Commuting Observables，简称 CSCO）。以原子轨道为例，在不考虑自旋的情况下，氢原子中能量算符$\hat{H}$只与主量子数n相关。在其它原子体系，它也只和n与l相关。但空间自由度为三，我们需要三个量子数才能确定量子态。因此，$\hat{H}$的本征态都不够成完备基。在不考虑自旋的情况下，我们需要的 CSCO 是$\{\hat{H},\hat{L}^2,\hat{L}_z\}$[47]。

现在，我们理解了可观测量需要用算符表达，理解了本征值与本征态的物理意义。*第三点我们需要理解的，是可观测量算符必须是个厄米算符*。这个要求由“可观测量对于任何量子态的期待值必须是实数”决定的。对于任意的$|\psi\rangle$，对于任意一个可观测量$\hat{A}$，它的期待值等于 $\langle\psi|\hat{A}|\psi\rangle$。由于这个期待值是实数，那必然有：

$$\langle\psi|\hat{A}|\psi\rangle = \langle\psi|\hat{A}|\psi\rangle^* = \langle\psi|\hat{A}^\dagger|\psi\rangle$$

公式 1.68

对任意的$|\psi\rangle$都成立，这样一个性质。这也就要求$\hat{A}^\dagger$必须等于$\hat{A}$。这样，可观测量算符必须是厄米算符这个性质背后原因也就清楚了。

第四点，我们来讨论一下测量的非厄米本质。在量子力学中，态矢量的演化

[47] 至于具体为什么会是$\{\hat{H},\hat{L}^2,\hat{L}_z\}$这三个，我们在后面的对称性部分会讲。

遵循两种截然不同的规则：一种是薛定谔方程描述的连续、确定、幺正的演化（第三条公设中我们会详细讨论）；另一种是测量发生时发生的概率性的、非幺正的投影。前者统治着系统在无观测时的演化，后者描述了观测者介入时观测对系统进行的投影操作。测量过程的非幺正本质，正是量子理论与经典物理最深刻的分歧所在，也是测量问题的核心来源。它看起来非常的反常识，但我们需要强调这种所谓的常识实际上是我们接触的宏观世界中的规律，它并不适用于微观世界，因为我们一般接触不到微观世界。在微观世界中，这些反常识的东西反而应该成为常识。这没有任何不合理的地方[48]。

这里，测量非厄米性的体现是它彻底打乱了系统的幺正演化。测量后的态依赖于测量结果，而不同初态可能以不同概率投影到同一末态。这也就意味着测量这个操作严格意义上应该标记为：

$$\hat{A}|\psi\rangle \to |a_n\rangle$$

公式 1.69

它的意思就是当我用$\hat{A}$这个算符对$|\psi\rangle$这个态进行测量的时候，得到某个投影$|a_n\rangle$。如果$|\psi\rangle$本身就是某个$|a_n\rangle$，那与上式对应，有：

$$\hat{A}|a_n\rangle = |a_n\rangle$$

公式 1.70

这里测量对量子态就不存在任何影响。换句话讲，公式 1.69 与 1.70 在量子力学里面存在本质的区别，一个存在测量带来的量子态的变化（公式 1.69），人们习惯地描述为塌缩，但或许用投影来描述更合适；一个因为这个态本身就是算符$\hat{A}$的本征态，测量过程就是一个等号。

[48]学习物理学，本质上就是用理性去理解客观世界。这也是 rational、rationalize、rationalization 这些词在我们的学科中很重要的原因。

在最一般的情况中，$|\psi\rangle$并不对应$\hat{A}$的本征态，测量过程应该用公式 1.69 来描述。这时，如果$|\psi\rangle$的展开形式如果满足公式 1.66，那测量后系统落在$|a_n\rangle$上的几率就是$|c_n|^2$。同时，一旦如公式 1.69 所标识的测量发生，系统是无法从投影后的$|a_n\rangle$唯一地恢复到测量前的态$|\psi\rangle$的。这也就意味着测量前量子态的信息丢失了。这与幺正演化（信息守恒）正相反。这里，非幺正性并非来自数学疏忽，而是来自我们描述测量行为时不可回避的物理现实：信息被记录并固化为一个结果，同时丢失了关于未实现可能性的相位信息。这一特性既是量子理论成功解释微观世界随机性的来源，也是其基础诠释争议的根源。理解这一本质，是把握量子力学哲学深度与数学结构之间张力的关键一步。

最后，在理解了上面四点的基础上，我们把*算符之间的对易与不对易关系作为落脚点来结束本小节关于算符公设与测量公设的讨论*。在算符公设中，我们的核心思想是可观测量(经典极限下就是经典力学中的力学量)对应线性厄米算符。而在测量公设中，我们的核心思想是测量后系统将投影到相应的本征态。初看之下，这两条公设似乎各自独立：一个构建数学工具，一个规定物理操作。然而，深入分析我们会意识到它们之间并非松散的并列关系，而是通过算符之间的对易与不对易关系，被有机地焊接成一个逻辑自洽、预言精准的整体。可以说，对易关系正是这两大公设的共同落脚点和逻辑枢纽。看似不合理的包含不确定性的测量（后面我们会讲），在线性代数的数学语言体系下，是一个自然的结果。

两个算符对易，是它们能够被同时对角化、拥有完备的共同本征态的充要条件。这个定理是构建可观测量完备集、理解测量理论的数学基石。关于这一点，我们可以*分两部分来进行证明*。

首先是充分性，也就是从：

$$\left[\hat{A},\hat{B}\right]=0$$

公式 1.71

出发，我们要推出：存在一组完备的态矢量$\{|\psi_i\rangle\}$，它们同时是$\hat{A}$和$\hat{B}$的本征态（从互易出发来证明它们具有可以同时被对角化的完备的本征态）。

这部分的证明过程可以从算符$\hat{A}$对应的本征态出发，简单分为四步。

1. 算符$\hat{A}$的本征态构成这个量子系统的完备基，我们按其本征值分组，每个本征值a_n对应的g_n维的简并子空间V_n。
2. 每个子空间，在算符$\hat{B}$的作用下，因为$\hat{A}$和$\hat{B}$互易，对$\hat{A}$一定是封闭的。这是因为我们可以取V_n中的任何一个向量$|\psi\rangle$，把$\hat{B}$作用上去，得到$|\phi\rangle$。如果我们可以得到$\hat{A}|\phi\rangle=a_n|\phi\rangle$，就可知这个子空间在$\hat{B}$的作用下形成的空间对$\hat{A}$是封闭的。而这一点很容易证明，因为：$\hat{A}|\phi\rangle=\hat{A}\hat{B}|\psi\rangle=\hat{B}\hat{A}|\psi\rangle=\hat{B}a_n|\psi\rangle=a_n\hat{B}|\psi\rangle=a_n|\phi\rangle$。
3. 第三小步，在这个V_n内，我们又可以利用算符$\hat{B}$将其对角化，进而将这里面的基组整理为：$\hat{B}|v,n\rangle=b_v|v,n\rangle$。对每个$|v,n\rangle$，因为上一小步证明了它对$\hat{A}$是封闭的。因此，自然有：$\hat{A}|v,n\rangle=a_n|v,n\rangle$。
4. 最后，对$\hat{A}$的每个本征值，重复第三小步的操作，就可以得到对$\hat{A}$和$\hat{B}$这个算符组而言一组正交完备的基$|v,n\rangle$。它们具有可以同时被对角化的完备的本征态。

之后，我们证明必要性。也就是说从存在一组对$\hat{A}$和$\hat{B}$同时对角化的基$\{|\psi_i\rangle\}$，满足$\hat{A}|\psi_i\rangle=a_i|\psi_i\rangle$和$\hat{B}|\psi_i\rangle=b_i|\psi_i\rangle$出发，证明：$\left[\hat{A},\hat{B}\right]=0$。

要证明它们互易，我们需要说明的是$\hat{A}\hat{B}=\hat{B}\hat{A}$。换句话讲，就是$\hat{A}\hat{B}$和$\hat{B}\hat{A}$作用到这个量子系统的希尔伯特空间中的任何一个向量上$|\phi\rangle$上，都有：$\hat{A}\hat{B}|\phi\rangle=$

$\hat{B}\hat{A}|\phi\rangle$。这个很直接，因为这个$|\phi\rangle$总是可以写为：$|\phi\rangle = \sum_i c_i |\psi_i\rangle$。这样的话，就有：

$$\hat{A}\hat{B}|\phi\rangle = \sum_i c_i \hat{A}\hat{B}|\psi_i\rangle = \sum_i c_i \hat{A}b_i|\psi_i\rangle = \sum_i c_i b_i\hat{A}|\psi_i\rangle = \sum_i c_i b_i a_i|\psi_i\rangle$$

公式 1.72

和：

$$\hat{B}\hat{A}|\phi\rangle = \sum_i c_i \hat{B}\hat{A}|\psi_i\rangle = \sum_i c_i \hat{B}a_i|\psi_i\rangle = \sum_i c_i a_i\hat{B}|\psi_i\rangle = \sum_i c_i a_i b_i|\psi_i\rangle$$

公式 1.73

两者显然是相等的。因此，有$[\hat{A}, \hat{B}] = 0$。

现在，利用“两个算符对易，是它们能够被同时对角化、拥有完备的共同本征态的充要条件”这个性质，我们就可以*基于互易的算符，对整个希尔伯特空间进行对角化*。这样，我们就可以确定前面讲的完备对易可观测量集（CSCO）。这样做的好处是每次测量都不会破坏之前测量的本征态。在所有测量完成后，这个量子态所有的自由度就都被探测到了。量子态的存在形式，也就被彻底确定了。

到这里，我们把最后一点“对易与不对易”中的“对易”讲完了。但我们需要知道的是量子力学中，可观测量不对易的情况实际上更加普遍。它是线性代数这个量子力学的数学结构的结果。毕竟，矩阵乘法在多数情况下是不对易的。这里，最简单的例子就是位置与动量，它们的不对易形式是公式 1.16。为了方便大家阅读，我们再写一些：

$$[\hat{x}, \hat{p}] = i\hbar$$

公式 1.74

对于这样的不对易的可观测量算符，根据前面证明的“*两个算符对易，是它们能*

够被同时对角化、拥有完备的共同本征态的充要条件”的逆否命题，我们知道它们不对易的话就不能被同时对角化。因此，*算符不对易这个数学性质在量子力学中对应的物理结果是测量之间相互影响的情况*，后面跟我们讲的测不准关系，实际上也是它的一种体现。

至此，通过前面列的五点，即 1）使用算符的物理必然性、2）本征值与本征态的物理意义、3）厄米性的要求、4）测量过程的非幺正本质、5）算符的对易与不对易关系，我们对算符公设与测量公设的讨论也就结束了。

很多教材上单独拿出来讲的波函数的统计诠释在这里被我们归结为第二点（本征值与本征态的物理意义)。它是测量公设的基本内容。在量子力学的理论架构中，由于它属于公设层面的内容。重要性自然是不容被低估的!

1.3.1.3 演化公设

现在，我们通过第一公设搭建了量子力学的“舞台”（希尔伯特空间）并定义了“演员”(态矢量)，通过第二、第三条公设就是定义了“演员身上可被观察的属性（用可观测量定义算符公设中的算符)”以及“我们如何从演员身上读取信息（测量公设相关内容)”。关于量子力学这个物理学的基本理论中“物质的存在形式以及如何被描述和感知”的部分的内容就结束了。第四条公设叫演化公设，它描述的量子世界中“物质的运动规律”[49]。前四条公设都不牵扯到多体系统。针对多体系统的公设，由第五条来完成。在笔者看来，从十九世纪末的庞加莱开

[49]但实际上基于演化方程我们是可以推出定态方程的，定态方程对应的又是存在形式。考虑到运动是绝对的，存在本身不能忽略运动，大家在理解我们这个展开逻辑的时候也不要太机械。笔者选择这样的表述是因为对于初学者，这样有比较好的带入感，笔者觉得读起来会友善一些。

始，多体问题就一直是物理学中最核心的挑战。笔者最近几年的工作也集中于此。这个我们后面会细讲。本小节，我们先看演化公设。

演化公设描述的核心内容，是在没有观测者干预时，一个孤立量子系统的态矢量应该如何随时间连续、确定性地演变？我们也可以把这个内容称之为量子动力学。因为我们熟悉的表达形式就是含时薛定谔方程，有时，它也被称为薛定谔方程公设。后面第二章，在我们讲量子动力学的时候，我们会讲到一个量子系统的演化本身是可以由几个*相互等价*的绘景来描述。因此，我们要知道演化公设的内容不一定必须用含时薛定谔方程的形式来表达。但我们在建立理论基础的时候，考虑到含时薛定谔的形式比较简单，我们还是选择避免用这种等价性将事情说得太复杂。我们采取的策略是从含时薛定谔方程出发来讲一个量子力学系统的演化。重点，是对量子态的演化所蕴含的物理内容的讨论。

一个孤立量子系统的态矢量$|\psi(t)\rangle$随时间的演化遵循薛定谔方程：

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = \hat{H}(t)|\psi(t)\rangle$$

公式 1.74

这个公式中，我们把哈密顿量写成了含时的形式，用以表达最广泛的情形。但它可以是一个不含时的算符。不含时的时候，公式 1.74 描述的是一个保守系统的演化，演化算符可以直接用一个指数函数来表达：

$$\hat{U}(t,t_0) = e^{-i\hat{H}(t-t_0)/\hbar}$$

公式 1.75

这里的$\hat{U}(t,t_0)$是通过：

$$|\psi(t)\rangle = \hat{U}(t,t_0)|\psi(t_0)\rangle$$

公式 1.76

定义的。它用以描述一个系统的量子态从t_0时刻向t时刻的演化。

当然，公式 1.74 中的$\hat{H}(t)$可以含时。这种含时的情况又可以被分为两种，一种是含时项是一个经典的外场，另一种是含时项是个量子化的外场（用场算符描述）。这些都比较复杂，特别是后者，不适合在本节讲。在第二章，我们讲量子动力学的时候，会针对这两种情况进行更为详细的介绍。这里，我们先以不含时这个相对简单的情况为例，把演化公设中一些非常核心的概念（比如*确定性演化*、*幺正演化*、*定态*）讲清楚。

先看*确定性演化*。在测量公设中，我们看到了量子世界的不确定性：测量结果概率性，态矢量的投影是非幺正的、非连续的。演化公设则揭示了一个完全相反的图景。在没有测量打扰的孤立系统中，态矢量的演化是完全确定性的。这意味着，只要我们知道系统在某一初始时刻t_0的态$|\psi(t)\rangle$和系统的哈密顿量$\hat{H}$，我们就可以通过求解薛定谔方程，唯一地预言未来任意时刻t的态$|\psi(t)\rangle$。这种演化是连续的、可逆的，信息不会丢失。这体现了量子力学与经典力学在精神上的继承：两者在最底层的动力学规律上都是决定论的。概率性仅仅在我们进行“测量”这一动作时才会涌现。

除了确定性，演化公设的*第二条*核心内容是这种演化必须满足一个非常重要的限制：量子态的归一化不能被破坏。如果一个态在初始时刻满足：

$$\langle\psi(t_0)|\psi(t_0)\rangle = 1$$

公式 1.77

那么，在之后的任意一个时刻t，因为态矢量的模平方最终要被解释为概率，量子态$\psi(t)$必满足：

$$\langle\psi(t)|\psi(t)\rangle = 1$$

公式 1.78

如果演化过程中归一化条件不再成立，那么概率解释就会失效。这就带来了一个后果，即*演化算符本身必须一个幺正算符*。具体，可以写为：

$$\hat{U}^{\dagger}(t,t_0)\hat{U}(t,t_0)=I$$

公式 1.79

这是一个物理的要求。

这个演化算符幺正性的物理的要求和我们算符公设中可观测量算符必须是厄米算符是相互呼应的。在那里，我们以物理学量的本征值必须为实数推出了其算符必须厄米的结论。这里，从时间演化算符必须幺正，我们也可以推出哈密顿量（作为可观测量的一个典型代表）必须厄米。这是因为公式 1.75 和 1.76 告诉我们：

$$|\psi(t)\rangle=e^{-i\hat{H}(t-t_0)/\hbar}|\psi(t_0)\rangle$$

公式 1.80

这样的话，它对应的左矢也就等于：

$$\langle\psi(t)|=\langle\psi(t_0)|e^{i\hat{H}^{\dagger}(t-t_0)/\hbar}$$

公式 1.81

为了让$\langle\psi(t)|\psi(t)\rangle=1$，就要有$\hat{H}^{\dagger}=\hat{H}$。

对于演化算符，除了前面提到的总概率守恒，它的另外一个物理意义是这个量子系统的任意两个态之间的内积不随时间改变。也就是说，如果两个初态之间的“相似程度”是某个数，那么在同一个孤立系统的演化下，这种相似程度不会被改变。从信息的角度看，幺正演化意味着信息不会凭空产生，也不会凭空消失。这也是为什么我们说孤立量子系统的演化是可逆的。

演化公设的*第三点*需要说明的内容是公式 1.74 中的$|\psi(t)\rangle$系统在时刻t的任

意态矢量，*并不要求它一定是哈密顿量的本征态，也不要求它一定是某个可观测量的本征态*。这里，对它的要求仅仅是它属于这个系统的希尔伯特空间，且归一。满足这两点，它就是系统的态矢量。它的演化就满足公式 1.74。如果哈密顿量不含时，在$t = 0$的时刻，我们*可以根据*:

$$\widehat{H}|\psi_n\rangle = E_n|\psi_n\rangle$$

公式 1.82

定出系统的一系列定态，它是这个希尔伯特空间中一组完备的基。这时，系统的任意一个态矢量都可以做下面这个展开:

$$|\psi\rangle = \sum\nolimits_n c_n|\psi_n\rangle$$

公式 1.83

演化也可以通过下式表达:

$$|\psi(t)\rangle = \sum\nolimits_n c_n e^{-iE_n t/\hbar}|\psi_n\rangle$$

公式 1.84

每个能量本征态分量都会获得自己的相位因子。不同能量对应不同相位演化速度，因此相对相位会改变.这种相对相位的改变会导致很多可观测物理量随时间变化，例如振荡、干涉、跃迁等（第二章我们还会详细讲解）。

最后，对于一般含时哈密顿量$\widehat{H}(t)$，我们也要知道它的演化算符仍然存在并且仍然是幺正的，只是通常不能简单写成普通指数函数。关于这一点，大家可以从一个无限小时间间隔的演化算符的具体形式理解。对于公式 1.76 定义的演化算符，当时间从t_0演化到$t_0 + \mathrm{d}t$后，它的形式应该满足:

$$\widehat{U}(t_0 + \mathrm{d}t, t_0)|\psi(t_0)\rangle = |\psi(t_0 + \mathrm{d}t)\rangle \approx |\psi(t_0)\rangle + \left.\frac{\partial|\psi(t)\rangle}{\partial t}\right|_{t=t_0} \mathrm{d}t$$

$$= |\psi(t_0)\rangle - \frac{i\widehat{H}(t_0)\mathrm{d}t}{\hbar}|\psi(t_0)\rangle \approx e^{-i\frac{\widehat{H}(t_0)\mathrm{d}t}{\hbar}}|\psi(t_0)\rangle$$

公式 1.85

因为$\widehat{H}(t_0)$厄米，$\widehat{U}(t_0+\mathrm{d}t,t_0)$自然幺正。进而，因为每一个小段都幺正，一个有限时长的演化算符自然也就幺正。它和前面的讨论唯一的区别就是由于不同时刻的哈密顿量算符不一定对易，总的演化算符需要用时间排序指数来表示。这在第二章我们会细讲。这里，我们只需要记住：无论哈密顿量是否含时，只要系统是封闭的，演化就应当由幺正算符描述。非幺正演化通常表明，我们讨论的已不是一个完整的孤立系统，而是某个与外界发生相互作用的子系统。这种情形常出现在测量、耗散、退相干等开放量子系统问题中。

1.3.1.4 多粒子对称公设

前面四条公设基本上完成了对一个单体的量子系统进行描述的理论框架：态矢量生活在希尔伯特空间中，可观测量由厄米算符表示，测量会给出本征值并伴随态矢量向相关可观测量在该本征态上的投影，孤立系统的连续演化由薛定谔方程或等价的幺正演化描述。

但是，在真实的自然界中，我们感兴趣的物理系统往往并不是单个粒子，而是由多个子系统组成的复合系统。例如，一个氢原子至少包含一个电子和一个质子；一个固体包含大量电子和原子核；一个量子计算机包含许多个量子比特。如何从单个量子系统的希尔伯特空间出发，构造多个量子系统组成的总系统，也是量子力学理论框架中非常关键的一部分。这就是第五条公设要回答的问题。

第五条公设可以表述为：如果一个复合量子系统由若干个子系统组成，其中第i个子系统的希尔伯特空间为$\mathcal{H}_i$。那么，总系统的希尔伯特空间是各个子系统

的希尔伯特空间的张量积。它可以表示为：

$$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_N$$

公式 1.86

其中，这个N是复合系统中子系统的个数。总系统的态矢量是这个张量积空间中的矢量。当复合系统中包含全同粒子的时候，多体态必须满足额外的对称性要求（对于费米子，总的态矢量满足交换反对称；对于玻色子，满足交换对称）。

这是多粒子量子力学的基本结构。它看起来只是一段很随意说出来的话，但实际上蕴含了量子理论与经典理论最深刻的差别之一，比如纠缠。为了理解这一点，我们可以先考虑两个子系统A和B。若子系统A的态为：

$$|\psi_A\rangle \in \mathcal{H}_A$$

公式 1.87

子系统B的态为：

$$|\psi_B\rangle \in \mathcal{H}_B$$

公式 1.88

那么复合系统的一个可能态就是：

$$|\Psi\rangle = |\psi_A\rangle \otimes |\psi_B\rangle$$

公式 1.89

这样的态称为直积态（product state），或者叫可分态。它的物理意义是若系统A有自己的态，系统B也有自己的态，则总的态就是把两者并列放在一起，由公式 1.89 表达。

以两个自旋 1/2 的粒子形成的双体系统为例[50]，如果两个粒子的自旋分别处在$|\uparrow\rangle$、$|\downarrow\rangle$，那么总系统一方面，可以处在：

[50]目前我们还没有讲自旋，暂且先用一下这个自由度。但是，从公式 1.87 到公式 1.92 的讨论暂时先不考虑全同性。之后的讨论才会考虑全同性。

$$|\uparrow\rangle \otimes |\downarrow\rangle$$

公式 1.90

（我们可以把它写为$|\uparrow\rangle|\downarrow\rangle$或者$|\uparrow\downarrow\rangle$）。这个就是直积态。但同时，由于希尔伯特空间是线性空间，直积态的线性叠加也是被允许的量子态，它可以具有：

$$\frac{1}{\sqrt{2}}(|\uparrow\rangle|\downarrow\rangle + |\downarrow\rangle|\uparrow\rangle)$$

公式 1.91

的形式。这个态，就不能写成某个单粒子态与另一个单粒子态的简单的张量积了。这样的态被称为纠缠态。基于纠缠态，我们就可以理解纠缠了。它是多体量子系统中最重要、也最反直觉的概念之一。它说明在量子力学中总系统的信息并不总是能够还原为各个子系统信息的简单相加。换句话说，整体不只是部分的总和。

这一点，和经典物理是完全不同的。在经典力学中，如果我们知道每个粒子的位置和动量，原则上就知道了整个系统的状态。多粒子系统的状态空间通常是各个粒子相空间的笛卡尔积。但在量子力学中，总希尔伯特空间是张量积空间。张量积带来的直接后果就是，复合系统可以存在纠缠态。纠缠使得两个空间上相距很远的子系统之间也可以具有非经典关联。爱因斯坦、波多尔斯基和罗森曾经通过这个提出了 EPR 佯谬来质疑这种结构的完备性。但后来的贝尔不等式及其实验检验则表明这种非经典关联并不是理论中的数学幻觉，而是自然界真实存在的物理。在第三章，我们会详细讲这些内容。

除了量子态，对于可观测量，在张量积空间中，它也有相应的结构。比如，若$\hat{A}$是作用在子系统A上的可观测量，而$\hat{B}$是作用在子系统B上的可观测量，那么在总空间中，它们应分别写成：

$$\hat{A} \otimes \hat{I}$$

公式 1.92

和

$$\hat{I} \otimes \hat{B}$$

公式 1.93

这里的$\hat{I}$是相应子系统希尔伯特空间上的单位算符。这个写法的意思是：如果我们只测量第一个子系统，那么这个测量操作只作用在第一个希尔伯特空间上，对第二个子系统不做任何操作；反之亦然。这时，如果我们同时考虑两个子系统的联合观测量，就可以出现：

$$\hat{A} \otimes \hat{B}$$

公式 1.94

这样的算符。

这里的可观测量最简单的例子就是哈密顿量。对于一个由两个子系统构成的复合系统，如果它们各自的哈密顿量分别为$\widehat{H}_A$和$\widehat{H}_B$，并且两者之间存在相互作用$\widehat{H}_{\text{int}}$，那么总哈密顿量常写成：

$$\widehat{H}_A \otimes \hat{I} + \hat{I} \otimes \widehat{H}_B + \widehat{H}_{\text{int}}$$

公式 1.95

其中前两项描述两个子系统各自的能量，第三项描述它们之间的相互作用。没有相互作用时，如果初态是直积态，那么在演化过程中通常会保持直积形式。但一旦存在相互作用，直积态往往会演化成纠缠态。因此，相互作用是产生纠缠的重要物理机制。

除了可区分的子系统构成的复合系统，多粒子公设还包含一个更特殊也更深刻的问题：全同粒子。所谓全同粒子，是指两个粒子不仅具有相同质量、电荷、自旋等物理属性，而且在原则上无法通过任何实验加以区分。例如，所有电子都是全同粒子，所有光子也是全同粒子。在经典物理中，即使两个粒子完全相同，

我们仍然可以想象给它们贴上标签：“粒子 1”和“粒子 2”。但在量子力学中，对于真正的全同粒子，这种标签不具有物理意义。

因此，全同粒子的多体态必须满足额外的对称性要求。对于两个全同粒子，交换它们的标签不应该改变可观测的物理结果。设交换算符为$\hat{P}_{12}$，它的作用是把粒子 1 和粒子 2 对调。量子态在交换下只能相差一个整体相位。进一步分析可以得到，三维空间中普通粒子只允许两种可能：

$$\hat{P}_{12}|\Psi\rangle = |\Psi\rangle$$

公式 1.96

以及：

$$\hat{P}_{12}|\Psi\rangle = -|\Psi\rangle$$

公式 1.97

第一种称为交换对称态，对应玻色子；第二种称为交换反对称态，对应费米子。

对于玻色子，多粒子波函数在交换任意两个粒子时保持不变。光子、声子、氦-4 原子等都是玻色子。玻色子可以有任意多个粒子占据同一个单粒子量子态，这导致了玻色-爱因斯坦凝聚、激光等重要现象。对于费米子，多粒子波函数在交换任意两个粒子时改变符号。电子、质子、中子等都是费米子。反对称性直接导致泡利不相容原理：两个完全相同的费米子不能占据同一个单粒子态。因为如果两个费米子处于同一个态，交换它们后态应当不变；但反对称性又要求态变号，于是只能得到零态。这一原理是原子壳层结构、元素周期表、固体能带结构乃至白矮星稳定性的基础。

这样的话，对于全同粒子，总希尔伯特空间就不再是简单地使用完整张量积空间，而是要取其中满足对称性要求的子空间。对于N个全同玻色子，物理态属于对称张量积空间；对于N个全同费米子，物理态属于反对称张量积空间。这个

事实最终会自然引向二次量子化（第三章内容）。在二次量子化语言中，我们不再追踪“第几个粒子处于什么态”，而是直接描述“某个单粒子态上有多少个粒子”。这正是多体量子理论中最方便、最深刻的表述方式。更多内容，我们会在第三章详细讲解。

最后，我们讨论一下第五条公设的意义。简单来讲，它使得量子力学不再只是描述一个粒子的理论，而成为描述原子、分子、凝聚态物质、量子场以及量子信息的基础。同时，我们要明白多粒子公设绝不是一个简单的“系统相加规则”。张量积结构带来了纠缠，全同粒子要求带来了交换对称性，而这两者共同构成了量子多体物理的核心。正是在这里，量子力学真正展现出与经典世界最深刻的差异。后面我们将看到，无论是原子结构、化学键、金属导电性、超导、超流，还是量子计算和量子通信，背后都离不开这一条公设所建立的基本框架。

1.3.2 测不准关系、互补性原理

有了前面建立的量子力学理论架构，我们现在讨论一下在教科书中我们经常看到的测不准关系和互补性原理。我们要知道测不准关系实际上是量子力学建立之后的一个推论，互补性原理是玻尔在 1927 年索尔维会议前把测不准关系上升到哲学高度的一个表达。在有些地方，关于测不准关系的介绍经常会和五条公设混起来讲。我们这里想强调的是按照公理化的理论的架构要求，它们都是量子力学理论架构建立后的结果，而非原因。

1.3.2.1 测不准关系

从公式 1.16 出发，以位置、动量算符为例，测不准关系就很容易被推出了。

我们在本小节进行这样一个推导。我们首先需要理解的是在针对位置、动量这些物理学量进行测量的时候，在量子力学的语言中我们完成的动作是:

$$\hat{x}|\alpha\rangle \to x|\alpha\rangle$$

公式 1.98

这里，$|\alpha\rangle$是我们要测量的量子态。$\hat{x}$是坐标算符。$\hat{x}$作用到$|\alpha\rangle$上，进行了一次位置测量。x是一个实数，是这次测量的测量值。在公式 1.98 中我们没有用等号，是因为不同次的测量可以给出不同的结果。我们用→来强调这是一次测量的结果。后面，我们不再做这种强调。当我进行了很多次测量之后，这些测量值可以有一个统计平均值$\bar{x}$，它是一个实数。

由于量子力学从本质上是统计性的，这个$\bar{x}$只能理解为好多次测量的测量值的平均。这个时候，如果我们稍微学过一些统计，我们就知道我们每次的测量值都是在这个平均值附近分布的。这个分布有个不确定度（均方差），它可以用$\sqrt{\langle(\Delta x)^2\rangle}$来表示。其中，

$$\Delta x = x - \bar{x}$$

公式 1.99

它代表的是我每次测量的值x对统计平均值$\bar{x}$的偏离。我们每进行一次测量，都有个Δx，它可正可负。我们对这个测量出来的Δx取平方，再对$|\alpha\rangle$这个量子态进行统计平均再求根号，对应的就是在针对这个态进行测量的时候的测量值的均方差。它应该用$\sqrt{\langle(\Delta x)^2\rangle}$表示，描述的就是对坐标进行测量的不确定度。

同样，对这个量子态，也可以测量动量，每次有个偏差:

$$\Delta p = p - \bar{p}$$

公式 1.100

其中p是每次的测量值，$\bar{p}$是多次测量的平均值。按统计的方法来描述，每次的测

量值都在$\bar{p}$附近分布，均方差为$\sqrt{\langle(\Delta p)^2\rangle}$。这些，和位置的测量都是对应的。

测不准关系，说的并不是某次测量的Δx和Δp之间的乘积的关系，而是对位置进行测量的均方差$\sqrt{\langle(\Delta x)^2\rangle}$和对动量进行测量的均方差$\sqrt{\langle(\Delta p)^2\rangle}$两者相乘，满足的关系。它可以表达为：

$$\sqrt{\langle(\Delta x)^2\rangle}\sqrt{\langle(\Delta p)^2\rangle} \geq \frac{\hbar}{2}$$

公式 1.101

这一点我们必须首先明确！在明确这一点后，我们就可以利用公式 8.45，来推公式 1.101 了。

起点，自然是由公式 1.99−1.100 给出的Δx和Δp来求$\sqrt{\langle(\Delta x)^2\rangle}$与$\sqrt{\langle(\Delta p)^2\rangle}$。而它们，又是基于$\langle(\Delta x)^2\rangle$与$\langle(\Delta p)^2\rangle$的。这个$\langle(\Delta x)^2\rangle$满足：

$$\langle(\Delta x)^2\rangle = \langle\alpha|(\hat{x}-\bar{x})^2|\alpha\rangle = \langle\alpha|(\hat{x}-\bar{x})(\hat{x}-\bar{x})|\alpha\rangle$$

公式 1.102

这里，由于$(\hat{x}-\bar{x})$是个厄米算符，我们知道：

$$\langle(\Delta x)^2\rangle = \langle(\hat{x}-\bar{x})\alpha|(\hat{x}-\bar{x})\alpha\rangle$$

公式 1.103

它代表的，是希尔伯特空间中一个向量$|(\hat{x}-\bar{x})\alpha\rangle$和它自身的内积，是一个大于等于零的实数。和公式 1.102 对应，我们还有：

$$\langle(\Delta p)^2\rangle = \langle\alpha|(\hat{p}-\bar{p})^2|\alpha\rangle = \langle\alpha|(\hat{p}-\bar{p})(\hat{p}-\bar{p})|\alpha\rangle$$

公式 1.104

同样，由于$(\hat{p}-\bar{p})$是个厄米算符，它继续等于：

$$\langle(\Delta p)^2\rangle = \langle(\hat{p}-\bar{p})\alpha|(\hat{p}-\bar{p})\alpha\rangle$$

公式 1.105

它代表的，也是希尔伯特空间中一个向量$|(\hat{p}-\bar{p})\alpha\rangle$和它自身的内积，也是一个大于等于零的实数。

这个时候，如果我们记：

$$|f\rangle = |(\hat{x} - \bar{x})\alpha\rangle$$

公式 1.106

以及

$$|g\rangle = |(\hat{p} - \bar{p})\alpha\rangle$$

公式 1.107

那么，很显然：

$$\langle(\Delta x)^2\rangle\langle(\Delta p)^2\rangle = \langle f|f\rangle\langle g|g\rangle$$

公式 1.108

这一步是我们理解测不准关系的一个非常重要的中间步骤。

之后，我们就可以利用希尔伯特空间中关于向量内积的柯西−施瓦茨不等式（Cauchy−Schwarz Inequility）：

$$\langle f|f\rangle\langle g|g\rangle \geq |\langle f|g\rangle|^2$$

公式 1.109

得到：

$$\langle(\Delta x)^2\rangle\langle(\Delta p)^2\rangle \geq |\langle f|g\rangle|^2$$

公式 1.110

这里，$\langle f|g\rangle$代表向量$|f\rangle$、$|g\rangle$之间的内积，它是一个复数。这个复数的模，是大于它的虚部的平方的（任何复数的模都等于其实部平方加虚部平方进而大于其虚部平方）。因此，继续有：

$$\langle(\Delta x)^2\rangle\langle(\Delta p)^2\rangle \geq |\langle f|g\rangle|^2 \geq \left[\frac{1}{2i}(\langle f|g\rangle - \langle g|f\rangle)\right]^2$$

公式 1.111

这个时候，我们再把$|f\rangle$、$|g\rangle$本身的定义代入（公式 1.106、公式 1.107），一切就会很明朗了。

其细节，可以描述为：

$$\langle f|g\rangle = \langle(\hat{x}-\bar{x})\alpha|(\hat{p}-\bar{p})\alpha\rangle = \langle\hat{x}\alpha|\hat{p}\alpha\rangle - \bar{p}\langle\hat{x}\alpha|\alpha\rangle - \bar{x}\langle\alpha|\hat{p}\alpha\rangle + \bar{x}\bar{p}\langle\alpha|\alpha\rangle$$

公式 1.112

由于$\hat{x}$、$\hat{p}$的厄米性，以及$|\alpha\rangle$的归一性，它进一步等于：

$$\langle f|g\rangle = \langle\alpha|\hat{x}\hat{p}|\alpha\rangle - \bar{x}\bar{p}$$

公式 1.113

与这个等式相应，也有：

$$\langle g|f\rangle = \langle\alpha|\hat{p}\hat{x}|\alpha\rangle - \bar{x}\bar{p}$$

公式 1.114

这样，把公式 1.113 和 1.114 代入公式 1.111 和 1.112，就有：

$$\langle(\Delta x)^2\rangle\langle(\Delta p)^2\rangle \geq \left[\left\langle\alpha\left|\frac{1}{2i}[\hat{x}\hat{p}-\hat{p}\hat{x}]\right|\alpha\right\rangle\right]^2$$

公式 1.115

这个时候，把公式 1.16 代入上式，就有：

$$\langle(\Delta x)^2\rangle\langle(\Delta p)^2\rangle \geq \left[\left\langle\alpha\left|\frac{1}{2i}i\hbar\right|\alpha\right\rangle\right]^2 = \left[\frac{\hbar}{2}\right]^2$$

公式 1.116

这样的话，就有公式 1.101，也就是我们常说的海森堡不确定性关系了。1927 年，海森堡在他的文章 Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik（英文对应是：On the Perceptual Content of Quantum Theoretical Kinematics and Mechanics）中将这个关于测量的不确定关系明确提出【Heisenberg, 1927】。之后，它也成了量子力学中一个极其关键的内容。它是量子力学统计性的一个重要体现。

1.3.2.2 互补性原理

历史上，海森堡的这个工作出来后，马上发生的事件就是著名的 1927 年索尔维会议了。在这个会议上，量子力学的统计性解释是核心话题。相关争论，既有玻恩的波函数统计性解释的部分，也有海森堡的测不准关系的部分。当然，我们知道玻尔在这个讨论中也扮演了一个非常重要的角色。他把这个测不准关系再次上升到了一个哲学的层面，将其描述为“互补原理”[51]。但客观的讲，当我利用线性代数的方法来描述微观粒子的运动（具体形式可以是矩阵力学、可以是波动力学、可以是路径积分），当我承认在量子力学中的观测需要基于统计来进行解释之后，测不准关系是在数学上可以推导出来的确定的结果。这里，我们无需再把问题虚化（或者说复杂化），上述到哲学的高度[52]。

1.3.3 电子自旋

目前，在仅仅考虑空间自由度的情况下，量子力学的理论架构已经完整了。它的一些推论我们也已经做了一些讨论。但是对于任何一个学习过《量子力学》的学生而言，看到这里我们一定会觉得这个理论架构是不完整的。其中最明显的，就是自旋这个在量子力学中涌现出来的全新的自由度我们基本上没有提。本小节按这个目的来架构。但我们一定要知道想要把自旋彻底讲清的话，需要对称性的

[51]它的正式提出实际上是在当年的索尔维会议之前在科莫会议上（Como Conference, Italy）。一定程度上，我们也可以说当年的索尔维会议玻尔是有备而来。这种哲学层面的论述往往是很难反驳的。今天，当我们重新整理量子力学的理论体系，我们可以很清楚的认识到五条公设是可以建立严格的量子力学理论架构的，互补性原理并不需要。

[52]整体而言，海森堡和玻尔比较擅长进行哲学上的思考。玻恩、狄拉克、薛定谔这些人比较善于把这些哲学上的思考整理为数学表达，进而建立系统的物理学理论。对于物理学革命中量子力学的诞生而言，这两种风格缺一不可！

知识、相对论量子力学的知识。因此，这个概念我们也不可能在本章彻底讲清。本着循序渐进的原则，我们在本小节先以“电子自旋的新的运动学内容”和“这个新的运动学内容的数学表达”为着手点，把电子自旋最基本的内容讲一下。这样，在后面要用到自旋的时候，我们可以先用。之后，在第五章和第十章，我们再针对自旋的物理本质进行更深入的讨论。

1.3.3.1 电子自旋的新的运动学内容

我们以历史为切入点来展开本小节介绍，这样轻松一些。应该说人类对自旋自由度的认识和人类对空间自由度在量子力学层面的认识的历史进程是有很大相似性的[53]，都和原子光谱方面的研究密切相关。但不同之处，是在自旋这里人们必须关注磁场对原子光谱的影响。与之相关的效应叫塞曼效应。它是 1896 年由荷兰物理学家塞曼（Pieter Zeeman，1865−1943 年）发现的。具体的内容，就是塞曼通过观测，发现磁场下钠原子的 D 谱线不再是之前的一条，而是劈裂为三条【Zeeman，1896】。

关于磁场影响原子光谱这一点现在看起来没什么，但只要我们把自己放到塞曼所处的环境，我们是应该可以意识到它的非平凡性的。在塞曼之前，我们要知道类似谱线和原子类型是一一对应的。从 1860 年基尔霍夫认识到这一点开始，人们就开始基于这一现象进行元素分析【李新征，2026】。很大程度上讲，这是当时人们基于原子论认识世界的基础。基尔霍夫的学生门捷列夫甚至还基于这种研

[53]人类对空间自由度的运动很早就有认识了，从古希腊到经典力学。但人类对空间自由度的运动在量子力学层面的认识，则是从上世纪初开始的。关于自旋自由度，它本质就是量子力学的内容，在经典力学中根本不存在。

究建立了元素周期表来帮人们理解不同元素的物理与化学性质。这塞曼这里，他认识到即使是同样的元素，只要加磁场，谱线也会发生变化。这是超出了传统的元素分析的内容。自然，这个现象本身也会蕴涵新的物理[54]。

在塞曼发现塞曼效应之后，他的老师洛伦兹（Hendrik Antoon Lorentz，1853–1928 年）马上基于量子化的电子轨道磁矩对其在理论上进行了一定程度的解释。他们也因此一起获得了 1902 年的诺贝尔物理奖[55]。但需要说明的是在这个奖被颁布之前，在 1897 年，爱尔兰物理学家普雷斯顿（Thomas Preston，1860－1900 年）就已经发现塞曼最早看到的钠原子 D 谱线的劈裂并非三条，而是很多条，谱线的结构也比塞曼最早的报道复杂很多【Preston，1898】。洛伦兹的理论可以解释塞曼观测到的的结果，因此人们把那个现象叫做正常塞曼效应。对于洛伦兹的理论解释不了的更为复杂的原子光谱劈裂现象，人们称之为反常塞曼效应。

这个疑问困扰了人们很久，直至二十多年后电子自旋自由度的出现，才被最终解决。期间，人们尝试了各种原子模型，甚至也会借助原子实而非价电子的角动量来进行解释。在这个过程中，玻尔－索末菲模型扮演了非常重要的角色【Sommerfeld，1916a–1916b】。玻尔－索末菲模型是玻尔模型的升级版。玻尔模型解释氢原子非常成功，但在解释其它原子的时候，就遇到了不少困难。其根本原因与对称性相关。在氢原子模型中，除了转动对称性，原子核与电子的相互作用

[54]实验最大的意义就在于发现当时的“理性”之外的物理。

[55]他们的获奖词是：in recognition of the extraordinary service they rendered by their researches into the influence of magnetism upon radiation phenomena。应该说，原子光谱的外场的响应这种现象一定是蕴藏新物理的。另一个粒子就是斯塔克效应，它由德国物理学家斯塔克（Johannes Stark，1874－1957 年）在 1913 年发现，他也因此获得了 1919 年的诺贝尔物理奖。斯塔克的获奖词是：for his discovery of the Doppler effect in canal rays and the splitting of spectral lines in electric fields。从这些奖项的确定，读者也可以感受到“场与物质的相互作用”在物理学中的地位。

势满足$-1/r$这样一个特殊的数学形式，与电子原子核的距离成反比。这个数学形式会让电子轨道的对称性从转动群的 SO(3)变成一个对称性更高的叫做 SO(4)的对称群（图 1.3）。这也让相对简单的玻尔模型在这个体系中非常适用。但其它原子的外层电子和原子核的相互作用因为收到了内层电子的屏蔽，不再满足$-1/r$这样一个距离依赖关系。与之相应，电子能级的能量不光依赖于主量子数n，还依赖于角量子数l。

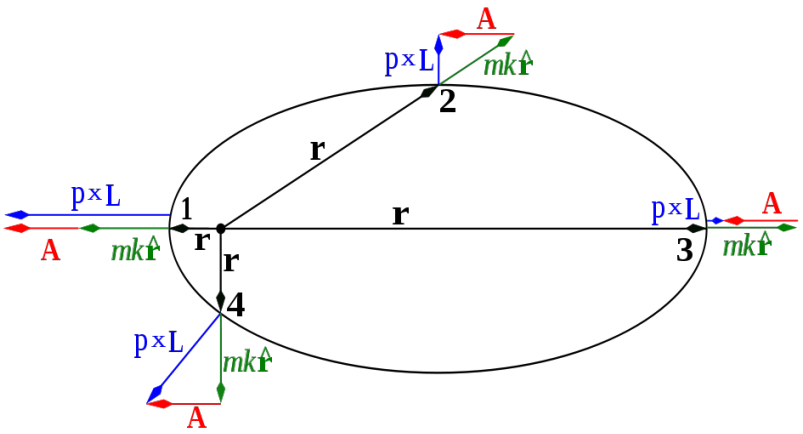


图 1.3 在势能满足$-1/r$这样一个特殊形式的中心力场体系，系统的对称群不是简单的转动群 SO(3)。在具有转动群 SO(3)对称性的体系，物理上的守恒量仅仅是能量、总角动量、以及总角动量在某个方向的分量。因此，在玻尔-索末菲的原子模型中，有 n、l、m 三个量子数。在具备$-1/r$这样一个势能项的中心力场体系，除了这些守恒量，本图中的红色箭头标出的矢量$\vec{A}$也守恒。人们认识到这个守恒量的存在经历了一百多年，从 18 世纪末、19 世纪初的拉普拉斯（Pierre-Simon Laplace，1749-1827 年）开始，到 20 世纪出的龙格（Carl David Tolmé Runge，1856－1927 年）和楞次（Wilhelm Lenz，1888−1957 年）。因此，这个守恒量也叫做 Laplace−Runge−Lenz 守恒量。具有这个守恒量的体系就是氢原子，更高的对称性也保障了更高的能量简并度，这也是为什么比玻尔-索末菲的原子模型更为简单的玻尔原子模型可以在氢原子中获得成功的原因。

这时，如果再加上外磁场，角量子数在某个方向的分量 m 也会起作用。因此，磁场中即使只考虑电子的空间自由度，其能量都会依赖于 n、l、m 这三个量

子数。这三个量子数恰恰来自于空间的三个自由度。在玻尔-索末菲的原子模型中，它们可以被考虑。因为这个原因，在二十世纪第二个十年的后半部分和 20 年代开始的时候，以索末菲课题组为中心，人们进行了很多的关于反常塞曼效应的研究【Debye，1916】【Lande，1921】【Lande，1923】【Sommerfeld，1923】。比如，有德拜、朗德（Alfred Landé，1888－1976 年）、泡利、海森堡，都是他的学生。

在所有这些工作中，突破性的认识，目前人们认为是来自于 1922-1923 年期间索末菲、朗德、泡利的合作[56]。这些合作让他们开始认识到辅助量子数对理解反常塞曼光谱的重要性，也就是之前的 n、l、m 不够了。但开始，他们认为这个辅助量子数与原子实有关。1924 年底，泡利意识到原子的角动量只能来源于外层电子，否则塞曼效应分叉的宽度就将依赖于原子序数，而这与事实不符。但泡利没有往自旋这个方向进行思考。

1925 年，来自美国的年轻学生克罗尼格（Ralph Kronig，1904－1995 年）访问已经在图宾根大学任职的朗德[57]。朗德向他展示了泡利给自己的信件，其中提

[56]泡利当时刚刚从索末菲组里拿到博士学位毕业，海森堡还是学生。朗德毕业的要早一些。德拜比他们大很多，这个时候已经在哥廷根做过教授转到苏黎世理工学院的教授了。因此，在 1922-1923 这两年，主要是朗德、泡利、海森堡在做。

[57]这里有两点可以说明。第一点，在上世纪的 20 年代，美国最好的博士生在毕业之后到欧洲游学是一个时尚。像前面提到的鲍林、奥本海默、拉比，以及这里的克罗尼格，都是其中的代表。他们从欧洲游学回去之后，将最先进的量子力学、统计物理带回美国。加之一些欧洲科学家的加盟（比如爱因斯坦，以及后面要提到的乌伦贝克、高德施密特）、二战对欧洲的毁灭性打击，美国在短短十几年时间内一跃成为全世界的科学中心。第二点，朗德在 1921 年曾经被帕刑（Friedrich Paschen，1865-1947 年）热情的邀请到图宾根大学工作，朗德也接受了邀请，并在那里工作了一段时间。但据朗德本人的回忆，那段经历很痛苦。图宾根大学的教授委员会中，有一帮非常反对聘用他的人。这些人认为他既是犹太人（当时的德国是反犹的）又是共产主义者（这个也是当时的一个社会现象）。朗德承认自己的犹太血统，但他早已放弃了犹太教（类似的事情也发生在玻恩身上，这是当时一批德国的犹太科学家的统一问题，他们更认同自己是德国人，但仍因为血缘被排斥），同时他也否认自己是共产主义者。上世纪 20 年代末，朗德访问美国后，迅速选择在

到在解释反常塞曼效应的时候，n、l、m 之外的第四个量子数不能归于原子实这样一个观点。既然来自于价电子，克罗尼格马上想到如果不把电子当作点电荷，它自己是可以有一个自旋自由度的。只要将这个自由度量子化，并通过以电子为中心的坐标系与以原子核为中心的坐标系进行坐标变换，就可以引入一个自旋–轨道耦合项，来对反常塞曼效应进行一定程度的解释。当然，这个解释并不完美。比如，克罗尼格算出来的劈裂值是实验值的一半。并且这个差别是系统性的，不依赖于到底是哪个具体的能级。

当克罗尼格满怀期待找到泡利进行讨论的时候，泡利对他这个想法又提出了更多的质疑。泡利的理由非常简单，可总结为两个。一个是前面提到的理论值是实验值的一半的问题。另一个，是当时人们知道电子的半径大概的估计是10^{-15}米。对于这样一个尺寸的带点球体，如果想产生解释反常塞曼效应所需要的磁矩，那么这个球的边缘的转动速率需要比光速大一到两个数量级。这个是不符合狭义相对论的基本假设的。后来，有人说泡利在当时就因为这两点意识到了这个自由度意义非凡，并且认为它本质上一定是没有经典对应的，只是他的追求完美的性格让他不能赞同克罗尼格当时对电子自旋的略显经典的描述。但作为一个历史事实，就是克罗尼格身为一个 21 岁的小伙，在当时确实很受挫。

之后，克罗尼格结束了对图宾根的访问，来到哥本哈根。在这里，克罗尼格的观点同样遭到了玻尔、海森堡、克拉默斯（Hans Kramers，1894－1952 年）等人的冷遇。前面提到过，这个时间点恰恰也是海森堡提出矩阵力学的关键想法的时间点。考虑到克罗尼格只是一个从美国过来交流学习的小伙儿，之前并没有在

那里生活。因此，很多文献介绍朗德的时候，就是 German-American Physicist。这个，和玻恩、海森堡、泡利他们的介绍方式是很不一样的。

这个群体中建立所谓的学术信誉，玻尔、海森堡、泡利、克拉默斯等人或许在心理上有轻视其想法的倾向[58]。同时，克罗尼格本人也还是太年轻，因此没有坚持，这也导致他最终没有将他的这个想法写成文章并发表。这无疑，是一个非常令人遗憾的决定。

仅仅几个月之后，1925 年夏，在荷兰的莱顿大学，埃伦费斯特的两个学生乌伦贝克（George Eugene Uhlenbeck，1900−1988 年）、高德施密特（Samuel Abraham Goudsmit，1902−1978 年）关注了同样的问题。他们也注意到了泡利的关于第四个自由度的理论。之前，他们对量子理论就有“每个量子数都对应一个自由度”这样一个理解。因此，他们认为这第四个量子数既然和价电子相关，那么价电子必然有个人们之前忽略掉的自由度。对于点电荷，欧氏空间的三个自由度就已经对应了 n、l、m 这三个量子数。那么第四个量子数一定是要超越“电子是点电荷”这样一个“运动学描述”的。很自然，他们也想到了自旋。只要将自旋自由度量子化，反常塞曼效应就有可能被更好地解释。基于这样一个理解，他们在 1925 年完成了关于电子自旋的理论工作，并于 11 月发表在一个德文杂志<Naturwissenschaften >（我们翻译为《自然科学》）上[59]【Uhlenbeck，1925】。

1925 年 12 月，莱顿大学举办了一个庆祝洛伦兹获得博士学位 50 周年的活动。在这个活动中，爱因斯坦对乌伦贝克、高德施密特的工作进行了鼓励【傅海

[58]毕竟很多科学的思想在提出的那一刻都不是完美的，认可思想远比认可结果重要，就像他们认可海森堡的矩阵力学理论，在提出的时候也不是完美的，但他们认可海森堡的思想。

[59]这个过程中，埃伦费斯特对于这两个年轻人的新想法始终是支持的。他也带着这两个年轻人找了洛伦兹。洛伦兹在很认真地听取了他们的想法后，并没有马上给答复，而是跟他们说自己要想一段时间。在这个过程中，乌伦贝克和高德施密特完成了第一篇文章并进行了投稿。之后，洛伦兹对他们进行了回复，从经典电动力学的角度提出了不少问题。这两个人当时也想过撤稿。但埃伦费斯特对他们说的是“你们年轻，犯些错误没什么”。类似这种包容性的态度，在笔者看来是好导师的一个重要标准。

辉，2002】。有趣的是，在玻尔、海森堡又一次听到与克罗尼格类似的想法的时候，他们都选择了一定程度的接受，虽然之前克罗尼格的理论中的两个问题依然存在。因此，玻尔也与乌伦贝克、高德施密特进行了进一步的交流。但遗憾的是，泡利依然坚持之前的想法。1926 年 2 月，Nature 杂志刊登了乌伦贝克、高德施密特与玻尔就自旋概念进行讨论的邮件及评论【UhlenBeck，1926】。

同样在这个时间，来自英国的托马斯（Llewellyn Hilleth Thomas，1903－1992 年）通过引入相对论效应，解决了理论值与实验值相差一倍的问题【Thomas，1926】。至于经典图像中电子球边界的速度依然超过光速，那与量子力学当时的问题一样，属于“运动学”方面的新内容，并不影响观测。至此，电子自旋这个概念被物理学界接受。人们也开始针对它进行各种理论与实验的深入研究，包括 1927 年泡利通过泡利矩阵将薛定谔方程变为两分量方程进而描述自旋-轨道耦合的工作【Pauli，1927a】、同期达尔文的类似工作【Darwin，1927】、1928 年狄拉克的相对论波动方程【Dirac，1928】。这些内容，加上之前已经存在的 1924 与 1926 年费米和狄拉克完成的关于费米－狄拉克分布的工作【Fermi，1924】【Dirac，1926】、1925 年泡利完成的不相容原理的研究【Pauli，1925a－1925b】，我们教科书中经常呈现的关于电子自旋的基本内容就基本成型了。

最后，需要说明的是在很多材料上人们会通过介绍 1921－1922 年的施特恩－格拉赫实验（Stern－Gerlach Experiment）来引入电子自旋【Stern，1921】【Gerlach，1922a－1922c】。但实际上，在这个实验最初的理论解释中，并不牵扯到电子的自旋自由度，而是基于量子化的轨道角动量来进行的。现在，人们习惯于从这个实验出发开始解释电子自旋自由度与量子力学基本原理的【Sakurai，2017】【孙昌璞，2024】【曹庆宏、杨李林，2025】。从知识传播的角度，这当然是非常高效

的。但为了能够让读者更加准确的了解这段历史并在以后的类似环境中进行相同的思考，我们在接受这种利于知识传播的做法时，也必须把这些细节交代清楚。

同时，我们也需要说明这个施特恩–格拉赫实验本身也没有获得诺奖。斯特恩（Otto Stern，1888 – 1969 年）在 1944 年获得了 1943 年的诺贝尔物理奖，他的获奖词是：for his contribution to the development of the molecular ray method and his discovery of the magnetic moment of the proton。而格拉赫，则始终没有获得诺奖。上面这些在电子自旋发现过程中起到关键作用的物理学家们，他们有不少是获得过诺奖的，但得奖理由都不是自旋。乌伦贝克、高德施密特也一直都没有获得来自诺奖委员会的认可。作为一个从事物理学研究若干年的一个科研人员，笔者能够感受到我们的民众对诺奖的渴望。如果将这种渴望看作是对科学精神的渴望，笔者肯定是非常认同的。但很多时候，笔者也能感受到很多时候我们的实际情况并不是这样，对世俗荣誉的渴望或许占据了更大的比重。这就使我们的科学研究完全偏离了本书强调的自然哲学、物理学的本质。电子自旋，在笔者看来是一个极其深刻的物理学内容。本小节提到的这些工作，也都是意义远大于相关荣誉的工作。值得我们慢慢品味。

1.3.3.2 这个新的运动学内容的数学表达

在自旋自由度的数学表述方面，我们以电子为例，把重点放到*泡利矩阵*的引入、它对*二阶零迹厄米矩阵*遍历、以及旋量波函数上。这样做的一个好处，是通过泡利矩阵，读者应该能够在一定程度上理解自旋自由度在“运动学”上为微观粒子赋予的量子力学新内容。同时，我们也不至于过多的陷入在描述类似空间转

动这样的操作对电子自旋自由度的影响时必须引入的二维特殊酉群这样的偏数学的语言的理解中。这些内容我们会在第四章讲解了对称性之后进行更为系统的介绍。本章，笔者仅仅希望能让读者理解泡利在 1924 年为什么要引入泡利矩阵？以及泡利矩阵带给了我们什么样的比仅考虑空间自由度的薛定谔方程更多的量子力学的新内容上。

为了达到这个目的，我们要把自己放到泡利的位置。我们需要明确的第一点，就是他是索末菲的学生，一直关注反常塞曼效应的研究。我们还需要知道他的师弟海森堡在 1921 年的时候就引入了角动量半量子数的概念来理解反常塞曼效应（当时这个角动量还是轨道角动量）。之前，像玻尔、索末菲等人使用的都是整数量子数【Bohr，1913a−1913c】【Sommerfeld，1916】。当然，在 1921 年这个时间节点，量子力学远远没有建立，自旋的概念也没有被提出。因此，海森堡的轨道角量子数为半整数的概念虽然获得了包括其导师索末菲在内的一些人的认可，但也同样备受争议。有一种说法是这种争议也使得他的博士论文不得不选择另外一个关于湍流的题目来完成[60]。

有趣的是，泡利在海森堡提出这个半量子数的概念的时候（1921 年），是这个概念的一个反对者。他的理由很简单：既然有 1/2，为什么没有其它分数？但是到了 1924、1925 年这个时间节点，当泡利认识到第四个量子数必须来自于价电子，而实验上在加入外磁场后，最简单的反常塞曼效应就是把原来一个峰的谱线劈裂成了两个，1/2 对应的 $2L+1$ 恰恰等于 2。他也就在一定程度上接受海森堡关于存在半量子数的主张。这就为泡利矩阵的诞生奠定了最重要的基础。

[60]更糟糕的是，他的博士论文答辩表现也不好。维恩是主考官，对他很不满意。这也给年轻的海森堡造成了一定的打击。好在导师索末菲对他还是足够认可的，使得他可以顺利毕业。

泡利矩阵对应的就是下面三个简单的2×2的矩阵：

$$\sigma_x=\begin{pmatrix}0&1\\1&0\end{pmatrix}、\sigma_y=\begin{pmatrix}0&-i\\i&0\end{pmatrix}、\sigma_z=\begin{pmatrix}1&0\\0&-1\end{pmatrix}$$

公式 1.117

它们都是二阶、零迹、厄米的。电子自旋算符，最为一个可观测量算符，有三个性质。首先，它是一个矢量，就像角动量。第二，它要能够和我们空间进行的操作进行一定程度的联系（进而带了自旋与磁场之间的耦合、自旋与轨道运动之间的耦合这样的性质）。第三，它要厄米。如果我们把它的形式设为：

$$\hat{S}=\frac{\hbar}{2}\vec{\sigma}=\frac{\hbar}{2}\left[\sigma_x\vec{\imath}+\sigma_y\vec{\jmath}+\sigma_z\vec{k}\right]=\frac{\hbar}{2}\left[\begin{pmatrix}0&1\\1&0\end{pmatrix}\vec{\imath}+\begin{pmatrix}0&-i\\i&0\end{pmatrix}\vec{\jmath}+\begin{pmatrix}1&0\\0&-1\end{pmatrix}\vec{k}\right]$$

公式 1.118

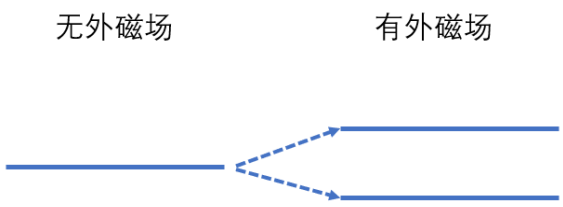
它恰恰符合这些要求。

之后，当公式 1.118 表达的自旋算符与外磁场相互作用的时候，我们参考的数学表达就是 1.3.1.4 小节我们有两个系统的时候哈密顿量的处理方式。按照公式 1.95 的指导，我们可以把其中的$\widehat{H}_{\mathrm{int}}$写为：

$$-\frac{e}{m_e}\hat{S}\cdot\vec{B}=-\frac{e\hbar}{2m_e}\left[\sigma_xB_x+\sigma_yB_y+\sigma_zB_z\right]=-\frac{e\hbar}{2m_e}\begin{pmatrix}B_z&B_x-iB_y\\B_x+iB_y&-B_z\end{pmatrix}$$

公式 1.119

从这个公式，我们可以看到泡利矩阵包含了二阶零迹厄米矩阵的所有三个维度。这是因为既然厄米零迹，就说明对角项必须为实数且合为零，而非对角要就必须实部相等、虚部相反。当这样一个自旋算符通过其自旋角动量与外场的耦合的时候，通过外场参数B_x、B_y、B_z的变化，人们可以遍历所有的二阶、零迹、厄米矩阵。这也给了实验观测所有的磁场变化的自由度。自旋自由度引入后，人们就可以理解反常塞曼效应了。它对应的劈裂的能级数是偶数，对应的角动量量子数是半整数。其中，关键的一分为二的劈裂，可参考图 1.4 来理解。



$$-\frac{e}{m_e}\hat{S}\cdot\vec{B}=-\frac{e\hbar}{2m_e}\left[\sigma_x B_x+\sigma_y B_y+\sigma_z B_z\right]=-\frac{e\hbar}{2m_e}\begin{pmatrix}B_z & B_x-iB_y\\ B_x+iB_y & -B_z\end{pmatrix}$$

图 1.4 自旋 1/2 引入后，能级在外磁场下从 1 劈裂为 2 的示意图。这里，泡利矩阵既简洁又准确。它完美地刻画出了一个可观测量在这里应该有的性质（二阶、零迹、厄米）。与外场的耦合，通过外场参数的变化，人们也可以通过泡利矩阵来遍历所有的二阶、零迹、厄米矩阵。这也给了实验观测所有需要的自由度。

至此，关于自旋的运动学的描述就清楚了。这里，基于泡利矩阵，人们可以引入自旋算符，进而遍历所有的二阶、零迹、厄米矩阵。后面，我们在讲两能级系统的时候，也会强调这三个矩阵。从数学结构上来讲，自旋 1/2 系统的希尔伯特空间与两能级系统的希尔伯特空间是一致的。这也是我们在做科研的时候，遇到两能级系统，总是把它往自旋 1/2 体系做类比的根本原因。

同时，和单纯的两能级系统不同，这里的自旋自由度可以与粒子运动的空间自由度进行耦合。由于人们针对空间自由度，在球对称模型中习惯于使用角量子数来描述其空间运动，自旋–轨道耦合（简称旋–轨耦合）也成了我们避不开的话题。这里，我们用最简单的单电子体系的问题来先进行一个简单的说明。耦合后，它的总的角动量$\hat{\mathrm{J}}$是其轨道角动量$\hat{\mathrm{L}}$与自旋角动量$\hat{\mathrm{S}}$的矢量和：

$$\hat{\mathrm{J}}=\hat{\mathrm{L}}+\hat{\mathrm{S}}$$

公式 1.120

由于电子自旋角动量在任何一个方向的分量就是量子化的$\hbar/2$与$-\hbar/2$，其波函数也会变成下面这种所谓“旋量（spinor）波函数”的形式：

$$\Psi_{jm}(\vec{\mathrm{r}},t)=\begin{pmatrix}\Psi_{jm}(\vec{\mathrm{r}},\hbar/2,t)\\ \Psi_{jm}(\vec{\mathrm{r}},-\hbar/2,t)\end{pmatrix}$$

公式 1.121

这里，我们把这个方向选为 z 轴。公式 1.121 中的$\Psi_{jm}(\vec{\mathrm{r}},\hbar/2,t)$就是该本征态自旋在 z 轴投影为$\hbar/2$的的空间依赖部分，$\Psi_{jm}(\vec{\mathrm{r}},-\hbar/2,t)$为该本征态自旋在 z 轴投影为$-\hbar/2$的空间依赖部分。大家可以理解为：考虑了电子自旋的波函数，是在原来的不考虑电子自旋的波函数的基础上，维度又增加了一倍即可。

至此，关于物理学革命中人们就电子自旋这个自由度的探索过程和关键的关于“运动学”的量子力学新内容我们也都介绍完毕。至于如何在物理学研究中运用这些新内容，感兴趣的读者请参考相关专业文献来进行进一步的学习【Sakurai，2017】。对于高自旋维度的体系，人们只需把公式 1.121 描述的旋量波函数改为一个$2S+1$维的旋量波函数即可[61]。相应的$2S+1$维度的自旋与外场的相互作用，人们也需要将2×2的泡利矩阵替换为这些$2S+1$维度的自旋空间中的角动量生成元所对应的矩阵来进行相似处理。物理学革命中人们在量子力学层面对微观粒子的新的“运动学描述”，在考虑了实空间的矩阵以及自旋空间的类似于泡利矩阵（自旋 1/2 粒子体系）的新内容之后，也基本完整。剩下的，就是对这些物理学新思想的细化与进一步应用了。

[61]对单电子，$S=1/2$，因此$2S+1=2$。对高自旋的体系，$2S+1$就是更大的整数。